American Society of Human Genetics 67th Annual Meeting October 17–21, 2017 in Orlando, Florida

POSTER ABSTRACTS

Session Topic/Title	Abstract/Poster Board Numbers	
	Start #	End #
Genome Structure and Function	373	438
Prenatal, Perinatal, and Reproductive Genetics	439	498
Genetic Counseling, ELSI, Education, and Health Services Research	499	593
Cancer Genetics	594	845
Mendelian Phenotypes	846	1180
Bioinformatics and Computational Approaches	1181	1457
Omics Technologies	1458	1541
Epigenetics and Gene Regulation	1542	1691
Developmental Genetics and Gene Function	1692	1795
Complex Traits and Polygenic Disorders	1796	2280
Evolution and Population Genetics	2281	2369
Molecular and Cytogenetic Diagnostics	2370	2601
Cardiovascular Phenotypes	2602	2760
Statistical Genetics and Genetic Epidemiology	2761	3020

Posters should remain on the board for all three days (Wednesday through Friday)

The program and abstract/poster board number next to each listing is followed by an **W** (Wednesday), **T** (Thursday), or **F** (Friday) to indicate the day on which authors must be present at their poster boards.

373W

The effect on the infection of Plasmodium falciparum by the increased Young's Modulus of G6PD deficient erythrocyte membrane. *W. Jiang[1], Z. Zhang[1,2], C. Jiang[1], X. Chen[3], Z. Fang[1], Y. Feng[1].* 1) Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen Univer, Guang Zhou, Guangdong, China; 2) Reproductive Medicine Research Center, Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou, 510275, China; 3) Department of Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou, 510623, China.

We screened more than 40,000 patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency and found that the G6PD Kaiping allele was under the most positive selection for fighting against malaria in the Chinese population. However, the mechanism is unknown. The current study was designed to investigate the anti-malarial effect and mechanism of G6PD deficiency. Dehydroepiandrosterone (DHEA) was utilised for inhibiting the G6PD activity of erythrocytes. Giemsa staining of blood smears were used for the detection of *Plasmodium falciparum* infection. An atomic force microscopy (AFM) was used for the analyses of morphology, roughness and Young's Modulus of the infective erythrocyte membrane. When G6PD activity was inhibited by DHEA, the infection rate of *P. falciparum* decreased and the Young's Modulus of the erythrocyte membrane increased with increasing DHEA concentrations. On the other hand, we found the number of knobs on the infective red blood cell surface decreased with increasing of DHEA concentrations by AFM. In the erythrocytic stage, the merozoites attached to the receptors on the surface of erythrocyte membrane through site specific identification. This will cause erythrocyte to generate extensive deformation. A parasitophorous vacuole will be formed around the merozoite recessed, merozoite can invade to erythrocyte through this parasitophorous vacuole, and it will be sealed after the invasion. We used spherical tip to detect the Young's Modulus of the erythrocyte membrane, and found it increased with increasing of DHEA concentrations. Tip: with the decrease of G6PD activity, under the same stress, the erythrocyte membrane deformability and elasticity decreased, stiffness increased , that is not conducive to the performance of the merozoite invasion of erythrocytes. So, the infection rate of *P. falciparum* declined. Understanding these points will have a major effect on searching for new anti-malarial drug targets.

374T

High-throughput functional genomic assessment of known and novel *HNF1A* missense variants to support clinical diagnostic interpretation. *S. Althari[1], L.A. Najmi[2,3], A.J. Bennett[1], M. van de Bunt[1,4], P.R. Njølstad[3,5], A.R. Majithia[6,7,8,9], M.I. McCarthy[1,4,10], A.L. Gloyn[1,4,10].* 1) Oxford Centre for Diabetes, Endocrinology and Metabolism, University, of Oxford, Oxford, United Kingdom; 2) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 3) Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 5) Department of Pediatrics, Haukeland University Hospital, Bergen, Norway; 6) Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 7) Diabetes Research Center, Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 8) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 9) Department of Medicine, Harvard Medical School, Boston, MA, USA; 10) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

Depending on frequency, penetrance and functional impact, heterozygous alleles in the transcription factor-encoding gene, *HNF1A*, cause Maturity Onset Diabetes of the Young (MODY), and/or increase Type 2 Diabetes (T2D) risk. A reliable molecular diagnosis for *HNF1A* variant carriers is a) crucial for effective disease management and treatment, and, b) contingent on an accurate understanding of the functional consequences of known and novel sequence variants. To mitigate this challenge in diagnostic interpretation, we have adopted an experimental strategy to interrogate all possible missense variants in *HNF1A* (11,989 alleles) *en masse*, using transcriptional targets as discriminators of molecular dysfunction. Markers of HNF-1a transactivation have not been identified for the purpose of massively parallel allelic screening; therefore, we conducted a systematic search for HNF-1a regulated transcripts in a high-throughput-compatible, physiologically relevant cell system. Based on cytogenetic, transcript and proteomic characterisation, the human liver cell line, HepG2, was selected. HepG2 cells were transduced with doxycycline (dox)-inducible *HNF1A* sequences. RNA from samples that showed dose-dependent induction of *HNF1A* was sequenced to identify differentially regulated targets. A total of 226 transcripts were differentially expressed (FDR ≤ 1%, $\log_2$FC ≥ 0.5 or ≤ -0.5) across dox doses (0-5 µg/ml), with highest exogenous *HNF1A* induction at 5 µg/ml (P=1.98x10^{-9}). We tested cell surface expressed targets, *CUBN* (P=1.48x10^{-7}), *CLDN2* (P=2.60x10^{-6}), *TMEM27* (P=3.07x10^{-3}), and *TM4SF4* (P=2.59x10^{-3}), for their ability to sensitively discriminate a series of *HNF1A* alleles (with strong genetic, clinical, and functional evidence) on the basis of molecular function. We achieved robust allelic segregation using TM4SF4 as a readout in flow cytometric analyses. The allelic series displayed a range of functional effects reflected by activity-dependant distributions of TM4SF4-positive cell populations: 2-8% from MODY-causal loss-of-function alleles, 30% from a low frequency allele associated with T2D risk, 60-70% from neutral alleles, and >80% from wild-type. In conclusion, we have built a robust and scalable assay that will allow pooled functional interrogation of an *HNF1A* saturation mutagenesis library, using endogenous transcriptional targets as markers of *HNF1A*-deficiency. The emerging data will aid in the interpretation of variants in the clinical diagnostic setting.

375F

A novel IDS gene mutation in two Japanese patients with severe mucopolysaccharidosis type II and correlation between developmental outcomes. *S. Kasuga[1], C. Kadono[2], S. Kudou[3], K. Fujita[3,4], N. Hikita[1], S. Nishigaki[3], T. Seto[1], T. Hamazaki[3], H. Shintaku[3].* 1) Department of Pediatrics, Osaka City University Graduate School of Medicine, Osaka, Japan,MD; 2) Department of Pediatrics, Osaka City University Graduate School of Medicine, Osaka, Japan,PhD; 3) Department of Pediatrics, Osaka City University Graduate School of Medicine, Osaka, Japan; 4) ATOX Co.,Ltd.

   Background: Mucopolysaccharidosis type II (MPSII, Hunter syndrome) is an X-linked recessive disorder caused by a deficiency of the lysosomal enzyme iduronate-2-sulphatase (IDS). Since not all the reported nonsense mutations of IDS gene caused severe phenotype, it is important to describe natural histories of patients with a specific mutation. Here we report a novel nonsense IDS gene mutation caused severe phenotypes in two unrelated patients. We further performed comparative structural analysis on this novel mutation with previously reported mutations and tentatively identified a critical region to manifest severe phenotype. **Patients and Methods:** We investigated two unrelated Japanese patients with the same unreported MPS II gene mutation, Q272X. They were diagnosed with high urinary glycosaminoglycans, low IDS enzyme activity in lymphocytes, and IDS gene mutation. We used Developmental Quotient (DQ) score by the Kyoto Scale of Psychological Development. Genomic DNA were isolated from their peripheral blood and PCR amplified each exon of *IDS* gene was analyzed by direct sequencing. The tertiary structure imaging of IDS was constructed using arylsulfatase as a template by homology modeling. **Results:** The first patient is 18 years old boy. He was diagnosed as MPSII at 1 year old by hepatomegaly and language delay. Head control was showed by 4 months after birth, walked alone by 14 months, and talked words by 18 months. Severe developmental retardation had manifested with constant declining in Developmental Quotient(DQ) score after 2 years old. He lost the ability of walking alone at 6 years old. Enzyme replacement therapy (ERT) was introduced since he was 9 years old, but now he is bedridden with enteral nutrition. The second patient is 4 years old boy. Limbs joint contraction was pointed out at the 3 years and ERT was initiated soon after the diagnosis. Although he had good development and growth history until by 3 years old, DQ score started declining thereafter. Based on our IDS structural analysis in silico, we found that severe phenotype was associated with surroundings of catalytic domain of IDS enzyme. **Conclusions:** In cases of nonsense or frame shift mutation of IDS, we may be able to predict the phenotypes of MPSII patients at early stage of the disease. It will be beneficial for the patients to intervene early when novel brain targeted treatments become available.

376W

Novel frame-shift mutation in the extracellular domain of WNT coreceptor, low-density lipoprotein receptor-related protein 6 in a Japanese family with autosomal dominant oligodontia and early onset metabolic syndrome. *H. Goto[1,9], M. Kimura[1,3,4], M. Nakashima[4], N. Tsuchida[4,5], J. Machida[1,2], T. Tatematsu[1,9], H. Miyachi[1], K. Takahashi[6], H. Izumi[7], A. Nakayama[8], Y. Higashi[9], K. Shimozato[1], N. Matsumoto[4], Y. Tokita[1,9].* 1) Maxillofacial Surgery, Aichi-Gakuin University, aichi ken nagoya shi, aichi, Japan; 2) Department of Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota, Japan; 3) Department of Oral and Maxillofacial Surgery, Ogaki Municipal Hospital, Ogaki, Gifu, Japan; 4) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan; 5) Department of Stem Cell and Immune Regulation, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan; 6) Department of Oral and Maxillofacial Surgery, School of Medicine, Kyoto University, Kyoto, Japan; 7) Department of Occupational Pneumology, Institute of Industrial Ecological Science, University of Occupational and Environmental Health, Kitakyushu, Japan; 8) Department of Embryology, and 9Department of Perinatology, Institute for Developmental Research, Aichi-Human Service Center, Kasugai, Japan; 9) Department of Perinatology, Institute for Developmental Research, Aichi-Human Service Center, Kasugai, Japan.

   Purpose: Congenital tooth agenesis is one of the most common anomalies in human development. The etiology of tooth agenesis, however, still remains unknown in many patients. We thus investigated the genetic cause of tooth agenesis in Japanese patients. **Methods:** We performed whole exome sequence analysis of a number of Japanese families with congenital tooth agenesis and their relatives. **Results:** A novel single nucleotide insertion in the *LRP6* gene (NM_002336.2:c.1924dup) was identified. This gene is responsible for WNT/β-catenin signaling as a coreceptor for WNT ligands. The single nucleotide insertion causes a frame-shift and introduces a premature stop codon in the extracellular region (p.(Ile642Asnfs*11)). Interestingly, tooth agenesis patients with this mutation had been diagnosed as hyperlipidemia. **Conclusions:** It has been demonstrated that null mutations in the *LRP6* gene cause selective tooth agenesis (STHAG7: MIM 616724), while other mutations in the *LRP6* gene leading single amino acid substitution at the extracellular domain are reported as causes for autosomal dominant early coronary artery disease and metabolic syndrome (ADCAD2: MIM 610947) to date. Our current study, however, shows that the gene mutation of LRP6 truncated form can cause a syndromic disease of familial hypercholesterolemia with tooth agenesis.

377T

Primary bilateral macronodular adrenocortical hyperplasia due to mutations in *ARMC5*: New mutations in humans and modeling in zebrafish.
A.G. Maria[1], F.R. Faucz[1], A. Berthon[1], F. Hannah-Shmouni[1], B. Feldman[2], J. Bertherat[3], C.A. Stratakis[1]. 1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, Maryland, USA; 2) Division of Developmental Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, Maryland, USA; 3) Institut National de la Santé et de la Recherche Médicale (INSERM) U1016, Centre National de la Recherche Scientifique (CNRS) UMR 8104, Institut Cochin, Paris, France; Department of Endocrinology, Referral Center for Rare Adrenal Diseases, Assistance Publique Hôpitaux de Paris, Hôpital Cochin, Paris.

Primary bilateral macronodular adrenocortical hyperplasia (PBMAH), a rare cause of ACTH-independent Cushing's syndrome (CS), is characterized by massive enlargement of the adrenal glands with multiple nodules (>1cm) with or without internodular atrophy and increased cortisol and/or aldosterone production. Armadillo Repeat Containing 5 (*ARMC5*) is a putative tumor suppressor gene whose mutations have recently been identified in about half of the patients with PBMAH. We sequenced blood DNA from 80 subjects with PBMAH for *ARMC5* variants. The functional role of ARMC5 was determined using a zebrafish model. Knockdown of the zebrafish ortholog *ARMC5* was performed using an antisense morpholino oligonucleotide (MO). The RNA from 100 3dpf (days post fertilization) embryos (50 from each group) was extracted and RNAseq was performed in those two groups and in the RNA extracted from an uninjected control group. Eighteen subjects (22.5%) had a damaging variant in *ARMC5* as predicted by an *in silico* analysis, including three not previously described: two missence (p.W476X and p.R362W) and one frameshift (p.C5795fs*112) variant. Two of eleven subjects with primary aldosteronism had damaging variants in *ARMC5*. All predicted damaging variants were located in regions that are shared by the three most frequent human *ARMC5* isoforms. Zebrafish RNAseq data suggest a role for Armc5 in apoptotic processes and steroid biosynthesis. Taken together, these data confirm our previous findings that damaging variants in *ARMC5* are frequent in PBMAH. Furthermore, the zebrafish model agrees with previously published human *in vitro* experiments and demonstrates a conservation of ARMC5 function in vertebrates. Pathway analysis highlights new candidates involved in post-translational modifications that may be involved in tumor formation initiated by *ARMC5* mutations.

378F

Cargo of miRNAs from synovial fluid exosomes contributes to pathogenesis of primary osteoarthritis.
M. Czarny-Ratajczak[1], V. Dasa[2], J. Li[3], J. Eastwood[1], C. Monjure[1], S.M. Jazwinski[1]. 1) Department of Medicine, Center for Aging, Tulane University, School of Medicine, New Orleans, LA; 2) Department of Orthopedic Surgery, Louisiana State University Health Sciences Center, New Orleans, LA; 3) Department of Global Biostatistics and Data Science, Tulane School of Public Health, New Orleans, LA.

Osteoarthritis (OA) is a chronic degenerative joint disorder and despite the high frequency of OA, the cause of the primary form of this disorder is still not elucidated. We have investigated the content of exosomes from synovial fluid and whether these vesicles released by synoviocytes into synovial fluid, transport the unique set of miRNAs that are involved in initiation and progression of osteoarthritis. The current knowledge about the exosomal miRNAs from synovial fluid is limited and a full profile of these molecules is not known. Exosomes were isolated from synovial fluid collected during joint replacement surgery in patients with osteoarthritis. We confirmed via electron microscopy and immuno-gold labeling the presence of exosomes in an isolated fraction. Expression of exosomal markers was confirmed via Western blot. RNA-Seq of miRNAs from nine patients was completed on the Ion Proton system and results were validated via qPCR in additional ten patients. NGS data were analyzed with small RNA analysis plugin (Ion Torrent Suite software) and edgeR. The analysis of potential targets of identified miRNAs and networks was conducted with the following databases: TargetScan, DIANA LAB, PICTAR, MIRANDA miRDB and IPA. We observed that the top ranked miRNAs are part of the networks, which include *TP53, IL-6,* NFKB complex and *ERK1/2,* according to IPA. Five miRNAs from this list are already linked to osteoarthritis. Target genes of top ranked miRNAs that overlap in four databases include: *TP53, MDM4, AFF4, NFIB etc.* We also detected TERRA transcripts in synovial fluid exosomes. Based on the current literature, this is the first study that shows the full profile of miRNAs from synovial fluid exosomes of patients with osteoarthritis and indicates that they are actively involved in the initiation and progression of this disorder. Targets regulated by these miRNAs are in control of cell death, cell cycle and inflammatory responses, which play significant role in OA pathogenesis. In addition, TERRA transcripts, which were identified by us in synovial fluid exosomes are known to stimulate inflammatory cytokines.

379W

Capture Hi-C identifies chromatin interactions between psoriasis-associated genetic loci and disease candidate genes. *H.F. Ray-Jones[1,2], A. McGovern[1], P. Martin[1], K. Duffus[1], S. Eyre[1], R.B. Warren[1,2].* 1) Arthritis Research UK Centre for Genetics and Genomics, Division of Musculoskeletal and Dermatological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom; 2) The Dermatology Centre, Salford Royal NHS Foundation Trust, University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom.

Purpose: Psoriasis is a common, complex autoimmune condition of the skin. Genome-wide association studies (GWAS) in psoriasis have identified many risk loci, but most lead variants are non-coding with an unknown function. It is hypothesised that these variants regulate gene function through enhancer-promoter chromatin interactions. Capture Hi-C (CHi-C) is a method combining chromosome conformation capture technology with a targeted capture step to discover interactions in specific regions of the genome. This work aimed to apply CHi-C in psoriasis risk loci using relevant cell types and conditions. **Methods**: CHi-C libraries were generated in duplicate using relevant cell lines: MyLa (CD8+ T cells) and HaCaT (keratinocytes). Two libraries were also generated from HaCaT cells stimulated with interferon gamma (IFN-γ). RNA capture baits were designed to target all known psoriasis-associated GWAS loci; each locus was defined by a set of variants in linkage disequilibrium with the lead SNP ($r^2 > 0.8$). The libraries underwent next-generation sequencing generating 75 bp paired ends. The sequences were analysed using the bioinformatics pipelines HiCUP and CHiCAGO. **Results:** Interaction data was obtained for psoriasis risk loci across the genome represented by 107 lead variants (59 in Europeans, 42 in Chinese and 6 in both populations). CHiCAGO identified 13393, 6737 and 6956 significant interactions in MyLa, unstimulated HaCaT and stimulated HaCaT cells, respectively (CHiCAGO score > 5). In approximately 25% of cases, interactions overlapped with gene promoter regions defined by fragments within 500bp of a transcription start site, corresponding to 1075 genes and non-coding RNAs in total. The data suggested gene targets in several loci; for example *KLF4* in 9q31.2 and *NFKBIZ* in 3q12.3. The intergenic locus at 1p36.23 harbours several potential gene targets including *TNFRSF9* and *SLC45A1*; however in HaCaT cells the psoriasis-associated variation interacted with the promoter of nearby *ERRFI1*. This gene encodes a protein that inhibits epidermal growth factor receptor signalling and is thought to be required for normal keratinocyte proliferation. **Conclusions**: This is the first use of CHi-C across all psoriasis risk loci and demonstrates how intergenic variation may contribute towards the pathogenicity of psoriasis. Continued analysis, such as integration of gene expression data, will help to further identify gene candidates for downstream functional investigation.

380T

Exome-wide copy number association study for arsenic-induced skin lesions: A prospective study. *M.G. Kibriya[1], F. Jasmine[1], F. Parzez[2], T. Islam[3], A. Ahmed[3], M. Rakibuz-Zaman[3], R. Rahaman[1], J.H. Graziano[2], H. Ahsan[1].* 1) Public Health Sciences, University of Chicago Medicine, Chicago, IL. USA; 2) Environmental Health Sciences, Mailman School of Public Health, Columbia University, NY, USA; 3) University of Chicago Research Bangladesh, Dhaka, Bangladesh.

Background: Exposure to arsenic in drinking water is a global public health problem and arsenic-induced skin lesions (ASL) are hallmark of chronic arsenic toxicity. Previously we and others have reported some SNPs as risk factors for ASL. The role of copy number variation (CNV) in the germline DNA in this regard is unknown. Conventional GWAS SNP-chips typically covers mainly intronic or inter-genic regions. Recently, using conventional GWAS chip, for the first time, we have identified some CNVs mainly in intronic regions and lincRNA regions to be associated with higher risk of development of ASL. In this study, we used microarray based Exome chip to detect CNVs in the exome regions and explore if any exonic CNVs are associated with higher risk of arsenic toxicity in subjects exposed to different level of arsenic through drinking water. **Material & Methods:** From a large and long-term (96 months) prospectively followed-up cohort in Bangladesh, exposed to arsenic, we randomly selected 1874 subjects (m=821, f=1053) without any ASL at enrollment and genotyped their whole blood DNA samples at baseline on Illumina HumanExome-12v1-1 SNP chips to generate DNA copy number. Participants were followed up every 2 years for a total of 8 years, especially for the development of ASL. In Cox regression models, each CNV segment was used as a predictor, accounting for other potential covariates, for incidence of skin lesions. **Results:** A total of 373 subjects (m=270, f=103) developed ASL. Male gender, older age, and high urinary arsenic creatinine ratio (UACR) were independently associated with higher risk of developing ASL. We identified multiple CNV regions with amplification/gain to have significant associations (FDR 5%) with the development of ASL, independent of gender, age and arsenic exposure level. The significant top most high-risk CNVs include amplification of oncogene *MAFK* [adjusted HR 1.72 (95% CI 1.27 – 2.33), p=4.1×10^{-4}] with a frequency of 7.4% and among the top most protective CNVs is amplification of tumor suppressor gene *PRDM2* [adjusted HR 0.26 (95% CI 0.14 – 0.49), p=3.1×10^{-8}] with a frequency of 8.9%. **Conclusion:** This first Exome-wide Copy Number Variant (CNV) analysis in arsenic exposed cohort, prospectively followed up for incidence of ASL, reveals that amplification of an oncogene *MAFK* in germline DNA may predispose an individual to a higher risk of development of ASL; and amplification of a tumor suppressor gene *PRDM2* provides protection against ASL.

381F

Gene correction of epidermolysis bullosa simplex mutations using CRISPR/Cas9 technology. *M. Bchetnia[1], L. Girard[1], J. Powell[2], C. McCuaig[2], A. Dupérée[3], C. Morin[3], L. Germain[4], J.P. Tremblay[5], C. Laprise[1].* 1) Université du Québec à Chicoutimi, Saguenay, Canada; 2) Hôpital Sainte-Justine, Montréal, Canada; 3) Centre de santé et de services sociaux de Chicoutimi, Saguenay, Canada; 4) Centre LOEX de l'Université Laval, Génie Tissulaire et Régénération, Québec, Québec, Canada; 5) Centre de Recherche du CHU de Québec, Université Laval, Québec, Canada.

Epidermolysis Bullosa Simplex (EBS) is a rare autosomal dominant skin disease characterized by non-scarring blisters and erosions caused by minor mechanical trauma. EBS is caused by different mutations within the *KRT5* or the *KRT14* gene and resulted in mis-folded proteins thus impaired normal assembly of the keratin intermediate filaments. *In vitro*, this is reflected by the formation of keratin aggregates in the mutant patient cells. To date, there is no efficient therapy for this hereditary disease and current therapies are palliative, aimed at treating infections and maintaining an acceptable quality of life. Recently, the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 technology emerged as a potent tool to edit DNA in a fast, cost-effective and precise method. As our goal is to develop an effective and permanent therapy for EBS, we propose to use the CRISPR/Cas9 technology to correct the genomic mutations that cause the EBS phenotype and engineer autologous gene corrected skin tissue grafts. This would be a viable option to alleviate blistering apparent and improve survival of EBS patients. Here, we assess the potential of CRISPR-Cas9 genome editing tool to repair two single point mutations in *KRT5* and *KRT14* genes respectively and responsible of the severe EBS form (EBS-gen sev) by enhancing the levels of homologous recombination in the presence of a wild type DNA donor template. We selected three single guide RNAs sequences (sgRNA) specific for each mutation and we show the high efficiencies of these sgRNAs to guide the Cas9 nuclease to create DNA double-strand breaks (DSBs) at the precise location close to the site of the mutation both in HEK293T cells and in primary fibroblasts derived from the two EBS patients carrying the two studied mutations. Successful induced homologous recombination was observed in HEK293T cells either when using single-stranded oligonucleotides (ssDNA) as donor template or a plasmid donor template. Experiments are ongoing in EBS fibroblasts in order to validate phenotypic correction. Gene repaired cells will serve to engineer skin tissue that will be transplanted to patients as autografts, which will potentially avoid the patient's immune system graft rejection. Our work supports our hypothesis that the CRISPR/Cas9 system may be used to treat EBS and other genetic dermatological pathologies.

382W

***ATP7B* gene variants for Alzheimer's disease affects ATP7B structure and markers of copper status.** *M. Rongioletti[1], R. Squitti[1,2], C. McCan[3], M. Siotto[4], S. Lutsenko[3].* 1) Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy; 2) IRCCS Istituto Centro San Giovanni di Dio- Fatebenefratelli, Italy Molecular Markers Laboratory; 3) Department of Physiology, Johns Hopkins University, Baltimore, MD, USA; 4) Don Carlo Gnocchi ONLUS Foundation, Milan, Italy.

Background: ATP7B is a copper chaperone that loads the metal into the serum copper-protein ceruloplasmin during its synthesis, escorts superfluous copper into bile, and represents a key protein in copper failure. Impaired function of this ATPase is associated with a well-known inborn error of copper metabolism, Wilson's disease (WD). Four specific gene variants of the ATP7B gene and a WD rare mutation have a statistical association with AD. These gene variants are informative on the ATP7B gene structure. **Methods:** We have examined the effect of these AD-related gene variants (allele rs1061472 and 732774) on the expression and localization of ATP7B. Four ATP7B protein variants R9R8 for R952/R832; R9K8 for R952/K832; K9K8 for K952/K832; and K9R8 for K952/R832 were generated and expressed in HEK293 cells. Effects of the corresponding gene variants in 81 healthy controls were analyzed also in blood stream markers of copper status, namely, serum copper, ceruloplasmin, copper not bound to ceruloplasmin (non-Cp Cu) and the Cu to Cp ratio (Cu:Cp), which is a surrogate of the stoichiometry between copper and ceruloplasmin. In particular we stratified the sample in carries of at least one ATP7B R832 variant (corresponding to GG+AG rs1061472 genotype) and carriers of at least at one ATP7B K952 (corresponding to TT +CT rs732774). **Results:** Following transient expression by Western blot analysis we found that protein abundance in cells differed despite similar mRNA levels. The ATP7B variant with both risk gene variants, K9R8, had the lowest protein level and the variant with neither risk allele, R9K8, has the highest protein level (p-value=0.0075). We have also examined the localization and trafficking of the variants and observed differences in their response to Cu-elevation. Furthermore, healthy subjects carriers of R832, carriers of K952 and carriers of both R832/K952 ATP7B variants had values of non-Cp Cu and Cu:Cp higher than non-carriers subjects. Moreover carriers of K952 and carriers of both R832/K952 ATP7B variants showed also ceruloplasmin values lower than non-carriers group. **Conclusions:** The two alleles investigated affect ATP7B stability, with a consequent decrease of ceruloplasmin concentrations, and increase of non-Cp Cu concentrations and of Cu:Cp ratio in blood stream. These results demonstrate a causative link of ATP7B variants and copper imbalance in AD etiology.

383T

Age-related structural changes of the olfactory receptor subgenome in human blood cells and autologous brain regions. *K. Szigeti, B. Birkaya, A. Ouf, L. Yan.* Dept Neurology, University at Buffalo, Bufalo, NY.

Olfactory identification deficit is associated with aging and neurodegenerative diseases including Alzheimer's disease (AD). To study whether *de novo* somatic structural variants in the olfactory receptor (OR) subgenome contribute to the mechanism of OID we developed an OR specific microarray (Agilent, coverage 90%). Using a cohort of 841 subjects (234 normal controls (NC), 192 amnestic mild cognitive impairment (aMCI) and 415 AD), age range 55-94, we found that structural variants in the OR subgenome are distributed differentially in younger (age 1st quartile) and older subjects (age 4th quartile), older subjects harboring more losses than expected. There was no difference between control and disease groups (p=0.8). Genome wide age correlation identified several regions contributing to the signal, chr 15 q11.2 and chr 11 q11.1 with the largest effect. Megabase - and Kilobase – range variants were identified, ranging from 3-187 CNVs per sample, none of the samples were outliers. The frequency of events at CNV loci ranged 10.6%-59% (mean 17.4%). The majority of high frequency associated regions were within 2Mb of the telomeres or centromeres yielding 77% of CNV for centromeric region and 13% and 10 % for p and q telomeric regions respectively. The associated CNVs were validated with TaqMan assay. aCGH of autologous olfactory bulb and temporal lobe DNA from eight autopsy specimens detected structural variants with 90% Variant correlation detected in the blood cohort suggesting existence of recombination hotspots. The same individual's brain region specific CNV is a result of postmitotic tissue specific *de novo* events with various levels of mosaicism. CNV content gene annotation revealed an enrichment for genes involved in innate and acquired immunity. Age related structural changes, specifically haploinsufficiency, may contribute to the immunosenescence observed during the aging process. Replication studies in the ADNI, NIA-LOAD, TGEN, TARCC, MIRAGE and ADC GWAS datasets is ongoing. .

384F

Molecular mechanisms in C9orf72 ALS/FTD using the BAC transgenic mouse model with behavioral, neuropathological and molecular features of disease. *A. Pattamatta, H.R Olafson, Y. Liu, M.M Scotti, L. Nguyen, T. Reid, T. Zu, J. Richardson, A. Berglund, E.T. Wang, L.P.W. Ranum.* University of Florida, Gainesville, FL.

Hexanucleotide repeat expansions in the first intron of *C9orf72* gene are the most common genetic cause of familial ALS and FTD. To understand the disease mechanisms, we generated a BAC transgenic mouse model of the disease (C9-500), which contains the entire *C9orf72* gene, 500 GGGGCC repeats and substantial flanking sequence. These mice show behavioral, molecular and neuropathological features of ALS and FTD including paralysis, anxiety-like behavior, decreased survival and widespread neurodegeneration of the entire motor unit, hippocampus and layers II/III of the cortex. Repeat associated non-ATG (RAN) protein aggregates increase with age and disease severity in these mice. Similar to human patients, some C9-500 mice show rapidly progressing phenotypes, others have a more slowly progressing disease and 17% show no overt phenotypes at 1 year of age. To better understand this phenotypic variability, we performed RNA-Seq on ten pre-symptomatic C9-500 females to identify transcriptome changes that occur early in disease and may explain the variability of disease presentation and penetrance. Interestingly, 30% of the C9-500 mice clustered together on a principal component analysis plot compared to non-transgenic or the other 70% of C9-500 animals. We are currently characterizing the gene expression and splicing differences in these mice to identify triggers of disease. Additionally, due to repeat instability, we have established a transgenic sub-line from the C9-500 line, in which the expansion grew to ~800 repeats. These C9-800 mice show decreased survival and increased disease penetrance at an earlier age of onset compared to the C9-500 mice. These data demonstrate that disease penetrance and severity are increased with increases in the GGGGCC repeat tract. We are currently using this mouse model to understand the molecular underpinnings of disease and to test therapeutic strategies that can increase survival and mitigate disease.

385W

Genomic architecture predisposes the *PLP1* region to the formation of complex genomic rearrangements. *H. Hijazi[1], V. Bahrambeigi[1,2], F.S. Coelho[3,4], J.A. Lee[5], C. Gonzaga-Jauregui[6], F. Zhang[7,8], L. Bernardini[9], C.R. Beck[1], K. Inoue[10], G.M. Hobson[11], C.M.B. Carvalho[1], J.R. Lupski[1,12,13,14].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Diagnostic Genetics, School of Health Professions, University of Texas MD Anderson Cancer Center, Houston, TX; 3) Programa de Pós-Graduação em Genética Departmento de Biologia Geral – UFMG, Belo Horizonte, MinasGerais, Brazil; 4) Centro de Pesquisas René Rachou – FIOCRUZ, Belo Horizonte, MG, Brazil; 5) Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, SC; 6) Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc., Tarrytown, NY; 7) State Key Laboratory of Genetic Engineering, Institute of Reproduction and Development, Obstetrics and Gynecology Hospital of Fudan University, Shanghai 200011, China; 8) Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China; 9) Cytogenetics Unit; "Casa Sollievo della Sofferenza" Foundation, San Giovanni Rotondo (FG), Italy; 10) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan. (Dept. MR & BD Res, Nat'l Inst of Neurosci, NCNP); 11) Nemours Biomedical Research, Alfred I. duPont Hospital for Children, Wilmington, DE 19803, USA; 12) Human Genome Sequencing Center, BCM, Houston, TX; 13) Department of Pediatrics, BCM, Houston, TX; 14) Texas Children's Hospital, Houston, TX.

Pelizaeus–Merzbacher disease (PMD) is a rare X-linked dysmyelinating leukodystrophy of the central nervous system. PMD manifests with a wide spectrum of clinical severity caused by different classes of mutations affecting the dosage-sensitive gene, *PLP1*. *PLP1* copy-number gain is the predominant cause of PMD; however, cataloging each of the observed patterns of complex structural variations (SVs) at this locus remains to be performed. We have molecularly studied 148 PMD samples with non-recurrent SVs involving *PLP1*; 138 were of copy-number gains and ten were of copy-number losses. Herein we characterize the novel *PLP1* SVs (61/138) and compare these data to our previous findings. Consistently, many of the novel copy-number gains have one or more breakpoints overlapping critical low-copy repeats (LCRs) in the region, termed PMD-LCRs. Additionally, we find that 56% of all described copy-number gain events appear as simple duplications, whereas 44% present as complex genomic rearrangements (CGRs). Duplication followed by triplication and subsequent duplication, DUP-TRP-DUP, is the predominantly observed pattern (43%) of all examined copy-number gain CGR cases. Another commonly observed CGR pattern is interspersed duplication separated by a copy-neutral segment, DUP-NML-DUP (33%). Importantly, we report additional samples of known but less frequently observed patterns. Our analyses further support the contention that PMD patients predominantly harbor copy number gains, many with complexities and LCR involvement. These data indicate that replicative DNA repair processes are often involved in PMD rearrangements. Furthermore, our study has shown that deletions of *PLP1* range from 6.7 kb to 5.6 Mb in size and in contrast to gain events, deletions of *PLP1* were rarely associated with complexities; wherein 90% appear as simple events. Sequence-mapped breakpoints of deletions overlap with distinctive genomic architecture in each of the studied cases including *Alu* elements, LINEs, as well as LCRs. Finally, we investigated potential molecular mechanisms for the penetrance and variable phenotypic expressivity in deletion carrier females by exploring X-inactivation patterns as well as possible bi-allelic variation in the *PLP1* region. This study expands our knowledge of CGR patterns at a discrete locus, their mechanisms of formation, and the molecular pathogenicity of PMD.

386T

Immunotherapy development for *C9ORF72* ALS/FTD using a BAC transgenic mouse model and human antibodies targeting RAN proteins. *L. Nguyen[1], F. Montrasio[2], A. Pattamatta[1], N. Cavegn[2], K. Nakamura[1], O. Bhardi[1], M. Laughrey[1], T. Zu[1], J. Grimm[2], L.P.W Ranum[1].* 1) Center for NeuroGenetics, Molecular Genetics and Microbiology, University of Florida, College of Medicine, Gainesville, FL; 2) Neurimmune, Zurich, Switzerland.

There is considerable interest in the development of immunotherapy for neurodegenerative diseases including Alzheimer's disease, Parkinson's disease and SOD1 ALS. The overall approach is to use human or humanized antibodies to target disease-specific misfolded proteins for neutralization and clearance by the immune system. While antibody therapies for neurodegenerative diseases to date have had mixed results in clinical studies, striking interim results have been reported for Aducanumab, a recombinant human antibody drug that targets Aβ aggregates and is currently in phase III trials. We recently developed a BAC transgenic mouse model of *C9ORF72* ALS/FTD that recapitulates key phenotypic and molecular features of the disease including motor neuron loss and the accumulation of dipeptide RAN proteins. Due to mounting evidence that RAN proteins contribute to *C9orf72* ALS/FTD and other expansion disorders, tools that target these proteins are increasingly important for the development of therapeutic strategies. We have generated a panel of human-derived recombinant monoclonal antibodies and selected lead candidates based on their binding profiles to RAN proteins expressed in vitro and in vivo. To explore blood-brain-barrier penetration and target engagement of these antibodies we performed peripheral administration of antibodies by i.p. dosing. Interestingly, IHC and IF analyses using a human secondary antibody showed that injected antibodies crossed the blood brain barrier and localized within the brain as punctate intracellular aggregates in a pattern similar to the RAN protein aggregates previously detected by IHC. These aggregates were found using both frozen and fixed paraffin embedded tissues. To determine if the peripherally injected human antibodies find the C9 RAN protein targets in vivo, we performed double labeling IF studies and show that the human antibodies co-localize with the intracellular RAN protein signal from a mouse monoclonal GA antibody. No similar signal was present in isotype control or uninjected animals or when the secondary human antibody was omitted. Taken together, these data show that peripherally injected human antibodies that recognize C9 RAN proteins enter the brain and colocalize with intracellular RAN protein aggregates. Murinized versions of these human antibodies are currently being tested in these C9 BAC mice for therapeutic efficacy.

387F

Prolonged pentylenetetrazole exposure modulates the Rho Family GT-Pases pathway in the zebrafish brain. *M.C.S. Nunes[1], J.F.de Vasconcellos[2], V.S. Zago[1], A.S. Vieira[3], C.S. Rocha[3], B.S. Carvalho[4], I. Lopes-Cendes[3], C.V. Maurer-Morelli[1].* 1) Zebrafish Laboratory, Department of Medical Genetics, School of Medical Sciences – UNICAMP, Campinas, SP- Brazil; 2) Genetics of Development and Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health; 3) Molecular Genetics Laboratory, Department of Medical Genetics, School of Medical Sciences – UNICAMP, Campinas, SP- Brazil; 4) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, UNICAMP, Campinas, SP-Brazil.

   Purpose: To investigate the molecular mechanisms underlying seizures, we analyzed the response of different seizure-induced protocols in the zebrafish brain by transcriptome analysis. For a comprehensive understanding of the biological mechanisms affected by the transcripts that were counter-regulated by the treatments, interactive networks and pathways analyses were performed using a systems biology approach. **Methods:** Wild-type zebrafish larvae at 7dpf were separated into three groups: control (CG, n=3), acute seizure (AS, n=2) and *status epilepticus*-like (SE, n=3). Larvae from groups AS and SE were exposed to PTZ 15mM for 20 minutes and 3 hours, respectively. Messenger RNA libraries were achieved after using Illumina's Sample Prep Kit and validated libraries were sequenced in the Illumina HiSeq 2500 System. Gene interactions and correlation networks were identified with the Ingenuity Pathway Analysis (IPA) software content version March/2017 (Ingenuity Systems, Mountain View, CA). The fold change threshold was >3 or ≤-3 and the signaling pathways were considered statistically significant with a $p<0.05$. All IPA available databases at the time of the analysis were used in this study. **Results**: The most activated signaling pathway for AS *vs.* CG and SE *vs.* CG was the *AMPK Signaling*, but in opposite directions, since in the SE it was activated (positive z-score), and in the AS inhibited (negative z-score). However, comparisons between AS *vs.* SE showed that *RhoGDI Signaling* (p=1.06E-07) and *Signaling by Rho Family of GTPases* (p=1.20E-07) are the top-pathways inhibited (negative z-score) and activated (positive z-score), respectively, in the SE group. **Discussion**: Transcriptome studies followed by signaling pathway analysis showed that AS and SE have distinct molecular responses. We also found that RhoGDI pathway is down-regulated and the Rho Family of GTPases is up-regulated in the SE compared to AS group. The RhoGDI family of genes regulate signaling through RhoGTPases by inhibiting the disassociation of Rho members from GDP, thus keeping these factors in an inactivated state. Also, it is known that the Rho family of GTPases have an important role in the morphogenesis of neurons' dendritic spines and contribute to the synaptic plasticity that is observed in epilepsy. This study supports the relevance of the zebrafish as a model for investigating seizures. **Support:** FAPESP 14/15640-8, CEPID-BRAINN FAPESP 13/07559-3 and Intramural Research Program of NIDDK/NIH. .

388W

Chromosome 16q22-q24 uniparental disomy unmasks a rare recessive cause of early infantile onset epileptic encephalopathy 28. *M. Davids[1,2], S. Thomas[1,2], C. Adams[1,2], L.A. Wolfe[1,2], T. Markello[1,2], C.J. Tifft[1,2], W.A. Gahl[1,2,3], M.C.V. Malicdan[1,2,3].* 1) Undiagnosed Diseases Program, Common Fund, NHGRI, National Institutes of Health, Bethesda, MD; 2) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Section of Human Biochemical Genetics, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

   Introduction. The genetic etiologies of many rare disorders, including infantile onset epileptic encephalopathies, are largely undiagnosed. The recent decade has seen a surge in the availability of next generation sequencing; several cases remain unsolved, likely due to unusual inheritance patterns that are not routinely evaluated. We report the discovery of the gene causing early onset epileptic encephalopathy by careful analysis of next generation sequencing data. **Methods.** The patient was evaluated at the NIH Undiagnosed Diseases Program (UDP) under the protocol 76-HG-0238. Whole exome sequencing (WES) and single nucleotide polymorphism (SNP) genotyping arrays were performed on DNA from all members of the immediate family. **Results.** A 5-year-old girl presented to the UDP with a history of neurodegenerative disorder, profound intellectual disability, infantile onset seizures, chronic respiratory failure, dysmorphisms, and an atrial septum defect. She also had neutropenia, thrombocytopenia and hypoglycemia. WES did not reveal gene candidates that could explain the phenotype in full, but on SNP-array analysis a large region of homozygosity was discovered on chromosome 16, spanning q22-q24. Upon closer inspection, it was discovered that the homozygosity was caused by maternal uniparental disomy (UPD) that included an exon 6 deletion in *WWOX* (NM_016373). *WWOX* is thought to play a role in apoptosis as a mediator of p53 and is potentially involved in steroid metabolism, skeletal morphogenesis and central nervous system development. Digital droplet PCR confirmed that the homozygous region was caused by uniparental disomy and was not a result of mosaicism. mRNA expression analysis revealed nonsense mediated decay of this transcript. Exon 6 of an alternative transcript (NM_130791) is utilized; that transcript lacks the short chain dehydrogenase domain. **Conclusion.** The deletion of exon 6 in *WWOX*, which is associated with early infantile onset epileptic encephalopathy 28, explains most of the phenotype in our patient. Whether the atrial septum defect, skeletal and hematological abnormalities represent an expansion of the *WWOX* phenotype or are attributable to other variants in the disomic region, or 3 identified variants elsewhere in the genome that may contribute to the phenotype, is under current investigation. Our report underscores the importance of UPD as an underlying mechanism of rare autosomal recessive disorders.

389T

Gene discoveries in autism are biased towards intellectual disability.
M. Jensen[1], L. Pizzo[1], V. Kumar[1], M.D. Singh[1], F. Hormozdiari[2], S. Girirajan[1]. 1) Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 2) Department of Biochemistry and Molecular Medicine, University of California, Davis, Davis, CA.

While dozens of genes have been linked to autism spectrum disorders (ASD), definitively connecting these genes to autism phenotypes has been difficult due to extensive comorbidities between ASD and other psychiatric disorders. Here, we show that the ASD gene discoveries to date have been biased towards genes that contribute to ASD comorbid with intellectual disability (ID). We analyzed 2,297 probands with autism from the Simons Simplex Collection (SSC), and found that probands with rare pathogenic copy-number variants (p=0.0310) and *de novo* loss-of-function (LOF) variants (p=0.00197) have significantly lower IQ than SSC probands without pathogenic variants. Further, we found no change in autism severity, as measured by SRS T-scores, between the two groups (p>0.05), indicating that pathogenic mutations used to identify autism genes in fact capture genes that contribute mainly to ID phenotypes. In addition, SSC probands with *de novo* mutations in 76 previously well-described ASD genes not only manifested more severe ASD (p=0.0232), but also a significantly lower IQ (p=2.65×10[-6]), showing that the current set of autism genes are not specific to autism only. We next isolated 397 probands with high-functioning autism (SRS>75 and IQ>100), and found they were less likely to carry rare pathogenic CNVs (p=0.0497) or *de novo* LOF mutations (p=0.0437) than the rest of the SSC cohort. These probands were also less likely to carry *de novo* LOF mutations in the 76 known ASD genes (p=0.0068). Additionally, we found that 42 genes with *de novo* LOF variants in individuals with high-functioning autism are less resistant to variation, with higher RVIS scores than genes with *de novo* LOF variants in probands with IQ<70 (p=0.0121). Mouse knockout models of these genes also have less severe developmental phenotypes, including neurological, growth and behavioral features, than those of known ASD genes (p<0.05). We also used computational methods to cluster variants into modules based on proband phenotypes, and found clustering of modules within neurodevelopmental pathways for *de novo* LOF mutations in probands with ASD and ID. However, significant clustering within these pathways were only observed for missense mutations in probands with ASD-only phenotypes. Overall, our results suggest that pathogenic rare CNVs and *de novo* LOF mutations associated with autism actually contribute to both ASD and ID, and subtler genetic changes contribute to pathogenicity in high-functioning autism.

390F

Marker chromosome architecture and temporal origin revealed in a family with pleiotropic psychiatric phenotypes. *C.M. Grochowski[1], S. Gu[1], K.J. Brennand[2], B. Yuan[1], J. Sebat[3], D. Malhotra[4], S. McCarthy[5], U. Rudolph[6,7], A. Lindstrand[8], Z. Chong[9], D.L. Levy[6], J.R. Lupski[1,10,11,12], C.M.B Carvalho[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States; 2) Icahn School of Medicine at Mount Sinai, NY, NY, United States; 3) University of California at San Diego, San Diego, CA, United States; 4) Hoffmann-La Roche Ltd, Basel, Switzerland; 5) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, United States; 6) McLean Hospital, Belmont, MA, United States; 7) Department of Psychiatry, Harvard Medical School, Boston, MA, United States; 8) Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 9) Department of Genetics and the Informatics Institute, the University of Alabama at Birmingham, Birmingham, AL, United States; 10) Human Genome Sequencing Center, BCM, Houston, TX, United States; 11) Department of Pediatrics, BCM, Houston, TX, United States; 12) Texas Children's Hospital, Houston, TX, United States.

Small supernumerary marker chromosomes (sSMC) are chromosomal fragments that are inherently difficult to genetically characterize due to their small size and abnormal genetic structure. Here we detail a mother and proband who present with neuropsychiatric disorders and a marker chromosome that segregates with disease. We explored the genomic architecture of this marker and investigated its temporal origin in the family. A customized array comparative genomic hybridization (aCGH) revealed 4 duplications and 2 triplications that spanned the short arm of chromosome 9, suggesting a chromoanasynthesis event. Subsequent FISH analysis showed the DUP/TRP segments were present on the marker and droplet digital PCR (ddPCR) revealed the presence of mosaicism. The discovery that the mother had two different subsets of cells, one carrying the marker (blood DNA) and one not carrying the marker (lymphoblastoid cell line (LCL)) provided a unique opportunity to distinguish between these two genotypes. A SNP array was performed in the distinct cell types; and using SNPs that were homozygous in the mother's LCLs, but heterozygous in her blood, we were able to delineate the genotype of the marker. Examining those same positions across the family, we were able to infer that the chromosomal fragments of the marker likely were inherited from the grandmother, who does not carry the marker itself, suggesting it was formed during a germline level event. Additional ddPCR assays were used to interrogate the junctions present and showed a duplication of breakpoints. To understand the exact architecture of the fragments on the marker, we performed whole genome sequencing (WGS) of the blood DNA from the mother, proband and grandmother. Using genomic locations derived from the aCGH, we located those regions in the WGS data and found reads with mate-pairs matched to regions on the other side of a breakpoint. This nucleotide resolution allowed for Sanger-sequencing of 5 out of 6 breakpoints and facilitated the creation of a proposed genomic architecture for the marker. The combination of multiple copy number gains as well as the complex rearrangement is indicative of chromoanasynthesis via a DNA replicative repair mechanism such as FoSTeS/MMBIR. While CNVs have been previously associated with neuropsychiatric disorders, here we detail their precise architecture on a marker. A better understanding of how these changes can contribute to disease may provide avenues for targeted therapies.

391W

Local and global chromatin interactions are altered by large genomic deletions associated with human brain development. *A. Urban[1], Y. Zhang[2], X. Zhang[1], X. Zhu[1], C. Purmann[1], M. Haney[1], T. Ward[1], J. Yao[3], S. Weissman[4]*. 1) Psychiatry and Genetics, Stanford University, Palo Alto, CA; 2) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Cell Biology, Yale University, New Haven, CT; 4) Genetics, Yale University, New Haven, CT.

   Large copy number variants (CNVs) in the human genome are strongly associated with common neurodevelopmental, neuropsychiatric disorders such as schizophrenia and autism. Using Hi-C for the genome-wide analysis of long-range chromosome interactions and ChIP-Seq for the analysis of regulatory histone marks we studied the epigenomic effects of the prominent large deletion CNV on chromosome 22q11.2, in a cohort of lymphoblastoid cell lines (LCLs), and also replicated a subset of the findings in LCLs with the common large deletion CNV on chromosome 1q21.1. We found that, in addition to local and global gene expression changes, there are pronounced and multilayered effects on chromatin states, chromosome folding and topological domains of the chromatin that emanate from the large CNV locus. Regulatory histone marks are altered in the deletion proximal regions, and in opposing directions for activating and repressing marks. There are also significant changes of histone marks elsewhere along chromosome 22q and genome wide. Chromosome interaction patterns are weakened within the deletion boundaries and strengthened between the deletion proximal regions. We detected a change in the manner in which chromosome 22q folds onto itself, namely by increasing the long-range contacts between the telomeric end and the deletion proximal region. Further, the large CNV affects the topological domain signal that is spanning its genomic region. Finally, there is a widespread and complex effect on chromosome interactions genome-wide, i.e. involving all other autosomes, with some of the effect directly tied to the deletion region on 22q11.2. These findings indicate novel principles of how such large genomic deletions can alter nuclear organization and affect genomic molecular activity. [manuscript under review] [no conflicts of interest to report].

392T

Clinical features in a pediatric population due to chromosome deletions at a third level pediatric Mexican hospital in 24 years period of time: Five case reports. *M.D. Hurtado-Hernandez[1], J.M. Aparicio-Rodriguez[2,5], R. Zamudio-Meneses[3], S. Chatelain-Mercado[4]*. 1) Cytogenetics,; 2) Genetics; 3) Cardiology, Hospital Para el Nino Poblano, Puebla Pue; 4) Biotechnology, Universidad AutÓnoma Metropolitana,; 5) Estomatology, Benemerita Universidad Autonoma de Puebla, Mexico.

   Chromosome aberrations are considered as alterations in the chromosome number or structure. They are mainly due to gametogenesis inborn error (meiosis) or occur during the zygote first cellular divisions where DNA repair processes are deficient. Two Wolf Syndrome patients and two criduchat patients or patients with deletion of the short arm of chromosome 4 and 5, respectively, and deletion of chromosome 9 in one patient, were observed among 4617 karyotype studies performed from 1992 to 2016, at the Hospital Para El Niño Poblano (Pediatric Hospital) in Mexico. These chromosome structural alterations or deletions at chromosomes 4, 5 and 9, observed among five patients from three different families were chosen to analyze their clinical characteristics, medical or surgical treatments and their medical evolution according to the genetic disease. Key words: Chromosome, deletion, chromosome aberration, structural changes, karyotype.

393F

Loss of Kctd13 in mice causes short-term memory deficiency. *T. Arbogast[1], R. Rodriguiz[2], W. Wetsel[2], P. Razaz[3], T. Aneichyk[3], J. Ellegood[4], M. Henkelman[4], M. Talkowski[3], N. Katsanis[1], C. Golzio[5].* 1) Duke University, Cell Biology, Center for Human Disease Modeling, Durham, NC, USA; 2) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA; 3) Molecular Neurogenetics Unit and Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 4) Mouse Imaging Center, The Hospital for Sick Children Toronto, Ontario, Canada; 5) Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France.

   Molecular cytogenetic approaches have identified a substantial number of copy number variants (CNVs) underlying human diseases, including a CNV at the 16p11.2 locus. Deletion and duplication of this genomic region have been associated with both intellectual disability and autism spectrum disorders. Additionally, a reciprocal effect of 16p11.2 gene dosage on body mass index and head size has been noted. The most common CNV at this locus corresponds to a 600 kb fragment of 31 genes including *KCTD13* (Potassium Channel Tetramerization Domain containing 13). We showed previously that overexpression of *KCTD13* human transcript in zebrafish embryos induces the microcephaly phenotype associated with 16p11.2 duplication, whereas suppression of the *kctd13* zebrafish ortholog yields the macrocephalic phenotype associated with 16p11.2 deletion, capturing the mirror phenotypes observed in humans. Recently, KCTD13 has been shown to regulate RhoA ubiquitination and degradation which may, in turn, lead to disrupted cell migration and influence brain size during prenatal development. To evaluate further the candidacy of *KCTD13* in neurodevelopmental phenotypes, we generated a ubiquitous knockout (KO) of *Kctd13* in the mouse and asked whether loss of *Kctd13* leads to behavioral and anatomical phenotypes. Mutant mice were tested in various behavioral assays to evaluate activity, anxiety, memory, and sociability. By performing novel object recognition, novel location recognition, and social transmission of food preference tests, we observed short-term memory deficits in both heterozygous (Het) and homozygous (Hom) KO mice. While MRI analysis indicated no anomalies in the brain architecture of mutant mice, spine density and morphology analysis on hippocampal pyramidal neurons revealed a reduction in density of mature spines in both Het and Hom KO animals. RNAseq analysis of wild-type, Het and Hom mice show a stepwise decrease in *Kctd13* expression for both cortex and hippocampal tissues, which is consistent with the expression changes of *Kctd13* in the cortex of mice carrying a deletion for the full 16p11.2 segment. Furthermore, a significant number of genes differentially expressed in *Kctd13* KO cortex has been found to be enriched in the 16p11.2 mouse coexpression module. Taken together, these results tentatively implicate *KCTD13* in the intellectual disability phenotype observed in patients carrying the 16p11.2 deletion.

394W

The use of NGS 10X Genomics linked-reads to solve complex breakpoints mapping in individuals with balanced translocation. *F. Mafra[1], M. Moisés[2], M. Gonzalez[1], R. Pellegrino[1], C. Kao[1], G. Monteiro[3], L. Kulikowski[3], J. Garifallou[1], K. Kaminski[1], M. Melaragno[2], H. Hakonarson[1,4].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Disciplina de Genética, Departamento de Morfologia e Genética, UNIFESP, São Paulo, SP, Brazil; 3) Faculdade de Medicina de São Paulo, São Paulo, Brazil; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

   Introduction: Chromosomal rearrangements can cause genetic diseases by disrupting or inactivating specific genes rendering the characterization of breakpoints of critical importance in the molecular elucidation of a wide variety of complex phenotypes. Despite the technological advance, current whole exome and targeted NGS sequencing still represent a challenge due the short reads. To overcome these limitations, we utilized a NGS-based method to resolve breakpoints using linked reads generated by the 10x Genomics Chromium technology in combination with WES using custom Agilent SureSelect baits. Methods: We performed experiments on six well represented samples that were previously investigated by karyotyping and microarray. High molecular weight DNA was extracted and the size verified by Genomic Tape Station (60kb required). We used the Chromium platform developed by 10x Genomics to generate barcoded reads. The Gel Bead-In-Emulsions were submitted to an isothermal reaction, followed by library preparation and target enrichment using a premium bait design (Agilent SureSelect) containing "bridging baits" were spaced within intronic and intergenic regions to preserve long-range information following capture. We subsequently sequenced individual samples using the Illumina Hiseq2500 at ~85X coverage. Results: We were able to confirm five of the six cases translocation events, including 46,X,t(X;22)(p22.31;q11.2), 46,X,t(X;2)(q24;p16.2), 46,XY,t(10;18)(p11.2;q11.32)dn, 46,XX,t(2;7)(p13;q22)dn and 46,XX,t(4;7)(q27;p22). The one sample that we were not able to solve (karyotype 46,XY, t(16;17)(p13.3;p11.2)dn) has both breakpoints in repetitive regions, 16p subtelomeric and 17p pericentromeric respectively, suggesting that this strategy is not capable of calling structural variations in highly repetitive regions since there are no baits in these genomic regions due the high repetitive nature. However, we successfully narrowed down the breakpoints for the other five different chromosomal rearrangements all of which were complex events. Conclusions: Using the Chromium platform to generate barcoded reads we have successfully mapped breakpoints for five complex chromosomal rearrangements, demonstrating that the improvement inbait design along with the linked-reads technology can facilitate the "one-test" approach for unresolved balanced translocation events not detected by microarrays, a notable advancement for the cytogenomics field and clinical diagnostics in general.

395T

Identification of a novel frameshift mutation in the *MCPH1* gene causes primary microcephaly in a Saudi family. *M.I. Naseer[1], O.Y. Muthaffar[2], M. Rasool[1], A.J. Sabbagh[3,4], A.G. Chaudhary[1], M.H. Al-Qahtani[1].* 1) Center of Excellence in Genomic Medicine Research, King Abdulaziz University , Jeddah, Makkah, Saudi Arabia; 2) Consultant of Pediatric Neurology Department of Pediatric, King Abdulaziz University Jeddah, 21589 Saudi Arabia; 3) Research and Development, Clinical Skill Simulation Center, King Abdulaziz University Jeddah, 21589 Saudi Arabia; 4) Pediatric Neurosurgeon, College of Medicine, King Abdulaziz University Jeddah, 21589 Saudi Arabia.

Primary microcephaly (MCPH) is a rare developmental defect characterized by impaired cognitive functions, retarded neurodevelopment and reduced brain size. It is genetically heterogeneous and so far more than seventeen genes associated with this disease have been identified. Primary microcephaly type 1 (*MCPH1*) gene encodes a protein called microcephalin, which is implicated in chromosome condensation and DNA damage induced cellular responses. It is suggested to play a role in neurogenesis and regulation of the size of the cerebral cortex. In our study whole exome sequencing revealed a novel, homozygous frameshift mutation (c. 373_374delAA) in *MCPH1*gene in exon 5 resulting in frameshift change from p.Lys125Glusfs*7). Whole exome sequencing data was further validated by using Sanger sequencing by designing the primers of the identified mutation. Moreover this frameshift mutation was not found in 100 healthy unrelated controls, while parents and one sister were heterozygous and one sister were normal as healthy controls. Our report presents the results of the simultaneous analysis of the trio exome data of both unaffected parents and their affected son with the phenotye of primary microcephaly. A homozygous frameshift variant in the *MCPH1* gene was identified as a plausible candidate causal variant for the clinical phenotype in this Saudi family.

396F

Single cell allele specific expression (ASE) in T21 and common trisomies: A novel approach to understand Down syndrome and gene dosage effects in aneuploidies. *G. Stamoulis[1], F. Santoni[2], P. Makrythanassis[2], A. Letourneau[1], M. Guiponni[2], M. Garier[1], N. Panouis[1], D. Popovic[4], L. Pelkmans[4], E. Falconnet[1], P. Ribaux[1], C. Borel[1], S.E. Antonarakis[1,2,3].* 1) University of Geneva, Geneva, Switzerland; 2) Geneva University Hospitals-HUG; 3) iGE3 Institute of Genetics and Genomics of Geneva; 4) University of Zurich, Zurich, Switzerland.

Trisomy 21(T21), is a model disorder of altered gene expression. We have previously used monozygotic-twins discordant for T21 to study the dysregulation of gene expression, (*Nature:*508;345-350;2014). The majority of previous studies on aneuploidies were conducted on cell-populations or tissues. Studies on gene and allelic-expression of single-cells (SC) may reveal important biological insights regarding the cellular impact of aneuploidy elucidating the fundamental mechanisms of gene dosage. We have studied the allele-specific-expression(ASE) from RNAseq of ~1000 single-cell fibroblasts in different aneuploidies: 352 SC (172 Normal-179 T21 cells) from the pair of monozygotic-twins discordant for T21, 166 from a mosaicT21, 176 mosaicT18, 151 mosaicT8, and 146 from mosaicT13. In the monozygotic-twins, a considerable number of observations for heterozygous sites genome-wide were expressed monoallelically (Normal:73.5%-564,668 observations, and T21:78.7%-549,799 observations). There was considerable monoallelic-expression for chromosome 21 sites in Normal and, surprisingly, in T21 cells as well (Normal:63,3%-5,009 observations, T21:72.8%-6,456 observations). We classified chr21 genes in three classes according to the level of the aggregate monoallelic-expression of their corresponding sites (9-monoallelic, 29-intermediate, 2-biallelic). We hypothesize that each class of genes contributes in a specific way to the phenotypic variability of Down-Syndrome (DS). Similar results, (i.e. extensive monoallelic-expression of genes on the supernumerary chromosomes), were also observed in the other aneuploidies. Our analysis demonstrated that, for genes with monoallelic-expression, the altered gene dosage induced by the aneuploid chromosome is due to a higher fraction of cells expressing the gene on the supernumerary chromosome. This difference in the fraction of expressing cells could contribute to the development and variability of phenotypes in aneuploidies. Indicatively, the fold difference between Trisomic and Normal and cells T/N regarding the fraction of cells which express the monoallelelic genes ranging from 1.2-1.6 fold in trisomies studied. This finding was also supported by sm-RNA-FISH showing higher fraction of expressed cells for specific monoalllelic genes in the trisomic group compared to the normal. This study provides a new fundamental understanding of gene dosage effects in aneuploidies and propose a new gene dosage imbalance mechanism in trisomies.

397W

A sequential screening strategy for efficient rare gene discovery in small families. *X.Z. Liu[1,2], S. Blanton[1,2], M. Tekin[1,2], D. Yan[1].* 1) Otolaryngology, University of Miami Miller School of Medicine, Miami , FL; 2) Dr. John T. Macdonald Department of Human Genetics and John T. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL 33136, USA.

The extreme heterogeneity of nonsyndromic sensorineural hearing loss (NSHL) makes serial sequencing approaches unfavorable in terms of efficiency and cost. However, the discovery that genes at only a few loci account for a major component of human deafness in most populations suggested that sequential screening of probands in multiplex sibships for mutations would be a cost effective strategy. Although next-generation sequencing (NGS) has expedited the discovery of genetic variants, the large number of variants naturally present in each individual makes it challenging to pinpoint the causative mutations. We aim to identify the genetic cause of NSHL through an integrated paradigm combining microarray, copy number variations (CNVs) analysis, whole exome sequencing (WES) and whole genome sequencing (WGS), and a hearing-centric database. We are using a DNA microarray panel (Miami-CapitalBio) as the initial screening to simultaneously detect the most common deafness-causative mutations in four genes in MiamiOtogenetic repository (*GJB2*, *GJB6*, *SLC26A4* and 12S rRNA). Samples are also subjected to a custom capture/ NGS gene panel (MiamiOtoGenes) composed of 180 known deafness genes. Patients who screen negative with both panels undergo WES for detection of single-nucleotide variants (SNVs), insertion/deletions (Indels) and CNVs. Genesis 2.0 (https://www.genesis-app.com/) is used for data analysis. Computational functional prediction algorithms and conservation scores are also applied. Causative mutations are stored in our HL Genotype Database, GeneHeal. In this study we performed WES on five families that screened negative on both panels. Approximately, 92,000 SNVs and 9,206 INDELS per sample were obtained before variant filtering. Coverage of targeted exons for >10 reads ranged from 90.3% to 93.5% and >20 reads from 80% to 83.5%. With our filtering strategy, we have rapidly identified homozygous and heterozygous variants in the five families presenting with autosomal recessive and dominant NSHL, respectively. Sanger sequencing confirmed co-segregation of the variants with the disease phenotype in each family. We have identified several novel NSHL genes. Our study shows that the integrated screening strategy is an efficient approach to rare HL gene discovery in small families not suitable for linkage analysis. If variants are not found in the genes included on Miami-CapitalBio and MiamiOtoGenes, WES should be considered in small multiplex families.

398T

Using a combined approach of comparative genomic analysis, luciferase assay and CRISPR/Cas9 to identify and characterize cis-regulatory elements in the critical region of the 9p deletion syndrome. *X. Hauge[1], D. Fortech-ngochi[1], J. Sloan[2], K. Traver[3].* 1) Dept. of Molecular and Cellular Biology, Kennesaw State University, Kennesaw, GA; 2) EGL Genetics Eurofins Clinical Diagnostics, Tucker, GA; 3) Oncology Cytogenetics, Emory University School of Medicine, Atlanta, GA.

Deletions of the terminal region of the short arm of chromosome 9 are associated with trigonocephaly, dysmorphic facial features and mental retardation. The deletion of Cerberus 1 (*CER1*) gene located at 14.7 Mb from the 9p terminus has been postulated as the cause of trigonocephaly. However, several independent studies showed that patients with typical 9p deletion phenotype possess smaller than 12.4 Mb deletions. Therefore, these patients have two, intact copies of the *CER1* gene. It raises the possibility that cis-regulatory elements for the *CER1* gene could be located in the first 12.4 Mb region of 9p. Our group has previously identified 8 conserved non coding sequences in the deletion breakpoint region between 10 Mb and 12.4 Mb of 9p. We tested the regulatory activity of these sequences in dual luciferase assays. Of these, we found a strong enhancer element 5008 and a strong repressor 3362. Enhancer element 5008 showed a cell-type specific activity. It increased the luciferase activity 14 fold in a human mesenchymal progenitor cell line derived from H9 hESCs as compared to a 3.5-fold increase of the luciferase activity in the HEK293T cells. We also detected an increase in the *CER1* gene expression in the mesenchymal progenitor cell line after the treatment with osteogenic differentiating medium. We utilized S. aureus Cas9 and 4 guide RNAs to excise the enhancer element 5008. We obtained 4 different constructs with deletions ranging from 500 bp to 2000 bp, which correspond to the core DNA sequence of the enhancer element 5008 and the surrounding genomic region, respectively. Screening of the transfected human cell lines is in progress, including the human mesenchymal progenitor cells. In addition, we are planning to investigate the effect of allelic deletion of the enhancer element 5008 on the *CER1* gene expression. In summary, we integrated comparative genomic analysis, the dual luciferase assay, and CRISPR/Cas 9 technology to investigate the genetic mechanism of the key phenotype of the 9p deletion syndrome.

399F

Trisomy 21 and chromatin. *S.E. Antonarakis[1], F. Bezrukov[1], C. Borel[1], O. Delaneau[1], B. Ren[2], SysGenetiX project consortium.* 1) Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Ludwig Institute for Cancer Research, UCSD School of Medicine, CA, USA.

We hypothesize that the presence of a supernumerary ~45 Mb chromosome 21 may substantially disturb the chromatin structure, so that some of the phenotypes of Down Syndrome may result from this disturbance. Our previous study using monozygotic twins discordant for Trisomy 21 (T21) has shown genome-wide dysregulation of gene expression that is organized in large genomic domains (GEDDs Gene Expression Dysregulation Domains, Letourneau et al Nature 2014). Here we use *insitu* Hi-C experiments to study the chromatin interactions in T21 versus normal, using fibroblasts from the pair of discordant monozygotic twins and 4 additional samples (3 T21 and 1 N). These experiments indicated: i/ there is no difference in the TAD structure in T21 suggesting that the short range regulation is unchanged; ii/ there is considerable quantitative difference of long range chromatin interactions within a chromosome (interchromosomal interactions) seen in the samples from the twins; iii/ There are more transchromosomal interactions with chromosome 21 sequences as expected; moreover the intensity of interactions of chromosome 21 (particularly the distal half of the long arm) with the other chromosomes follows a pattern that is compatible with the GEDDs. Remarkably there is an inverse correlation between the strength of transchromosomal contacts of chromosome 21 with the fold difference of gene expression T21/N (rho 0.4). We conclude that the chromatin interaction differences due to the extra chromosome 21 may mediate the genomewide expression differences in T21. Using data from 320 LCL and 80 fibroblasts from normal individuals (genome variants, gene expression, ChIP-seq for H3K27ac, H3K4me1 & H3K4me3) as part of the SysGenetiX project we were able to identify Chromatin Regulatory Domains (CRD) on chromosome 21. These CRDs largely correspond to TADs and could be interpreted as the functional chromatin domains as opposed to the structural domains represented by TADs. The transchromomal interactions of CRDs are currently under investigation in order to further understand the functional disturbances of the chromatin due to the presence of an additional unbalanced genomic material as the whole chromosome 21.

400W

A comprehensive portrait of human somatic mosaicism. *S. Vattathil[1,2], J.M. Akey[1,2].* 1) Dept. of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ; 2) 2Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ.

An adult human body contains on the order of 10^{15} cells, all of which can be traced back to a single zygote. The large number of cell divisions that occur during an individual's lifetime provides many opportunities for somatic mutations to arise and accumulate. Somatic mutations have been found to drive many forms of cancer, play an important role in ageing phenotypes, and contribute to an increasingly large catalog of human diseases. Theoretical and empirical studies suggest that the burden of somatic mutation in humans is high, but few systematic and comprehensive studies have been conducted to date to characterize human somatic variability in apparently healthy individuals. To address this question, as part of the enhanced Genes, Tissues, and Expression (eGTEx) Project, we generated high-coverage (>150x) exome sequence data from 14 individuals sampled for a median of 18 tissues (ranging from 11 to 32 tissues per subject). We developed a novel mixed model framework to take advantage of the unique data structure to jointly call variants across tissues from the same subject. The multi-tissue sampling design also allows us to infer the developmental timing of mutations and compare somatic mutation profiles across tissue types. We also leveraged our data to characterize large somatic chromosomal imbalances (deletions, duplications, and copy-neutral loss of heterozygosity), and assess the relationship between the burden of somatic mutations and covariates such as age and sex. Collectively, our data provides substantial new insights into the landscape and determinants of human somatic variability, and we anticipate that our comprehensive catalog of somatic mutations will enable new insights into development, mutational processes and mechanisms, and provide guidance for interpreting mutations observed in cancer samples.

401T

The commitment complex *in vivo* over long intron genes. *Y. Leader, G. Lev Maor, G. Ast.* Department of Human Molecular Genetics & Biochemis, Tel Aviv University, Tel Aviv, Israel.

The first and critical step in mRNA splicing is identification of the two splice sites delimiting an intron and bring them into proximity – termed the commitment complex. Within the commitment complex U1 small nuclear ribonucleo-protein particle (snRNP) binds the 5' splice site and form a protein bridge with U2AF65 (also called U2AF2) that binds the 3' end of that intron. Although the commitment complex was study extensively *in vitro*, how the same pairing occurs *in vivo* for splice sites separated by long introns (>400 bases) is still obscure. Our laboratory proposed a new model, suggesting a dynamic pairing between the two splice sites involving interplay of RNA polymerase II (RNAPII), splicing factors and chromatin organization. We propose that U1 snRNP and U2AF65, associate with the CTD of RNAP II when that travels from one end of an intron downstream. We propose that U1 binds to 5' splice site while being associates to the CTD of RNAPII. When the 3' end of the intron is synthesize, U2AF65 binds the 3' splice site and the commitment complex is formed. In my research I will examine the validity of this model. I already demonstrated that *in vivo* U1 and U2AF65 are associated with RNAPII when the polymerase is in the elongating and pausing phases. I will further explore how the interaction of U1 and U2AF65 with RNAPII facilitate the formation of the commitment complex using CRISPR method.

402F

Exon-intron architecture in high and low GC-content genes affects alternative splicing. *L. Tamer, A. Savchinko, D. Dollander, G. Ast.* Human Molecular Genetics & Biochemistry, Tel Aviv University, Haim Lebanon 30, Tel Aviv 6997801, Israel.

The splicing machinery recognizes exons and introns by using multiple signals. Two splicing signals that delineate proper exon-intron boundaries are the 5' and 3' splice sites. Alternative splicing is a major mechanism that enhances transcriptomic diversity. Our previous studies revealed that gene architecture had evolved from an ancestral state of low GC-content, in exons flanked by short introns of lower GC-content. One group of genes maintained the low GC-content and their introns became longer. In another group mutations accumulated, leading to elevation of the GC-content, which revoked the differential GC-content between exons and introns. In this group introns are under selection to remain short. Our research hypothesis is that the splicing machinery selects exons and introns differently between the two groups of exon-intron architecture. This hypothesis is based on my lab's previous findings that exon skipping predominantly associates with the low GC-content genes and intron retention with the higher GC-content genes. To examine this hypothesis we reduced the affinity of the splicing machinery to the 5' splice-site using site-specific mutations. According to the results we concluded that the splicing machinery recognizes the exons and introns of the two GC-content gene groups differently. Also, our bioinformatics analysis revealed that the high GC-content genes harbor an even higher GC content region downstream of the 5' splice-site. we examined the functionality of that sequence by replacing it with even higher or lower GC-content segments. Our results show that this region is important for the inclusion of the upstream exon. Furthermore, we checked how the splicing machinery can switch between intron definition and exon definition through the formation of exon-intron differential GC-content. And how intron lengthening impacts intron definition in high GC-content genes. Overall, our research shed light on genomic elements that contribute to the diversity in splicing recognition units in the high and low GC-content families.

403W

Chitinases mRNA levels quantified by qPCR in crab-eating monkey tissues. *M. Uehara, E. Tabata, M. Ohno, M. Sakaguchi, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hchioji, Japan.

Crab-eating monkey (*Macaca fascicularis*) is one of the most important nonhuman primate animal models in basic and applied biomedical researches. Chitin is a polymer of *N*-acetyl-D-glucosamine. Although mammals do not contain endogenous chitin, mammals express two active chitinases, acidic mammalian chitinase (AMCase) and chitotriosidase (CHIT1). Because the level of expression of these chitinases is increased in many inflammatory conditions including Gaucher disease and mouse models of asthma, both chitinases may play important roles in the pathophysiological conditions. Here, we performed gene expression analysis of crab-eating monkey chitinases by quantitative real-time PCR (qPCR) using a single standard DNA molecule to quantify the expression levels of chitinases and reference genes. Our analysis showed that AMCase mRNA was expressed at exceptionally high level in the stomach than other tissues, whereas CHIT1 mRNA was predominantly expressed in the lung.

404T

Comparison of chitinolytic activities among mammalian chitinases and bacterial chitinase. *M. Kimura, T. Umeyama, S. Wakita, K. Okawa, M. Sakaguchi, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo 192-0015, Japan.

Humans and mice express two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase). Chit1 and AMCase have attracted considerable attentions due to their increased expression in individuals with different pathological conditions. Chit1 is increased in individuals with Gaucher disease, chronic obstructive pulmonary disease (COPD) and Alzheimer's disease. AMCase expression and activity are also upregulated during allergic airway responses in mouse models of asthma. In addition, polymeric chitin induces the AMCase expression and the recruitment of immune cells associated with allergy and asthma. These results strongly suggest that chitinolytic enzymes play important roles in many pathophysiological conditions. In this study, we clarify the functional differences between the mammalian chitinase. Furthermore, their functions were compared with *Serratia marcescens* chitinase B (Sm_ChiB) which is a well-studied bacterial chitinase and structurally similar to the mammalian chitinases. AMCase possesses maximal activity at pH 2.0, whereas Chit1 and Sm_ChiB exhibit at around pH 5.0. AMCase had the highest activity at around pH 2.0 among tested enzymes. At pH 5.0-7.0 Sm_ChiB showed higher activity, followed by Chit1 and AMCase. Our results indicate that mammalian chitinase can function under various pH conditions and possess similar chitinolytic activity comparable to bacterial chitinase.

405F

Revisiting the mouse reference genome: Single molecule sequencing of C57BL/6J "Eve". *A. Srivastava[1], V.K. Sarsani[1], I. Fiddes[2], V.A. Schneider[3], L.G. Reinholdt[1]. 1) The Jackson Laboratory, Bar Harbor, ME; 2) University of California, Santa Cruz, Santa Cruz, CA; 3) DHHS/NIH/NLM/NCBI, 45 Center Drive-Natcher Bldg, Bethesda, MD.*

The mouse reference genome, GRCm38, represents the most widely used inbred laboratory mouse strain, C57BL/6J; a strain that is maintained and distributed by The Jackson Laboratory. The currently available C57BL/6J mice sourced from the The Jackson Laboratory source are up to 26 inbreeding generations removed from the samples used to generate the mouse reference genome. Moreover, the strain is now maintained through periodic reintroduction of mice from cryopreserved embryo stock, a husbandry practice that is designed to maximize genetic homogeneity by effectively limiting the number of inbreeding generations over time. The embryo stock is derived from a single breeder pair, named C57BL/6J Adam and Eve. This genetic divergence explains a sub-set of true positive variants that are found when comparing currently available C57BL/6J genomes to the mouse reference genome. Additionally, while GRCm38 is one of the highest quality and well-characterized published mammalian genome assemblies, it has some features that may contribute to increased false positive rates in variant discovery. There are 187 gaps and 211 problematic regions reported by Genome Research Consortium in Mouse Genome Build 38, patch 5 and the genome does not represent a single individual. To better represent C57BL/6J mice that are currently in use by the research community and to understand/resolve the problematic regions in the existing reference genome, we have generated a whole genome de novo assembly of C57BL/6J Eve using ~60X PacBio single molecule sequencing and optical map data generated from Bionano Genomics platform. This hybrid assembly approach yielded 12,690 scaffolds with N50 scaffold size of 1.5MB and went under comprehensive annotation using comparative annotation toolkit from UCSC. Using this data, we resolved 1,788 repeats in reference genome and corrected 134 recurring C57BL/6J exome variants. Further, we were able to find the B6Eve scaffolds that span four of the reported gaps completely and span seven gaps partially, implying the utility, albeit limited, of gap closures with our assembly. Finally, with the help of PacBio brain transcriptome (Iso-Seq data) of mouse and assembled B6Eve genome, we were able to identify novel isoforms of two important mouse genes. This B6Eve assembly will provide an extremely useful resource to scientific community working with C57BL6/J mouse in genomics research as well as genetic engineering/mouse model development.   .

406W

Heat regulation of germ cell transcriptome: A bid to identify novel targets for contraception. *A. Pandey, S.K. Yadav, L. Kumar, R. Vishvkarma, J.P. Maikhuri, R. Singh, G. Gupta. Endocrinology, CSIR-Central Drug Research Institute, Lucknow, India.*

Mammalian spermatogenesis is extremely sensitive to body temperature and fails rapidly; but occurs normally at 3-6°C lower temperature. The scrotum is nature's uniquely designed organ to maintain testes at a lower temperature than the body. Cryptorchidism (abdominal testes) is a common congenital abnormality affecting 1–5% of full-term male births and is a risk factor for sub-fertility. Limited clinical studies have reported that transient testicular heating of adult human males results in reversible spermatogenic arrest, hence could be used as a method of contraception. The spermatogenesis is regulated at transcriptional, post-transcriptional, and epigenetic levels by the integrated expressions of an array of testicular genes in a precise temporal sequence. The meiotic (pachytene/diplotene spermatocytes) and post-meiotic (round spermatids) germ cell types are most vulnerable to testicular heating. The heat sensitive changes in germ cell transcriptome reflecting early changes in these cells, if known, could identify novel targets that could be exploited to achieve 'molecular heating' in testis leading to contraception. We have employed centrifugal elutriation to isolate >90% pure germ cells from normal and surgically-cryptorchid rat testis, and NGS technology to analayse their transcriptome and miRNA profile. Some significant changes have been noted in genes of different biological processes like metabolism (*MCT2, MCT4, Glut3, LDHc*), lipid biogenesis (*Acly*), ROS and Ca++ signaling (*Daxx, Camk2d*), and apoptosis (*Daxx*), that were further validation by RT-PCR. Functional validation studies would reveal their application in transient arrest of spermatogenesis for contraception.

407T

Breakpoint analysis of chromosomes having inverted duplication with terminal deletion by NGS. *H. Inagaki[1,2], K. Kanyama[1], T. Kato[1], Y. Ouchi[1,2], T. Yamamoto[3], H. Kurahashi[1,2].* 1) Division of Molecular Genetics, ICMS, Fujita Health University, Toyoake, Aichi, Japan; 2) Genome and Transcriptome Analysis Center, Fujita Health University, Toyoake, Aichi, Japan; 3) Institute of Medical Genetics, Tokyo Women's Medical University, Shinjuku, Tokyo, Japan.

   Inverted duplication with terminal deletion (inv dup del) is one of the complex constitutional structural variations of chromosomes. The most frequent inv dup del known to date is located at 8p. In 8p inv dup del, a non-allelic homologous recombination (NAHR) between segmental duplications in inverted orientation produces a dicentric chromosome, and it induces breakage between the two centromeres during mitotic division. In this study, we investigated other types of inv dup dels as well as inverted triplication with terminal deletions (inv trip dels) of 4p, 9p, 10q, 11q, 18p and 21q to find a common mechanism by using cytogenetic microarray and whole genome sequencing techniques. By means of mate pair libraries of approximately 9 kb in average, we identified the fusion structures of breakpoint regions by NGS. Although complete palindromic structures at the center of the inverted duplications had been predicted, the whole genome sequencing data showed asymmetric spacer region with several thousand bps in size in all cases. Sequencing of these regions by PCR amplification reveals the size of the spacer was from 1744 to 3547 bp. The junctions appeared to utilize microhomology. This structure could be generated in two possible mechanisms: formation of complete palindromic sequence at the breakpoint of dicentric chromosome, followed by the deletion of the central region due to instability of the palindromic sequences in living cells, or formation of dicentric chromosome with asymmetrical inverted breakpoint regions with several thousand bp gaps. Possibility of replication-based mechanism generating the similar deletion sizes around 2 kb is discussed.

408F

MethylHiC reveals long-range genetic-epigenetic and epigenetic-epigenetic interactions within the same single molecule. *Y. Liu[1,2], G. Li[3], B. Ren[3,4], M. Kellis[1,2].* 1) CSAIL, MIT, Cambridge, MA; 2) The Broad institute of MIT and Harvard, Cambridge, MA; 3) Ludwig Institute for Cancer Research, La Jolla, CA; 4) Department of Cellular and Molecular Medicine, School of Medicine, University of California San Diego, La Jolla, CA.

   DNA methylation is the most extensively studied epigenetic marker that plays direct role in mammalian gene regulation, such as gene silencing and imprinting. Recent studies on allelic specific methylation (ASM) have revealed the direct effect of genetic variants on DNA methylation at local region. Association studies of methylation quantitative trait loci (meQTLs) further suggest the genetic variants may influence DNA methylation level across large genomic intervals. Epigenetic gene silencing and activation has also long been envisaged as a local event. Very recent investigations indicate that large regions of chromosomes can be co-coordinately suppressed or activated. The direct genome wide experimental approach to explore the long-range genetic-epigenetic and epigenetic-epigenetic interactions, however, is limited especially for the current whole genome bisulfite sequencing (WGBS) based on short reads platform. Long read sequencing technology (SMRT or MinION) may potentially reveal DNA methylation coordination in longer region, but the distance is still limited at kilobases level. Chromosome Conformation Capture followed by massively parallel sequencing (Hi-C) provides long-range interaction between regulatory elements and linear organization of genome sequences along the entire chromosomes that had been used to assemble haplotype. Here, we developed MethylHiC, which applies *in situ* Hi-C followed by bisulfite treatment to understand the long-range genetic-epigenetic and epigenetic-epigenetic interactions within the same single DNA molecule.   First, we validated that MethylHiC is globally consistent with both of *in situ* Hi-C and WGBS data generated from the same cell line. Second, we found that DNA methylation is highly concordant at distal interacted regulatory regions within the same DNA molecule. The concordance level varies and shows different patterns between regions at different chromatin states. We also detected long-range allelic specific methylation for the first time. Further, we showed that these long-range genetic-epigenetic and epigenetic-epigenetic interactions can be disrupted by Ten-eleven Translocation (TET) knockout. Moreover, by incorporating additional DNA methylation concordance information into the analysis, we can largely improve the Hi-C resolution. Our method here paves the road to evaluate the direct long-range effect of genetic and epigenetic alterations at different pathological conditions within the same DNA molecule.

409W

Complementary long and short-read sequencing techniques identify rearrangement structures and concurrent single nucleotide variant formation on chromosome 17p11.2. *C.R. Beck[1], C.M.B. Carvalho[1], Z.C. Akdemir[1], Z. Chong[2], E.S. Chen[1], P.C. Thornton[1], P. Liu[1], B. Yuan[1], M. Withers[1], S.N. Jhangiani[3], A.C. English[3], D.M. Muzny[3], R.A. Gibbs[3], C.A. Shaw[1], P.J. Hastings[1,6], J.R. Lupski[1,3,4,5,6].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Genetics and the Informatics Institute, the University of Alabama at Birmingham, Birmingham, AL; 3) Human Genome Sequencing Center, BCM, Houston, TX; 4) Department of Pediatrics, BCM, Houston, TX; 5) Texas Children's Hospital, Houston, TX; 6) Dan L Duncan Comprehensive Cancer Center, BCM, Houston, TX.

   A majority of Smith-Magenis Syndrome (SMS MIM #182290) and Potocki-Lupski Syndrome (PTLS MIM #610883) patients have *de novo* deletion or duplication of chromosome 17p11.2, respectively. These copy number variants (CNVs) affect the dosage-sensitive gene *RAI1*. Approximately 80% of rearrangements at this locus are due to recurrent non-allelic homologous recombination events between low copy repeats (LCRs) leading to 3.7Mb CNV, and ~15% of the deletions and duplications are non-recurrent. Repeats comprise almost a quarter of 17p, and ~60% of non-recurrent 17p11.2 rearrangements contain one breakpoint in an LCR, making characterization of breakpoint junctions a challenge. Further complexities at non-recurrent CNV junctions such as inversions and/or evidence for additional template switches (TS) and single nucleotide variants (SNVs) that can accompany complex genomic rearrangements formed by replicative repair (*e.g.* FoSTeS/MMBIR) necessitate comprehensive molecular and genomic analyses of individual rearrangements. To gain mechanistic insights into the formation of rearrangements at 17p11.2, we combined array comparative genomic hybridization (aCGH) with a targeted 7Mb Illumina and Pacific Biosciences SMRT sequencing approach. Preliminary data from 19 subjects with non-recurrent rearrangements identified junctions in 16 cases and found complexities not seen by high-density aCGH. Additionally, 32 *de novo* SNVs were revealed within the 7Mb capture region; only one would have been identified by breakpoint PCR. This yields a mutation rate of 2.4×10^{-7} or 10-fold higher when compared to the intergenerational germline spontaneous *de novo* rate (~2×10^{-8}/bp/generation). Whole exome sequencing of the 19 trios underscored the association of the CNV with the *de novo* point mutations, as SNV mutation rates were not globally higher. The majority (~70%) of observed SNVs were transversion substitutions and ranged from 202bp to ~1.1Mb away from the nearest junction. Our findings are consistent with the use of an error prone polymerase with reduced fidelity and processivity (reflected by iterative TS at the breakpoint junction) that switches to a more processive polymerase with greater fidelity as the replisome proceeds. These data show that complementary approaches can resolve rearrangements in complex genomic regions, identify point mutations far from breakpoint junctions, and indicate a higher local mutation rate accompanying CNV as predicted for replication-based mechanisms.

410T

Method for deciding detection limit of droplet digital PCR for rare mutations. *T. Takahashi, M. Umekage, M. Nomura, T. Hanatani, H. Dohi, Y. Sato, N. Takasu, N. Amano.* Center for iPS Cell Research and Application, Kyoto University, Kyoto, Japan.

   In order to ensure safety of induced pluripotent stem cells (iPSCs) and differentiated cells, we are investigating droplet digital PCR (ddPCR) to assess the quantity for rare mutations. Especially, we are checking cancer related mutations in the following two gene lists, cosmic census gene list and Japan pharmaceuticals and medical devices agency (PMDA) gene list for medical application of iPSCs. In this study, we discussed how to decide a detection limit of rare mutant allele frequencies when false positive droplets were detected using ddPCR. The iPSCs were differentiated, cultured and evaluated a certain mutation in the differentiated cells. As the result, the mutant allele frequencies of the differentiated cells and negative control were 0.41% and 0.27%, respectively. We faced an issue to determine whether the 0.41% was noise or not. To solve this problem, we decided the detection limit of this mutation by considering dilution series of positive samples and three independent negative samples. The droplet counts of the three negative samples were merged and adjusted using Poisson statistics. We conducted the experiments 3 times, and decided the detection limit using the averages and standard deviations. The detection limit was 0.5% for the specific probe. We concluded the allele frequency 0.41% of the differentiated cells was lower than the detection limit of this mutation. In addition, we repeatedly measured 10 times, and confirmed that all measurements of the differentiated cells were below the detection limit. In the near future, we will successfully apply ddPCR to measure rare structure variant and mosaic for medical application of iPSC and differentiated cells.

411F

Targeted next-generation sequencing for identifying genes related to horse temperament. *S. Song[1], D. Oh[2], G. Cho[3], D. Kim[4], Y. Park[5], K. Han[1].* 1) Nanobiomedical science, Dankook University, Cheonan, South Korea; 2) Livestock Research institute, Gyeongsangbuk-Do, Yeongju, South Korea; 3) College of Veterinary Medicine, Kyungpook National University, Daegu, South Korea; 4) Department of Anesthesiology and Pain Manangement, College of Medicine, Dankook University, Cheonan, South Korea; 5) Department of Equine Industry, Korea National College of Agriculture and Fisheries, South Korea.

It is a fundamental challenge to discover the association of genetic traits with phynophic traits. In this study, we aimed to identify possible genetic traits related to horse temperament. Based on previous findings, we selected 71 candidate genes related to temperamental trait and examined them in the human and horse reference genomes (hg38 and equCab2, respectively). We found 16 orthologous genes and, by comparing with the human reference genome, 17 homologous genes in the horse reference genome. We designed probes specific for the 33 horse genes. Using the probes, we built sequencing libraries of the genomic DNA samples from 8 aggressive and 8 docile horses, and sequenced the constructed libraries using the Illumina Hiseq2500 platform. Through the analysis of the targeted exome sequences, we identified single nucleotide polymorphisms (SNPs) in the genes. SNPs could be served as genetic markers to evaluate aggressive or docile levels of horses. To examine whether any generic variants are associated with horse temperament, we performed genome-wide association study (GWAS) using the SNP data. GWAS analysis identified 10 variants (*p*-value <0.05) which could be related to horse temperament. We validated the variants using Sanger sequencing. The most significant variants were found in *MAOA* (c.1164+41T>C) and *AR* (c.1047+27G>T) genes with 8.09 x 10^{-4} *p*-value. We suggest that the variants might be used to assess horse temperament and to determine superior horses for riding or racing.

412W

Cross-tissue protein expression and genetic regulation of transcription factors and cell signaling genes in enhancing Genotype-Tissue Expression (eGTEx) samples. *M. Oliva[1,2], M. Fernando[1,2], F. Wu[1,2], C. Linke[1,2], A.D. Skol[1,2,3], B.E. Stranger[1,2,3], Genotype-Tissue Expression (GTEx) Consortium.* 1) Section of Genetic Medicine, Department of Medicine, University of Chicago, IL; 2) The Institute for Genomics and Systems Biology, University of Chicago , IL; 3) Center for Data Intensive Science, University of Chicago , IL.

The human transcriptome has been studied extensively but investigations of the proteome have been limited by the lack of robust technologies for large-scale protein quantification. However, the proteome comprises a vital element in the causal network linking genetic variation to higher order organismal phenotypes. As part of the enhancing Genotype-Tissue Expression program (eGTEx), in which additional –omics assays are performed on human tissue samples that have been deeply characterized at the transcriptomic and genetic level, we aim to quantify the cross-tissue protein expression levels within and between GTEx tissues and to characterize the effect of genetic variation on protein expression in humans. We have adapted microwestern arrays to quantify the expression of 353 transcription factors (TFs) and cell signaling proteins in 203 GTEx samples of 33 tissues from 14 individuals. Tissue clustering profile derived from the protein expression data recapitulates tissue similarity relationships derived from transcriptome data. Relative protein abundance levels reveal strong tissue specificity: proteins are expressed in a tissue-enriched (39%) or –enhanced (36%) manner. Brain, muscle and skin are enriched for tissue-specific TFs relative to other tissues. For the majority of the proteins assayed, cross-tissue protein and mRNA abundances do not correlate, possibly indicating buffering or a "translation on demand" mechanism. We observe a subset of proteins (e.g. ZNF35, HMGXB4, SHPRH) that are strongly anti-correlated with mRNA levels, suggesting negative feedback regulation. We are quantifying the same 353 proteins in N=165-192 samples from ten GTEx tissues to better characterize the protein expression landscape, assess the genetic basis of protein expression variation and build protein-mRNA regulatory networks. We have quantified this set of proteins in 174 skeletal muscle GTEx samples and have characterized local and distal protein expression quantitative trait loci (cis- and trans- pQTLs). Preliminary results indicate that mRNA-protein correlations within tissues are weak and cis-pQTLs are both less prevalent and have significantly smaller effect sizes compared to cis-eQTLs, suggesting cellular mechanisms that buffer the effects that genetic variation exerts at a transcript level. In addition, a fraction of significant cis-pQTLs are not eQTLs, suggesting genetic effects on post-transcriptional regulation.

413T

Hanwoo-specific structural variations generate genomic diversity in the cattle genome. *J. Park, W. Shin, S. Mun, K. Han.* Nanobiomedical science, Dankook university, Cheonan, South Korea.

The total length of the bovine genome is approximately three billion base pairs, which has been containing 29 autosomes and sex chromosomes. Among them, as much as about a half of the bovine genome (~46%) is composed of transposable elements (TEs). TEs insertions directly affect host genome diversity caused by the structural variations. In addition, they can change the gene expression and function related to nearby their regions. Therefore, it is generally believed to be difference of the interspecies properties in details. We anticipate that TEs-mediated insertion and deletion (INDEL) has a strong influence on the bovine genome to geographic speciation since the divergence of Korean native cattle (Hanwoo) and other cattle species. We investigated TEs-mediated INDELs by comparing *Bos taurus*(UMD3.1.1, Reference), Jersey, Holstein, Angus, and Hanwoo whole genome data, respectively. Among them, we collected the Hanwoo-specific TEs-mediated INDEL candidates through a bioinformatics approach. To make certain the structural variations in the Hanwoo genome, we conducted the amplification of targeted-sites for Sanger sequencing and comparatively analyzed the sequencing data. In this study, we found the Hanwoo-specific TEs-mediated INDEL regions and inspected how to make structural changes related to TEs in the Hanwoo genome. In addition, we suggest that Hanwoo-specific TE-mediated INDELs could be helpful to understand the genomic variations and the species differentiation in the bovine genome as well as in Hanwoo genome.

414F

Study of G2 phase cell cycle arrest and its application in mediating *SOX9* mutagenesis in pluripotent stem cells (PSCs). *T.Y. Ha, H.H. Cheung, W.Y. Chan.* School of Biomedical Sciences, Faculty of Medicine, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR.

Pluripotent Stem Cells (PSCs) have been widely used to study early developmental events, together with genome engineering techniques such as CRISPR/Cas9. However, mutagenesis is a challenging task since both DNA repair pathways, Homology Directed Repair (HDR) and Non-homologous End Joining (NHEJ) have their own limitations. HDR, although regarded as a precise repair mechanism, has extremely low efficiency in PSCs. On the other hand, NHEJ is an error-prone mechanism which generates insertion-and-deletions (indels) and in-frame mutations, which may lead to formation of inactive proteins. Between these two pathways, HDR, although being inefficient, is the preferred DNA repair pathway for PSCs. In the search for methods to increase HDR efficiency, several groups suggested that HDR efficiency could be increased by synchronization of G2/M phase in cell cycle. Nonetheless, optimizations and benchmarking of this method is required. In an attempt to increase HDR efficiency, we used siRNA against a cell cycle related gene to induce cell cycle arrest in 293T and HepG2 cells. Results showed that G2 phase arrest could be achieved in both cell lines and there is a potential to achieve similar results in pluripotent stem cells. To demonstrate in principle that the efficiency of HDR could be increased using cell cycle arrest, a GFP rescue assay in 293T cells was performed. In addition, different methods reported in literature such as using NHEJ inhibitor in PSCs will be compared with the cell cycle arrest approach proposed above. Two parameters, apoptosis and aneuploidy after manipulations of cell cycle arrest and the addition of NHEJ inhibitor will be assessed by cell cycle analysis. Moreover, to demonstrate that the method is applicable to PSCs, mutagenesis of *SOX9* with CRISPR/Cas9 in PSCs will be created and HDR efficiency will be assessed.In conclusion, we propose that cell cycle arrest could be an efficient method to increase HDR efficiency. The method is applicable to PSCs in order to achieve precise genome editing and this method could be a potential application for the study of early developmental events.

415W

Detection of novel L1 insertions in the chimpanzee (*Pan troglodytes*) genome. *S. Lee[1], W. Lee[1], S. Mun[1], K. Han[1], W. Tang[2], P. Liang[2].* 1) Dankook University, Cheonan, Chungnam, South Korea; 2) Brock University, St. Catharines, Ontario, Canada.

Long Interspersed Element-1 (LINE-1 or L1) is known as a most successful retrotransposon in the primate genomes and has contributed to their genome diversity and variations during the primate evolution. With complete chimpanzee genome sequencing, it was demonstrated that L1s are powerful genetic elements as well as responsible for more than 5.0 MB of genomic differences between human and chimpanzee. Thus, we here aimed to identify novel L1 insertions, which can generate genome difference between chimpanzee individuals, by using an advanced target-enrichment sequencing method and bioinformatics analysis. Targeted enrichment is an appropriate system to detect regions of interest, generally where genetic variants such as gene fusions and copy number variations occurred. First, to design the chimpanzee L1-specific probe, we collected ~2,000 chimpanzee-specific L1s from the chimpanzee reference genome (UCSC panTro5) and sorted out full-length of chimpanzee-specific L1s (> 6kb) that have conserved 5' UTR sequences. Then, the probe for sequence enrichment containing a part of chimpanzee-specific L1s was designed in the conserved region of the L1 consensus sequence. We produced family-based sequencing data with DNA samples from father, mother, and offspring chimpanzees by using target-enrichment sequencing. The results of this study suggest that L1 target-enrichment sequencing is useful to identify novel L1 insertions.

416T

Detection of non-reference L1 insertions in the human genome by using a target enrichment method. *W. Shin, S. Mun, K. Han.* Nanobiomedical Science, Dankook University, Cheonan, South Korea.

Long interspersed element-1s (L1s) have been successfully expanding in the mammalian genomes during the evolutionary time of their genome structure. L1s are a common large family of retrotransposons which constitute ~17% of the human genome and have resided in the mammalian genomes for >150 million years. Previous studies suggested that a retrotranposition activity in the human genome and the majority of these elements have driven human genomic variations. Some of the new L1 elements have recently inserted into the human genome that are polymorphic in human individuals. A few human-specific L1s (L1Hs) have been widely known to be capable of extremely high levels of mobilizing activity and result in several human genetic diseases. Here, we describe a developed method for detecting non-reference L1 insertions using target enrichment method. We suggest the development of a NGS system that advanced an approach for detection non-reference L1 insertion in the human genome through targeted-probe method. In conclusion, we identified the candidate regions of 566 non-reference novel L1 insertions in the human genome through bioinformatics approach and targeted-sequencing. Based on a variety of analysis, we suggest that a target enrichment method is efficient for detecting specific elements such as transposable elements. Thus, our method could be useful for developing a new polymorphic genetic marker across human individuals.

417F

Gene expression analysis and enzyme function of pig mammalian chitinase. *E. Tabata, A. Kashimura, S. Wakita, M. Ohno, M. Sakaguchi, Y. Sugahara, F. Oyama. Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan.*

Pigs are an ideal biomedical model for human diseases and conditions bridging the gaps between mouse models and humans. Chitin, a linear polymer of β-1, 4-linked *N*-acetyl-D-glucosamine (GlcNAc), is the second most abundant natural polysaccharide in nature and functions as a major structural component in fungi, crustaceans and insects. Although mammals do not produce chitin, humans and mice express two active chitinases. Firstly, chitotriosidase (Chit1) was identified in macrophages of Gaucher disease patients. Acidic mammalian chitinase (AMCase) was discovered later and has attracted substantial attention due to its expression fluctuations under certain pathological conditions such as asthma, allergic inflammation, ocular allergy, dry eye syndrome and gastric cancer. Recently we showed that mouse AMCase can function as a protease-resistant major glycosidase under the conditions of stomach and intestine. However, the physiological roles of the AMCase in other mammals remain unknown. Here, we show that AMCase mRNA was predominantly expressed in pig stomach using quantitative real-time RT-PCR system and it was much higher than that of Chit1, housekeeping genes and gastric mucosa genes. Porcine AMCase has highest chitinolytic activity at pH of around 2-4. Also, it was resistant to endogenous pepsin digestion and degraded polymeric chitin substrates under mimicking stomach condition. Our study suggests that AMCase functions as a part of the host defense against chitin-containing pathogens or as a digestive enzyme that breaks down chitin in pig stomach tissues.

418W

Gene expression signature as a potential treatment monitoring biomarker for active tuberculosis in Thailand. *R. Miyahara[1], N. Satproedprai[2], N. Wichukchinda[2], H. Yanai[3], L. Toyo-oka[1], K. Tokunaga[1], S. Mahasirimongkol[2]. 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Medical Genetics Center, Medical Life Science Institute, Department of Medical Sciences, Ministry of Public Health, Thailand; 3) Fukujuji Hospital, Japan Anti-Tuberculosis Association (JATA), Kiyose, Japan.*

[Introduction] Recent research suggested that blood transcription signatures by a set of 7 genes could descriminate between the active TB patients and healthy controls. This 7 genes set is potentially useful as potential biomarkers for monitoring TB patients after their treatment. Further, we described the influence of blood cell counts on each genes expression to identify the blood cell populations influencing these genes expression in active TB patients. [Methods] 12 active TB patients confirmed by TB culture positive were invited to Chaingrai Prachanukroh Hospital in Northern Thailand and collected blood samples at enrollment, 2weeks, 2months and 6months after the TB treatment initiation. Using the whole blood samples, the gene expression levels of the 7 genes (*FCGR1A, FCGR1B* variant1, *FCGR1B* variant2, APOL1, GBP5, STAT1 and KCNJ15) and 2 novel candidates (*MAFB* and *KAZN*) were measured by real-time PCR. The number of each types of white blood cell was counted by standard flowcytometry. The gene expression levels were compared at two time points (first visit vs 6months) by paired t-test. In order to predict each gene expression level by white blood cell counts and the type of blood cell counts, we fit a multivaiate regression model by backwards stepwise selection. [Results] All 9 gene expression levels were decreased at 6 months after TB treatment initiation compared to the enrollment. Especially, *FCGR1A, FCGR1B* variant1 and *FCGR1B* variant2 showed statistically significant decreased by 1.74, 1.49 and 1.33 (log scale), respectively (p<0.001 by paired t-test). Absolute lymphocyte, monocyte and neutrophil counts were correlated with *FCGR1A, FCGR1B* variant1, *FCGR1B* variant2 gene expression especially at enrollment and 2 weeks after the treatment start. The regression model only including neutrophil counts could fit to *FCGR1A* gene and *FCGR1B* variant2 gene expression at enrollment (R-squared=43% and 54%). [Conclusion] *FCGR1A* and *FCGR1B* gene expression levels decreased after TB treatment. Among lymphocyte, monocyte and neutrophils counts, neutrophil counts were the main influencing factor contributed to their expression levels. Previous studies implied that *FCGR1A (CD64)* and *FCGR1B* is associated with bacteria infection and TB infection. *FCGR1A (CD64)* and *FCGR1B* might be an effective and inexpensive biomarker not only identify active TB cases but also monitoring the TB treatment. .

419T

Heterogeneity of human ribosomes inferred from rDNA and rRNA sequencing. *R. Nagaraja[1], J.H Kim[2], A.T. Dilthey[3], H.S. Lee[2], S. Koren[4], D. Dudekula[1], W.H. Wood III[1], Y. Piao[1], A.Y. Ogurtsov[4], K. Utani[1], V.N. Noskov[2], S.A. Shabalina[4], D. Schlessinger[1], A.M. Phillippy[3], V. Larionov[2].* 1) Laboratory of Genetics and Genomics, NIA/NIH, Baltimore, MD; 2) National Cancer Institute, Developmental Therapeutics Branch, Bethesda, MD; 3) National Human Genome Research Institute, Computational and Statistical Genomics Branch, Bethesda, MD; 4) National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD.

    Human cells contain several hundred ribosomal DNA (rDNA) genes clustered in nucleolar organizer regions (NORs) on the short arms of five acrocentric chromosomes. Each NOR is a tandemly repeated unit of 43 kb transcribed into 13.3 kb 45S, encoding 18S, 5.8S and 28S rRNAs, and a 29.7 kb intergenic spacer (IGS). However, little is known about their precise structure and variation. Prior attempts to characterize NOR sequences have failed due to their highly repetitive structure, and these regions are missing from the human reference genome. We targeted chromosome 21 NOR sequences using transformation-associated recombination in yeast, starting from a rodent/human hybrid cell line. After conversion to BACs they were sequenced on PacBio and Illumina platforms, enabling assembly. Structures were validated using an Oxford Nanopore protocol capable of ~90 kb of single-read sequence. Thirteen clones, (0.4-fold coverage ,~0.8 Mb) of the chr. 21 NOR, revealed variants of standard rDNA reference. We identified 216 variant alleles in the 45S region, 90 in mature rRNAs sequences, as well as 7 inversion breakpoints in the IGS resulting in palindromic structures. The 90 sequence variations, compared to predicted 18S and 28S rRNA 2D structures, showed most located in species-specific expansion segments and rather than universal functional cores of 18S and 28S rRNA. Therefore, these variants are less likely to influence the catalytic core function of ribosome. However, variations in RNA 2D structure of the 5' ETS would likely affect structural stability, consistent with a possible role in fine-tuning rRNA transcription. In addition, a predicted candidate gene was found in the IGS for a subset of clones. The candidate gene, IGS breakpoints, and 61% of variant alleles were also seen in independent whole-genome PacBio and Illumina data. Thus, we have constructed a new ~45 kb rDNA reference sequence that has improved support from whole-genome data. This also provides an approach to full analysis of NORs and reagents for the study of human NOR function.

420F

Discovery of rare, diagnostic *Alu*Yb8/9 elements in diverse human populations. *J. Feusier, D.J. Witherspoon, W.S. Watkins, C. Goubert, T.A. Sasani, L.B. Jorde.* Human Genetics, University of Utah, Salt Lake City, UT.

    Polymorphic human *Alu* elements are excellent tools for assessing population structure, and new retrotransposition events can contribute to disease. Next-generation sequencing has greatly increased the potential to discover *Alu* elements in human populations, and various sequencing and bioinformatics methods have been designed to tackle the problem of detecting these highly repetitive elements. However, current techniques for *Alu* discovery may miss rare, polymorphic *Alu* elements. Combining multiple discovery approaches may provide a better profile of the polymorphic *Alu* mobilome. *Alu*Yb8/9 elements have been a focus of our recent studies as they are young subfamilies (~2.3 million years old) that contribute ~30% of recent polymorphic *Alu* retrotransposition events. We have modified and improved our ME-Scan methods for detecting *Alu* elements and have applied these methods to discover new insertions in a large set of individuals with diverse ancestral backgrounds. We identified 5,288 putative *Alu* insertion events, including 313 putative novel *Alu*Yb8/9 elements from 213 individuals from 18 diverse human populations. Hundreds of these loci were specific to continental populations, and 23 non-reference population-specific loci were validated by PCR. We PCR validated 68 rare *Alu*Yb8/9 elements, of which 11 have hallmarks of an active source element. The subfamily distribution of these rare *Alu*Yb8/9 elements is consistent with previous datasets, and may be representative of rare loci. We also find that while ME-Scan and low-coverage, whole-genome sequencing detect different *Alu* elements in 41 1000 Genomes individuals, the two methods yield similar population structure results. Current *in-silico* methods for *Alu* discovery may miss rare, polymorphic *Alu* elements. Therefore, using multiple techniques can provide a more accurate profile of *Alu* elements in individuals and populations. We improved our false-negative rate as an indicator of sample quality for future ME-Scan experiments. In conclusion, we demonstrate that ME-Scan is a good supplement for next-generation sequencing methods and is well-suited for population-level analyses.

421W

Identification of active LINE-1 retrotransposons in the baboon genome.
W. Lee[1,2], W. Tang[3], P. Liang[3], K. Han[1,2]. 1) Dankook University, Cheonan, South Korea; 2) 2DKU-Theragen institute for NGS analysis (DTiNa), South Korea; 3) Brock University, St. Catharines, Ontario, Canada.

Long INterspersed Element-1 (LINE-1 or L1) retrotransposons and *Alu* elements belonging to Short INterspersed Element (SINE) are most successful retrotransposons which have generated genetic diversity in primate genome evolution and they are accounted for around 27.7% of the primate genome. In the mechanism of transposition, L1 and *Alu* copies have spread into a new location in the genome by RNA intermediate transposition, known as "copy-and-paste" mechanism. In the previous baboon genome study, the nine times higher expansion of race-specific *Alu*Y subfamilies were observed in the baboon genome compared with the human genome. Thus, we speculate that the huge expansion of *Alu*Y subfamilies during baboon genome evolution might be correlated with L1 activity at the same moment, because the expansion of *Alu* copies is dependent on L1-encoded protein activities such as endonuclease and reverse transcriptase. Here, we identified 493 baboon-specific full-length L1 candidates from the baboon reference genome and observed functionality and phylogenetic relationship of active L1 copies through comprehensive analysis to understand correlation between the active L1 copies and the expansion of *Alu*Y subfamilies in baboon.

422T

Extending and improving GENCODE gene annotation. *J.E. Loveland[1], The. GENCODE Consortium[2,3,4,5,6,7,8], A. Frankish[1,2], P. Flicek[1,2].* 1) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Trust Campus, Hinxton, Cambridge CB10 1SA, United Kingdom; 3) University of California, Santa Cruz, California 95064, USA; 4) Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; 5) Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; 6) Centre for Genomic Regulation (CRG) and UPF, 08003 Barcelona, Catalonia, Spain; 7) Yale University, New Haven, Connecticut 06520-8047, USA; 8) Spanish National Cancer Research Centre (CNIO), E-28029 Madrid, Spain.

High quality gene models are essential for annotation of sequence variation in clinical samples. Incomplete and inaccurate annotation of gene structure and function has the potential to introduce errors in identification of causal variants identified from genome and exome sequence. As part of the GENCODE project, we are responsible for producing detailed reference annotation of all human and mouse protein-coding genes, pseudogenes, long non-coding RNAs and small RNAs. GENCODE gene annotation is available from gencodegenes.org and via the Ensembl and UCSC genome browsers. Updates to annotation are made continuously. To allow access to inter-release gene annotation changes we have created a track hub that is updated every 24 hours and available at ftp://ngs.sanger.ac.uk/production/gencode/update_trackhub/hub.txt. We will describe how GENCODE is improving the annotation of the human reference genome at the level of gene loci and alternatively spliced transcripts. More that 90 completely novel protein-coding genes have been identified in our analysis of un-annotated regions which have evolutionarily conserved coding potential by phyloCSF. Integrating multiple orthogonal datasets, especially RNAseq and CAGE libraries, has made it possible to annotate genes that were previously missed because of their restricted expression. In order to produce the most complete gene set possible we are integrating novel transcriptomic data into our analysis pipeline. These data suggest that we have missing genes from GENCODE as well as many missing exons and alternatively spliced transcripts. A study using public transcriptional evidence on next-generation sequence data from human brain was undertaken to re-annotate 140 genes from a reference diagnostic panel for Early Infantile Epileptic Encephalopathies (EIEE). CAGE was used to confirm transcriptional start sites followed by, RNAseq, SLRseq and PacBio reads to find new alternative splicing events including novel exons and splice site shifting. Despite previous annotation by GENOCODE these new data enabled us to annotate many hundreds of alternatively spliced transcripts, novel exons, shifted splice sites and over 150kb of additional genomic coverage. These efforts suggest that despite the current GENCODE human gene set (v26) being the most comprehensive to date, it is still far from complete. The challenge now is to utilise new data effectively and efficiently to provide the most complete gene set for the clinical community.

423F

Updates to the human reference genome assembly (GRCh38). *T. Rezaie[1], V.A. Schneider[1], T. Graves-Lindsay[2], K. Howe[3], P. Flicek[4] for the Genome Reference Consortium.* 1) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 2) The McDonnell Genome Institute at Washington University, St. Louis, MO; 3) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 4) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK.

The Genome Reference Consortium (GRC) provides users with updates to the human reference genome assembly. In the era of high quality genome assemblies and the accessibility of clinical genome sequencing, the reference genome provides the coordinate system for genomic-based analyses and gene annotations, facilitates identification of disease-associated variants and represents the diversity of the human population to both basic and clinical research communities. Despite providing the highest quality model of the human genome, reference assembly representations of some highly complex and medically important genomic regions, as well as population variation, remain problematic or incomplete and thus still need improvement. Since the 2013 release of GRCh38, the latest coordinate-changing update to the human reference, the GRC has continued to provide public assembly updates in the form of non-coordinate changing patch releases. As of GRCh38.p10 (GCA_000001405.25), the reference assembly includes 53 fix and 49 novel patch scaffolds, which represent chromosome path changes and alternate representations of chromosome sequences, respectively. These patches, which are stand-alone accessioned scaffolds, are given chromosome context via alignments that are available for download with the rest of the assembly. The current suite of patches covers 32 Mb (~1% of the total chr. length), of which 1.42 Mb is novel sequence, and includes coding regions. We will show examples of recent patches that close assembly gaps, improve reference representations of genes, including clinically important loci, as well as those that add genomic diversity, and discuss their usage in basic and clinical analyses. We will describe how GRC curation efforts have been impacted by the public availability of new, highly contiguous, genome assemblies from different populations, optical maps, as well as genomic studies that provide allele frequency data, and how these data are contributing to discussions regarding the future of the reference assembly. The GRC is committed to making its curation efforts publicly accessible. These can be viewed at the GRC website (https://www.genomereference.org) and in browser tracks that are accessible via track hub. We will review the tracks and how to access them in browsers at NCBI, EMBL-EBI and elsewhere. We will also discuss the resources available at the GRC website, including mechanisms for contacting the GRC with general questions and assembly curation requests.

424W

Deep characterization of the contribution of short tandem repeats to gene expression across tissues. *S. Feudjio Feupe[1,2], M. Gymrek[1,3].* 1) Department of Medicine, University of California San Diego, La Jolla, CA; 2) Department of Biomedical Informatics, UC San Diego, La Jolla, CA; 3) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA.

Recent studies have identified thousands of genetic variants contributing to variation in gene expression across dozens of tissues and cell types. However, these efforts have mainly focused on the contribution of single nucleotide polymorphisms (SNPs), and have largely ignored more complex variants such as repetitive regions and structural variants. This is due both to bioinformatic limitations in genotyping complex variants from short reads and low coverage sequencing data, and to the challenge of performing traditional association tests at highly multiallelic loci. STRs have been implicated in dozens of Mendelian diseases such as Huntington's disease and Fragile X Syndrome and have been shown to affect human phenotypes through a variety of mechanisms. Recent evidence suggests an important role of STRs in gene regulation. For instance, a genome-wide survey of the effects of STRs on gene expression identified more than 2,000 STRs whose lengths were linearly associated with expression of nearby genes (eSTRs). These STRs explained 10-15% of gene expression heritability due to cis variants, suggesting an important role for STRs in complex traits. Previous studies of the contribution of STRs to gene expression variability faced several important limitations. First, short read lengths and low coverage sequencing resulted in low quality genotypes, reducing power to detect true associations. Second, analyses were restricted to a single cell type, limiting the biological insight and generalizability of the results in different contexts.Here we look at impact of STRs on gene expression across multiple tissue types. First, we used HipSTR, a new and accurate STR profiling algorithm, to genotype STRs in 147 high coverage (30x) whole genome sequencing datasets from the Genotype-Tissue Expression (GTEx) project. We performed association tests relating STR lengths and expression levels of nearby genes, and identified 2,448 genes with eSTRs across seven tissues at 5% FDR; including 694 genes with eSTRs in two or more tissues. Finally, using the extensive GTEx resource we fine-map the contribution of different variant classes to gene expression, identify putative causal regulatory variants overlapping GWAS loci, and characterize tissue-specific features of predicted causal eSTRs. Overall, our analysis provides a valuable resource for future efforts to integrate STRs into complex trait analyses.

425T

The effect and mechanism of inhibiting G6PD activity on the proliferation of *Plasmodium falciparum*. *Z. Zhang[1,2], W. Jiang[2], X. Chen[3], C. Jiang[2], Z. Fang[2], Y. Feng[2].* 1) Reproductive Medicine Research Center, Sixth Affiliated Hospital, Sun Yat-Sen University, Guangzhou, China; 2) Department of Medical Genetics, Zhongshan Medicine School, Sun Yat-sen University, Guangzhou, China; 3) Department of Endocrinology, Guangzhou Women and Children's Medical Center, Guangzhou, China.

Glucose-6-phosphate dehydrogenase deficiency (G6PD deficiency, OMIM 305900), an X-linked inherited disease, is the most common human enzyme deficiency which affects over 400 million people worldwide. This disease often occurs in malaria endemic areas. The global distribution of G6PD deficiency was highly consistent with the geographic distribution of malaria, and the correlation between them have been confirmed by epidemiology, but the effect and mechanism were not clear. Therefore, we investigated the effect and mechanism of G6PD deficiency against malaria, hope to reveal the anti-malaria selection mechanism. Dehydroepiandrosterone (DHEA) as a potent uncompetitive inhibitor of mammalian gluG6PD was utilized for inhibiting the G6PD activity of erythrocytes. Quantitative real-time PCR were used for the detection and quantification of *Plasmodium falciparum* infection. A transmission electron microscope was used to observe the structural changes of *P. falciparum*. We found that with the increase of DHEA concentrations, the proliferation of *P. falciparum* decreased by qPCR. We further observed the cell nucleus of *P. falciparum* shrank, the cell organelles and metabolites were reduced gradually by transmission electron microscope. Because the *P. falciparum* in erythrocytic stage need to synthesize large amounts of nucleic acid to complete proliferation. Therefore, *P. falciparum* multiplication would be inhibited in G6PD deficient erythrocytes because the *P. falciparum* organelles could not obtain enough electron source (NADPH) and the raw material(ribose-5-phosphate). So, the cell nucleus of *P. falciparum* shrank, the cell organelles and metabolites were reduced gradually and eventually died. These results will play an important role in understanding the mechanism of antimalarial resistance selection of G6PD deficiency.

426F

Integration of Hi-C chromatin loop calls across multiple resolutions identifies loops that are consistent across cell types and functionally associated. *H. Li[1], WW. Greenwald[2], P. Benaglio[3], A. Schmitt[4], Y. Qiu[2], B. Ren[1,5,6,7], M. D'Antonio[1], EN. Smith[3], KA. Frazer[1,3].* 1) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 2) Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA; 3) Department of Pediatrics and Rady Children's hospital, University of California San Diego, La Jolla, CA; 4) Arima Genomics, San Diego, CA; 5) Ludwig Institute for Cancer Research, La Jolla, CA; 6) Moores Cancer Center, University of California San Diego, La Jolla, CA; 7) Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA.

Hi-C allows for the characterization of 3D chromatin architecture by identifying regions of the genome that spatially colocalize through chromatin looping. However, identification of high confidence chromatin loops is challenging as the sparseness of data requires binning the genome to identify statistically significant regions, and loop calling algorithms are sensitive to their bin size parameter. Here, we investigated if integrating loop calls across multiple resolutions from two algorithms results in larger and better-quality loop sets that show consistency between cell types and are enriched for functional ROADMAP chromatin states. We performed in-situ Hi-C on paired iPSC and iPSC-derived cardiomyocyte (iPSC-CM) lines, generated ~6 billion reads, and constructed 2kb resolution contact maps for each cell type. We called loops at 21 resolutions using Fit-Hi-C (10-30 restriction enzyme fragments) and HICCUPS (5kb-25kb), and identified a total of 50,612 distinct Fit-Hi-C and 17,782 distinct HICCUPS loops in iPSC, and 61,315 distinct Fit-Hi-C and 20,609 distinct HICCUPS loops in iPSC-CMs, many of which were only called at one resolution. Visual inspection of Hi-C contact frequency heatmaps indicated loops unique to a single resolution were often not supported by the underlying data, whereas loops present in at least 7 Fit-Hi-C or 3 HICCUPS resolutions tended to be supported, suggesting that aggregate loop sets are more accurate than single resolution loop sets. We therefore merged and filtered the loops based on the above aggregation criteria, resulting in 13,314 Fit-Hi-C loops and 12,362 HICCUPS loops in iPSC, and 13,118 Fit-Hi-C loops and 15,138 HICCUPS loops in iPSC-CM. As chromatin loops are often conserved across multiple cell types, we calculated the proportion of loops shared between iPSC and iPSC-CM for each single-resolution loop set and for the aggregate loop sets. We found that the aggregate loop sets contained considerably more loops than the single resolution loop sets (up to 2.3 times more) and that they were more consistent between the two cell types than any single resolution set (up to 12% more). Additionally, we observed that the aggregate loops contained more functional chromatin state bases than the average across the single resolution sets. These results suggest that the integration of chromatin loops at multiple resolutions is an effective method to identify the greatest number of functionally relevant chromatin interaction intervals.

427W

Moving into the darkness: Improving variant analysis with linked-reads.
A.W. Xu, S. Garcia, S. Williams, C. Catalanotti, M. Schnall-Levin, D.M. Church, P. Marks. 10x Genomics, Pleasanton, CA., United States.

   Structural variants (SVs) are important contributors to disease, but are notoriously difficult to detect using short read (SR) data. Balanced and complex events are particularly challenging because they do not result in variation of read coverage and can span long genomic distances, making them intractable with SR data alone. Targeted approaches pose additional challenges as stochastic and systematic coverage variation between samples and probes can resemble single exon deletion or duplication events and yield false diagnostic results.   To address these shortcomings, we partition limiting amounts of high molecular weight DNA such that unique barcodes can be added as part of library generation. This approach allows us to couple long-range information with high-throughput, accurate SR sequencing, generating a data type known as Linked-Reads. Our reference-based pipeline, Long Ranger™ , leverages the unique properties of Linked-Reads to call a broader range of variant types. Linking heterozygous variants to distinct barcodes enables haplotype reconstruction, providing increased power for variant calling across all variant types, but particularly for balanced events and single exon deletions and duplications. Algorithms assessing barcode overlap between distant regions enable large balanced and unbalanced SV detection. Assessing barcode coverage, rather than read coverage, provides increased sensitivity for exome/targeted applications, obviating the need for a pool of samples to detect coverage variations- the unaffected haplotype serves as the internal sample control. Using NA12878 as a truth sample, we demonstrate the ability to call and haplotype challenging variants including long insertions, heterozygous deletions, and SNPs on distinct paralogous loci. For full depth (median 30x coverage) whole genome sequencing (WGS), the PPVs/sensitivities are 1/0.94 for large deletions (>=30kb) and 0.8/0.86 for short deletions (50bp - 30kb). For exome sequencing (median 75X coverage), these values are 0.73/0.72 for short heterozygous deletions and 0.83/1 for short homozygous deletions. For lower coverage (1-10x coverage) WGS, all large structural events, including balanced and unbalanced events (>=200kb), were detected across multiple samples. Linked-Reads are critical for enabling the analysis and discovery of SVs in the human genome, particularly for cost-effective targeted exome sequencing approaches that canonically are unable to detect structural variation.

428T

Detection of rare copy number variants (CNV) in Kaiser cohort using DNA microarray data from Affymetrix axiom array. *T. Haldar, K.N. Mark, T.J. Hoffmann, N. Risch.* Institute for Human Genetics, University of California, San Francisco (UCSF).

   Affymetrix axiom platform is not optimized for copy number variant (CNV) detection: the amplification step is believed to reduce difference in total intensity due to copy number changes. Moreover, a majority of the SNPs within a common CNV region deviate from Hardy Weinberg equilibrium (HWE) and are removed from the array. We explored the SNP content of the Affymetrix axiom array customized for European population and observed that the number of SNPs typed in the chromosomal regions containing common deletions are significantly smaller compared to other regions. As expected, most of the SNPs typed within the chromosomal regions containing common deletions significantly deviate from HWE and in general, an additional cluster consisting of off-target variants (OTV) is clearly visible in the log-intensity plot of these SNPs. We analyzed the data of 69,508 European individuals in the Kaiser GERA cohort who were genotyped on the Affymetrix axiom array customized for European population. Using the exome sequence data of 194 individuals, we found that these OTV clusters correspond to the individuals with null copy number at a locus. So it is evident that some CNV signals are present in the Affymetrix axiom data. Even though it may be difficult to identify single deletions in a common CNV region due to small number of typed SNPs, we hypothesize that the axiom data can be used to detect long rare CNVs. We observed in our data that the presence of a single long deletion at a locus generally results in a long run of homozygosity (ROH) by the design of the Affymetrix genotype calling algorithm. So we propose a new approach combining the ROH and intensity data to identify long rare deletions. However several confounding factors should be adjusted while using the intensity data, e.g., variation across chips, waving pattern of intensity with GC content, length of deletions, etc. Using this method, we have identified a large number of CNVs which include a significant number of deletions at 10q21.3 (a known hotspot for long rare deletions). We did not observe any significant difference in the number of rare deletions detected among males and females. There is also no significant change in rare CNV loads with age.

429F

Genomic and structural integrity of human induced pluripotent stem cells. *K. Kanchan[1], C. Malley[1], L.R. Yanek[1], L. Cheng[1], Z. Wang[1], D. Becker[1], L. Becker[1], I. Ruczinski[2], R.A. Mathias[1].* 1) Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Biostatistics, Bloomberg School of Public health, Johns Hopkins University, Baltimore, MD.

Human induced pluripotent stem cells (hiPSCs) constitute model systems with enormous capability to shed light on the pathogenesis and mechanisms for a wide range of human diseases. Their application as disease models depends on the genomic integrity and stability through their generation. Recent studies confirm various genomic instabilities, such as, chromosomal aberrations, copy number variations (CNVs) and differences in single nucleotide polymorphisms (SNPs) in hiPSCs. We analysed genome-wide genotype array data (Illumina Infinium MEGA Chip) on 1,321,228 genetic variants to evaluate genomic instability of 132 hiPSC lines generated from their matched peripheral blood mononuclear cells using non-integrating episomal vectors. We observed very low levels of genotype mismatch between genetic variants available on the genotype array between the 132 donor and hiPSC pairs (mean mismatch error rate = 0.004%). Processed SNP array data were subjected to automated CNV calling followed by manual inspection of all CNVs. For automated analysis, we applied Hidden Markov Models on the genotype data to identify CNVs in the samples using R package '*Vanilla*ICE'. Given prior documentation of high rates of false positive CNV calls through automated pipelines, sub-chromosomal plots were generated from the log R ratios and B-allele frequencies for manual inspection and confirmation of the automated calls. Our automated pipeline detected no differences in average numbers of CNVs between the hiPSC and donor DNA. There was an average of 7 deletions and 3 amplifications per hiPSC line, and 6 deletions and 3 amplifications per donor sample. All called CNVs were then compared within the hiPSC and donor pair to identify non-overlapping CNVs (i.e. those called in the hiPSC but not the donor, or those that overlap but were of differing lengths between the hiPSC and donor). Manual inspection of the non-overlapping CNVs revealed that most pairs (N=124) had no detectable difference in CNVs called in the hiPSC from those in the donor DNA. Eight hiPSC lines had 1 or more differences compared to the donor: 6 lines had 1 deletion, one line had 2, and one had 3 deletions. The deletions ranged in size from 3.8-181.2kb, with an average size of 35.5kb. Our results suggest that much of the called CNVs in the hiPSC are pre-existing in the donor DNA. Further, there are only low levels of genomic instability in our lines as quantified by CNV *differences* between the hiPSC and its paired donor DNA.

430W

FusorSV: An algorithm for optimally combining data from multiple structural variation detection methods. *A. Malhotra[1], T. Becker[2], W.P. Lee[1], J. Leone[1], Q. Zhu[1], C. Zhang[1], J. Sargent[1], K. Shanker[1], A. Mil-Homens[1], E. Cerveira[1], M. Ryan[1], J. Cha[1], F.C.P Navarro[3], T. Galeev[3], M. Gerstein[3], R. Mills[4,5], DG. Shin[2], C. Lee[1].* 1) Jackson Laboratory for Genome Medicine, Farmington, CT; 2) Department of Computer Science and Engineering, University of Connecticut, Storrs, CT; 3) Yale University Medical School, Computational Biology and Bioinformatics Program, New Haven, CT; 4) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 5) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Comprehensive and accurate discovery of Structural Variations (SVs) from next generation sequencing data remains a major challenge. Existing popular approaches to overcome performance limitations of any one SV identification method are to use multiple complementary methods to determine the SV loci and naively merge them based on coordinates with no consideration of their individual biases. These approaches do not take into account the strengths and weaknesses of individual method and hence either under or over merge the variant loci resulting in missing and/or false SV calls. Here, we present FusorSV, an open source algorithmic framework that uses a truth set to train a fusion model by comparing calls from different SV calling algorithms to the truth set. FusorSV uses discriminating features (such as SV Type and Size) to promote subsets of algorithms that are complementary to each other. For example, DELLY and LUMPY appear similar in methodology and produce very similar deletion calls but diverge significantly for inversion calls. Hence, we empirically determine the similarity/dissimilarity of each combination of SV calling algorithms for each SV type. Integration of individual algorithm performance, along with the pairwise similarity/dissimilarity of algorithms across SV types, allows FusorSV to select subsets of algorithms for each SV type that are more comprehensive and balanced (maximum discovery with minimal false positives) in nature. We built a SV fusion model using an ensemble of nine SV calling algorithms to analyze 27 deep coverage (50X) human genomes that were investigated as part of the 1000 Genomes Project (1000GP). Our model identified an additional 562 (~10%) novel SV calls that were not reported in these 27 samples by the 1000GP. Experimental validation of a subset of these novel SV calls yielded a validation rate of 86.15%. These results highlight the power of the FusorSV approach and we believe that it has the potential to become a new gold standard for SV calling from short read whole genome sequencing datasets.

431T

Pervasive transcriptional dosage compensation buffers impact of autosomal structural variation. *R.C. McCoy, J.M. Akey.* Department of Genome Sciences, University of Washington, Seattle, WA.

Structural variation (SV) is known to contribute to human disease through its impact on gene dosage. Yet population-scale maps of SV have revealed that many genomic regions can be duplicated or deleted with no apparent phenotypic consequences. We sought to reconcile these observations by characterizing the genome-wide prevalence and molecular mechanisms of autosomal dosage compensation. To this end, we developed a generalized linear mixed model approach to combine expression data from multiple individuals and tissues to systematically test hypotheses about dosage compensation at 1,906 autosomal genes fully encompassed by an SV. We applied this approach to data from the GTEx Consortium, comprising SV detected in whole-genome sequences from 147 individuals and corresponding RNA-seq data from a total of 53 tissues (median 15 tissues per individual). We found that both duplicated and deleted genes show strong evidence of dosage compensation, with expression distributions significantly deviating from dosage-proportionate expectations in the direction of diploid expression levels. Substantially more deleted (34.9%) than duplicated (9.5%) genes exhibited dosage compensation (10% FDR), consistent with the known predominance of haploinsufficiency over triplosensitivity. Interestingly, dosage compensation was extremely common among highly expressed deleted genes (69.3%), but extremely rare among highly expressed duplicated genes (4.5%). These findings suggest that directionally-biased buffering mechanisms have evolved under stabilizing selection to maintain high or low gene expression. Furthermore, dosage compensation of duplications, but not deletions, varied substantially according to gene class, with duplicated short noncoding RNAs showing stronger dosage compensation than protein-coding, long noncoding, and pseudogenes (β = -0.332; P = 9.15 x 10^{-5}). This result suggests that low-expressed regulatory RNAs may be under strong expression constraint. Intriguingly, duplicated protein-coding genes with greater density of miRNA binding sites exhibited stronger dosage compensation (β = -0.0957; P = 9.56 x 10^{-5}), indicating that miRNA may mediate dosage compensation via mRNA degradation of duplicated genes. Together, our results reveal functional and evolutionary characteristics of mechanisms buffering the transcriptional effects of autosomal dosage alteration, with implications for dissecting the causal basis of pathogenic SV.

432F

Germline *de novo* mutation clusters arise during oocyte aging in genomic regions with increased double-strand break incidence. *C. Gilissen[1], J.M. Goldmann[2], V.B. Seplyarskiy[3,4], W.S.W. Wong[5], T. Vilboux[5], D.L. Bodian[5], B.D. Solomon[5,6,7], J.A. Veltman[1,8], J.F. Deeken[5], J.E. Niederhuber[5,9].* 1) Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, the Netherlands; 3) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 4) Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoi Karetny pereulok 19, Moscow 127994, Russia; 5) Inova Translational Medicine Institute (ITMI), Inova Health Systems, Falls Church, VA, USA; 6) Department of Pediatrics, Inova Children's Hospital, Inova Health System, Falls Church, VA, USA; 7) Department of Pediatrics, Virginia Commonwealth University School of Medicine, 1201 E Marshall St, Richmond, VA, USA; 8) Institute of Genetic Medicine, International Centre for Life, Newcastle University, Newcastle upon Tyne, United Kingdom; 9) Johns Hopkins University School of Medicine, 733 North Broadway Street, Baltimore, MD, USA.

Clustering of mutations has been found both in somatic mutations from cancer genomes and in germline de novo mutations (DNMs). To investigate the underlying mutational mechanisms of clustered DNMs (cDNMs) we used whole-genome sequencing data from 1,291 parent-offspring trios and identified 1,755 cDNMs across 799 clusters. By performing read-phasing, we successfully identified the parent-of-origin for 700 cDNMs across 400 clusters. In addition, we created a replication dataset of published studies, resulting in a total of 1,643 cDNMs across 745 clusters. We found that although the number of paternal cDNMs did not show a significant correlation with the paternal age (p=0.087), we found a highly significant correlation of maternal cDNMs with maternal age (p<10^{-10}), accounting for 23% (c.i. 7-38%) of the maternal age effect. This maternal age effect of clusters stems mostly from clusters with intra-mutational distances greater than 1kb. Moreover, the maximum number of DNMs in the clusters of an individual, correlates positively with maternal age (p<10^{-10}), but only marginally significant with paternal age (p=0.050). Interestingly, 58.4% of maternal cDNMs localize to chromosomes 8, 9 and 16 (p<10^{-16}, Chi-square test) whereas the number of paternal correlates with chromosome length (R^2=0.72, p=6*10^{-7}). The maternal cDNMs on these three chromosomes occur specifically in regions that are also enriched for maternal unclustered DNMs and their mutation spectrum is strongly enriched for C>G substitutions compared to other maternal cDNMs (bootstrapping p=0.022). Finally, we investigated whether maternal clusters associate with processes involving double-stranded-breaks (DSBs). We found that sites with meiotic gene conversions (MGC) co-localize with maternal cDNMs significantly more often than expected (p<0.004, bootstrapping). In addition, we identified 45 *de novo* CNVs, of which 5 have a total of 17 DNMs within 100kb flanking the breakpoints. Exactly 15 of these 17 DNMs are cDNMs (p<2.2*10^{-16}, Fisher's exact test). All 5 CNVs and all phased cDNMs arose on the maternal allele (p=0.03, Fisher's exact test). Lastly, we found that maternal recombination scores at maternal clusters are significantly higher than paternal recombination scores at paternal cluster locations (p=0.004, bootstrapping) Overall these findings suggest accumulation of DSB-induced mutations throughout oocyte aging as an underlying mechanism leading to maternal mutation clusters.

433W

Linked-reads for high resolution individual genome analysis via haplotype reconstruction. *S. Williams, S. Garcia, C. Catalanotti, A.W. Xu, P. Marks, M. Schnall-Levin, D.M. Church.* 10x Genomics, Pleasanton, CA.

Next-generation sequencing is a standard tool for identifying the molecular underpinnings of disease. This has been driven by cost-efficient and accurate short reads which lack long-range information, limiting long-range haplotype reconstruction. Some regions of the genome are inaccessible to short reads and many variant types, particularly structural variants, are missed with standard approaches. Efficient and effective interpretation of these genomes has not been realized and the diagnostic yield for both genome and exome approaches is lower than anticipated. To advance the utility of NGS sequencing assays, we developed a technology that retains long-range information while maintaining the power, accuracy, and scalability of short read sequencing. At its core, haplotype-level dilution of long input molecules into >1 million barcoded partitions creates a novel data type called 'Linked-Reads' that enables high-resolution genome analysis with minimal DNA input (~1 ng). Using Linked-Reads, we show improved performance in regions of the genome typically inaccessible with traditional short read technology due to the presence of paralogous sequence/highly repetitive regions. Highly homologous sequences which traditionally lead to ambiguous mapping can now be analyzed when associated with distinct barcodes. We rescued ~50 Mb of previously inaccessible sequence, enabling the identification of tens of thousands of novel single nucleotide variants (SNV) per genome. Clinically-relevant variants were identified in genes within these previously inaccessible regions including CYP2D6, SMN1, and STRC. Including phase information in consequence prediction not only disambiguates local variation such as compound and exon spanning variation, but because of the long distance phase information captured by Linked-Reads we are able to decipher the true impact of genetic variation at megabase range. Next, we tested variant effect prediction tools tuned for phased genotypes which have shown promise in local variation but currently struggle to fully take advantage of data which phases variation at great distance. Improvements in this area will be critical for improved genome interpretation.Linked-Reads, coupled with improved variant classification tools has the potential to be a first line assay for identification, classification, and clinical interpretation of how genetics impact the pathology of disease.

434T

Characterisation and genotyping structural variation at the malaria-associated human glycophorin A-B-E cluster. *W. Algady[1], P. Brajer[1], D. Mateja[1], S. Gomes[2], E. Weyell[1], F. Yang[2], E. Hollox[1].* 1) Department of Genetics and Genome Biology, University of Leicester, Leicester, UK; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

Human glycophorins A and B are proteins expressed on the surface of erythrocytes, and are receptors for invasion of the malaria-causing *Plasmodium falciparum* parasite. The proteins are encoded by the genes *GYPA* and *GYPB* which, together with *GYPE*, reside on a tandemly-duplicated repeat region on chromosome 4q31.21. Recent genomewide analysis has suggested that a structural variant within this region (DUP4) is common is East Africans and is strongly protective against severe malaria. DUP4 and other structural variants can be genotyped by detailed analysis of whole genome Illumina sequencing, but a suite of rapid, cost-effective assays to reliably genotype structural variants on small degraded DNA samples is lacking. Here, we use the paralogue ratio test (PRT), a comparative PCR method which has been shown to be more robust that real-time quantitative PCR for typing genomic CNV, and has been previously used on a wide variety of CNV loci. It requires <20ng genomic DNA and has a low cost per sample, making it ideal for genotyping large DNA cohorts to test for association with malaria severity, for example. We also present detailed characterisation of structural variation across this region using fiber-FISH, long PCR and sequencing. This allows exact mapping of structural variant breakpoints and development of junction-fragment PCRs for particular structural variants, using paralog-specific PCR based on linked nucleic acids. These junction-fragment PCRs enable validation of PRT results, or rapid genotyping across a very large cohort for a particular structural variant, in, for example, basic molecular biology labs. These assays can be used in budget-limited settings for population-based screening of malaria susceptibility variants at the glycophorin locus.

435F

Low coverage sequencing of inbred animal backcrosses to check and correct genome assemblies. *G.W. Nelson[1], C. McIntosh[2], J. Lautenberger[2], S. Burkett[3], V. David[4], A.O. Perantoni[5].* 1) Leidos Biomedical Research, Frederick National Laboratory, Frederick, MD; 2) Genomics Analysis Unit, CCR, NCI, Frederick, MD 21702; 3) Molecular Cytogenetics Core Facility (MCGP), CCR, NCI, Frederick, MD 21702; 4) Basic Research Laboratory, CCR, NCI, Frederick, MD 21702; 5) 5Cancer and Developmental Biology Laboratory, CCR, NCI, Frederick, MD 21702.

Different inbred animal strains of the same species are distinguishable by essentially fixed SNP differences. To map a locus for the Noble rat's susceptibility to a nephroblastoma analogous to the human Wilms tumor we sequenced 71 rats, Noble-Fischer hybrid backcrossed to Fischer, at an average of 2x coverage. This provided calls for ~70% of the ~3 million SNP differences between these strains, yielding a recombination map at ~2 Kbp resolution. This resolution revealed genomic regions with sharply defined boundaries within which the genomic ancestry of multiple rats was uncorrelated with the genomic ancestry of the surrounded regions, suggesting errors in the RN6 reference sequence. Of these we have clearly delineated 20, with sizes ranging from 120 Kbp to 4.4 Mbp. Comparison of the ancestry within the region to ancestry across the genome indicated, with high statistical confidence, correct genomic locations for 7 of these. We performed fluorescence in situ hybridization (FISH) to verify the proposed remapping of two of these regions. We designed probes to hybridize to reference sequence inside the anomalous regions, and with a different color label to hybridize to reference sequence in regions flanking these regions. If the RN6 assembly were correct, the probes would colocalize. The FISH results were inconsistent with the mapping to RN6, but consistent with the proposed corrections. For a region mapped to the reference sequence near the center of chr. 2, but predicted close to the chromosome end, the flanking probes appeared at the center of the chromosome, but the probes for the anomalous region localized to an end of the chromosome. For a region mapped to the reference sequence near the start of chr. 8, but predicted to occur on chr. 1, the flanking probes appeared at an end of a small chromosome, but the probes for the anomalous region hybridized to the largest chromosome. We have used the recombination map to assign locations to unmapped RN6 fragments. Subtracting chr. Y fragments, there are 240 fragments carrying informative SNPs, of which we can assign definite locations to 81. Demonstrating the power of this approach, all fragments with identified chromosome mapped to the named chromosome, and we have assigned locations to 32 fragments with unknown chromosome. Low coverage whole genome sequencing of backcrosses to one parental strain, of hybrids of two strains, are a cost-effective way to obtain accurate recombination maps. .

436W

The mechanism and function of targeting lincRNAs by NMD in mammals. *L. Hu[1], X. Kong[1], Z. Zhang[2], M. Jin[1].* 1) Inst Hlth Sci , Shanghai, Shanghai, China; 2) Department of Biology, University of Rochester, Rochester, New York 14627, United States of America.

Thousands of long noncoding RNAs (lncRNAs) have been identified in a mammalian genome, but their functions remain elusive. Recent studies showed that lncRNAs' expression is affected by nonsense-mediated mRNA decay (NMD), providing an avenue to infer lncRNAs' functions because RNA stability was reportedly correlated with functions. In this study, we evaluate the properties of human intergenic lncRNAs, i.e. lincRNAs, in order to answer two questions: how can NMD affect lincRNAs' expression given that NMD relies on translation while lincRNAs are generally untranslated, and what is the function of NMD-targeted lincRNAs? First, we find that 11.27% of lincRNAs are affected by NMD. Second, we provide evidence for that NMD can directly target lincRNAs, including bound by the NMD factor UPF1 and exon-exon junction complex (EJC), cleaved by the NMD factor SMG6, and scanned by ribosomes. Third, we find that NMD-affected lincRNAs are more lowly expressed and less conserved than the other lincRNAs, suggesting that NMD-affected lincRNAs are more likely expression noise than functional products. Notably, all these properties resemble those of NMD-targeted mRNAs. Taken together, lincRNAs can be directly targeted by NMD, and such targeted lincRNAs are probably noise stemming from erroneous or leaky transcription.

437T

Human germline mutation hotspots are characterized by a transversion-rich mutation signature. *J. Carlson[1], J.Z. Li[1,2], S. Zöllner[3,4], The BRIDGES Consortium.* 1) Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Human Genetics, University of Michigan, Ann Arbor, MI; 3) Biostatistics, University of Michigan, Ann Arbor, MI; 4) Psychiatry, University of Michigan, Ann Arbor, MI.

   The germline mutation rate in humans is highly variable across the genome, and is influenced by many factors including meiotic recombination rate, replication timing, and the nucleotide sequence flanking each mutation site. Despite these well-documented effects, the factors that contribute to the spatial heterogeneity of mutation rates are still largely unknown. Previous studies have applied parametric models to identify potential drivers of spatial mutation rate heterogeneity, but these models are subject to specification bias and measurement error among the considered features. Moreover, such models are generally trained on small datasets of *de novo* mutations, rendering them underpowered and limited in scope. To overcome these constraints, we present a nonparametric approach which leverages a high-resolution dataset of ~36 million extremely rare variants observed in a sample of 3,716 whole-genome sequences. With these data, we apply non-negative matrix factorization to summarize spatial mutation rate heterogeneity as a composite of three distinct mutation signatures. We show that this model captures cryptic patterns of spatial heterogeneity among *de novo* mutations with greater accuracy than previously proposed parametric models. The clear majority of the genome (94.5% of 1-megabase windows) is dominated by a balanced contribution of the three mutation signatures, suggesting that broad regional perturbations to the germline mutation rate are generally uncommon. The remaining 5.5% of the genome, however, is characterized by an exceptionally high concentration of transversion mutations, with transversion rates per 1Mb window up to 6 times higher than the genome-wide average. This transversion-rich mutation signature occurs predominantly in subtelomeric regions of chromosomes 7, 8, 9, and 16. Importantly, these loci overlap known mutation hotspots (e.g., 8p21-23), suggesting that regimes of germline hypermutability are not simply an artifact of uniform variability in mutation rates but rather correspond to asymmetric shifts in the frequencies of point mutation types. Accounting for a set of known mutagenic genomic features explains only a fraction of the observed excess of transversion mutations in these regions, implying that additional causative mechanisms have yet to be identified. Our findings thus warrant further investigation into the relationship between the germline mutation rate and corresponding features of the genomic landscape.

438F

Predicting hotspots of *Alu/Alu*-mediated rearrangements in the human genome. *X. Song[1], C.R. Beck[1], R. Du[1], I.M. Campbell[1], Z.C. Akdemir[1], S. Gu[1], A.M. Breman[1,2], C.A. Shaw[1,2], J.R. Lupski[1,3,4,5].* 1) Molecular and Human Genetics, BCM, Houston, TX; 2) Baylor Genetics, Houston, TX; 3) Department of Pediatrics, BCM, Houston, TX; 4) Human Genome Sequencing Center, BCM, Houston, TX; 5) Texas Children's Hospital, Houston, TX.

   The short interspersed element *Alu* comprises ~11% of the human genome and numbers >1 million copies per haploid genome. *Alu*s are enriched at the breakpoint junctions of genomic rearrangements, and can mediate the formation of copy number variants (CNVs) between pairs. These *Alu/Alu*-mediated rearrangements (AAMR) can cause Mendelian diseases including spastic paraplegia 4 and Fanconi anemia. Further understanding of the genome-wide impact of AAMR on variation and human health is limited by current experimental technologies. We have developed a bioinformatic approach to investigate characteristics of *Alu*s involved in mediating CNVs (CNV-*Alu*s) and have predicted genome-wide AAMR hotspot loci. First, we collected and analyzed 219 CNV-*Alu* pairs with experimental evidence supporting their role in deletion formation. Second, we characterized *Alu* sequence features and genomic architectural features enriched in CNV-*Alu*s relative to locus-specific control *Alu*s. CNV-*Alu*s tend to be longer, have better pairwise alignment, contain internal pol III promoters that are more likely to be active, and have higher GC percentages and a longer poly-A tail. Next, we developed a classification model to predict CNV-*Alu* pairs based on these distinct properties. The test dataset was built by choosing *Alu* pairs that are in the same orientation, span at least one exon, and are located <250 Kb from each other. As a result, ~18% of the 78 million *Alu* pairs in the test dataset were predicted as high risk for AAMR. We further determined the relative risk of AAMR in 12,075 human genes using the density of predicted CNV-*Alu* pairs. Finally, we experimentally validated this prediction with 89 samples selected by correlating hotspot genes with a database of ~54,000 chromosomal microarrays (CMA). We fine-mapped 47 duplications, 40 deletions and 2 complex rearrangements with a custom designed array comparative genomic hybridization (aCGH). We then mapped 50 of these breakpoint junctions at nucleotide-level resolution with PCR and Sanger sequencing. 74% of the junctions are mediated by *Alu* pairs, 20% have one end in an *Alu* and only 6% are mediated by non-*Alu* sequences. Remarkably, ~80% of the *Alu* pairs were predicted by our classifier as being at high risk for AAMR. These results demonstrate the utility of our predictive model, give insights into the genomic features and mechanisms underlying AAMR, and provide a general framework for improved assessment of AAMR and its role in human disease.

439W

Chromosomal microarray analysis in fetuses with congenital heart disease: 5 years of clinical experience. *P. Hu, Y. Wang, ZF. Xu.* Obstetrics and Gynecology Hospital Affiliated to Nanjing Medical University, Nanjing, Jiangsu, China.

Purpose: Currently, chromosomal microarray analysis (CMA) is considered as the first-tier test in pediatric care and prenatal diagnosis. However, the diagnostic yield of CMA for prenatal diagnosis of congenital heart disease (CHD) has not been evaluated based on a large cohort. Our aim was to evaluate the clinical utility of CMA as the first-tier test for chromosomal abnormalities in fetuses with CHD. **Methods:** In this prospective study, 602 prenatal cases of CHD were investigated using SNP array over a 5-year period. **Results**: Overall, pathogenic chromosomal abnormalities were identified in 130 (21.6%) cases, with 50% of them being numerical chromosomal abnormalities. The detection rates of likely pathogenic copy number variations (CNVs) and variants of uncertain significance were 2.8% and 5.1%, respectively. The detection rate in CHD plus extra cardiac anomalies group was significantly higher than that in isolated CHD group (48.9% vs. 15.4%, P < 0.05). There was no significant difference between the detection rates in CHD plus nonstructural anomalies and isolated CHD groups (21.8% vs. 15.4%). The detection rate in fetuses with CHD plus mild ventriculomegaly was significantly higher than those with other types of soft markers (50.0% vs. 15.6%, P < 0.05). Notably, three recurrent likely pathogenic CNVs were identified, including microduplications at 18p11.23, 15q11.2 and 22q11.22, of which, 18p11.23 duplication involving *PTPRM* was likely to be the candidate locus for CHD. Our study suggests CMA is a reliable and high-resolution technology and should be used as the first-tier test for prenatal diagnosis of CHD in clinical practice.

440T

Do fetal endothelial nitric oxide synthase (eNOS) gene haplotypes influence prolonged preterm rupture of fetal membranes (PPROM)? *K. Yanamandra[1], D. Napper[1], J.A. Bocchini Jr.[1], R. Dhanireddy[2].* 1) Pediatrics, LSU Health Science Center, Shreveport, LA; 2) Neonatology section, UT Health Science Center, Memphis, TN.

PPROM is a multifactorial etiology leading to growth restriction of the fetus in utero. It could risk severe prematurity, small for gestation, and fetal distress leading to respiratory failure after birth. Animal studies suggest that NO in placental vasculature is required to facilitate fetal growth. Earlier studies showed that preterm infants with severe respiratory failure respond favorably to inhaled nitric oxide. These infants may have a deficiency in the inflammatory response, including a defect in nitric oxide generation in airspaces. We have previously reported that the mutant eNOS genotypes, leading to decreased Nitric oxide, increased the risk of respiratory distress, choreo-amnionitis, and IUGR, in premature infants. To study the potential association of infant eNOS genotypes and PPROM, we analyzed eNOS gene haplotypes (T786C, G894T, intron 4 VNTR b/a genotypes), in 161 preterm infants (62 PPROM and 99 Non-PPROM controls), using microplate PCR genotyping methods. Our data on eNOS double marker (T786C and G894T) analysis revealed that PPROM infants showed 31% mutant C/T haplotypes compared to 16% without PPROM history among the Caucasians. Our eNOS triple marker (T786C, G894T, and intron 4 VNTR b/a) analysis in Caucasians infants with PPROM showed 27% vs. those without PPROM history of 15%. Among the African American (AA) infants our analysis revealed 11% vs. 6%; 19% vs. 13% in PPROM infants compared to infants without PPROM history in double and triple marker analyses respectively. Table: eNOS gene haplotypes in Caucasian and AA PPROM & Non-PPROM premature infants.

eNOS Diplo & Haplotype	PPROM Caucasians	Freq	Non-PPROM Caucasians	Freq	OR pValue Caucasians	PPROM AA	Freq	Non-PPROM AA	Freq	OR (95% CI) pValue (AA)
CC-TT	1	0.06	2	0.02		0	0	0	0	
TC-GT	8	0.5	23	0.27		24	0.22	14	0.12	
TT-GG	7	0.44	61	0.71		86	0.78	102	0.88	
C-T	10	0.31	27	0.16	2.5(1-6.2)0.036	24	0.11	14	0.06	1.91(0.9-4)0.06
T-G	22	0.69	145	0.84		196	0.89	218	0.94	
CC-aa-TT	1	0.04	3	0.02		6	0.04	3	0.02	
TC-ba-GT	11	0.46	31	0.24		51	0.31	42	0.24	
TT-bb-GG	12	0.5	94	0.73		106	0.65	128	0.74	
C-a-T	13	0.27	37	0.15	2.2(1-4.8)0.03	63	0.19	44	0.13	1.6(1-2.5)0.027
T-b-G	35	0.73	219	0.85		263	0.81	294	0.87	

Our data suggest a significant association with the eNOS mutant haplotype only with triple marker (C-a-T) in both Caucasians and AA infants, but not with double marker in AA. Infants with PPROM have higher frequencies of mutant haplotypes, reducing the endogenous NO; this is the first report to suggest the influence of eNOS gene haplotypes in the etiology of PPROM in premature newborn infants.

441F

Alteration of the expression and methylation modification of RAS in cardiomuscular tissue of the mice conceived by in-vitro fertilization. *Q. Wang[1,2], F. Zhang[1,2], Y. Zhang[1,2], Y. Lou[1,2], L. Wang[1,2], M. Yuan[1,2], M. Hu[1,2], F. Jin[1,2].* 1) Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou,Zhejiang, Ch, Hangzhou, Zhejiang , China; 2) Key Laboratory of Reproductive Genetics, National Ministry of Education (Zhejiang University), Women's ReproductiveHealth Laboratory of Zhejiang Province, Hangzhou, Zhejiang, China.

Epidemiological studies have found that newborns conceived by in vitro fertilization (IVF) had an increased risk of cardiovascular malformations and cardiovascular dysfunction, suggesting that IVF may disturb the process of embryonic and fetal development. However, the exact mechanisms have not been investigated and its long-term effects on the health of IVF offspring remain unclear. In the present study, the expressions of renin-angiotensin system (RAS) in the cardiomuscular tissue of IVF mice from childhood to old age have been measured. Meanwhile the methylation levels of CpG island of *Ctgf* and *Col1a1* and the regulatory functions of microRNAs and DNA methyltransferases (*Dnmts*) were investigated. The altered expression levels of COLI/COLIII and RAS-COLI/COLIII related epigenetic regulators were also observed, which indicates that IVF process may cause abnormal expression of RAS-COLI/COLIII in the offspring through microRNA-mediated pathway. The results showed that IVF could cause long-lasting alterations in RAS expression of cardiomuscular tissue of IVF mice.

442W

Ancestral disparities in genetic architecture of life course correlations between early growth and adulthood cardiometabolic disorders. *F. Tekola Ayele[1], T. Workalemahu[1], A. Amare[2].* 1) NICHD, National Institutes of Health (NIH), Bethesda, MD; 2) School of Medicine, University of Adelaide, Adelaide, SA, Australia.

Aberrant fetal growth has been consistently associated with cardiovascular diseases and diabetes in adulthood. Recent genome-wide association studies found genetic loci that influence both growth in early life and cadiometabolic diseases in later life. However, whether and to what extent genetic variations contribute to the enigma of disparities in fetal growth and cardiometabolic diseases among diverse ancestral populations is poorly understood. In the present study, we (1) tested whether the burden of genetic variants that increase the risk of reduced birthweight and cardiometabolic diseases varies by ancestry, (2) evaluated the distributions of the genetic risk alleles with reference to the "common-disease common-variant hypothesis" in diverse ancestral populations, and (3) determined whether genetic selection contributed to enrichment of rare variants associated with increased burden of reduced growth and cardiometabolic disorders. Genotype data were extracted from phase 3 of the 1000 genomes project for 2,504 participants from 26 global populations grouped into five super-populations, namely Europeans, Africans, Admixed Americans, East Asians, and South Asians. Genetic risk burden (GRB) was calculated as the weighted sum of the number of genetic variants associated with increased risk for low birthweight, type 2 diabetes, obesity, and dyslipidemia (n=416 SNPs). We found differences between Europeans and non-Europeans in the allelic architecture and cumulative burden of genetic risk of early growth and cardiometabolic abnormalities. Specifically, GRB was significantly higher in Africans, Admixed Americans, East Asians, and South Asians compared to Europeans for type 2 diabetes, dyslipidemia, and low birthweight, but not for obesity. In all traits, risk genetic variants in Africans and East Asians were enriched for rare alleles ($P<0.001$), displaying the greatest deviation from the expectation of the "common disease-common variant" hypothesis. GRB of ancestral alleles was significantly higher and that of derived alleles was significantly lower in non-Europeans compared with Europeans ($P<0.001$). We also found that negative selection did not explain the relatively higher proportion of rare genetic risk loci. Our findings suggest that genetic variations, interacting with environmental factors, contribute to observed population disparities in the life course correlations between early growth and adulthood cardiometabolic disorders.

443T

PheWAS analysis of 13,000 individuals detects a common genetic variant that tags the Rh D blood group system in the European population. *J. Fadista[1], L. Skotte[1], J. Bybjerg-Grauholm[2], F. Geller[1], H. Boyd[1], H. Hjalgrim[1,3], D.M. Hougaard[2], M. Melbye[1,4,5], B. Feenstra[1].* 1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Danish Centre for Neonatal Screening, Department of Clinical Biochemistry, Immunology and Genetics, Statens Serum Institut, Copenhagen, Denmark; 3) Department of Haematology, Rigshospitalet, Denmark; 4) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 5) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.

The Rh D blood group system is the second most important blood group system after ABO, and besides its role in transfusion medicine, the D antigen is of primary importance in obstetrics due to its causal role in hemolytic disease of the newborn. By using existing genotype data on 3500 mothers of European ancestry (GWAS and HRC imputed) and performing association testing against all their ICD-8 and ICD-10 codes from the Danish National Patient Registry we were able to detect a PheWAS significant (P<5 × 10^{-8}/626 ICD codes) association in the *RHD* gene for the ICD-10 code "Prophylactic Immunotherapy – Z29.1" (OR=33, P=1.5 × 10^{-73}, 272 cases). The top SNP rs72660908 perfectly tags the *RHD* gene deletion as detected by copy number variation analysis. Moreover, for a subset of mothers for which we had blood group information, we confirmed that homozygous GG individuals for rs72660908 were all Rhesus D negative. The fact that the ICD-10 code "Prophylactic Immunotherapy – Z29.1" was always registered within 6 months before the second pregnancy delivery date, further confirms that this ICD-code tags the known anti-D prophylaxis in pregnancy for preventing Rhesus alloimmunisation for Rh-negative women who developed Rh-negative antibodies during their first pregnancies with Rh-positive fetuses. This result was replicated in an independent cohort of 9500 individuals, also GWAS and HRC imputed, from the Michigan Genomes Initiative (P=3.2 × 10^{-18}, <50 cases). In conclusion, by using a common variant that perfectly tags the Rh D blood group system, it is possible to impute this blood group in populations of European ancestry. Rh D status can then be tested against clinically relevant health outcomes.

444F

Compromised DNA repair and genomic imbalances in human male infertility. *V. Singh[1], K. Singh[1], S. Rajender[2].* 1) Molecular and Human Genetics, Banaras Hindu University, Varanasi, India; 2) CSIR-Central Drug Research Institute, Lucknow, India.

Introduction DNA repair mechanisms are important to maintain the fidelity and genomic stability of developing germ cells. Defects in DNA repair during spermatogenesis underlie a major cause of testicular failure. **Objective** To investigate the expression of strong candidate genes of DNA repair pathway and whole genome imbalances in infertile male individuals. **Materials and Methods Study Population** The study was approved by the Institutional Human Ethics Committee, Faculty of Science, Banaras Hindu University, Varanasi, India (Dean/2011-12/119 & Dean/2012-13/546). Informed written consent was obtained from every participant of each group. **Histological and Cytogenetic analysis** Histological analysis of testicular biopsies was performed by routine histopathology stain. Cytogenetic analysis was carried out using standard PHA stimulated whole blood culture followed by karyotyping using standard human chromosome nomenclature. Y-Chromosome deletion mapping using genomic DNA was performed by standard methods. **Expression analysis of DNA repair pathway genes** Relative expression of genes in testicular biopsies of infertile patients (n=4) vs control (n=1) was done using Human DNA Repair RT² *Profiler*™ PCR Array (SABiosciences, Frederick, MD, USA). Data analysis was performed using web-bases automated analysis methods provided by SA Biosciences. Results of PCR array were validated in more number of samples (n=52) using Quantitative real-time PCR analysis of selected genes followed by data analysis using comparative CT Method (ΔΔCT Method). **Genomic imbalances in infertile patients** To identify the genomic imbalances in infertile patients (n=23) vs fertile controls (n=10), we employed CytoScan™ 750K Array (Affymetrix, USA). The data were analyzed using Chromosome Analysis Suite (ChAS) software package (Affymetrix, USA). **Results** Expression analysis of DNA repair pathway genes showed a significant down-regulation of majority of the genes. The cytogenetic array analysis detected a common gain on chromosome 19p13.3 region in 4 (17.3%) cases. The analysis also identified & gt5Mb Loss of Heterozygosity on chromosome 3p21.31 region in 6 (26%) cases. **Conclusion** The results clearly indicate compromised DNA repair machinery and genomic imbalances in infertile patients and opens new horizons for further investigation of the role of these genes in spermatogenesis. **Support** This study was funded by Board of Research in Nuclear Sciences (BRNS), Mumbai, India.

445W

Validation of a novel copy number variant detection algorithm for CFTR from targeted next-generation sequencing data. *K. Kosheleva, K. Robinson, N. Faulkner, M. Umbarger.* Good Start Genetics, Cambridge, MA.

Cystic fibrosis (CF) is a severe, recessive disorder resulting from the inheritance of two null copies of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. While most causative mutations for CF are single nucleotide variants, insertions, and deletions, it is estimated that CFTR copy number variants (CNVs) represent about 1-2% of pathogenic mutations underlying CF. Here, we outline the validation of a new strategy for utilizing targeted NGS data to call copy number variants in the CFTR gene. Specifically, we have developed a read-count based CNV calling algorithm to detect deletions and duplications in the CFTR gene at single-exon resolution. The algorithm employs a log odds ratio statistic derived from the number of reads mapping to each exon, which are first corrected for batch-, sample- and exon-specific sources of variability, to assess the relative probability of different copy number states. To evaluate the performance of the method, we estimated specificity and sensitivity by applying our algorithm to a panel of 25,203 patient DNA samples. This panel included 16 samples identified as containing CFTR CNVs via genotyping or multiplex ligation dependent probe amplification, as well as 140 cell-line-derived controls carrying a deletion of CFTR exons 2 and 3. Next, to evaluate sensitivity across a broader array of CNV sizes and positions, we simulated single- and multi-exon duplications and deletions of varying lengths and assessed the rate at which these simulated CNVs were called via our analysis methodology. Our method correctly identified all 156 samples known to contain CNVs. Among the remaining 25,187 samples, we estimated a sample-level specificity of 99.85% +/- 0.05% and, using simulation data, an average sensitivity of 99.998% +/- 0.001% for deletions (100% whole-gene; 99.988% single-exon) and 99.978% +/- 0.001% for duplications (99.98% whole-gene; 99.97% single-exon). Of note, these statistics include the historically challenging paralogous region of exon 10 (99.97% single-exon sensitivity). In summary, we have developed and validated a CNV calling algorithm that is able to detect single exon to whole gene deletions and duplications at high sensitivity and specificity, thereby further enhancing the clinical sensitivity of our NGS-based CF carrier screening test.

446T

Genetic diagnosis of disorders of sex development (DSD): A national registry for disease-specific precision health. *E.C. Delot[1,2], J.C. Papp[1], H. Barseghyan[1,2], D.E. Sandberg[3], E. Vilain[1,2], The DSD-TRN Genetics Workgroup.* 1) Dept Human Genetics, Gonda Center, UCLA, David Geffen School of Medicine, Los Angeles, CA, USA; 2) Center for Genetic Medicine Research, Children's National Medical Center, 111 Michigan avenue, Washington DC, NW, USA; 3) Dept. Pediatrics & Communicable Diseases & the Child Health Evaluation and Research Center, University of Michigan Medical School, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA.

The 2005 DSD Consensus Conference called for the establishment of an infrastructure for collaborative interdisciplinary clinical practice and research, to integrate scientific understanding of DSD with real-time standardization and improvement in clinical practice. In response, the DSD-Translational Research Network (DSD-TRN) was created, the first such North American infrastructure, with the collaboration of patient advocates convened by Accord Alliance, a network of 4 (now expanded to 12) US research and clinical sites and a central registry. To address variability within and across medical, surgical, and behavioral health aspects of care, the DSD-TRN is dedicated to scientific discovery and the standardization of diagnostic and treatment protocols, including an early and comprehensive diagnostic process, associated with extensive standardized phenotyping and psychosocial screening of patients and families. An accurate diagnosis is critical to predicting occurrence of life-threatening crises, response to hormone-replacement therapy, stability of gender identity, fertility, recurrence risk, cancer risk, and patient quality–of-life outcomes. Genomic technologies, chromosomal microarrays, and next-generation sequencing are revolutionizing the approach to DSD diagnosis. With exome sequencing and DSD-specific panels now routinely available in the clinical realm, DSD-TRN best practice guidelines recommend early, comprehensive genetic testing as a means to improve the path to an accurate diagnosis and optimized clinical management. Longitudinal data collection and monthly clinical activity reports allow tracking of the effort, success, completion, and timeline of the diagnostic process for each clinical team. Relative diagnostic efficacies of genetic testing methods, frequencies of the different DSD conditions, and rates of definitive genetic diagnosis were analyzed. Analysis of data for 144 probands at 9 clinical sites showed that genetic diagnostic efforts by DSD-TRN teams were rewarded by a substantial increase in patients with a firm diagnosis: from 24% to 46%. The rate of successful diagnosis can however be vastly improved by increased adherence to DSD-TRN diagnostic guidelines: for almost all (97%) of the patients who remain without a diagnosis, currently available diagnostic methods (such as trio exome sequencing and chromosomal microarray) have not been exhausted.

447F

Transethnic meta-analyses from genome-wide association studies of fibroid characteristics in African and European American women. *M.J. Bray[1], M.F. Wellons[2], T.L. Edwards[1,2,3,4], D.R. Velez Edwards[1,4,5,6].* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Department of Medicine, Vanderbilt University, Nashville, TN; 3) Division of Epidemiology, Vanderbilt University, Nashville, TN; 4) Institute of Medicine and Public Health, Vanderbilt University, Nashville, TN; 5) Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN; 6) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN.

The majority of women in the US will develop at least one uterine fibroid by menopause. Fibroids are heterogeneous varying in size and number, leading to a range of symptoms from chronic pelvic pain to heavy, painful periods. While the heritability of fibroid characteristics such as size and number is unknown, the estimated heritability for fibroid risk ranges between 26-69%. It is possible that specific fibroid characteristics are heritable and that this heritability is driven by genetic loci. To address this, we performed transethnic meta-analyses from genome wide association studies (GWAS) on fibroid characteristics on African and European American women. Using BioVU, a clinical cohort from the Vanderbilt University Medical Center, and the Coronary Artery Risk Development in Young Adults (CARDIA) cohort, a prospective cohort, we identified 1,520 women with documented fibroid characteristics. We imputed ungenotyped SNPs using the 1000 Genomes Cosmopolitan reference panel phase 1 version 3. We performed single SNP association analyses on $\log_{10}$ transformed volume of largest fibroid, $\log_{10}$ transformed largest fibroid dimension, and fibroid number (1 vs >1) for common SNPs (minor allele frequency >5%) and adjusted for age, BMI, and five principal components. Finally, we performed fixed effects inverse-variance weighted meta-analyses. We then evaluated associations between genetically predicted gene expression and fibroid characteristics using MetaXcan using results from GWAS meta-analyses. We identified nominal genetic signals ($p < 1 \times 10^{-5}$) on chromosomes 1q31.1 and 4q35.1 for fibroid volume, on chromosomes 6q27, 4p16.1, and 10p11.22 for largest fibroid dimension, and on chromosomes 15q15.3, 7p14.3, and 1q24.1 for fibroid number. The strongest signal from the meta-analyses was in 1q31.1 with flanking genes phospholipase A2 group IVA (*PLA2G4A*) and BMP/retinoic acid inducible neural specific 3 (*BRINP3*) (rs6605005, β = 0.24; 95% confidence interval: 0.15, 0.33; $p = 7.68 \times 10^{-8}$). MetaXcan results showed an association between the predicated gene expression of ER degradation enhancing alpha-mannosidase like protein 2 (*EDEM2*) in thyroid tissue for fibroid number ($p = 6.34 \times 10^{-5}$). In summary, this study found many associations between genetic loci as well as gene expression and fibroid size and number. This work will help lay the foundation for understanding the link between genetic variation and fibroid characteristics.

448W

Non-immune hydrops fetalis with long bones fragility: A new subtype of osteogenesis imperfecta type II or a new form of skeletal dysplasia? *G.A. Molfetta[1,2,3,5], A. Piram[4,5], A.J. Machry[2], T.O. Anjos[3], A.A. Marques[3], W.A. Silva Jr[1,2,3].* 1) Department of Genetics, Ribeirao Preto Medical School, University of São Paulo, Ribeirao Preto, SP, Brazil; 2) Center for Medical Genomics at General Hospital of the Ribeirao Preto Medical School, University of São Paulo, Ribeirao Preto, SP, Brazil; 3) Regional Blood Center at General Hospital of the Ribeirao Preto Medical School, University of São Paulo, Ribeirao Preto, SP, Brazil; 4) Gemina Women's Diagnostic Clinic, Londrina, PR; Brazil; 5) Authors Equally Contributed to the Work.

A 29-year-old primigravida was referred at the 22nd week of gestation for second trimester morphology ultrasound because of hydrops fetalis. She and her husband were no consanguineous. During the assessment at the fetal medicine clinic, the finding of massive hydrops markedly in face and skull on a phenotypically female fetus was confirmed and additional anomalies were detected such as small nose, slight rhizomelic shortening of upper extremities, radial hands deviation, slight ankle varus deformity, important shortening of femurs and unilateral angulated ulna suggesting ulnar fracture, small ventricular septal defects. Amniocentesis procedure was scheduled for the 23rd week of gestation when it was found the fetus was already dead. Postmortem examination of the spontaneous expulsed fetus weighing 540 grams confirmed the ultrasound findings. X-rays images showed osteopenia throughout the skeleton with multiple bone fractures and deformities of the long bones more accentuated on lower limbs and left upper limb resembling Osteogenesis Imperfecta (OI). Both fetal blood and amniotic fluid samples for cytogenetic and molecular studies were collected. Based on the radiological images indicating a major clinical suspicion of OI, a NGS panel was employed to screen for mutations the fourteen genes coding for OI. No pathogenic mutations were detected. Regarding the in utero differential diagnosis, we ruled out the presence of Campomelic Dysplasia and Thanatophoric Dysplasia due to no pathogenic mutations in SOX9 and FGFR3, respectively. In addition, we ruled out Achondrogenesis type 1B and Hyphophosfatasia by the clinical criteria. Severe forms of OI are frequently diagnosed prenatally as consequence of in utero ultrasound scanning. The main classical sonographic features of OI type II include diffuse hypomineralization of the skeleton, shortening and bowing of long bones, multiple fractures. The three subtypes of OI type II (A, B and C) are characterized by different radiological features. Hydrops fetalis can be seen in association with the subtype IIA. In the present case, it was not observed undermineralization of the skull and ribs abnormalities. Whole Exome Sequencing is being executed and might help to elucidate this case. Depending on the findings of a mutated gene other than COL1A1/COl1A2, whose protein product is involved with the collagen pathway or the findings of a non-related to collagen gene, we will be able to better describe this clinical phenotype.

449T

Prenatal diagnosis of diastrophic dysplasia: Importance of prenatal approach with NGS panel. *M. Tamayo, M. Garcia-Acero, L. Mora.* Instituto de Genetica Humana, Facultad de Medicina, Bogota DC, Colombia.

Introduction: Diastrophic Dysplasia (DD) is a congenital osteochondrodysplasia with an autosomal recessive inheritance pattern. The most common clinical findings are short stature, cleft palate, cysts pinna with calcified cartilage and micromelia, associated to scoliosis, cervical kyphosis, hypoplasia of the cervical vertebrae, lumbar lordosis, clubfoot, joint contractures with progressive degenerative changes, brachydactyly, abductees and hypermobile thumbs and symphalangism. During the neonatal period mortality can reach up to 25% mainly due to airway obstruction and aspiration pneumonia. For those who survive this period the prognosis for life is good. **Case presentation:** We present a male patient of 1 year old, son of non-consanguineous parents but of the same location, with prenatal diagnosis at 13 weeks of short long bones, bilateral clubfoot and abducted hitchhiker thumb, with history of sister with unknown skeletal dysplasia who dies due to pulmonary hypoplasia. Was requested molecular diagnostic through NGS panel of skeletal dysplasia in sample of amniocentesis, which identified two heterozygous pathogenic variants in *SCL26A2* gene (c.532C>T and c.835C>T), changes were demonstrated in each parent. At birth with evidence of rhizomelic shortening, clubfoot, passive and active reduced mobility, abduction and earlier implementation of the thumb. Radiologic studies showed shortening of long bones with metaphyseal widening, short tubular bones, short and wide first metacarpal and bilateral clubfoot. Currently with clubfoot corrected, he has a sitting position and starts crawling and is in multidisciplinary monitoring. **Discussion.** It was made molecular prenatal diagnostics of DD suspected by the ultrasonographic findings, ruling out the presence of a lethal skeletal dysplasia given the history of the affected sister. DD is marked by short stature with short extremities and joint malformations can lead to either limitation of joint movement or hyperlaxity. The severity of the clinical manifestations is variable, ranging from very severe to moderate forms that may be diagnosed very late as clinical heterogeneity of skeletal dysplasia. **Conclusion.** The availability of NGS panel for Skeletal dysplasia lead stablish an early diagnosis to determine the prognosis and to prepare the delivery team. Modifying the course of an initial suspicion of lethal skeletal dysplasia.

450F

Neurodevelopment in Japanese singletons, aged 4—6 years, conceived by assisted reproductive technologies. *T. Shimada[1], T. Kawamoto[2,3], A. Yoshida[4], Y. Takehara[5], T. Kuroda[5], N. Kawasaki[5], K. Kato[5], K. Shimodaira[6], Y. Osuga[7], Y. Kamei[7], T. Okai[6], N. Kato[8], T. Sasaki[8,9].* 1) Office for Mental Health Support and Communication Support Room, Division for Counseling and Support, the University of Tokyo, Tokyo, Japan; 2) Faculty of Letters, Keio University, Tokyo, Japan; 3) Research fellow of Japan Society for the Promotion of Science; 4) Reproduction Center, Kiba Park Clinic, Tokyo, Japan; 5) Kato Ladies Clinic, Tokyo, Japan; 6) Department of Obstetrics and Gynecology, Showa University School of Medicine, Tokyo, Japan; 7) Department of Obstetrics and Gynecology, Faculty of Medicine, the University of Tokyo, Tokyo, Japan; 8) Department of Neuropsychiatry, Showa University School of Medicine, Tokyo, Japan; 9) Division of Physical and Health Education, Graduate School of Education, the University of Tokyo, Tokyo, Japan.

Background: Assisted reproductive technologies (ART) are infertility treatments which handle gametes. The number of children born using ART sharply increases from 1990s. Similarly, the prevalence of neurodevelopmental disorders (NDD) increases rapidly over the last several decades. On the other hand, recent studies suggest that *de novo* mutations and epimutations in the genome play a role in NDD. ART may have effects on NDD and neurodevelopment of children through these processes. A number of studies have examined effects of ART on child neurodevelopment. Although most of the studies found no significant effect, the conclusion remains to be reached, due to several confounders including parental age and traits of neurodevelopment which were not fully or at all adjusted for. Differences in the effects according to ART methods (i.e. *in vitro* fertilization (IVF), vs. intracytoplasmic sperm injection (ICSI), and fresh vs. frozen embryo transfer) also remain to be studied. **Methods:** We conducted clinic-based multicenter retrospective cohort study, to investigate effects of ART on neurodevelopment, with careful adjustment for confounders. Neurodevelopment was compared between Japanese singletons, with maternal age at birth < 36 years, born after ART (n = 300) and natural conception (NC) (n = 181), at age of 4–6 years using parent-reported Child Behavioural Checklist (CBCL) and Social Responsivity Scale (SRS). Effects according to ART subtype (IVF (without ICSI, n = 168) vs ICSI (n= 130), and fresh (n = 136) vs frozen embryo transfer (n = 158)) were examined. Propensity score with doubly robust estimation was employed to control for potential confounders, including parental neurodevelopmental traits assessed using Autism Spectrum Quotient score and Adult ADHD Self-Report Scale. **Results:** Total problems, externalizing and internalizing scores of CBCL were significantly better in the ART than the NC group (standardized difference (95% CI): -0.29 (-0.48 — -0.10), -0.20 (-0.39 — -0.10) and -0.20 (-0.39 — -0.02), respectively). No significant difference was observed between IVF and ICSI. The scores, however, were significantly better in the fresh than the frozen embryo subgroup. The SRS total score was not different between NC and ART or ART subgroups, while some SRS subscores were significantly better in the ART group/subgroups than NC group. **Conclusions:** A negative effect of ART on child neurodevelopment was not suggested in a Japanese population.

451W

Perinatal outcomes have little influence on FSIQ in children with 22q11.2DS. *T. Crowley[1], R. Sharkus[1,2], M. Unolt[1,3], A. Silverman[1], B. Emanuel[1], E. Zackai[1], D. Mcdonald-McGinn[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) Rowan University School of Osteopathic Medicine, Stratford, NJ; 3) Sapienza University, Rome, Italy.

BACKGROUND: Perinatal outcomes, such as prematurity and low birth weight have been reported in association with lower FSIQ scores and higher risk of psychosis in adults with 22q11.2DS, however, to our knowledge, no similar study has been conducted in a pediatric cohort. **METHODS:** We reviewed birth weight (BW), gestational age (GA), Apgar Scores and early (<72 hours) neonatal complications in 368 patients with 22q11.2DS and FSIQ measures. **RESULTS:** Overall mean age at assessment was 9.8 years. FSIQ was 76.22. Prematurity (~34 to <37 weeks) was noted in 53/350 (15%). 11% had VLBW; 4% had ELBW. No correlation (R^2=.01) was found between FSIQ and GA or BW (R^2=.01). Of 297 full term (FT) infants 10% were SGA and 6% were LGA with no correlation (R^2<.01) between FSIQ and GA or BW (R^2<.01). Thus, regardless of BW, when FSIQ of the FT group (76.7) was compared with the premature group (73.5) no significant difference was found (p=.16). In contrast, 7/138 patients (5%) had low Apgar Scores (< 7 at 5 minutes) associated with significantly lower FSIQ, though the small sample size was small. Early neonatal complications of any type (RDS, cyanosis, seizures, jaundice) were described in 45% while the remainder had an uncomplicated perinatal course (discharged to home on DOL 2 or 3). Surprisingly, the presence of 1 or more perinatal complication did not effect FSIQ. **CONCLUSIONS:** We found no significant correlation between perinatal outcomes and FSIQ in our pediatric cohort apart from low Apgar Scores at 5 minutes in a small subset, which may reflect an ascertainment bias related to the lack of decline in FISQ observed in older patients or rather that the effect of the 22q11.2 deletion supercedes all other factors except a very severe secondary insult. Lastly, we found a lower prevalence of SGA full term births than previously reported (10% vs 20.3%) disputing the suggestion that SGA is a common feature of 22q11.2DS.

452T

Independent head-to-head comparative diagnostic accuracy of NIPT methods in a prospective Canadian cohort of high-risk and low-risk pregnant women: The PEGASUS study. *F. Rousseau[1,2,5], S. Langlois[3], F. Audibert[4], J. Gekas[5], J. Johnson[6], M. Walker[7], S. Giroux[5], A. Caron[5], V. Clément[5], T. MacLeod[8], R. Moore[8], J. Gauthier[9], L. Jouan[9], A. Laporte[9], O. Diallo[9], J. Parker[9], L. Swanson[8], Y. Zhao[8], Y. Labelle[5], J.C. Forest[1,2,5], Y. Giguère[1,2,5], J. Little[11], A. Karsan[8,12], G. Rouleau[9,10], PEGASUS group.* 1) Dept Laboratory medicine, CHU de Québec - Université Laval, Quebec City, Quebec, Canada; 2) Dept molecular biology, medical biochemistry and pathology, Faculty of Medicine, Université Laval, Québec, QC, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver (BC), Canada; 4) Département d'obstétrique-gynécologie, Faculté de médecine, Université de Montréal, Montréal (QC), Canada; 5) Centre de recherche du Centre hospitalier universitaire de Québec, Québec (QC), Canada; 6) Head Section of Maternal Fetal Medicine Foothills Medical Center, Department of Obstetrics and Gynecology Cumming School of Medicine, University of Calgary; 7) Ontario Hospital Research Institute, Ottawa, Canada; 8) Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC, Canada; 9) CHU Ste-Justine, Université de Montréal, Quebec, Canada; 10) Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; 11) School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa (ON), Canada; 12) Department of Pathology & Laboratory Medicine, BC Cancer Agency, Vancouver, BC, Canada.

Genomics-based Non-invasive prenatal screening (NIPS) is recognized as a second-tier prenatal screening test for common major aneuploidies (T13, T18 and T21). However, NIPS assays are laboratory developed tests and there is little data on the comparative performance of different NIPS assays. Further, most of the NIPS test offer is currently available through commercial laboratories. We aimed at independently comparing head-to-head, in a prospective cohort, the screening performance of two NIPS platforms implemented in Canadian public laboratories for detecting T21, T18, T13 and 45,X. **Methods** We recruited prospectively 2012 high-risk and 1807 low-risk pregnant women (NCT01925742). Clinical outcomes collected were: prenatal screening results, cytogenetic analysis, newborn exam and follow-up outcome at age 6 weeks. 69 cases were lost to follow up. 3616 samples were randomly selected for analysis. Two untargeted (MPSS) NIPS assays were implemented in three Canadian public labs (Illumina's HiSeq-2500 and LifeTech's Proton). Samples were randomized between labs and tested by both assays sequencing 4-8 million random maternal ccfDNA fragments to compute chromosomal ratio's z-scores. Chromosomal-ratios mean CVs were XX0.47 for T21, 0.33 for T18 and 0.31 for T13. Fetal fraction was estimated with SeqFF. **Results** The cohort included 155 cases of T21, 50 T18, 8 T13, 22 45,X. Detection rates(DR) and false positive rates(FPR) for T21, T18, T13 and Turner were not statistically different between platforms: mean DR 99% (T21), 100% (T18, T13); 70% (Turner); FPR 0.3% (all). The no-result rate was not significant (between assays, nor between high-risk and low-risk : mean 3.7% (Chi2 1df<0.9, p>0.7). For both methods in high-risk pregnancies, PPV were 97%(T21), 89% (T18), 75% (T13) and 70% (Turner). NPV>99.9 for aneuploidies except Turner (99.5% vs 99.8%). **Conclusions** Our results suggest that both MPSS methods compared (semiconductor-based NGS and optical-based NGS) showed equivalent and excellent test performance. Although slight differences between labs and methods were observed, overall the two assays and the three public laboratories performed well in this independent study. NIPS could be implemented and covered by the public health care system as a second tier screening test for high-risk women identified by traditional prenatal screening at higher risk of aneuploidy. Implementation of this approach as a first-tier screening test remains to be investigated.

453F

Single-cell RNA sequencing in sperm from fathers of autistic children.
J.A. Rosenfeld[1], D. Tomoiaga[2], S. Shrestha[3], S. Levy[3], C.E. Mason[2]. 1) Pathology, Rutgers, NJ; 2) Weill Cornell Medical College, New York, NY; 3) Hudson Alpha Institute for Biotechnology, Huntsville, AL.

The incidence of autism has been shown by numerous studies to be correlated an increase in the age of the father at the time of conception. This connection is thought to be related to the increase in de novo mutations or copy number variations in sperm. In addition to autism, there are many other disorders associated with advanced paternal age that are collectively known as paternal age effect (PAE) disorders. There is currently no effective method to test sperm for the presence of mutations related to PAE disorders. The current guidance from clinical geneticists is for patients to avoid having children when the male is older, but there is no clear way to screen out those younger men who could have mutations or dys-regulated genes in their sperm. Until recently, the only way to look for rare mutations in sperm would be to take a bulk sample of cells, sequence them to a very high depth and then to look for variants identified by a very small number of reads. This type of assay is expensive, and it is difficult to differentiate the very low frequency variants from those that are a result of a sequencing errors or alleleic drop-out (ADO). This problem has been alleviated with the introduction of systems such as the 10X Genomics Chromium instrument. Using an innovative molecular bar-coding scheme and the separation of individual cells, the Chromium allows for the direct sequencing of single cells. For each cell, the amount of RNA for each gene is identified and the cells can be grouped by expression patterns. We have used this technology, for the first time, to examine the individual sperm of donors who have children with autism as compared to control sperm from donors who do not have autistic children. We have sequenced >10,000 sperm cells in the donors and patients and determined that germline-specific gene expression patterns are specifically enriched in these samples. Moreover, we observe significant differences in the developmental pathways between the cases and controls, indicating this may be a method to reveal baseline functional genomics differences in sperm donors. Using innovative computational methods, we have been able to call SNP variants from the RNA-seq data. It is hoped that this work will lead to a screen for sperm to determine which patients are at a higher risk of having a child affected with autism or another PAE disorder.

454W Unprogrammed presentation number

455T

The Dutch TRIDENT studies: Implementing NIPT as part of the national prenatal screening program. *E.A. Sistermans[1], L. Henneman[1], H. Scheffer[2], M.J. Pieters[3], M. Macville[2], M.N. Bekker[4], C.J. Bax[5], A.R.M. van Opstal[7], F.T.J. Sleutels[7], M.M. Weiss[1], A.J.E.M. van der Ven[6], R-J.H. Galjaard[7], the Dutch NIPT Consortium.* 1) Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands; 2) Department of Clinical Genetics, Maastricht UMC+;, Maastricht, The Netherlands; 3) Department of Obstetrics and Gynecology, Grow, School for Oncology and Developmental Biology, Maastricht University Medical Center (MUMC), Maastricht, The Netherlands; 4) Department of Obstetrics, University Medical Center Utrecht, The Netherlands; 5) Department of Obstetrics, Academic Medical Center, Amsterdam, The Netherlands; 6) Verloskundigenpraktijk Velp e.o., Velp, The Netherlands; 7) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.

Non-Invasive Prenatal Testing (NIPT) has been rapidly introduced worldwide. In many countries this has been done by companies and/or (academic) medical centers, without any governmental guidance. In the Netherlands, a license from the Minister of Health must be obtained before offering prenatal screening for Down syndrome, because of the Population Screening Act. In 2014 the Dutch NIPT Consortium, which includes all stakeholders that are involved in prenatal care, obtained a license for the TRIDENT-1 study, allowing NIPT for women with high-risk pregnancies based on first trimester combined test (FCT) results or medical history. TRIDENT-2, which aims at implementing NIPT for all pregnant women within the framework of the national prenatal screening program, started April 1ˢᵗ 2017. In TRIDENT-2, NIPT will become available for all pregnant women in the Netherlands (on average 180,000 women/year). The Minister decided to invest €26 million/year to ensure that NIPT analysis would be affordable to all women. Women who want to have prenatal screening can now choose which test they prefer (NIPT or FCT). As in TRIDENT-1, a whole genome sequencing approach is used for NIPT. Women have the choice between receiving results for chromosomes 21, 13 and 18 only, or for all autosomes. Sex chromosomal abnormalities will not be analyzed and reported. As the numbers of participants in TRIDENT-2 are much higher than for TRIDENT-1, the entire chain from counselor to laboratory and back needed to be revised. The set-up of the project was closely supervised by the Ministry of Health during monthly meetings and included: training of ~3000 counselors; development of informational booklets and websites for pregnant women and professionals; increase of the laboratory capacity; contracting organizations for blood withdrawal; adapting the national prenatal screening database for ordering the NIPT test and reporting results; writing a grant for questionnaire and interview studies to study the women's perspective. During the first 7 weeks approximately 10,000 NIPT analyses were requested. The turnaround time is 6-7 working days for most of the samples. No major problems were encountered during the first weeks. Detailed information on the results will be presented at the meeting. We expect that this 3-year study will provide us with all the results needed for the final decision on the responsible introduction of NIPT in the Dutch National prenatal screening program.

456F

Large-scale cytogenomic analysis of samples from conception to childhood: A comprehensive assessment of the landscape of unbalanced genomic abnormalities. *T. Sahoo, M.N. Strecker, S.B. Commander, M.K. Travis, N. Dzidic, C.E. Doherty, L. Drugan, K. Hovanes.* CombiMatrix Diagnostics, Irvine, CA.

Background: No large studies have cataloged chromosomal imbalances and their phenotypic consequences over the course of human development, from conception through early childhood. **Methods**: We undertook a multi-year analysis of ≈32,000 samples submitted for cytogenomic evaluation at different stages of human development, including embryo biopsies (N=6883), miscarriage (N=12324), prenatal (N=4176), neonatal (N=1564), and pediatric (N=7047) samples. Embryo biopsy samples were analyzed by arrayCGH or NGS and all other samples were analyzed by chromosomal microarray analysis (CMA). **Results**: Clinically significant chromosome abnormalities (CSCAs) were identified in 46% (3105/6683) of embryo biopsies. The majority (70%) were autosomal whole chromosome aneuploidies (WCAs): 32% monosomies, 29% trisomies and 8% with monosomy and trisomy. Another 17% samples had complex abnormalities. In miscarriage samples, 54% (6665/12324) demonstrated CSCAs, the most common being autosomal trisomies (63%), triploidy (12%) and monosomy X (11%). Autosomal monosomies were rare (<1%). CSCAs were present in 13% (542/4176) of prenatal samples, with trisomies (41%) and segmental deletions/duplications (44%) being the most common. In neonatal samples 18% (273/1564) had CSCAs; WCAs (34%) and segmental abnormalities (65%) were the most frequent. For pediatric samples, CSCAs were present in 12% cases (854/7047); most frequent were deletions (60%) or duplications (23%). **Conclusions**: This study provides a comprehensive view of the evolution of unbalanced genomic abnormalities with adverse effects on development from conception up to early childhood. Across all groups studied, the most severe chromosome abnormalities were detected in embryos and early pregnancy loss, resulting in failed implantation and/or non-viable pregnancy. The high frequency of autosomal monosomies in embryonic samples, rarely seen at later developmental stages, reflects the extreme vulnerability of cellular mechanisms that help maintain normal and stable genomic content. In contrast, the genomic alterations detected prenatally and neonatally are not necessarily incompatible with life, and the pediatric samples demonstrated a preponderance of less drastic deletions/duplications with significant diversity in terms of type and chromosomal involvement. These data reveal an interesting pattern of genomic alterations of decreasing severity as human growth progresses from conception through childhood.

457W

Perinatal features and genotype-phenotype correlations in a large cohort of 355 patients with Prader-Willi syndrome. *V. Kimonis.* UC Irvine, Irvine, CA.

Background: Prader Willi Syndrome (PWS) is caused by lack of expression of genes on the paternal chromosome 15q11.2 - 13 . There is an evolving phenotype for those patients with either a Deletion, Uniparental Disomy (UPD) or an Imprinting Center Defect (ICD).   Objectives: We evaluated the early manifestations of PWS in pregnancy and early neonatal period to identify clinical features that would lead to early detection of PWS. We also looked for genotype-phenotype correlations for the genetic subtypes.   Methods: Data from 355 patients from the Rare Diseases Clinical Research Network (RDCRN) PWS registry were used to analyze multiple clinical maternal and neonatal factors.   Results: Out of 355 patients 61% had deletion, 36% had UPD and 3% had an IC defect. 54% were born by C-section (population 32%). Fetal movements were decreased in 72%. Mean birth weight was 2.71 kg. All babies were hypotonic, 72% needed gavage feeding and 23% needed G-tube placement, these values being significantly higher than the general population. Maternal age was significantly higher in the UPD group as was pre-pregnancy weight.   Conclusions: There is a higher incidence of babies born with UPD to older mothers with higher pre-pregnancy weights. Although few significant differences were found in the variables between the different genetics subgroups, the incidence of C-section, low fetal movements, lower birth weight and other variables was significantly higher than the general population. The goal of this study is to increase awareness of the perinatal and neonatal features of PWS thus enable earlier diagnosis, treatment and prevention of morbid obesity and co-morbidities.   .

458T

Next generation sequencing based carrier screening study in Chinese population. *G. Chen, Y. Xu, Q. Zhang, Q. Zhao, Y. Bai, D. Wang, Y. Meng, S. Chang, X. Ma, Y. Xie, L. Wei, P. Fang, L. Liu.* WuXi NextCODE Genomics.

Obtaining carrier status information for severe recessive diseases can help individuals to make reproductive decisions. The current next generation sequencing (NGS) technique allows development of carrier screening test containing large number of genes with affordable price. Simultaneously analyze large number of genes can identify variants of significant clinical impact and facilitate establishing detection rate and reproductive risk of associated recessive Mendelian diseases for diverse populations.   We have developed a NGS based expanded carrier screening test containing about 200 genes associated with 135 recessive diseases specifically for Chinese population. The selected genes are associated with early onset, severe inherited disease with known relatively high incidence in Chinese or East Asia population. After NGS sequencing and bioinformatics analysis, all variants are manually curated based on published The American College of Medical Genetics and Genomics (ACMG) guidelines. Pathogenic and likely pathogenic variants are reported as positive carrier status while variants of uncertain significance (VUS) are reassessed annually and reported out if they are upgraded to pathogenic or likely pathogenic. In total, 614 healthy Chinese adults were tested. 245 individuals were identified to carry at least one pathogenic or likely pathogenic variant, accounts for nearly 40% of the total tested individuals. Top five genes with highest carrier frequencies are *SLC25A13* (4.72%) , *GJB2* (2.92%) , *ATP7B* (2.76%), *USH2A* (2.60%) and *GALC* (2.12%), and the corresponding top five diseases are Citrin deficiency, autosomal recessive hearing loss type 1A, Wilson disease, Usher syndrome type 2A and Krabbe disease, respectively. Our results demonstrate unique gene/disease spectrum in Chinese population. This information would help us to develop a more focused basic panel suitable for a nationwide carrier screening customized to Chinese population. In addition, data obtained from this study would facilitate improved genetic counseling in China.

459F

Carrier screening for 316 monogenic recessive diseases revealed high carrier frequency of rare known pathogenic mutations. *D. Bercovich[1,2], S. Horn-Saban[2], E. Kellerman[2], J. Ronen[3], R. Gershoni[2].* 1) Biotechnology, Tel Hai College, Tel Hai, Gelil Elyon, Israel; 2) Galil Genetic Analysis Center, Galile, Israel; 3) KaryoBank, Ramat Hasharon, Israel.

Genetic mutations can deter genes from functioning properly and disrupt the proteins which they encode. Being a carrier for an autosomal recessive genetic disease does not usually affect one's health. However, if both biological parents happen to carry mutations in the same gene, their offsprings are at increased risk of incurring the related genetic disease. In most countries carrier screening addresses only common mutations based purely on domestic ethnicity. Consequently, babies are still born with genetic diseases that could have been avoided. Using NGS technology we screened several hundred healthy sperm and egg donors for mutations along entire coding regions and exon-intron boundaries of 316 genes related to common monogenic diseases. To avoid reporting VUS (variants of unknown significance), we reported only Indels, stop codons, splicing and missense mutations with known pathogenic effect according to HGMD® Professional, ClinVar, NCBI (PubMed & OMIM), Alamut (v.2.7.1) or in our database . The presence of these mutations was confirmed by Sanger sequencing. The carrier rates amounted to 2 to 7 mutations per DNA sample with an average of 2.8 mutations. 80% of mutations were known rare mutations which are not screened by genetics labs on regular basis. Missense mutations which do not appear in these databases and had a population frequency of less than 1% were defined as VUS and were not reported. 0.001% of couples were both carriers of mutations in the same gene which can lead to approximately 8.3×10^{-5} unhealthy offsprings out of all births per year in Israel. These results indicate the importance of full gene sequencing for common monogenic diseases for the detection of rare recessive mutations in the general population and prevention of sick babies being born. We advise that such full gene sequencing be applied as routine pre-conceptional screening, followed by genetic counseling.

460W

The utility of exome sequencing in prenatal diagnosis. *E.A. Normand[1], A. Braxton[2], P. Liu[1,2], F. Xia[1,2], W. Bi[1,2], R. Xiao[1,2], M. Leduc[1], J. Zhang[1], X. Wang[1], L. Meng[1], C. Qu[2], W. He[2], F. Vetrini[2], A.V. Dharmadhikari[2], P. Ward[1,2], S. Narayanan[1], S. Nassef[1], S. Plon[1], D. Muzny[1,3], J.R. Lupski[1], R. Gibbs[1,3], I. Van den Veyver[1,4], M. Walkiewicz[1,2], Y. Yang[1,2], C. Eng[1,2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 2) Baylor Genetics Laboratories, Houston, TX 77021; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 4) Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX 77030.

Fetal anomalies occur in approximately 3% of pregnancies, many of which have a genetic etiology. While karyotyping and chromosomal microarray (CMA) are invaluable first-tier tests for prenatal genetic diagnosis, single gene defects cannot be detected by these methods. With the continuous improvement of exome sequencing, including updated instrumentation, shorter turnaround time (TAT), and improved analysis, clinical exome now provides a valuable diagnostic option in such cases. Here, we report on 116 consecutive prenatal samples referred to our diagnostic laboratory for exome sequencing from March 2012 to April 2017. We define a prenatal sample as one obtained through a diagnostic procedure or a product of conception from a fetus with at least one structural anomaly detected by fetal imaging. Next-generation sequencing was performed on an Illumina HiSeq platform and 97% of targeted exome regions were sequenced at 20X depth. Sequencing data were analyzed for small nucleotide changes and large CNVs using the Illumina HumanExome-12v1 array. Among 116 prenatal samples, exome yielded a molecular diagnosis in 35% of cases. This diagnostic rate did not vary appreciably based on whether brain and/or cardiac anomalies were present. The diagnostic rate also did not differ based on whether there was a significant family history, but the distribution of *de novo* versus inherited variants did differ significantly between these cohorts. Seven families used their prenatal exome results for targeted mutation analysis in subsequent pregnancies. Three genes were causative for multiple unrelated samples: *KMT2D* (3 samples), *COL1A1* (3 samples), and *COL1A2* (2 samples). Three test options were available to ordering physicians: proband exome (12 week TAT, parental samples for Sanger-based inheritance studies, proband report only); standard trio exome (8 week TAT, trio-based variant interpretation, proband and parental reports issued); and prenatal trio exome (introduced in April 2015, 2-3 week TAT excluding cell culture time, trio-based variant interpretation, fetal and parental reports). Prenatal trio exome was ordered for 44 of 116 cases. The mean TAT was 2.0 weeks and 43% received a molecular diagnosis. Overall, our results demonstrate that exome sequencing is a valuable diagnostic tool and may be considered in prenatal cases where structural anomalies suggest a genetic etiology, but for which other genetic tests such as karyotyping and CMA are unlikely to provide a diagnosis.

461T

A genomic autopsy of perinatal death: Diagnosis and discovery by whole exome and whole genome sequencing. *A.B. Byrne[1,6,8], J. Feng[2,9], A.W. Schreiber[2,9], P.J. Brautigan[1], M. Babic[1], W. Waters[3], S. Yu[3], T.Y. Khong[4], L. Moore[4,10], M.E. Dinger[5], D.G. MacArthur[6], C.N. Hahn[1,3,8,10], K.S. Kassahn[3,9], C.P. Barnett[7,10], H.S. Scott[1,2,3,8,9,10].* 1) Genetics and Molecular Pathology Research Laboratory, Centre for Cancer Biology, Adelaide, South Australia, Australia; 2) ACRF Cancer Genomics Facility, Centre for Cancer Biology, Adelaide, South Australia, Australia; 3) Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia; 4) Department of Surgical Pathology, Women's and Children's Hospital/SA Pathology, North Adelaide, South Australia, Australia; 5) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, New South Wales, Australia; 6) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) SA Clinical Genetics Service, Women's and Children's Hospital/SA Pathology, North Adelaide, South Australia, Australia; 8) School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, South Australia, Australia; 9) School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia; 10) School of Medicine, University of Adelaide, Adelaide, South Australia, Australia.

Perinatal death (PD) refers to the stillbirth of a foetus from 20 weeks gestation or the neonatal death of a liveborn baby within 28 days of birth. Current autopsy procedures have identified congenital abnormalities as the leading cause of death, accounting for more than 33% of cases, yet the underlying cause of the malformation is rarely determined. A further ~15% of PD is due to or is complicated by infectious causes. In this study, we examine the utility of a 'genomic autopsy' in not only elucidating the genetic causes underlying congenital abnormalities but also in identifying causative infectious agents. Whole exome sequencing (WES) and whole genome sequencing (WGS) are being performed using either an Illumina HiSeq 2500 (WES) or X Ten System (WGS). For cases of congenital abnormality, major phenotypes investigated include neurologic abnormalities, ciliopathy-like disorders, urinary tract malformations, non-immune hydrops, and musculoskeletal abnormalities. An analysis pipeline incorporating bioinformatic and experimental laboratory techniques has been developed to identify and causally link novel variants and genes to disease. Of 14 families investigated to date, 6 have resulted in the identification of causative genetic variants; 2 in known disease genes, 2 expanding the phenotypic spectrum of known disease genes (*TPI1, B3GAT3)*, and 2 in novel disease genes (*FGFR2, TMEM212)*. 4/6 families are using this information for preimplantation genetic diagnosis, already leading to one successful reproductive outcome. A further 3 cases likely representing novel discoveries are currently under investigation. For cases of perinatal infection, samples with both known and suspected infectious causes are being investigated. Standard nucleic acid preparation protocols are being compared to microbial enrichment approaches to assess the feasibility of a single WGS test to detect multiple causes of PD. Preliminary results from non-microbial enriched WGS data show ~5% of reads do not map to the human genome, but many can be mapped at high quality to viral or bacterial genomes. This ongoing work demonstrates the power of a genomic autopsy in identifying genetic variation causal of congenital abnormalities and in characterising novel genes critical in embryonic developmental pathways. It also suggests that a genomic autopsy approach has the potential to identify infectious agents, together, allowing a precise cause of PD to be established in up to 50% of cases.

462F

The Genomic Autopsy Study: Data from the first 50 cases indicates that whole exome sequencing/whole genome sequencing is a powerful adjunct to standard autopsy in the investigation of complex fetal and neonatal presentations. *C.P. Barnett[1,8], A. Byrne[2,3], L. Moore[4], Y. Khong[4], N. Manton[4], J. Lipsett[4], S. Yu[2], M. Dinger[5], M. Babic[2,3], P.J. Brautigan[2,3], Q. Schwarz[3], P.Q. Thomas[6], J. Liebelt[1,8], C.N. Hahn[2,3], F. Feng[6,7], A.W. Schreiber[6,7], K. Kassahn[2,3], H.S. Scott[2,3,6,7].* 1) Pediatric & Reproductive Genetics Unit, Women's and Children's Hospital, North Adelaide, South Australia, Australia; 2) Department of Genetics and Molecular Pathology, SA Pathology; 3) Centre for Cancer Biology, an alliance between SA Pathology and University of South Australia; 4) Department of Anatomical Pathology, SA Pathology at the Women's and Children's Hospital; 5) Kinghorn Centre for Clinical Genomics, the Garvan Institute; 6) School of Biological Sciences, University of Adelaide; 7) ACRF Cancer Genomics Facility, Centre for Cancer Biology, SA Pathology; 8) University of Adelaide, Faculty of Health and Medical Sciences.

Background: In 2016, the Genomic Autopsy Study was introduced at ASHG and in this paper we present our updated data. Congenital abnormalities are the most frequent reason for mid-term stillbirth and termination of pregnancy. Unexplained late fetal deaths in utero, or in the newborn period, affect ~3.5/1000 births. Despite a high rate of formal post-mortem examination in South Australia, a definitive cause of the congenital abnormalities or death is only identified in 25-50% of cases. Aim: To use whole exome sequencing (WES) and whole genome sequencing (WGS) to identify genetic causes of fetal and newborn abnormalities that result in termination of pregnancy, death due to congenital abnormalities, death in utero or death in the newborn period, in view to providing families with answers regarding cause and likelihood of recurrence of these congenital problems. Methods: WES and/or WGS are being performed on cases of perinatal death, using the Illumina HiSeq X 10 system. Samples are available for ~1200 cases. Ethical approval has been granted for prospective cases referred to the Pediatric and Reproductive Genetics unit (parent-fetus trios) and retrospective cases from stored autopsy samples (singletons). High priority cases are fetuses with congenital abnormalities with consanguineous parents, fetuses with multiple malformations; and unexplained fetal/newborn death. Statistical, bioinformatics and experimental laboratory techniques are being used to confirm causality of variants. Results: Data from the first 50 WES/WGS's will be presented. We will present data on new gene discovery, new syndrome identification, novel severe fetal presentations of existing rare pediatric disease, diagnosis of ultra-rare disorders that would not have been identified without a WES/WGS approach, missed diagnoses by commercial testing and successful IVF pregnancies following preimplantation genetic diagnosis based on the results of the Genomic Autopsy Study. Discussion: Our results indicate that genomic autopsy using WES/WGS should be a routine component of the investigation of multiple unexplained congenital malformations leading to termination of pregnancy or death. Further work will ascertain its utility as an adjunct to traditional autopsy for families who have experienced unexplained pregnancy loss, death in utero or in the newborn period.

463W

Novel pathogenic point mutation of KDM6A identified in a Chinese woman with Kabuki Syndrome type 2. *W. Shi[1], J. Zhang[1,2], C. Xu[1,2].* 1) The International Peace Maternity & Child Health Hospital Affiliated to Shanghai Jiao Tong University School of Medicine; 2) Department of Reproductive Genetics,The International Peace Maternity & Child Health Hospital Affiliated to Shanghai Jiao Tong University School of Medicine,.

  ABSTRACT Introduction: Kabuki syndrome (KS) is a rare congenital mental retardation condition with typical facial dysmorphia, visceral and skeletal malformations and developmental delay. KDM6A (NM_021140.2) is one of the causative gene of Kabuki syndrome type 2 (KS2), accounting for 3%-8% of KS cases. **Materials and methods:** We report a 24-year-old Chinese woman who came for pre-pregancy consultation, presenting moderate learning difficulties, typical facial features, short stature, skeletal malformations including brachydactyly and hip dysplasia. To clarify the diagnosis and provide pre-pregnancy genetic counseling, the multigene panel approach on the basis of next-generation sequencing was used. **Results:** Peculiar facial features and clinical examinations of the patient, a Chinese 24-year-old married woman, was in line with the typical manifestations of Kabuki Syndrome. Furthermore, it was identified that the patient harbored a de novo point mutation of KDM6A (c.3521G>A), inducing a premature termination codon (p.Trp1174*) of KDM6A, as the same locus in the patients' healthy parents who were confirmed by identity testing is normal. In view of pre-pregancy consultation, due to KS2 is a X-linked dominant inherited disease, the risk for the patient to give birth to an affected son or daughter is 50%, a prenatal genetic diagnosis or preimplantation genetic diagnosis (PGD) will be definitely necessary for the baby when she gets pregnant. **Discussion:** We discuss the sex-biased phenotypic severity in KS2 patients and analysis the published KDM6A variants, as well as the identified muation in our study, finding that variants in exon 6 are most frequent so that may be a mutational hot spot. X-linked dominant inherited KS2 has been reported only once in China, thus our study enrich the KDM6A mutation database for further investigations.

464T

Congenital infection-like phenotypes: Are there any specific particularities available for prenatal diagnosis of rare diseases? *N. Bourgon[1], M. Lefebvre[1,2,3], J. Thevenon[1,2,3], P. Kuentz[1,2,4], T. Jouan[1,2], Y. Duffourd[1,2], F. Tran Mau-Them[1,2,3], C. Philippe[1,2,5], T. Rousseau[6], N. Laurent[7], L. Faivre[1,2,3], C. Thauvin-Robinet[1,2,3].* 1) UMR1231 , Equipe Génétique des Anomalies du Développement , Dijon, Burgundy, France; 2) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement, CHU Dijon, Dijon, Burgundy, France; 3) Centre de Génétique Médicale, CHU Dijon, Dijon, Burgundy, France; 4) UF Diagnostic-Innovation génomique des maladies rares, Laboratoire de Cytogénétique et Génétique moléculaire, CHU Dijon, Dijon, Burgundy, France; 5) Génétique Biologique Histologie, CHRU de Besançon, Besançon, Franche-Comté, France; 6) Service de Gynécologie-Obstétrique, CHU Dijon, Dijon, Burgundy, France; 7) Service d'Anatomopathologie, CHU Dijon, Dijon, Burgundy, France.

  Some rare monogenic disorders could be considered differential diagnoses of congenital infections. Congenital Infection-Like Phenotypes (CLIP) include Aicardi-Goutières syndrome (AGS), pseudo-TORCH syndrome (PTS), *PCDH12*-linked phenotypes and *JAM3*-linked phenotypes. Even though whole-exome sequencing (WES) is a powerful tool for the genetic investigation of heterogeneous causal hypotheses, clinical indications, especially for the prenatal period, should be defined. We aimed to describe prenatal AGS cases to identify specific differentiating characteristics and to define clinical criteria to justify WES. We initially identified a fetal case of AGS presenting with microcephaly, hyperechoic bowel, hepatosplenomegaly, ascites and pericardial effusion. We have now collected data for five prenatal forms of AGS. Four homozygous causal variants were identified in the five AGS cases from four consanguineous families: 2 *TREX1* frameshift (p.Leu133Serfs*17 and p.Leu123Cysfs*27), 1 *RNASEH2A* missense (p.Leu106Pro) and 1 *RNASEH2B* missense (p.Lys162Tyr), all inherited from healthy heterozygous parents. Fetuses presented intracranial calcifications (3/5), mild ventriculomegaly (3/5), congenital microcephaly (2/5) and neuronal migration disorders (2/5), ascites (2/5), hepatosplenomegaly (2/5) and hyperechogenic bowel (2/5). AGS was suspected in three fetuses during the prenatal period, leading to termination of the pregnancy. We compared these features with the classical phenotype reported in PTS and *PCDH12*-linked phenotypes. These syndromes present high clinical heterogeneity and significant phenotypic overlap. Citerna Magma anomalies appear to be more frequent in PTS, whereas digestive anomalies and ventriculomegaly predominate in AGS. There is no difference for congenital microcephaly and intrauterine growth restriction. Even though AGS is considered the main syndrome that mimics fetal infections, other entities, such as *JAM3*-linked phenotypes, must also be considered even if no prenatal form has been reported to date. In conclusion, in the prenatal diagnosis of CILP, the phenotypic overlap between CILP and fetal infections does make it possible to define criteria to justify molecular analysis. WES should be discussed in CILP, because of the importance for genetic counseling. Indeed, the identification of molecular bases is essential to organize an early prenatal diagnosis for future pregnancies given the risk of recurrence of these autosomal recessive diseases. .

465F

Prenatal diagnosis of Roberts syndrome suspected on sonogram in a consanguineous couple: A case report and a review of the literature. *J.R. King, T. Goldwaser, J. Powers, S. Klugman, P. Dar.* Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY.

Introduction: Roberts syndrome is a rare autosomal recessive disorder with approximately 150 cases reported. Roberts syndrome is characterized by prenatal growth restriction, limb malformations, and craniofacial abnormalities. Mortality is high among severely affected pregnancies and newborns. Although there have been previous reports of prenatally suspected Roberts syndrome in the literature, few of these cases have been confirmed with molecular diagnostic testing. Case presentation: 27 yo G4P3 of Argentinian Jewish descent presented for genetics consultation at 21 weeks gestation after fetal anomalies were noted on sonogram. She had no prior prenatal aneuploidy screening. Family history was significant for consanguinity, mother and partner are first cousins. The patient declined expanded carrier screening. Sonogram demonstrated severe growth restriction, with head circumference, abdominal circumference, and femur length <3%. Other findings included absence or severe shortening of the distal long bones bilaterally (radius, ulna, fibula and tibia), the hands and feet were in an abnormal position, and bilateral abnormal thumbs were noted. The fetal profile was also abnormal with a sloping forehead, mild exopthalamus and suspicion of cleft palate. Based on these findings, there was concern for Roberts syndrome. The patient received genetic counseling and elected to proceed with amniocentesis and karyotype, microarray, and NGS single gene full sequencing of the ESCO2 gene. The karyotype resulted normal female, 46 XX. The microarray showed normal female dosage but with long contiguous regions of homzygosity in multiple chromosomes. The NGS single gene full sequencing detected a homozygous pathogenic splice site variant (c.1674-2A>G) in the ESCO2 gene. This variant is considered pathogenic and causative for Roberts syndrome. The patient received these results and additional genetic counseling. She elected to continue her pregnancy. Conclusion: ESCO2 is the only gene in which pathogenic variants are known to cause Roberts syndrome. The splice variant c.1674-2A>G mutation was first described in contributing to the pathogenesis of Roberts syndrome in 2008 by Gordillo et. al via familial DNA testing. To our knowledge, this is the first report of this mutation detected prenatally. Although not commonly obtained, molecular confirmation of ultrasound findings can aid in the management and counseling of patients with suspected skeletal dysplasias.

466W

Identifying the genetic causes underlying prenatally diagnosed structural congenital anomalies (SCAs) by whole exome sequencing (WES). *G.K.C. Leung[1], C.C.Y. Mak[1], S.L.C. Pei[1], M.H.Y. Tsang[1], M.H.C. Yu[1], K.S. Yeung[1], G.T.K. Mok[1], A.P.W. Hui[2], M.H.Y. Tang[2,3], K.Y.K. Chan[2,3], A.S.Y. Kan[2,3], B.H.Y. Chung[1,2,3].* 1) Department of Paediatrics and Adolescent Medicine, LKS Faculty of Medicine, The University of Hong Kong, HKSAR; 2) Department of Obstetrics and Gynaecology, Queen Mary Hospital, The University of Hong Kong, HKSAR; 3) Prenatal Diagnostic Laboratory, Department of Obstetrics and Gynaecology, Tsan Yuk Hospital, HKSAR.

Background: Despite WES being extensively used in the clinical diagnosis of pediatric genetic disorders, routine application of the technology in pregnancies with SCAs is yet to be widely accepted. We aim to identify the genetic causes underlying SCAs in fetuses without a definitive diagnosis after conventional genetic testing. **Methods**: Couples with SCAs of the fetus detected during pregnancy were recruited from the prenatal diagnostic service, Tsan Yuk Hospital. Their antenatal histories were reviewed and those with a known cause of SCA, including aneuploidy or clinically significant chromosomal abnormalities on karyotyping and microarray, were excluded. WES was performed as trios if possible. **Results**: Thus far from 25 fetuses (23 trios, 2 singleton) analyzed, we identified 3 plausible disease-causing mutations. One fetus with small cerebellum and flexed elbows carried a *de novo* mutation in *EEF1A2*. Pathogenic alternation in *EEF1A2* leads to an autosomal dominant neurodevelopmental disorder with infantile epileptic encephalopathy. Two inherited compound heterozygous mutations, in *DIS3L2* and *NPHP4* are known to cause Perlman syndrome and Senior-Loken syndrome respectively. They were plausible explanations for the phenotypes of agenesis of corpus callosum and dandy-walker malformation identified from prenatal ultrasound of these fetuses. Our analysis also identified mutations in a novel gene that may cause a fetal cerebellar malformation syndrome that will require further functional work to characterize. **Conclusion**: WES aids the identification of genetic causes of prenatal SCA in ~10% families with negative clinical testing. Hence we suggest that the use of WES for prenatal diagnosis is practically feasible in Hong Kong. The findings could make a major impact in genetic counselling and lead to novel genes discovery, or phenotypic expansion of known genetic syndromes. **Acknowledgement**: We would like to thanks the Health and Medical Research Fund (ref# 02131816) offered by the Government of HKSAR.

467T

Increased nuchal translucency and Noonan Spectrum Disorders – A Mount Sinai hospital experience. *P. Sinajon[1], H. Sroka[2], S. Carmona[3], M. Roifman[4], E. Kolomietz[5], A. Noor[5], K. Murphy[6], D. Chitayat[4], K. Chong[4].* 1) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) GeneDX, Whole Exome Sequencing Program; 3) Medical Informatics – Information Sciences, Mount Sinai Hospital, Joseph and Wolf Lebovic Health Complex; 4) Prenatal Diagnosis and Medical Genetics, Department of Obstetrics and Gynecology, Mount Sinai Hospital – Joseph and Wolf Lebovic Health Complex; 5) Division of Diagnostic Medical Genetics, Department of Pathology and Laboratory Medicine, Mount Sinai Hospital – Joseph and Wolf Lebovic Health Complex; 6) Maternal and Fetal Medicine, Department of Obstetrics and Gynecology, Mount Sinai Hospital – Joseph and Wolf Lebovic Complex.

Objective: To examine a cohort of patients seen at the Prenatal Diagnosis and Medical Genetics Program - Mount Sinai Hospital between 2013 – 2015, with fetal ultrasound findings of an increased nuchal translucency (NT ≥ 3.5mm), and to: 1. Correlate between the NT measurements and a diagnosis of Noonan Spectrum Disorders (NSD) 2. Conduct a systematic literature review to determine the correlation between NT measurement and a diagnosis of NSD 3. To create a clinical protocol to guide physicians in the investigation of an increased fetal NT **Methods:** A cohort of patients presenting between 2013 – 2015 with fetal ultrasound findings of increased NT. All patients were offered amniocentesis/CVS for QF-PCR as a first line test. If negative, patients proceeded to karyotype/microarray analysis and NSD panel testing. Patients were also offered fetal ultrasounds at 16 weeks and 18 – 22 weeks gestation (GA) along with fetal echocardiogram. A systematic review was conducted in accordance to PRISMA criteria. Pubmed, Embase, Ovid MEDLINE and Web of Science were searched from January 2005 – August 2016 for articles involving NSD and increased NT. Seventeen papers were included for analysis. **Results:** 226 patients with increased fetal NT were seen. In 116/226 patients, chromosomal aneuploidy was detected through QF-PCR. The remaining 110/226 patients had further testing. 8 had karyotype abnormalities, 13 had abnormal microarray findings and 5 had NSD findings through DNA analysis. **Discussion:** Based on the cohort findings and literature review the following guidelines were created regarding the best approach to a fetus with an increased NT: 1. QF-PCR and microarray should be performed for NT ≥ 3.5 mm 2. If QF-PCR/microarray analysis is normal, DNA analysis for NSD should be performed for NT ≥ 4.0 mm 3. Early anatomic ultrasounds at 16 weeks GA 4. Fetal echocardiogram at 18 – 22 weeks GA 5. Detailed fetal ultrasound at 18 – 22 weeks GA.

468F

WES identifies likely pathogenic *FANCG* variants in a fetus with multiple congenital anomalies. *B.D. Webb[1,2], L. Tambini-King[3], B. Rosenn[3], J. Liao[1,2], L. Edelmann[1,2], L. Mehta[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) Sema4, a Mount Sinai venture, Stamford, CT, USA; 3) Department of Obstetrics and Gynecology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Fetal whole exome sequencing (WES) may be used to identify a genetic etiology for fetal ultrasound anomalies after karyotype, microarray, and standard of care testing are unrevealing. Here we report the use of WES to identify the genetic etiology of a condition presenting with multiple congenital anomalies. A consanguineous, healthy Egyptian couple was seen for genetic counseling due to fetal ultrasound findings of enlarged fetal bladder and umbilical cord cyst at 11 weeks. The couple's three prior pregnancies resulted in perinatal losses. The first pregnancy was terminated at 20 weeks gestation due to bilateral renal agenesis resulting in severe oligohydramnios and Potter sequence. Autopsy of the female fetus revealed absent bladder, radial aplasia, and absent thumb. The second pregnancy was a male, 27 weeks gestation fetus that died shortly after birth; the pregnancy was complicated by severe oligohydramnios. The third pregnancy resulted in a first trimester loss of unknown cause. Chorionic villus sampling of the current pregnancy showed normal karyotype; chromosomal microarray showed a maternally inherited deletion of 552 kb on chromosome 2q12.3 of uncertain clinical significance. The pregnancy was terminated at 14 weeks, and autopsy revealed: renal cystic dysplasia; massively dilated, thin-walled bladder; pulmonary hypoplasia; cloacal dysgenesis with urethral atresia; high anorectal malformation; absence of external genitalia; and absent thumb of the right hand. The couple enrolled in a fetal WES study available through the genetic testing laboratory at Mount Sinai. WES was completed using DNAs isolated from the CVS sample and peripheral blood samples from the couple. A homozygous likely pathogenic variant in *FANCG* was identified in the fetal sample: c.1133C>A; p.S378* (NM_004629.1). Both members of the couple were heterozygous for this variant. Pathogenic variants in *FANCG* cause Fanconi anemia (FA), complementation group G (MIM 614082). Common malformations in this type of FA include renal, radial and gastrointestinal defects (Feben*, et al*, 2013); however, the diagnosis of FA is rarely suspected in the prenatal setting. *FANCL* mutations have been reported in perinatal lethal cases by NGS based testing (Vetro*, et al*, 2015). This case further highlights the utility of WES in the setting of prenatal fetal anomalies. A severe VACTERL-like presentation, particularly with radial and renal anomalies, should alert the clinician to the possibility of FA.

469W

Large genome-wide meta-analysis of age at menopause including X chromosome, gene–environment interactions and Mendelian randomization analysis. *F. Day[1,2].* 1) on behalf of The ReproGen Consortium; 2) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom.

Previous GWAS meta-analysis by the ReproGen consortium in 70,000 women identified 44 genetic loci containing 56 signals that together explain ~6% of population variance in age at natural menopause (ANM). We aimed to identify further low-frequency and common genetic variants associated with ANM, by using imputation to the denser 1000 Genomes reference panel and also for the first-time including chr X. We went on to use these genetic variants to explore gene–environment interactions and the impact of menopause timing on various health outcomes by Mendelian randomization (MR) analysis. We carried out a meta-analysis of 108,000 women from European populations in 35 studies. We identified 123 independent signals at 101 loci (MAF 1–49%; per allele effect -0.1 to -1.4yrs; smallest $P<5x10^{-8}$) including 60 new signals, multiple signals at 14 loci and 4 signals on chr X. Non-synonymous variants were found in LD with 25 signals ($r^2>0.8$). As in previous analyses, DNA repair genes were enriched, including new signals in *FANCA* (per allele effect -0.1yrs; $P=2\times10^{-9}$) and *FANCM* (per allele effect -0.2yrs; $P=1\times10^{-10}$). Other new signals included a non-synonymous variant in *CALCR* (per allele effect -0.1yrs; $P=8\times10^{-9}$) and an intronic variant in *FTO* (per allele effect -0.3 yrs; $P=3\times10^{-8}$) which was an eQTL for *FTO* ($P=2\times10^{-7}$) and was independent of the known BMI signal at this locus. The effect of a ANM genetic risk score (GRS) was higher in women whose mother smoked during pregnancy (P-interaction=9×10^{-4}). From MR analyses, later ANM in women ($P<1\times10^{-6}$) was causally related to: higher bone mineral density (BMD); lower risk of osteoporosis (women only); higher uterine fibroids and female specific cancers (all $P<5\times10^{-5}$). In conclusion, by increasing the sample size and using a denser imputation we have more than doubled the number of known genetic signals for ANM. The resulting GRS explains 8.5% of the population variance in ANM, and provides a powerful instrument for causal modelling analyses. Our results suggest a novel role for FTO in reproductive lifespan. Finally, gene–environment interactions highlight the importance of the uterine environment during oocyte formation, and suggest that *in utero* exposures can modify genetic susceptibility to reproductive lifespan. We anticipate that these analyses will be further reinforced by the imminent inclusion of the final phase of UK Biobank. .

470T

HDP associated gene analysis in Japanese pregnant women in Maternity Log Study. *Y. Tsunemoto[1], T. Yamauchi[1,2], D. Ochi[1,2], K. Misawa[2], K. Kojima[2], T. Mimori[2], F. Katsuoka[2], R. Yamashita[2], T. Shibata[2], N. Minegishi[2], N. Fuse[2], O. Tanabe[2], S. Hiyama[1], M. Nagasaki[2], J. Sugawara[2,3], Maternity Log Study Group.* 1) Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa,Japan 239-8536; 2) Tohoku Medical Megabank Organization, Tohoku University, 2-1,Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8573; 3) Department of Gynecology and Obstetrics, Tohoku University Graduate school of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8574.

Maternity Log Study is a Japanese prospective cohort study that aims to predict or prevent of pregnancy-related diseases by multi layered omics analysis. Approximately 300 pregnant women have been recruited in the first or second trimester of pregnancy. We collected peripheral blood samples of the participants as well as urine samples and many kinds of daily lifelog information. Hypertensive Disorders of Pregnancy (HDP) is a serious pregnancy-related disease that is major cause of maternal and fetal death. It is caused by a complex interaction of genetic factors and environmental factors such as lifestyle. In our study, the proportion of HDP cases was approximately 10%. In this study, we initially focused on the understanding to the genetic factors of HDP towards the future analysis of the complex interaction of both genetic factors and environmental factors. For this purpose, we have obtained hundreds whole-genome sequencing data from the DNA from the participants of Maternity Log Study by using a short-read sequencing technology. Usually, genetic functionalities were affected by variants on the gene. Thus it is conceivable that cases have more deleterious variants than controls on their HDP associated genes. Based on this hypothesis, we have gathered candidates of HDP associated genes from literatures and public databases, and applied Burden-Test to select significant genes among them. In detail, we gathered HDP associated single nucleotide variants (SNVs) from online databases and several researches on HDP meta-analysis. Next, we investigated the genes that contain those SNVs with reference to The Human Gene Mutation Database and approximately a hundred genes were chosen as HDP associated genes. For each gene, we counted all SNVs with high functional impacts, i.e. missense, nonsense, splicing, and frameshift mutation. We then performed Mann–Whitney U test between the total SNVs of cases and controls, and the p-value was adjusted by Benjamini–Hochberg procedure to correct multiple comparisons. Interestingly, almost all HDP associated genes in cases showed more mutations than controls. Additionally, cases have more missense mutations than controls. In conclusion, our study has suggested the significantly enriched genes and their causal relationship to the onset of HDP in Japanese. Towards a health care of pregnant women in Japanese, we are planning to combine these genetic factors and other omics data to construct more accurate prediction model of HDP.

471F

Prenatal evaluation of a fetal cystic hygroma: An unexpected finding of a de novo fetal BRCA1 deletion. *N.S. Seligman[1], S. Laniewski[1], M.A. Iqbal[2].* 1) Obstetrics and Gynecology, University of Rochester Medical Center, Rochester, NY; 2) Pathology & Laboratory Medicine, University of Rochester Medical Center, Rochester, NY.

Objective: Fetal cystic hygromas are associated with an increased risk of aneuploidy, non-chromosomal genetic syndromes, and congenital anomalies. The American College of Obstetricians and Gynecologists recommends chromosomal microarray in patients undergoing prenatal diagnosis for structural fetal anomalies. We report a de novo fetal BRCA1 deletion. Methods: Karyotype was produced using standard G banding techniques. Reflex aCGH was performed on the ISCA 4 × 180K plus SNP v2.0 platform (Agilent Technology Inc. Santa Clara, CA). FISH was performed using SureFISH (Agilent Technology Inc. Santa Clara, CA) and TelVysion (Abott Molecular Inc. Des Plaines, IL). Results: 33 year old G3P0020 at 13w0d was referred for a 6mm septated cystic hygroma. Following genetic counseling, the patient opted for chorionic villus sampling. G-banding revealed a normal 46, XX karyotype. Chromosomal microarray showed an 80kb loss at 17q21.31 from 41186542 to 41266359 basepairs encompassing the BRCA1 gene, confirmed by BRCA1 (41162433-41265378) SureFISH probe. FISH analysis of parental blood samples with BRCA1 probe showed no evidence of deletion 17q21.31 confirming the deletion is de novo in this fetus. Subsequent anatomic ultrasound and fetal echocardiogram revealed no gross anomalies. The pregnancy was complicated by the development of gestational diabetes which was controlled with Glyburide. The patient had an uncomplicated vaginal delivery at 38w4d. Neonatal evaluation was unremarkable. Conclusion: This is the first report of a de novo 80kb fetal BRCA1 deletion. The fetus is a heterozygous carrier of a pathogenic BRCA1 gene deletion and is anticipated to be at increased risk of developing breast, ovarian, and/or pancreatic cancer though the penetrance is still being actively studied. Increased understanding of the BRCA1 gene has shown that it is involved in multiple biological processes. BRCA1 gene deletions have also been implicated as a genetic factor in brain development and neural tube defects (NTD) based possibly though regulation of neurulation. While this fetus did not have spina bifida, NTDs are known to be multifactorial in origin. Close follow-up of neurodevelopment is needed.

472W

Detection of fetal subchromosomal aberration with cell-free DNA screening led to diagnosis of parental translocation: Review of 11344 consecutive cases in a university hospital. *Y.Q. Qian, X.Q. Wang, M. Chen, Y.Q Luo, K. Yan, Y.M. Yang, B. Liu, L.Y. Wang, Y.Z. Huang, H.G. Li, H.Y. Pan, F. Jin, M.Y. Dong.* Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China.

Purpose: Fetal chromosome aberrations and sub-chromosomal copy number variations (CNVs) are not rare. There are several ways to detect duplications and deletions; cell-free DNA screening (cfDNA screening) is nowadays an accurate and safe detection method. The purpose of this study is to report the feasibility of cfDNA screening as an indicator of parental balanced chromosome translocation. **Results:** From February 2015 to March 2016, cfDNA screening was offered to 11344 pregnant women. 137 out of 11344 individuals tested positive for aneuploidies using cfDNA screening were confirmed by karyotyping. 6 additional cases also tested positive for other deletion/duplication were confirmed by chromosomal microarray analysis (CMA). 11201 patients tested negative and 10342 of them were confirmed through interviews after delivery. Among the 137 cases that were screened positive in cfDNA screening, 91 were common trisomies (63 cases of trisomy 21, 25 cases of trisomy 18 and 3 cases of trisomy 13) and 46 cases were positive for sex-chromosomal abnormalities. In addition, 6 cases were positive for other deletion/duplication in which 2 were identified as terminal duplication and deletion on different chromosomes. The cfDNA screening findings were confirmed by CMA or karyotyping, and the origins of CNVs were validated afterward by karyotyping or fluorescence in situ hybridization (FISH) using parental blood samples. **Conclusion:** CfDNA screening may help identify deletions and duplications in fetus, which in some cases may indicate risk of a parent being a balanced rearrangement carrier, and that the diagnostic follow-up testing is necessary.

473T

Fetal cell-free DNA fraction in maternal plasma is affected by fetal trisomy. *N. Suzumori[1], T. Ebara[2], T. Yamada[3], O. Samura[4], J. Yotsumoto[5], M. Nishiyama[6], K. Miura[7], H. Masuzaki[7], Y. Kamei[8], H. Sago[6], J. Murotsuki[9], H. Sawai[10], J-S. Saldivar[11], N. Dharajiya[11], A. Sekizawa[12].* 1) Department of Obstetrics and Gynecology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Occupational and Environmental Health, Nagoya City University Graduate School of Medical Sciences, Nagoya; 3) Department of Obstetrics and Gynecology, Kyoto University Graduate School of Medicine, Kyoto; 4) Department of Obstetrics and Gynecology, The Jikei University School of Medicine, Tokyo; 5) Department of Genetic Counseling, Ochanomizu University, Tokyo; 6) Center of Maternal-Fetal, Neonatal and Reproductive Medicine, National Center for Child Health and Development, Tokyo; 7) Department of Obstetrics and Gynecology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki; 8) Department of Obstetrics and Gynecology, Saitama Medical University, Saitama; 9) Department of Obstetrics and Gynecology, Tohoku University Graduate School of Medicine, Sendai; 10) Department of Obstetrics and Gynecology, Hyogo College of Medicine, Nishinomiya; 11) Sequenom Laboratories, San Diego, CA, USA; 12) Department of Obstetrics and Gynecology, Showa University School of Medicine, Tokyo.

   The purpose of this noninvasive prenatal testing (NIPT) study was to compare the fetal fraction of singleton gestations by gestational age, maternal characteristics, and chromosome specific aneuploidies as indicated by *z*-scores. This study was a multicenter prospective cohort study. Test data were collected from women who underwent NIPT. Relationships between fetal fractions and gestational age, maternal weight and height, and *z*-scores for chromosomes 21, 18, and 13 were assessed. A total of 7,740 pregnant women enrolled in the study of which, 6,993 met the study criteria. Approximately 95.5% of this study cohort included women of 35 years age or older. As expected, fetal fraction was inversely correlated with maternal weight ($p<0.001$). The median fetal fraction of samples with euploid result (n=6,850) and trisomy 21(n=70) were 13.7% and 13.6% respectively. In contrast, median fetal fraction values for samples trisomy 18 (n=35) and 13 (n=9) were 11.0% and 8.0% respectively. The fetal fraction of samples with trisomy 21 NIPT result is comparable to that of samples with euploid result. However, the fetal fractions of samples with trisomy 13 and 18 is significantly lower compared to that of euploid result, which may make detecting these two trisomies more challenging.

474F

Uncovering novel cytogenetic and molecular etiologies for infertility. *S.L.P. Schilit[1,2], T. Kammin[3], E. Wilch[3], C. Hanscom[4], C. Redin[2,4,5], D.V.S. Sudhakar[6], K. Thangaraj[6], C.B. Gurumurthy[7], A.J. MacQueen[8], J.F. Gusella[2,4,5,9,10], M.E. Talkowski[2,4,5,10,11,12], C.C. Morton[2,3,10,13,14].* 1) Program in Genetics and Genomics and Certificate Program in Leder Human Biology and Translational Medicine, Biological and Biomedical Sciences Program, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA, USA; 2) Harvard Medical School, Boston, MA, USA; 3) Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Boston, MA, USA; 4) Molecular Neurogenetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 6) Centre for Cellular and Molecular Biology, Hyderabad, India; 7) Developmental Neuroscience, Munroe Meyer Institute, Mouse Genome Engineering Core Facility, University of Nebraska Medical Center, Omaha, NE, USA; 8) Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, CT, USA; 9) Department of Genetics, Harvard Medical School, Boston, MA, USA; 10) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 11) Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; 12) Department of Pathology, Massachusetts General Hospital, Boston, MA, USA; 13) Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; 14) Manchester Academic Health Science Centre, University of Manchester, Manchester, UK.

   Unexplained infertility affects 2-3% of reproductive aged couples. One approach to identifying genes involved in infertility is to study subjects with a clinical phenotype accompanied by a *de novo* balanced chromosomal aberration (BCA) through the Developmental Genome Anatomy Project (DGAP). DGAP230 has oligospermia and 46,XY,t(20;22)(q13.3;q11.2). While BCAs may reduce fertility by production of unbalanced gametes, a chromosomal rearrangement may also disrupt or dysregulate genes important in fertility. By large-insert whole genome sequencing, we refined the location of chromosomal breakpoints in DGAP230. A qPCR walk across all genes in the topologically associated domains (TADs) at the sites of rearrangement revealed exclusive dysregulation of *SYCP2* (overexpressed 20-fold by RNA [N=3; p<0.02] and five-fold by protein [N=3; p<0.0032]), which resides 1.5 Mb away from the der(20) breakpoint. By comparing sequence traces of DGAP230 DNA and RNA from a polymorphic region in *SYCP2*, we found that this misexpression derives from a single allele. We developed a novel chromatin conformation capture and nested PCR technique and found that the expressed allele resides in *cis* with the der(20) breakpoint. 4C-Seq is being used to identify putative enhancers from chr22 that may be responsible for the misexpression. *SYCP2* encodes synaptonemal complex protein 2 (MIM 604105), a member of the synaptonemal complex (SC) involved in homologous chromosome synapsis in meiosis I. We hypothesize that *SYCP2* misexpression may impair proper spermatogenesis by meiotic nondisjunction. To model the impact on meiosis, we misexpressed *RED1*, the functional axial element homolog of *SYCP2* in *S. cerevisiae*, using a β-estradiol-induced promoter system. By performing Red1 immunolocalization on surface-spread meiotic nuclei, we discovered that excess Red1 leads to formation of polycomplexes. This disrupts the structural integrity of the SC, leading to failed incorporation of the transverse filament Zip1 (N=50; p<0.0001). In support of the hypothesis that genetic variants causing infertility would be eliminated from the general population, we found a depletion of *SYCP2* loss of function variants (by 90%; pLI = 1.00) and coiled-coil domain missense mutations (by 85%; p<0.04), prompting ongoing genetic case-control human studies. In addition to pinpointing a putative genetic etiology for DGAP230's oligospermia, our study more generally illustrates an additional mechanism for BCA-mediated infertility.

475W

Differential miR-346 and miR-582-3p expression in association with selected maternal and fetal complications. *M. Su, P. Tsai, H. Tsai.* National Cheng-Kung University Hospital, Tainan City, TaiwanDepartment of Obstetrics and Gynecology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan.

    Introduction: Several miRNAs are expressed in human gestational tissue, and some have been shown to be associated with placental dysfunction and complicated pregnancy outcomes. **Objectives:** To investigate the roles of miR-346 and miR-582-3p in adverse obstetric events. **Patients & Methods:** We analyzed these 2 miRNAs in three samples (maternal blood, umbilical cord blood and placenta) obtained from pregnant women in five groups, including healthy control (n=60), preeclampsia (n=29), preterm delivery (n=38), preterm premature rupture of the membranes (n=12) and small for gestational age (n=19) patients. **Results:** The expression levels of miR-346 and miR-582-3p in all included adverse obstetric outcome groups were significantly higher in the maternal plasma samples but lower in the placenta samples (all p value <0.05). In addition, the miR-346 expression levels in fetal cord blood were also significantly lower in all of the included adverse obstetric outcome groups (all p<0.05). Subgroup analysis of the three specimens from healthy and compromised term pregnancies (>37 weeks) gave the same results. **Conclusion:** Aberrant miR-346 and miR-582-3p expression level in pregnancy was associated with multiple maternal and fetal complications. Their differential expression in maternal blood, umbilical cord blood and placenta could be potential biomarkers or therapeutic targets for adverse obstetric outcomes.

476T

Challenges associated with increasing the predictive power of AMH in controlled ovarian stimulation. *S.E. Parets, R. Shraga, M. Berliss, S. Munne, A. Bisignano, O. Puig.* Phosphorus, New York, NY.

    Reproductive endocrinologists rely on clinical factors such as age and anti-mullerian hormone (AMH) to assess ovarian reserve and counsel patients regarding expectations for controlled ovarian stimulation (COS). Although AMH has emerged as an accurate predictor of ovarian response, previous studies have identified genetic variants that influence COS outcomes. Our objective was to determine whether the addition of genotypes increases the power of AMH to predict ovarian response. We retrospectively enrolled from women between the ages of 18 and 42 undergoing COS (N=369) at five fertility centers. Patients with a history of polycystic ovarian syndrome, cancer, chemotherapy, or bone marrow transplant were excluded. We used an Illumina microarray panel comprised of genes previously associated with infertility. Response categories were defined as low response (<5 oocytes retrieved) and high response (>20 oocytes retrieved). A single genotype association analysis (N=217) was performed by transforming the number of oocytes retrieved into a binary variable and fitting to a series of logistic regression models, with AMH and individual genotypes as the dependent variables. To create a polygenic predictive model, samples were divided into discovery (70%) and validation (30%) sets. We used the discovery set to rank genotypes by predictive capacity in three ways: associated p-value from the single genotype analysis; calculated p-value from subsets of data using a bootstrapping procedure; and a random order as a negative control. Logistic regression models were trained on the discovery set using these ranked lists and model performance was measured on the validation set using the area under the ROC curve. The average clinical characteristics of participants reflected the general population seeking fertility treatment. Single genotype analysis yielded no significant results after multiple test correction (p >5.7x10[-3]). For the polygenic predictive model, genotypes failed to add predictive value to AMH alone. Therefore, the genotypes tested did not increase the predictive power of AMH in COS. Although we were unable to detect significance in this pilot study, previous research has identified genetic factors that influence ovarian response. Our results reflect decreased power related to heterogeneous phenotypes and a small sample size. Further investigation may clarify the role of genetics in infertility and improve COS outcomes.

477F

Association of multiple TLR6-TLR1-TLR10 gene cluster SNPs with preterm birth in a Wisconsin cohort suggests a potential evolutionary selection bias. *C. Hoffman[1], T. Ranade[1], W. Luo[2], J. Eickhoff[3], B.R. Pattnaik[2,4], M. Baker[2,5], D.-A.M. Pillers[2,6,7].* 1) University of Wisconsin, Madison, WI; 2) Department of Pediatrics, University of Wisconsin, Madison, WI; 3) Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, WI; 4) Department of Ophthalmology & Vision Sciences, University of Wisconsin, Madison, WI; 5) Wisconsin State Laboratory of Hygiene, Madison, WI; 6) Department of Medical Genetics, Laboratory of Genetics, Crow Institute for Evolution, University of Wisconsin, Madison, WI; 7) Department of Pediatrics, University of Illinois, Chicago, IL.

Background: Preterm birth (PTB) is the leading cause of neonatal and childhood mortality and affects 11% of live births. Infection is a major causative factor for PTB. The first line of defense against infection is the innate immune system, and the portal of entry is the Toll-like receptor (TLR). TLRs are pattern-recognition receptors that recognize pathogen-associated molecular patterns (PAMPs). Genetic variation including SNPs in TLRs is associated with susceptibility to infectious diseases. We speculated that SNPs in fetal TLR genes may be associated with PTB. We have previously reported results for SNPs in three TLR receptors (TLR1, 2, and 4). We found that *TLR1* SNPs (N248S, H305L, and S602I) in fetal DNA are associated with PTB in Wisconsin infants. Interestingly, TLR1 is part of the TLR6-TLR1-TLR10 gene cluster reported to be associated with inflammation and infection. Purpose: We sought to determine whether SNPs in other TLR genes in the TLR6-TLR1-TLR10 cluster are associated with preterm birth. We hypothesized that *TLR6* SNP rs5743810 (P249S) and *TLR10* SNP rs11096957 (N241H) also associate with an increased risk for preterm birth. Methods: We used DNA samples from residual newborn screening specimens from Wisconsin infants of gestational ages (GA), 23-42 wks. IRB approval was secured. We collected 3000 DNA samples. To date, 194 of these samples have been subjected to *TLR6* SNP (P249S) analysis, and 581 of them for *TLR10* SNP (N241H) analysis. The SNP genotype was determined using standard TaqMan SNP Genotyping Assays (Lifetechnologies, CA). An Odds ratio (OR) analysis was used to determine whether a SNP predicted preterm birth. Results: *TLR6* SNP (P249S) minor allele T was associated with PTB at GA < 34 wks with OR 2.33 (P < 0.00006), and at GA< 37 wks with OR 2.22 (P < 0.00016). In contrast, there was no association of *TLR10* SNP (N241H) with PTB. The OR for *TLR10* SNP (N241H) minor allele C at GA < 28 wks was 0.95 (P = 0.84), at GA < 34 wks was 0.97 (P = 0.83), and at GA < 37 wks was 0.83 (P = 0.13). Discussion: We found a significant association between *TLR6* SNP (P249S) and preterm birth in a Wisconsin newborn cohort. We have shown that multiple SNPs in the TLR6-TLR1-TLR10 gene cluster are associated with PTB, suggesting potential involvement of genetic variation within this cluster in the pathogenesis of preterm birth. The association with this linkage group also suggests that a potential evolutionary selection bias is contributing to preterm birth.

478W

Off the street phasing (OTSP): Free no hassle haplotype phasing for molecular PGD applications. *G. Altarescu[1], F. Zahdeh[1], Y. Kling[2], P. Renbaum[1], E. Levy-Lahad[1], S. Carmi[2], D.A. Zeevi[1].* 1) Preimplantation Genetic Unit, Medical Genetics, Shaare Zedek Medical Center, Jerusalem, Israel; 2) 2 Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel.

Introduction: Parental phasing of mutation-flanking haplotypes is an essential, yet time-consuming, laborious, and costly pre-requisite for preimplantation genetic diagnosis (PGD) of monogenic disorders. **Aim**: To validate rapid, low cost, population-assisted haplotype derivation in a pre-clinical setting. **Materials and Methods:** Targeted sequencing of *CFTR* variants and gene-flanking polymorphic SNPs in 38 Jewish individuals from 9 different PGD families was performed at the SZMC PGD lab. Heterozygous genotype calls were both trio-phased to obtain ground-truth haplotypes, and also population-phased using Shapeit software. Reference panels for population phasing were derived from either '1000 Genomes' or from 128 or 574 sample Ashkenazi Jewish whole genome sequences .Accuracy of resulting haplotypes was benchmarked against trio-phased haplotypes. The study population consisted of 4 subgroups of Jewish individuals: a) Non-Ashkenazi (NA; 10 negative controls); b) Full Ashkenazi (FA; 10 individuals); c) Partial Ashkenazi (13 individuals featuring at least 1 Ashkenazi and 1 non-Ashkenazi ancestor) without Ashkenazi *CFTR* founder mutation (PANM); and d) Partial Ashkenazi (5 individuals) with W1282X CFTR Ashkenazi founder mutation (PAWM). **Results:** The574 sample Ashkenazi genome reference was the most accurate and appropriate for population-based phasing. Virtually all phase errors in the FA group (and the PAWM group, along a subregion of 3Mb) were traceable to low-coverage sequencing errors in the ground-truth. **Discussion:** These results indicate that it ispossible to replace experimental haplotype phasing with clinical "OTSP," population-based phasing, provided that one has access to an appropriate population-matched reference dataset of sufficient size.

Study subgroup (Abbreviation; No. individuals)	1000 Genomes Avg. No. Heterozygous SNPs per individual (+/-SEM)	1000 Genomes Phasing Accuracy (+/-SEM)	128 Ashkenazi Ref. Avg. No. Heterozygous SNPs per individual (+/- SEM)	128 Ashkenazi Ref. Phasing Accuracy (+/-SEM)	574 Ashkenazi Ref. Avg. No. Heterozygous SNPs per individual (+/-SEM)	574 Ashkenazi Ref. Phasing Accuracy (+/-SEM)
NA;10	589.4+/-66.2	71.0%+/-4.2%	519.9+/-57.3	72.7%+/-4.2%	579.4+/-62.3	84.9%+/-3.3%
PANM; 13 No CFTR mutation	544.5+/-35.9	73.1%+/-4.3%	484.9+/-32.8	78.9%+/-4.9%	547.8+/-34.7	86.7%+/-5.2%
PAWM; No.5 CFTR mutation	741.2+/-48.7	70.9%+/-4.0%	665.6+/-40.3	94.3%+/-2.2%	732.8+/-47.0	**91.4%**+/-5.3%
FA; 10	626.5+/-48.5	65.2%+/-6.4%	559.4+/-42.0	88.4%+/-5.5%	621.4+/-47.4	**98.2%**+/-1.1%

479T

Novel genes for male infertility: Genetic and functional perspectives. *D.V.S. Sudhakar[1], S. Nizamuddin[1], R. Phanindranath[1], R. Singh[2], N.J. Gupta[3], R. Dada[4], K. Singh[5], M. Deenadayal[6], K. Thangaraj[1].* 1) Centre for Cellular and Molecular Biology, Hyderabad, India; 2) Central Drug Research Institute, Lucknow, India; 3) Institute of Reproductive Medicine, Kolkata, India; 4) All India Institute of Medical Sciences, New Delhi, India; 5) Banaras Hindu University, Varanasi, India; 6) Infertility Institute and Research Center, Hyderabad.

   Introduction: Approximately one out of every seven couples are infertile worldwide and the male factor infertility accounts for 30-50% of the infertility cases. Previous studies from our lab have shown that about 8.5% Indian men are infertile due to Y chromosome microdeletions. Further, analysis of a few autosomal genes and mitochondrial genome accounted for additional 21% of the genetic factors responsible for infertility among Indian men. However, etiology of large proportion (71%) of infertile men still remained unknown. Therefore, we have performed exome sequencing to identify novel autosomal genetic causes of male infertility. **Materials and methods:** We have performed exome sequencing of 44 idiopathic infertile men (azoospermia and severe oligozoospermia) using Illumina Hiseq-2000 platform with 100X coverage. Various bioinformatics tools and custom pipelines were used for data analysis and prioritizing the variants. We have identified 61 novel and rare variants (from 58 genes) and genotyped them in 740 cases and 700 ethnically matched controls. We have identified 14 variants (12 genes), which are strongly associated with infertility. Further, we have genotyped 14 additional damaging variants from the above 8 genes in 960 individuals. Expression pattern of all the 8 genes in various human tissues was studied using Western blotting. Of the novel genes identified, *CETN1* has been found to be a potential candidate; hence variants of this gene (one 5'UTR variant and one missense variant) were further genotyped using Taqman® probes in 552 samples, characterized using biophysical (CD, ITC and NMR etc.), cell biology approaches, luciferase reporter assays and tubulin polymerization assays. **Results:** We have identified a total of 21 novel, rare, nonsense and missense variants from 10 genes. Of which, rs189882379 (*ANKRD7*) and rs61734344 (*CETN1*) showed highly significant association with male infertility *(p=3.455e-34 and* p=1.658e-35 respectively)*. Multiple mutations in combination were found in azoospermic men. Some mutations were exclusvely found in infertile men. Functional studies have shown that rs61734344 (p. Met72Thr) alters calcium binding affinity, thermodynamic properties, surface hydrophobicity and disrupts spindle formation during cell division; suggesting its role in infertility. 5'UTR variant rs367716858 (*CETN1*) has abolished a methylation site and showed increased luciferase activity compared to wild type. .

480F

Modeled fetal disease risk of a 175 condition expanded carrier screening panel. *K.A. Beauchamp, K.E. Kaseniit, P. Grauman, G.J. Hogan, K.R. Haas, G.M. Gould, K.K. Wong, G. Lazarin, E.A. Evans, D. Muzzey.* Counsyl, Inc., South San Francisco, CA.

   Introduction: Expanded carrier screening (ECS) identifies couples whose future children are at increased risk of Mendelian conditions. Recent ACOG guidelines identified ECS as an acceptable screening strategy and proposed general principles for selecting panel diseases. Inspired by these guidelines, here we use modeled fetal disease risk (MFDR) to predict the clinical sensitivity of a new 175-condition NGS-based ECS panel that also incorporates panel-wide copy number variant (CNV) deletion calling. Methods: Pathogenic allele frequencies were estimated using a cohort of 9,150 consecutive anonymized patients. Allele frequencies also included CNV deletion calling on most genes; CNV duplications are also called in DMD and CFTR. Copy number calling was performed using a Hidden Markov Model on NGS depth data. The MFDR, defined as the probability that a random fetus is affected by a panel disease, was estimated for a US-census weighted population. Results: A previous 94 condition panel was found to assess 183 affected fetuses per 100,000 births, with 10% attributed to the 78 least-prevalent conditions and 29% attributed to technically challenging conditions (e.g., fragile X syndrome). The present panel assesses approximately 350±50 affected fetuses per 100,000 births. The 146 least prevalent conditions contribute approximately 11% of the total risk, while technically challenging conditions contribute 13% of the total risk. Furthermore, CNV deletion and duplication calling contributes 16% of the total risk, most of which is attributed to DMD. Conclusions: Using a 9,150 patient cohort, we evaluated the impact of augmenting an existing ECS panel to include panel-wide deletion CNV calling and 81 additional conditions. The expanded panel suggests that the long-tail of rarer diseases, technically challenging conditions, and copy number variants each contribute substantially to the clinical sensitivity of ECS.

481W

Analysis of blood plasma metabolomic profile analysis of pregnancy and the association analysis with pregnancy-related diseases in Maternity Log Study. *Y. Harada[1], D. Saigusa[1], R. Yamashita[1], D. Ochi[2,1], T. Yamauchi[2,1], Y. Tsunemoto[2], S. Koshiba[1], O. Tanabe[1], S. Hiyama[2], M. Nagasaki[1], J. Sugawara[1,3], G. Maternity Log Study Group[1].* 1) Tohoku University, Sendai, Japan; 2) NTT DOCOMO, INC., Yokosuka, Japan; 3) Tohoku University Graduate school of Medicine, Sendai, Japan.

In Japan, about 200,000 people are affected by pregnancy-related diseases including hypertensive disorders of pregnancy, gestational diabetes and premature deliveries each year. Pregnancy-related diseases are multifactorial diseases caused by a complex interaction of genetic factors and environmental factors such as lifestyle and living environments. Identification of substances in biospecimens such as blood and urine that reflect pathophysiological conditions during pregnancy is important to develop methods for early prediction of these diseases. For these objectives, we are conducting a prospective cohort study "Maternity Log Study" in which biospecimens and health information are collected from pregnant women. Three hundred participants have been recruited at Tohoku University Hospital, and have donated their blood, urine, saliva, and dental plaque samples for multi-omics analyses, i.e. genomic, transcriptomic, proteomic, metabolomic, and microbiomic analyses, but also reported daily health information through smartphones such as grades of morning sickness, abdominal pain, and uterine contractions, as well as physiological data such as body weight, blood pressure, heart rate, and body temperature. Metabolite analysis is a powerful methodology for biomarker discovery and elucidation of disease pathophysiologies. It is well known that maternal physiological conditions change drastically during pregnancy, but their relationship with genetic factors and changes in metabolomic conditions are still unclear. In this study, we focused on pregnant women's metabolic profile analysis. Blood plasma samples were collected twice during pregnancy and once after birth. 279 metabolomic profiles of 95 pregnant women's blood plasma were acquired using gas chromatography-mass spectrometry (GC-MS) employing chemical derivatization pretreatment. We detected 143 metabolites, of which 83 were detected from all subjected samples. Now we have been analyzing the relationship between these metabolites and the pregnancy-related diseases, and we have a plan to report it in our presentation.

482T

Newborn metabolomics: Accurate prediction of gestational age from cord blood. *E. Jasper[1], S. Oltman[2,3], B. Bedell[4], J. Dagle[4], P. Jagannathan[5], T. Clark[6], G. Dorsey[6], E. Rogers[7], L. Jelliffe-Pawlowski[2,3], K. Ryckman[1,4].* 1) Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA; 2) Department of Epidemiology and Biostatistics, School of Medicine, University of California San Francisco, San Francisco, CA; 3) University of California San Francisco (UCSF) Preterm Birth Initiative, San Francisco, CA; 4) Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA; 5) Department of Medicine (Infectious Diseases), Stanford University Medical Center, Stanford, CA; 6) Department of Medicine, School of Medicine, University of California San Francisco, San Francisco, CA; 7) Department of Pediatrics, School of Medicine, University of California San Francisco, San Francisco, CA.

Background: Accurate estimation of gestational age is critical for clinical care decision of newborns. While prenatal ultrasound dating is the most accurate method for estimating gestational age, it is not feasible in certain settings. Previous research has demonstrated that metabolic and endocrine markers, captured in routine newborn screening, may be an accurate alternative method for estimating gestational age after birth. This research aimed to replicate this finding using metabolites from banked cord blood. **Methods:** Between 2001 and 2016, 938 newborns were enrolled at the University of Iowa Hospitals and Clinics. Cord blood was obtained from the infants and banked before being spotted on 903 Whatman filter paper. These samples were then analyzed by the State Hygienic Laboratory in Iowa. Forty-seven metabolites, including biomarkers of amino acid and fatty acid metabolism, thyroid-stimulating hormone, and 17-hydroxyprogesterone, were examined. Multiple linear regression modeling was performed, with gestational age, in weeks, as the outcome measure. Coefficients of determination and root-mean-square error were utilized to evaluate models. The best model was subsequently used to predict gestational age of newborns in a small, independent, East African population. **Results:** An algorithm containing metabolites and birthweight was better at discriminating between preterm and term infants compared to a model that relied solely on birthweight (AUC: 0.9877 vs. 0.9679). The top prediction algorithm contained 37 variables, including 17 metabolites, 12 squared metabolite terms, 6 cubed metabolites, and birthweight. When the algorithm was applied to the East African population, it was able to predict gestational age within 2 weeks for 83% of infants. When differentiating between infants born prematurely versus those born at term, only one premature infant was misclassified; whereas no control (term) infant was incorrectly classified (sensitivity: 83.33%, specificity: 100%). **Conclusions:** The newborn metabolic profile, derived from cord blood, is an accurate method for estimating gestational age. The model utilizing metabolites and birthweight predicts gestational age to a better degree than newborn birthweight alone. Newborn metabolic screening via cord blood may be an effective method for determining the gestational age of infants at birth, particularly in settings where prenatal ultrasound measurements are infeasible.

483F

Bacterial taxonomic analysis of oral microbiome in spontaneous preterm birth for Maternity Log Study. *D. Ochi[1,2], T. Yamauchi[1,2], R. Yamashita[2], Y. Tsunemoto[2], M. Wagata[2], Y. Harada[2], S. Ogishima[2], O. Tanabe[2], S. Hiyama[1], M. Nagasaki[2], J. Sugawara[2,3], Maternity Log Study Group.* 1) Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536; 2) Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8573; 3) Department of Gynecology and Obstetrics, Tohoku University Graduate school of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8574.

Objective: Pregnancy related diseases, including preterm birth (PTB) are caused by complex interactions of genetic and environmental factors such as lifestyle and living environments. We are conducting a prospective cohort study named Maternity Log Study to investigate unrevealed mechanisms of those diseases. Three hundred pregnant women were recruited at Tohoku University Hospital. We collected 900 dental plaque samples and saliva samples from the participants at several time points during and after pregnancy, as well as blood and urine samples, and daily lifelogs of physiological data and symptoms to analyze the complex interaction of genetic and environmental factors. Several previous studies showed relationship between periodontitis and preterm birth. Despite the critical role of microbiota as a part of environmental factors, the dynamic composition during pregnancy is still unclear. We have focused on bacterial taxonomic analysis of maternal dental plaque and investigated the possible relationship with spontaneous PTB (SPTB). **Methods:** Maternal dental plaque specimens were collected twice during pregnancy (first between 12 and 23 weeks of gestation, and second between 24 and 34 weeks) and once after birth from 15 women who spontaneously delivered preterm and 130 women with term birth. Species in plaque were determined by sequencing the 16S rRNA genes. We compared the relative bacterial abundances between the SPTB case group and the term-birth control group. **Results**: In the analysis of dental plaque collected between 12 and 23 weeks of gestation, a set of species in the SPTB group showed significantly lower relative abundances compared to the control group. However, in the samples collected later between 24 and 34 weeks, the difference in the abundances of these species between the two groups was not significant, because of the higher relative abundances of the set of species in the SPTB group at this period compared to the earlier period. **Discussions:** The relative abundances of the set of species in the SPTB group changed dynamically, although those in the control group were rather stable. It may reflect dynamic changes in the maternal physiological condition during pregnancy. We will further investigate the physiological changes by multi-layered omics analyses of blood and urine samples as well as lifelog analysis in addition to the changes in oral microbiome including the set of species, and thereby establish a prediction model for SPTB.

484W

Maternal oral microbiota profile associated with hypertensive disorders of pregnancy. *T. Yamauchi[1,2], D. Ochi[1,2], R. Yamashita[2], M. Wagata[2], J. Kawashima[2], Y. Tsunemoto[1], Y. Harada[2], T. Takai[2], O. Tanabe[2], S. Hiyama[1], M. Nagasaki[1], J. Sugawara[2,3], Maternity Log Study Group.* 1) Research Laboratories, NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka, Kanagawa, Japan 239-8536; 2) Tohoku Medical Megabank Organization, Tohoku University, 2-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8573; 3) Department of Gynecology and Obstetrics, Tohoku University Graduate school of Medicine, 1-1, Seiryo-machi, Aoba-ku, Sendai, Miyagi, Japan 980-8574.

Objective: Despite the seriousness of pregnancy complications including hypertensive disorders of pregnancy (HDP), which often causes maternal and fetal death, the pathogenesis and progression have not been well elucidated. Pregnancy complications are multifactor diseases caused by a complex interaction of genetic and environmental factors. In addition, it has been reported that human microbiota is involved in the onset. Several studies have revealed causal relationship between vaginal microbiota during pregnancy and preterm birth. However, there have been few studies on other body sites and other pregnancy complications. Oral microbiota affects not only caries and periodontal diseases, but also systemic diseases such as cardiovascular diseases. Therefore, it is conceivable that oral bacteria may also affect pregnancy complications. In this research, we analyzed maternal oral microbiota profile to clear the association between the microbiota and HDP. **Methods**: We used the dental plaque collected in Maternity Log Study, which is a Japanese prospective cohort study for pregnant women. In this study, pregnant women were recruited in the first or second stage of pregnancy at Tohoku University Hospital. Study participants uploaded daily physiological data such as blood pressure and body weight. Furthermore, this study collected maternal dental plaque, blood, and urine. Plaque specimens gathered twice during pregnancy, and once in about one month after birth. Using approximately 900 plaque specimens from over 300 women, we estimated microbiota in plaque specimens by 16S rRNA gene sequencing. Based on relative abundances of genus or species per specimen, we performed principal component analysis (PCA). **Results**: PCA showed that cases and controls formed different genus clusters, especially, there was a significant difference between severe HDP cases and controls. *Rothia* was one of the genus greatly contributing to PC1. **Discussions**: We demonstrated new findings maternal oral microbiota profile associated HDP. *Rothia dentocariosa* is a common species of the human oral cavity. In the previous reports, *Rothia dentocariosa* was found in the blood culture of deceased fetuses and the amniotic fluid of pregnant women with preterm pre-labor rupture of membranes. Our results indicate a possible relationship between the oral microbiome and these pregnancy complications. Further research may lead to elucidate the detailed mechanism and prevent the onset of HDP.

485T

Prenatal screening for 22q11.2 deletions using a targeted microarray-based cell-free DNA (cfDNA) test. *J. Shabbeer[1], M. Schmid[1], E. Wang[1], P. Bogard[1], J. Zahn[1], C. Hacker[1], S. Wang[1], J. Doshi[1], K. White[1], J. Kaplan[1], A. Sparks[1], J. Jani[2], R. Stokowski[1].* 1) Ariosa Diagnostics Inc., Roche Sequencing Solutions Inc; 2) Department of Obstetrics and Gynecology, University Hospital Brugmann, Université Libre de Bruxelles, Brussels, Belgium.

Objective: Cell-free DNA (cfDNA) testing provides high sensitivity for common trisomies and a lower false positive rate than traditional prenatal screening methods. Recently, the scope of testing has expanded to include subchromosomal deletions. Here we aim to determine the performance of a targeted microarray-based cfDNA test (Harmony Prenatal Test®) to screen for pregnancies at increased risk for a 22q11.2 deletion. **Methods:** Test performance was determined in two steps including a total of 1,953 plasma samples. Analytical validation was performed in 1736 plasma samples. Clinical verification of performance was performed in an additional 217 prospectively ascertained samples from pregnancies with fetal deletion status determined by diagnostic testing. **Results:** Analytical sensitivity was 75.4% (95% CI: 67.1-82.2%) based on 122 samples with deletions ranging from 1.96 to 3.25 Mb. In 1614 presumed unaffected samples, specificity was determined to be at least 99.5% (95% CI: 99.0-99.7%). In the clinical cohort, 5 of 7 samples from pregnancies affected with 22q11.2 deletion were determined to have a high probability of deletion. There were no false positive results in the 210 unaffected samples in this cohort. This clinical data is consistent with the performance demonstrated in the analytical validation. **Conclusions:** Cell-free DNA testing using targeted microarray quantitation is able to identify pregnancies at increased risk for 22q11.2 deletions of 3.0 Mb and smaller while maintaining a low false positive rate.

486F

Correlation between nutrigenetic variants, adherence to the Mediterranean diet and obesity in male infertility. *M. La Rovere[1], M. Franzago[1], M. De Santo[2], GM. Tiboni[3], P. Piomboni[4], L. Stuppia[1].* 1) Laboratory of Molecular Genetics, Department of Science, Psychology of Health and Territory. "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy; 2) Private nursing home Synergo Pierangeli, Spatocco, Chieti, Italy; 3) Department of Medicine and Aging Sciences, "G. d'Annunzio" University of Chieti-Pescara, Chieti, Italy; 4) Department of Molecular Medicine and Development, University of Siena, Policlinico Le Scotte, Siena, Italy.

It is estimated that 15-20% of couples in industrialized countries have fertility problems. This percentage is expected to grow under the influence of genetic and environmental factors. Many studies have stressed the correlation between male nutrition and infertility, and obesity is also closely related to infertility, as several studies reported that males with high BMI have a reduced reproductive potential and a progressive decline in seminal parameters. The purpose of this study was to evaluate the relationship between seminal parameters and 9 nutrigenetic variants involved in food metabolism (namely: *PPARG2* rs1801282; *PPARGC1A* rs8192678; *TCF7L2* rs7903146; *LDLR* rs2228671; *MTHFR* rs1801133; *APOA5* rs662799; *GCKR* rs1260326; *FTO* rs9939609 and *MC4R* rs17782313) in 120 male subjects of infertile couple undergoing ART. In particular, 49 subjects had normal weight (BMI 18.5-24.99), 56 were overweight (BMI 25-29.99) and 15 were obese (BMI> 30). According to the sperm count and morphology, each group was further subdivided in i) subjects with altered spermatogenesis (mild oligo/astheno/teratozoospermia) and ii) subjects with a normal spermiogram. Adherence to the Mediterranean Diet has been evaluated through the MedDiet software. A significant association between obesity and variants in *FTO* gene has been identified. Indeed, the frequency of the AA genotype in *FTO* was higher in obese (46.7%) than in normal (30.6%) and overweight (17.9%) subjects (p<0.05). In addition, also the frequency of CC genotype in *MC4R* was higher in obese (13.3%) than in normal (6.1%) and overweight (5.4%) subjects, but this difference was not statistically significant. Moreover, a possible correlation between the aforementioned polymorphisms and a worsening of the phenotypic picture of the spermiogram can be hypothesized, since in the group of overweight and obese subjects a greater frequency of FTO AA and MC4R CC genotypes was observed in subjects with an altered spermiogram compared to those with normal spermiogram (AA: 25% vs. 20%, CC: 11.1% vs 2.9%). However, also in this case the difference was not statistically significant, likely due to the limited number of cases analyzed. In conclusion, our data supports an association between FTO and BMI. On the other hand, a direct relationship between FTO and MC4R variants with seminal parameters in obese subjects can not be confirmed, suggesting that the link between obesity and infertility is not related to common genes of susceptibility.

487W

Knowledge and attitudes on non-invasive prenatal pharmacogenetic testing among pregnant and preconception women. *M.J. Ross[1], H. Naik[1], S.A. Suckiel[1], E. Boniferro[2], E. Bazinet[2], Z. Nelson[2], D. Reed[3], S.A. Scott[1,4].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Department of Obstetrics and Gynecology, Mount Sinai West, New York, NY 10019; 3) Department of Obstetrics and Gynecology, Tufts Medical Center, Boston, MA 02111; 4) Sema4, a Mount Sinai venture, Stamford, CT 06902.

Drug metabolism and transport during pregnancy is an understudied area, as the potential teratogenic effects of most commonly prescribed medications are currently unknown. The safety profiles of these medications are likely influenced, in part, by maternal, placental and fetal pharmacogenetic variability, suggesting that pharmacogenetic-guided drug selection and/or dosing during pregnancy may be possible in the future. Moreover, fetal and maternal pharmacogenetic genotyping may be feasible with non-invasive prenatal screening (NIPS) using cell-free fetal DNA (cffDNA) derived from maternal peripheral blood. Although non-invasive prenatal pharmacogenetic testing (NIPPT) is not currently available, we sought to determine the interest and attitudes of pregnant and preconception women regarding NIPPT. Women presenting for prenatal and preconception counseling were recruited from two Mount Sinai reproductive genetic counseling clinics (n=114) and administered an anonymous survey. The majority of subjects (n=86; 75.4%) were previously unaware of the availability of pharmacogenetic testing; however, 91 participants (80%) reported an interest in NIPPT, which was comparable to their interest in NIPS for chromosomal abnormalities (86%). Greater interest in NIPPT was associated with previous awareness of NIPS (p=0.031), a history of adverse drug reactions (p=0.021), and general interest in pharmacogenetic testing (p<0.001). The motivation for wanting to pursue NIPPT was driven by potential benefits related to medication efficacy and reduced adverse drug effects for both mother and fetus. However, the main concerns with NIPPT were that results would impact health insurance eligibility and that physicians would be unable to identify medications suitable for both mother and fetus. Taken together, these results indicate that pregnant women are interested in pursuing NIPS for prenatal applications beyond aneuploidy screening, including pharmacogenetic testing, as NIPPT results could directly inform maternal medication use and fetal teratogenicity risk. However, as technical advances enable the future availability of NIPPT and related cffDNA-based testing, careful assessment of clinical utility and continued patient reassurance of genetic nondiscrimination is warranted.

488T

Association between *MTHFR*, *MTHFD1* and *RFC1* gene polymorphisms and unexplained spontaneous pregnancy loss in Korean women. *S. Shim[1], S. Sung[1], J. Kim[2], J. Park[1], K. Kang[1], Y. Shin[1], M. Chin[1], Y. Jung[2], S. Lyu[2], D. Cha[1,2].* 1) Genetic Lab, CHA Gangnam Medical Center, Seoul, South Korea; 2) Department of Obstetrics and Gynecology, CHA Gangnam Medical Center, CHA University Seoul, Korea.

The previous studies have been reported that relationship between polymorphisms in folate metabolism is closely associated with the homocysteine-folate balance, and recurrent pregnancy loss. This study aimed to investigate the association between unexplained spontaneous pregnancy loss and genetic polymorphism in mother -newborn and mother-POC matched sample. A total of 91 Korean women with spontaneous abortion and 92 women with normal delivery were recruited. Genomic DNAs were extracted from mothers, POC samples and newborn samples. This Study was approved by the Internal Review Board of CHA Gangnam Medical Center and written informed consent was obtained from all participants. The *RFC1* 80A>G, *MTHFD1* 1958 G>A, *MTHFR* 677C>T and *MTHFR* 1298A>C variation were examined. Genotyping for each variation was accomplished by polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP). Statistical significances of allele distribution in mother-POC group and mother-newborn group were evaluated by Fisher's extract test and logistic regression test. The frequency of the C allele of *MTHFR* 1298 was significantly higher in mother group with spontaneous abortion (SA) than in mother with newborn (P<0.05). The frequency of the AG genotype of *RFC1* was significantly higher in POC samples than in newborn (P<0.05). We subdivided POC samples based on karyotype results. The frequency of the GA genotype of *MTHFD1* was significantly higher in POC samples with normal karyotype than in newborn (P<0.05). On the other hand, the frequencies of the AG genotype of RFC1 80A>G and the CC genotype of *MTHFR*1298A>C were statistically significant in POC samples with abnormal karyotype compare with newborn. The allele combinations of *MTHFR* 677/*MTHFD1* 1958 and *RFC1* 80/*MTHFR*677/*MTHFD1* 1958 (T-A and G-T-A) was significantly higher in mother with POC samples than in mother with newborn compared with the reference combination. Our study is the first to elucidate the association between *MTHFR*, *MTHFD1* and *RFC1* polymorphisms and unexplained pregnancy loss in mother-newborn and mother-POC matched sample. However, since it was carried out on a small sample limited only Koreans, further studies are necessary to confirm our results more large and heterogeneous populations. .

489F

Lessons from prenatal CMAs in low- and high-risk pregnancies: Is the incidental truly incidental? *A.E. Eilat, V. Meiner, S. Shkedi-Rafid, H. Daum, A. Szmulewicz, N. Zvi, A. Frumkin, S. Yagel.* Hadassah Medical Center, Jerusalem, Israel.

Background: Chromosomal-microarray-analysis (CMA) is gradually replacing karyotyping in prenatal diagnosis. Reporting copy-number-variants (CNVs) conferring risk for adult onset and autosomal-recessive conditions is controversial. We report three cases that illustrate the complexities of disclosing such findings **Cases:** (1) A 38-year-old woman of Ashkenazi-Jewish descent underwent amniocentesis because of advanced maternal age. A paternally inherited 1250Kb deletion was identified on chromosome 13, which includes the *BRCA2* gene, in a female fetus. Negative result for Ashkenazi *BRCA2* founder mutation ruled out Fanconi anemia in the fetus and the pregnancy resulted in the birth of a healthy baby. (2) A 37 year-old woman, of Ashkenazi origin, underwent amniocentesis because of IUGR. A maternally inherited 76Kb deletion, which includes part of the BLM gene associated with Bloom syndrome on chromosome 15, was identified. Positive testing of the Ashkenazi father for the common *BLM* founder mutation led us to direct testing of the fetus and diagnosis of Bloom syndrome. (3) A 32-year-old woman, underwent amniocentesis following an ultrasonic finding of macrocephaly. *de novo* deletion of the *PTEN* gene was identified. The future risk of later onset malignancies and developmental delay was discussed with the parents who opted to continue the pregnancy. **Discussion:** The use non- targeted advanced genomic tests in pregnancy may reveal risks for adult-onset and autosomal-recessive conditions in the ongoing pregnancy. Although reporting of such findings is not unanimously recommended, we stress that at least in some cases, benefit of disclosure outweighs potential harms.

490W

Revealing transcriptome and methylome landscapes in a human oocyte by parallel sequencing. *T. Lee[1], Y. Qian[1], J. Liao[1], L. Chi[2], G. Kong[2], C. Chung[2], T. Leung[2], K. Yip[3], K. Chow[4], W. Chan[1], T. Li[2].* 1) School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong; 2) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong; 3) Department of Computer Science and Engineering, The Chinese University of Hong Kong; 4) Division of Life Science, Hong Kong University of Science and Technology.

Oocyte quality plays a critical role in fertility, embryonic development and fetal health. Previous studies have reported that morphological assessment is limited and was not associated with fertilization rate, embryo quality or implantation rate after intracytoplasmic sperm injection. Molecular assessment based on genetics and epigenetics signatures would be an alternative strategy. With the recent development of single cell sequencing technologies, several studies have established single oocyte transcriptome and methylome profiles separately. In this study, we aimed to profile the simultaneous transcriptome and methylome landscapes in a single human oocytes by adopting a modified single cell parallel methylome and transcriptome sequencing (scM&T-seq). We performed scM&T-seq on two oocytes donated from a 32-year-old in vitro fertilization patient through follicle induction approach. A total of 10317 genes were commonly expressed in the two oocytes, and the correlation of two transcriptomes was 0.984 using Pearson's correlation test. For RRBS libraries, the mapping efficiency was 25.1% and 26.9% respectively, whereas the bisulfite conversion ratios were 98.5% and 96.7% respectively. We detected 222,335-464,057 CpG sites covered by at least one reads and 146,040-322,323 CpG sites covered by at least five reads. Two samples shared 3450 CpG islands. Nearly half of detected CpG sites (43% and 45%) were located in gene promoters and the majority of CpG sites were annotated in CpG islands or shores. A high correlation (R=0.870) of methylation level between two single oocytes was observed in 1000-bp bin size window. Maternal imprinted gene, IGF2R, were fully methylated in both two oocytes. Furthermore, we also observed a positive correlation between gene expression and methylation level in gene body, but not in promoter regions. In summary, we are the first to perform single-cell parallel M&T sequencing on a single human oocyte. The technique allows delineating the dynamics between epigenetic modification and transcriptional activity in parallel fashion. We demonstrated that the quality of transcriptome and methylome profiles are comparable to conventional single-cell transcriptome or methylome approach. We also showed the two single oocytes from the same donor exhibited a positive correlation between gene expression and methylation level on gene body region. The study therefore provides a foundation for the development of molecular diagnosis on oocyte quality.

491T

Targeted next generation sequencing-based preimplantation genetic screening can enable detection of uniparental isodisomy, familial relationships, and polyploidy. *M. Umbarger, E. Boyden, N. Faulkner, M. Zhu, K. Robinson, D. Neitzel, G. Porreca.* Good Start Genetics, Cambridge, MA.

   We previously validated the FAST-SeqS technology for targeted next generation DNA sequencing-based detection of full chromosome and segmental aneuploidy in human embryos (Gole et al 2016, Umbarger et al 2016). While not leveraged in those validations, we observed that the loci captured by FAST-SeqS include thousands of polymorphic sites. The representation of alleles at such sites can be used to derive information beyond segmental and full chromosome aneuploidy. Here, we assessed the feasibility of using single nucleotide polymorphism (SNP) information captured by FAST-SeqS to detect uniparental isodisomy (UPD), familial relationships among samples, and polyploidy. First, to evaluate whether FAST-SeqS can detect UPD we ran cell line-derived samples containing isodisomic chromosomes through FAST-SeqS and assessed whether the captured SNP genotypes were consistent with the expected UPD. We observed an abnormally low frequency of heterozygous SNP calls within isodisomic chromosomes (1.2% vs 62% for control samples), providing a clear metric that could be utilized to call UPD. Next, to determine if FAST-SeqS can recapitulate expected familial relationships, samples from a 17-member multi-generation family were tested in replicate and per-sample SNP genotypes were called and subsequently used to cluster the samples. We observed that the resulting dendogram perfectly matched the expected family tree. Finally, to assess the ability of FAST-SeqS to identify polyploidy, we examined FAST-SeqS data for 200 embryo biopsies for which sex chromosome and autosome copy number calls were potentially consistent with polyploidy (69,XXY; 69,XYY; 92,XXXY) and evaluated the degree to which the observed allele fraction distributions at polymorphic loci supported such karyotypes. For the majority (74.2%) of these biopsies the observed allele fraction distributions displayed characteristic multi-modal patterns that could readily be identified using an automated kernel fitting and peak finding strategy. Collectively, our studies illustrate the feasibility of leveraging embryo-derived SNP information captured by FAST-SeqS to detect UPD, familial relationships among samples, and polyploidy, likely extending to karyotypes that are not currently detectable with most NGS-based preimplantation genetic screening technologies, including whole genome-amplified low-pass sequencing.

492F

The revelation of complex chromosomal rearrangements through genome-wide cfDNA testing. *T. Boomer, S. Caldwell, E. Almasri, J. Wardrop, S. Boshes, M. Hackbardt, P. Cacheris, R. McCullough.* Sequenom Laboratory, San Diego, CA.

   Background: Recent adoption of genome-wide cfDNA prenatal screening provides unique insight into placental findings not previously recognized. Here we present data from our first 28K clinical samples for expanded cfDNA screening, including genome-wide aneuploidy detection and subchromosomal copy number variants (CNVs) ≥ 7Mb, with specific attention to complex chromosomal rearrangements. **Methods:** Maternal blood samples submitted for genome-wide cfDNA testing were subjected to DNA extraction, library preparation, and whole-genome massively parallel sequencing as described by Jensen et al. Sequencing data were analyzed using a novel algorithm as described by Lefkowitz et al. **Results:** A total of 28,760 samples were submitted to the clinical laboratory, resulting in 1392 positives (4.8%) reported: • 53% Common trisomies (Trisomy 13, 18, 21) • 18% Sex chromosomal aneuploidies • 3% Microdeletions • 14% Esoteric chromosomal aneuploidies (e.g. Trisomy 7) • 7% Isolated CNVs • 4% Complex CNVs (≥2 segments, e.g. translocations) Complex CNV samples showed an enrichment of ultrasound findings (61%), high risk personal and/or family histories (34%) and multiple high risk indications (20%) when compared to our testing population as a whole (20%, 6%, and 12% respectively). CNV sizes range <10Mb to ~100Mb, with the majority 10-20Mb. Of the 64 complex CVNs reported, 49 were interpreted as possible translocation events between two chromosomes and 15 as possible recombinant events, isolated to one chromosome. Subsequent fetal confirmation was reported in the majority (56%), with 27% pending, 11% lost to follow-up, and 5% discordant. Prior knowledge of a parental rearrangement (e.g. translocation, insertion, inversion) was true for 22% of complex results, while 19% were identified post testing. A minority of cases (6%) were de novo, with the remaining subset pending full parental assessment (56%). Totals will be updated prior to final publication. **Conclusions**: Identification of complex chromosomal rearrangements via cfDNA prenatal screening marks a new era in prenatal testing. These findings tend to segregate with significant high risk prenatal indications. Many known familial rearrangements not previously amenable to cfDNA screening may now benefit from early identification or added reassurance. New discovery of families at risk of carrying a recombinant event often helps to explain past pregnancy complications, as well as clarify future reproductive risk.

493W

Comparing maternal malignancies and multiple aneuploidies on prenatal cell-free DNA (cfDNA). *E. Soster, B. Dyr, T. Boomer, S. Caldwell.* Integrated Genetics, 3400 Computer Drive, Westborough, MA 01581.

   Maternal malignancies are a rare cause of prenatal cfDNA testing results that are non-reportable or discordant for fetal status. Although prenatal cfDNA testing is not currently designed or validated to detect neoplasm, cases associated with maternal malignancies do rarely occur (~1 in 10000 samples in our laboratory), creating challenges for both the clinician and laboratory. With the advent of prenatal genome-wide cfDNA testing, a handful of cases have identified an increased risk for multiple aneuploidies in a pregnancy, which may raise concomitant concern for a neoplasm. However, data from these samples is often distinct from cases of known maternal malignancy. Maternal blood samples submitted for genome-wide cfDNA screening were subjected to DNA extraction, library preparation, and whole genome massively parallel sequencing.[1] Sequencing data were analyzed using an algorithm to detect trisomies and sub chromosomal, genome-wide events 7Mb and larger.[2] In data suggestive of multiple aneuploidy, events are generally confined to specific chromosomes, appear as full trisomies, and are consistent with the sample's expected fetal representation. These have been reported out by the laboratory as positive for multiple aneuploidy. Conversely, the case examples below with malignancy known or later confirmed were reported out as non-reportable (NR) due to their chaotic appearance and genome-wide aberrant sequencing data. Results suggestive of multiple aneuploidies in cfDNA testing are rare but challenging. Such results in an ongoing pregnancy are most likely associated with placental mosaicism[3] which is reflected in the laboratory reports. While malignancy cannot be ruled out with any result, cases in which maternal neoplasm is present often look distinctly different. Currently, cfDNA testing during pregnancy is not validated to detect or rule out cancer; however, this insight into some differences in laboratory data may help clinicians navigate these complex results.

Lab Call	Positive: Trisomy 13, 15, 20	Positive: Trisomy 8 & 22	Positive: Trisomy 8 & 14	NR	NR	NR
Referral	Maternal age (MA)	MA	Serum screen	Personal/ family history	MA	MA
Status	Ongoing pregnancy; Amniocentesis: normal karyotype; normal maternal karyotype; Normal first trimester analytes, nuchal translucency, anatomy scan	Ongoing pregnancy; Amniocentesis: normal karyotype	Ongoing pregnancy; Amniocentesis: normal Karyotype	Breast Cancer confirmed, diagnosed recent to cfDNA draw	Known advanced breast cancer	Known colon cancer

494T

Broad spectrum of prenatal cultural artifacts detected by microarray. *V. Potluri[1], A. Penton[2], K. Rudd[2], N. Leach[3], H. Risheg[4], J. Knops[5], G. Sekhon[5], R. Pasion[2], R. Burnside[2], P. Papenhausen[2].* 1) Cytogenetics, Lab Corp, Houston, TX 77040; 2) Laboratory Corporation of America, 1904 Alexander Drive, RTP, NC 27709; 3) Integrated Genetics, 3400 Computer Drive, Westborough, MA 01581; 4) Laboratory Corporation of America Holdings/Dynacare, Seattle, Washington; 5) Integrated Genetics, 2000 Vivigen Way Santa Fe, NM 87505.

   The importance of direct microarray analysis (CMA) of prenatal cells can be emphasized by a retrospective study of mosaic results detected in parallel flasks grown for f/u of cytogenetic analysis. The delayed CMA studies were usually part of reflex testing if the chromosome analysis was normal or if more detail was desired concerning an abnormal result. The cohort being cited includes 9 amniocyte cultures and 4 CVS cultures which were in culture from 16 to 43 days before CMA. Eight of the cultures were between 16 and 21 days in culture. The percentage of mosaicism primarily ranged from 8-25%, although a 40% loss of Y (35 days in culture) and a 30% copy neutral segmental UPD of 9q (21 days in culture) were also seen. The artifacts were primarily whole chromosome gain (isolated chromosome gain of 5, 7, 9,) with the one case of Y loss. More complicated artifacts included one triple aneusomy (+2,-4,+14) and another with four trisomies (+3,+4,+5,+10) and a dup(17q). Segmental changes with single alterations included deletions of 3q, 4q and 12q, along with isolated duplications of 1q, 14q and Xq. All these mosaic cases were visually confirmed as artifacts by separate original culture analyses in extended analysis of all available metaphases. Two cases required BAC FISH follow-up due to the small size of the CNV, but are not included in this report. Whenever cultured cells are necessary for CMA it is important to have a parallel culture of separate origin to properly interpret the many artifacts that arise from even a moderate time in vitro.

495F

Prenatal NGS testing confounded by low levels of maternal cell contamination. *Z. Wolf[1,2], S. Bandyadka[1], E. Murphy[1], M. Lebo[1,2], H. Mason-Suares[1,2].* 1) Laboratory for Molecular Medicine, Partners Personalized Medicine, Cambridge, MA; 2) Department of Pathology, Harvard Medical School and Brigham and Women's Hospital, Boston, MA.

   Next generation sequencing (NGS) is rapidly becoming the standard of care in prenatal diagnostics for high risk pregnancies, but working with prenatal samples requires particular care and considerations. Unlike other sample types used in molecular diagnostics, prenatal samples can be contaminated with maternal cells, which can introduce errors leading to misdiagnosis and pregnancy mismanagement. While the level of maternal admixture that affects the sensitivity and specificity for NGS is unknown, studies involving other assays suggest as little as 10% maternal cell contamination (MCC) can result in misdiagnoses. To define these levels for NGS, this study examines the consequence of maternal admixture on the interpretation of NGS results from prenatal samples with varying levels of MCC. MCC was simulated at levels of 2.5%, 5%, 7.5%, 10%, 20%, 30%, 40%, and 50% by spiking maternal DNA into fetal DNA samples previously confirmed to be negative for MCC. NGS was then performed using a targeted capture kit of 78 genes, and variants were called according to GATK best practices. Samples had an average coverage of 350X. Discordant variant calls between maternal and prenatal samples were examined in the spiked-in samples and analyzed based on clinically relevant scenarios. Simulation of disorders with an autosomal recessive inheritance, where the maternal sample was heterozygous and the fetal sample was homozygous for a given allele, demonstrated that as little as 5% MCC compromised test interpretation with 19% of variants being miscalled and passing quality control (QC) metrics. Simulation of disorders with a de novo autosomal dominant inheritance, where the maternal sample was homozygous for the reference allele and the fetal sample was heterozygous, showed that variants were miscalled starting at 20% MCC. In summary, here we report that very low-levels of MCC can confound NGS test results, leading to erroneous interpretation of clinical results and ultimately affecting pregnancy management. This study highlights that MCC must always be considered when interpreting results from prenatal samples, since levels as low as 5% may result in incorrect findings.

496W

FXPOI: Modifying factors may play a larger role among the most vulnerable mid-range premutation group of women. *E.G. Allen[1], K. Charen[1], H. Hipp[1], L. Shubeck[1], C. Trevino[1], W. He[1], A. Amin[1], A. Glicksman[2], N. Tortora[2], S.L. Nolin[2], H.R. Johnston[1], D.J. Cutler[1], M. Zwick[1], M.P. Epstein[1], P. Jin[1], S.T. Warren[1], S.L. Sherman[1].* 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) New York State Institute for Basic Research in Developmental Disabilities, Albany, NY.

   Fragile X-associated primary ovarian insufficiency (FXPOI) is seen in about 20% of women who carry a *FMR1* premutation allele (55-200 CGG repeats). These women develop hypergonadotropic hypogonadism with menstrual dysfunction before the age of 40. A non-linear association with repeat size and risk for FXPOI has been confirmed in multiple studies: those with a mid-range of CGG repeats are at highest risk (~70-100 CGG repeats). Importantly, not all female carriers with 70-100 CGG repeats experience FXPOI. We hypothesize that other factors modify the risk for FXPOI, including those associated with the characteristics of the repeat as well as other genetic and environmental factors. We have determined age at menopause on 415 women with a premutation based on self-report reproductive questionnaires and interviews with a reproductive endocrinologist. To identify genetic modifying factors, we have conducted whole genome sequencing (WGS) on 70 early-onset cases of FXPOI and 55 late-onset controls. From these results, we will identify candidate genes to investigate in a *Drosophila* model screen. In parallel, we are investigating *FMR1* molecular modifying factors, including repeat size, AGG interruption pattern, transcript levels, and X-inactivation ratio among subsets of this population and potential environmental modifying factors including smoking, body mass index, and autoimmune disorders. To start, we characterized the amount of variation in age at menopause explained by repeat size using a model that included the repeat size and the quadratic form of repeat size to test the non-linear relationship. This model was highly significant (p<0.0001); however, only 6% of the variation in age at menopause was explained ($r^2=0.06$). Neither the AGG interspersion pattern nor X-inactivation ratio was significantly associated with age at menopause. Interestingly, although the sample size was limited (N=66), adding transcript level and the product of repeat size and transcript level to this model was significant (full model p=0.0060) and explained 21% of the variation in age at menopause ($r^2=0.21$). In addition to these molecular findings, we will present potential modifying genes identified from our WGS and investigate environmental factors and their effect on the risk for FXPOI.

497T

Novel application of fragile X CGG repeat analysis in a clinical setting: Outcomes and data derived from 716 embryos undergoing preimplantation genetic diagnosis (PGD). *R. Cabey, D. Goldberg-Strassler, B. Chu, N. Goodall, A. Jordan, R. Prates, M. Konstantinidis.* CooperGenomics, a Cooper-Surgical Company, Livingston, NJ.

Analyses of fragile X mental retardation gene (*FMR1*) CGG repeat testing have always been done on samples derived from living individuals or prenatal testing. This is the first study with a significant amount of data derived from human preimplantation embryos undergoing preimplantation genetic diagnosis (PGD). Between 01/2014 and 04/2017, 98 patients, with average maternal age of 34.1±0.3 (standard error of the mean), underwent PGD for fragile X syndrome via a linkage-based technology (Karyomapping; Illumina, USA), and direct CGG repeat analysis. Samples from 716 embryos were received from 163 IVF cycles. 28 patients banked embryos from multiple IVF cycles. A total of 132 PGD tests were performed; 123 of these included aneuploidy screening via aCGH or NGS (Illumina, USA). Following exclusion of embryos with no diagnosis, sex aneuploidy, or that were untested, CGG repeat results were obtained in 641/653 samples (98.2%). Patients who banked embryos had more embryos tested per PGD cycle (6.8±0.7) compared to patients who did not (4.9±0.3). Analysis of expansion patterns revealed that the premutation allele expanded to a full mutation at the following rates per category group: 0% (0/59) in the maternal premutation group of 55-59 CGG repeats, 8.2% (6/73) in the 60-69 group, 9.5% (4/42) in the 70-79 group, 30.8% (8/26) in the 80-89 group, 91.7% (22/24) in the 90-99 group, and 100% (47/47) in the 100-199 group. Overall, 47.2% of embryos inherited the maternal normal allele. Of the embryos that inherited the mutant allele, only 32.1% expanded to a full mutation. Interestingly, 1.1% (7/641) of embryos contracted from the maternal allele size (ranging from 64 to >200). The average number of embryos available for transfer per PGD cycle, increased significantly from 1.4±0.1 (normal FMR1 alleles only) to 2.1±0.2 when including premutation embryos (P<0.001). Follow up from patients revealed that 53 normal allele embryos, 5 premutation, and 2 normal or premutation embryos were transferred; altogether resulting in 26 live births and 6 ongoing pregnancies. Subsequent confirmatory testing was pursued by 3 patients revealing concordant results with PGD. This unique data set provides a novel insight into the FMR1 CGG repeat patterns at the preimplantation embryo stage and clinical implications of such testing. The ability to distinguish premutation and full mutation preimplantation embryos provides more options and information for embryo transfer decisions.

498F

CarrierTest: The first experience with expanded preconception carrier screening. *M. Koudová, F. Lhota, F. Zembol, M. Bittóová, L. Dohnalová, Z. Vilímová, B. Honysová, D. Stejskal.* GENNET - Clinic of Genetics and Reproductive Medicine, Prague, Czech Republic.

Introduction Over 1300 recessive disorders have been described with estimated cumulative birth prevalence of recessive homozygotes/hemizygotes 3/1000. Preconception carrier screening testing asymptomatic individuals for single copy mutations and estimating risk of recessive disorders of their offspring has a potential to decrease this burden. **Methods:** We have designed a NGS (next-generation sequencing) amplicon-based panel testing 860 key mutations of 79 genes which can influence reproductive health of prospective parents or can cause recessive disorders in offspring. We have developed a bioinformatic pipeline using local installation of Ensembl genomic database for annotation and SQL server variant database for data handling and clinical reporting. To replace MLPA and fragmentation analysis methods we developed coverage analysis-based CNV detection of frequent large deletions of *SMN1* and *CFTR* genes. A software tool developed for this application generates report semi-automatically. **Results:** Results are grouped according to clinical impact: (1) mutations in genes associated with severe recessive disorders in offspring (e.g. *SMN1, CFTR, GJB2* genes), (2) low penetrant mutations in genes associated with disorders influenced by environment (*SERPINA1, HFE, BCHE*), (3) mutations in set of genes predisposing to blood hypercoagulation- trombophilic profile (*F2, F5, MTHFR, ANXA5*- M2 haplotype), (4) ovarian response to FSH (*FSHR* polymorphism). The report contains in addition to general information on found variants and associated disorders residual risk estimation and couple´s preconception compatibility. **Conclusion:** CarrierTest can contribute to individualization and therefore more effective treatment of infertility and prevention of rare genetic disorders. There is a clinical option of preimplantation or prenatal diagnostics or using compatible donor gametes for couples with significant risk of having affected offspring.

499W

Potential impact of predictive genetic testing among at-risk female relatives of *ATM* heterozygotes on breast cancer surveillance recommendations. *D. Almanza[1], C. Lewis, MS, CGC[2], K. Ray, MS, CGC[1], A. Cantor, MS, CGC[3], R. Hayes Morgan[4], K. McReynolds[4], G. L Wiesner, MD[4], T. Pal, MD[4].* 1) Moffitt Cancer Center, Tampa, FL; 2) St. Jude Children's Research Hospital, Memphis, TN; 3) NYU Langone Medical Center, New York, NY; 4) Vanderbilt University Medical Center/Vanderbilt-Ingram Cancer Center, Nashville, TN.

 Purpose: Widespread multi-gene panel testing for inherited cancer risk, inclusive of moderate penetrance genes, has led to the identification of many *ATM* heterozygotes. Heterozygous *ATM* pathogenic variants impart >20% lifetime female breast cancer (BC) risk, the threshold for high-risk breast surveillance. Among a registry-based sample of probands with a pathogenic (P) or likely pathogenic (LP) variant in *ATM*, we sought to assess the impact of *ATM* positivity on BC surveillance recommendations for at-risk female relatives. **Methods:** Using a registry-based sample of 47 *ATM* heterozygotes, lifetime BC risks were calculated for unaffected female first-degree relatives (FDRs) and second-degree relatives (SDRs) under age 80 using BOADICEA and summary statistics were generated for level of BC risk. **Results:** Among 90 eligible FDRs and 137 eligible SDRs, most (74.4% and 81.8%, respectively) had lifetime BC risk estimates <20% using risk modeling. **Conclusion:** Our results suggest that *ATM* positivity would alter BC surveillance recommendations for the majority of at-risk female relatives, as most had lifetime BC risks <20% thus would not qualify for high risk surveillance based on family history alone. Our data highlight the potential utility of risk modeling to inform clinical utility of predictive testing for moderate penetrance BC genes, such as *ATM*, among at-risk relatives.

500T

Genetic testing in adult cancer patients in palliative care: What they understand, want, and may need. *J. Bodurtha[1], M. Abusamann[2], L. Bailey[3], O. Emidio[2], J. Kang[3], B. Ma[3], O. Owudunni[2], J. Quillin[5], R. Razzak[4], B. Yu[3], T. Smith[4].* 1) Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Bloomberg School of Public Health, Baltimore, MD; 3) Johns Hopkins University, Baltimore, MD; 4) Johns Hopkins Hospital, Baltimore, MD; 5) Massey Cancer Center, VCU, Richmond, VA.

 Background: Hereditary cancer (CA) assessment and communication about family history (FH) risks in palliative care (PC) are underexplored and could be critical for surviving at-risk relatives. Yet, FH assessment and offer of appropriate GT may be suboptimal in PC. This study explored the need for genetics services among PC CA patients, and their related knowledge and attitudes. Methods: A cross-sectional study was conducted including interviews that assessed knowledge and attitudes toward GT, DNA banking, and FH knowledge and reviews of participants' medical records. **Results:** Seventy-five adult English-speaking CA patients enrolled (66% response rate). Forty-one (55%) self-identified as male; 20 (27%) as black, 49 (65%) white, and 6 (8%) as other race. Thirty-two (43%) thought it likely that their CA had a genetic or inherited component. Twenty-five (33%) answered that they had heard or read "a fair amount" or a "lot" about GT for CA risk. Eleven (15%) reported having had genetic testing for inherited cancer risk, although this could only be confirmed for two (3%) patients by systematic chart review. Participants may not have understood the difference between somatic vs. germline GT. Twenty-eight (37%) met ACMG-NSGC criteria for genetics referral, with minimal chart evidence by standardized review of consideration, referral, or discussion about GT by oncology services. Forty-eight (64%) thought that the PC setting was the right place to consider the implications of a CA FH. Fifty-seven (76%) wanted CA prevention measures, like smoking cessation, to be offered to relatives as part of PC services. Seventy-three (97%) participants were interested in receiving GT information. Seventy-four (99%) participants reported wanting to share results with relatives. Six (8%) patients had either banked DNA themselves or knew a relative who had banked DNA. 54(72%) thought Angelina Jolie had done the right thing by sharing her story. The mean comfort level in addressing the survey questions was 8.7 on a 10-point scale with 10 indicating extremely comfortable. **Conclusions:** Enhanced openness to these topics among PC providers, oncology specialists, and families, review of FH and access to GT might enhance CA prevention, especially among high-risk families. Patients are willing to discuss FH and risk in the palliative care setting. There appears to be room for improvement in how FH and GT are integrated in CA care and palliative care. .

501F

Educational and support needs of Lynch Syndrome probands and their relatives. *W. Kohlmann[1], J. Petersen[3], C. Koptiuch[1], K. Szczotka[2], M. Keener[1], S. Johnson[1], P. Kanth[4], A. Soisson[1,2], K. Kaphingst[1].* 1) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 2) Department of Obstetrics and Gynecology, University of Utah, Salt Lake City, UT; 3) Department of Human Genetics, University of Utah, Salt Lake City, UT; 4) Division of Gastroenterology, University of Utah, Salt Lake City, UT.

The benefit of identifying colorectal and endometrial cancer patients with Lynch syndrome (LS) is maximized when relatives are tested for the pathogenic variant (PV) and take screening and prevention actions. However, uptake of testing and adherence to screening remains suboptimal and educational interventions are needed. To assess the educational preferences of LS families, semi-structured telephone interviews were conducted with probands and their relatives. Participants were identified from registries, family referral and clinic records. 25 probands (mean age 59 yrs, 88% with prior cancer) and 49 relatives (mean age 46 yrs, 27% with prior cancer) enrolled. The majority of probands received the diagnosis of LS from a genetic counselor (GC) (52%) and received informational resources (76%). The diagnosis of LS was most commonly shared with siblings (84%) and children (76%) and less so with aunts/uncles (50%) and cousins (46%). No probands reported withholding the diagnosis from anyone, and the majority (75%) felt confident discussing LS. Relatives were notified of the diagnosis of LS in person (47%) and by phone (33%). Only 10% were informed via email/social media. Disclosure of the LS diagnosis in the family to relatives was less commonly accompanied by informational resources (49%). Of the 88% of relatives who had genetic testing, 55% met with a GC and 88% were positive for the familial PV. Withholding information from others was rare among relatives, with only 7% not disclosing results due to young age or lack of close relationship. When asked preferences for ongoing education, probands rated the following as very/extremely helpful: LS specialty clinic (84%), other family (77%), website (74%), primary care provider (PCP) (68%) and printed materials (64%). While relatives preferred: LS specialty clinic (84%), website (80%), talking with PCP (68%), and other family (64%). A minority indicated social media (probands 40%, relatives 30%) or in person support groups (probands 44%, relatives 33%) as preferred methods for ongoing education. These findings show that family members actively share the diagnosis of LS and are often significant sources of support and information for each other. New methods of communicating, such as social media, may be useful for some. However, traditional resources such as access to specialists and PCPs who are knowledgeable about LS and websites continue to be important, needed resources for LS families.

502W

***MSH6* and *PMS2* germline pathogenic variants implicated in Lynch syndrome are associated with breast cancer.** *M.E. Roberts[1], S.A. Jackson[1], L.R. Susswein[1], N. Zeinomar[2], X. Ma[2], M.L. Marshall[1], A.R. Stettner[1], B. Milewski[1], Z. Xu[1], B.D. Solomon[1], M.B. Terry[2], K.S. Hruska[1], R.T Klein[1], W.K. Chung[3].* 1) GeneDx, Gaithersburg, MD; 2) Department of Epidemiology, Columbia University, New York, NY; 3) Departments of Pediatrics and Medicine, Columbia University, New York, NY.

Statement of Purpose: An association of Lynch syndrome (LS) with breast cancer has been long suspected in the cancer genetics community but available data are conflicting. The aim of our study was to characterize the respective breast cancer risk for each LS-associated gene (*MLH1*, *MSH2*, *MSH6* and *PMS2*) in women with a single germline pathogenic or likely pathogenic variant (collectively, PV) identified via a multi-gene hereditary cancer panel. **Methods:** In a retrospective query of over 50,000 women who had panel testing, 423 were identified as harboring a single PV in *MLH1*, *MSH2*, *MSH6*, or *PMS2*. Women with a second PV in these or another gene were excluded. An in-depth review of personal medical and family histories was completed. The standard incidence ratio (SIR) of breast cancer was calculated by comparing the cancer frequencies in our study cohort with those in the general population (SEER: Surveillance, Epidemiology, and End Results Program) and a Chi-square test was used to obtain p-values. **Results:** Of the 423 women, the distribution of PVs was 33% (n=140) in *MSH6*, 29% (n=124) in *PMS2*, 22% (n=94) in *MSH2* and 15% (n=65) in *MLH1*. A personal history of breast cancer was reported in 25.3% (107/423) of women with PVs, with 1.4% (6/423) reporting a history of more than one primary breast cancer. Age-standardized breast cancer risk was approximately 2-fold higher in our cohort of women with a LS-associated PV compared to the general population (SIR = 1.96; 95% CI, 1.62-2.37). When evaluating risk by gene, *MSH6* (SIR = 2.11; 95% CI, 1.56-2.86) and *PMS2* (SIR = 2.92; 95% CI, 2.17-3.92) were associated with a statistically significant risk for breast cancer, while no association was observed for *MLH1* (SIR = 0.87; 95% CI, 0.42-1.83) or *MSH2* (SIR = 1.22; 95% CI, 0.72-2.06). Of the 423 women, those with PVs in *MSH6* or *PMS2* were more likely to have a personal or family history of breast cancer and no history of colon, endometrial, or ovarian cancers (*MSH6*: 8.0%, 11/137; *PMS2*: 24.6%, 30/122) compared to women with PVs in *MLH1* and *MSH2* (MLH1: 0%, 0/65; MSH2: 2.2%, 2/92) (p = <0.001). **Conclusions:** Our data demonstrate that *MSH6* and *PMS2* are associated with a modest but statistically significant increased risk for breast cancer, and should be considered when ordering genetic testing for individuals who have a personal and/or family history of breast cancer. Women with PVs in *MSH6* and *PMS2* may benefit from increased breast cancer screening.

503T

Could population-level genomic screening be cost effective? An economic analysis informed by a large sequencing study. *D.L. Veenstra[1], G.F. Guzauskas[1], M.F. Murray[2], M.S. Williams[2].* 1) Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, WA; 2) Genomic Medicine Institute, Geisinger Health System, Danville, PA.

   Background: The identification of clinically actionable, highly penetrant, and relatively common genomic variants in the general population, combined with low cost genomic assays, has brought consideration of population-level genomic screening into the realm of possibility. However, significant challenges and uncertainties remain. Our objective was to model the potential clinical and economic impacts of population-level genomic screening. **Methods:** We developed a decision-analytic policy model to project the quality-adjusted life years (QALYs) and lifetime costs associated with genomic screening for risk of breast and colon cancer, familial hypercholesterolemia, cardiomyopathy, and inherited cardiac arrhythmias. The model was informed by previously developed cost-effectiveness models, variant prevalence estimates from a large exome sequencing study (>92,000 sequenced participants) conducted in an integrated healthcare delivery system, and the published literature. We modeled the long-term clinical benefits and risks of interventions indicated by genomic findings, as well as costs associated with screening and downstream interventions. We assumed a range of testing costs. Value of information analyses were conducted to assess the value of conducting further research to reduce uncertainty in genomic and behavioral parameters. **Results**: An average of 81 (95% credible range [CR]: 53-125) QALYs were generated per 10,000 patients screened, At a testing cost of $250, the increase in total cost of approximately $6.9M (95% CR: $4.6M-$9.1M). The incremental cost-effectiveness ratio was $93,100/QALY, and screening was cost effective in 71% of simulations at a $100,000 willingness-to-pay per QALY threshold. Value of information simulations indicate improving estimates of penetrance and healthcare behavior could provide a maximal societal value of $5B. **Conclusion:** Population-level genomic screening may be effective and cost effective, but significant uncertainty remains. Large prospective and well controlled studies that improve our understanding of population-level gene penetrance and healthcare behaviors in response to both positive and negative genomic findings are needed to inform adult genomic screening policies.

504F

Clinical predictors of gene panel choice when selected by the patient. *E.K. Courtney[1], J.C. Allen[2], J. Ngeow[1,3,4].* 1) Cancer Genetics Service, National Cancer Centre Singapore, Singapore, Singapore; 2) Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Singapore; 3) Oncology Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore; 4) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore.

   PURPOSE The Cancer Genetics Service (CGS) at the National Cancer Centre Singapore has offered multi-gene panel testing for hereditary cancer syndromes to patients since it was established in 2014. Unlike other testing models where panel selection is entirely dependent on clinical indication, the CGS has adopted a 'menu' approach where patients are given the choice whether to opt for the clinically indicated panel or a broader panel covering a wider range of hereditary cancer syndromes. The present study sought to investigate clinical factors associated with panel choice by the patient. **METHODS** Female breast and ovarian cancer patients who were referred to the CGS for consideration of genetic testing for hereditary breast and ovarian cancer syndrome and who subsequently underwent panel testing were included. Demographic, clinical, panel (breast cancer panel, BCP or multi-cancer panel, MCP) and genetic result information were collected. Univariate, multiple and stepwise multiple logistic regression analyses were used to identify clinical factors predictive of panel choice. **RESULTS** A total of 265 breast and ovarian cancer patients were included in this study. Subjects who chose the broader MCP compared with the targeted BCP were significantly more likely to be aged ≥ 50 years (49 vs. 31%, $P < 0.05$), Chinese (76 vs. 47%, $P < 0.001$) and have a personal history of ovarian cancer (41 vs. 8%, $P < 0.001$). Various permutations of family history data were not significantly associated with panel choice, including (i) family history of any cancer, (ii) family history of breast cancer only and (iii) family history of colorectal and/or endometrial cancer. Multiple stepwise logistic regression identified race and personal history of ovarian cancer as the best predictors of panel choice. **CONCLUSION** Interestingly, our results indicate that personal factors, rather than family history, are more likely to predict whether patients opt for targeted versus broader gene panels. This study provides the foundations for deeper exploration of other personal factors that drive patient choice using a 'menu' approach to panel testing.

505W

BRCA1/BRCA2 population screening in Ashkenazi Jews: Long term impact and familial communication. *S. Lieberman[1,2], A. Tomer[1], A. Ben-Chetrit[3], O. Olsha[4], R. Beeri[1], A. Raz[5], A. Lahad[6,2], E. Levy-Lahad[1,2].* 1) Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; 2) Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel; 3) Department of Obstetrics and Gynecology, Shaare Zedek Medical Center, Jerusalem, Israel; 4) Breast Unit, Dept. of Surgery, Shaare Zedek Medical Center., Jerusalem; 5) Department of Sociology, Ben-Gurion University of the Negev, Beer Sheva, Israel; 6) Department of Family Medicine, Clalit Health Services, Jerusalem, Israel.

Background: In Ashkenazi Jews (AJ) population screening for common (2.5%) BRCA1/BRCA2 mutations could identify all carriers, including many (about half) lacking suggestive family history. Towards implementation, we examined the long-term impact of *BRCA* screening. **Methods:** Unaffected AJs, age ≥25 years were either self-referred (SR) or recruiter-enrolled (RE), received pre-test written information and self-reported family history (FH). Post-testing, non-carriers with significant FH and carriers received in-person genetic counseling. Psychosocial outcomes, familial communication and health behaviors were assessed using questionnaires at one week, 6 months and 2 years post-testing **Results:** We report the 2 year follow-up of 1771 participants, including 32 carriers. *Psychosocial outcomes:* RE and SR participants had similar rates of satisfaction (94%) and endorsement of population screening (91%), and similarly low stress (IES score= 4.1). Knowledge scores were higher in SR vs. RE (7.5 vs. 6.9/10, P<.001). Among carriers, 92% expressed satisfaction and 91% endorsed population screening. Stress was higher but declined over time (IES score 13.9 vs. 19.9 at 6 months, NS). Knowledge was greater than in non-carriers (8.7 vs. 7.15/10, p<.001). *Familial communication:* 88% of responders and 97% of carriers informed at least one relative. Satisfaction With Health Decision (SWHD) score was the only substantial predictor: communication was 6-fold higher in the highest vs. lowest score quartile (p<.001). Of carriers' relatives, 11-27% remained uninformed. 48% of carriers' first- and second-degree relatives were tested, with higher rates in first-degree (58% vs. 26%, p=.0002) and in female relatives (55% vs. 37%, p=0.02). *Health behavior:* All 25 women carriers had breast surveillance, 3/25 (12%) underwent risk-reducing bilateral mastectomy (similar to published rates in Israeli carriers). 15/16 (94%) carriers age >40 underwent Risk-reducing salpingo-oophorectomy (RRSO). Among non-carriers, mammography screening rates did not decline compared to pretest rates, and even increased in non-carriers >50 with non-suggestive FH (p=.003). **Conclusions:** Long-term, *BRCA* screening is highly acceptable. Familial communication rates and characteristics are similar to those reported in the clinical testing setting. Non-carriers do not demonstrate false reassurance, whereas carriers universally adopted increased surveillance with the vast majority undergoing age-appropriate RRSO.

506T

Cancer susceptibility mutations in very young Israeli women with breast cancer. *B. Nehoray[1], L. Van Tongeren[1], D. Behar[2], S. Dishon[2], F. Lejbkowicz[2], J. Herzog[1], T. Slavin[1], D. Castillo[1], S. Sand[1], J. Weitzel[1], G. Rennert[2].* 1) City of Hope, Division of Clinical Cancer Genomics, Duarte, CA; 2) Clalit National Israeli Cancer Control Center, Haifa, Israel.

Background: Germline *BRCA1* and *BRCA2* (*BRCA*) mutations are associated with an increased risk of early onset breast cancer (BC). While there is uniform testing of the *BRCA* founder mutations in Israel, extended panel testing is limited. We hypothesized that a percentage of young Israeli women who tested negative for *BRCA* Jewish founder mutations would have mutations in other breast cancer susceptibility genes, including *TP53*. **Methods:** All participants consented to an IRB-approved cancer genetics study in Israel. The study included the use of medical and family history data, and prospective biospecimens. DNA from 64 Israeli women with BC diagnosed before the age of 35 who previously tested negative for *BRCA* founder mutations was sequenced using a 767-gene custom Agilent SureSelect (Alameda, CA) kit on a HiSeq 2500. Sequencing was performed to an average of 100x read depth. All findings over 20% percent allele frequency were evaluated. American College of Medical Genetics guidelines were used to classify variants using Ingenuity Variant Analysis (Redwood City, CA). Copy number variation was evaluated by a combination of the exome hidden Markov model (XHMM) and COpy number Detection by EXome sequencing (CODEX) algorithms. **Results:** Ten participants (15%) had likely pathogenic or pathogenic variants (mutations) in known cancer predisposition genes, including non-founder mutations in *BRCA1* (2) and *BRCA2* (2), as well as *ATM* (2), *CHEK2, MSH2, MSH6,* and *MUTYH*. All mutations were within 45.7-54.4% allele fraction. One *TP53* mutation was identified. However, the allele fraction (20.4%) was lower than would be expected for a germline variant. Confirmatory Sanger sequencing did not identify the variant, indicating likely clonal hematopoiesis. Many other mutations were identified in candidate predisposition genes such as *RAD50* and *FANCF*. Copy number evaluation did not yield additional significant results. **Conclusions:** The identification of non-founder *BRCA* mutations and pathogenic variants in other susceptibility genes, supports the use of multigene panel testing in young Israeli women with breast cancer who test negative for the *BRCA* founder mutations. While we did not detect any germline *TP53* variants (1-3% expected), this may be due to regional differences and small sample size; testing is in process in a larger sample. Further studies are needed to detail the spectrum of hereditary cancer predisposition risk in young Israeli women with breast cancer.   .

507F

Anticipated responses of genetic specialists and non-genetic specialists to unsolicited genomic secondary findings. *K.D. Christensen[1], B.A. Bernhardt[2], G.P. Jarvik[3], L.A. Hindorff[4], J. Ou[3], S. Biswas[5], B.C. Powell[6], R.W. Grundmeier[7], K. Machini[8], D.J. Karavite[7], J.W. Pennington[7], I.D. Krantz[9], J.S. Berg[6], K.A.B. Goddard[10].* 1) Department of Medicine, Division of Genetics, Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 2) Division of Translational Medicine & Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Division of Medical Genetics, University of Washington, Seattle, WA; 4) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 5) Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA; 6) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, PA; 7) Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA; 8) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Department of Pathology, Brigham and Women's Hospital, and Harvard Medical School, Cambridge, MA; 9) Division of Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 10) Center for Health Research, Kaiser Permanente Northwest, Portland, OR.

Purpose As genomic sequencing use increases, secondary findings will become more common. We compared how health care providers with and without specialized genetics training anticipated responding to different types of secondary findings. **Methods** Fifty-four physicians and genetic counselors with genomic sequencing experience reviewed laboratory reports containing five secondary findings and reported attitudes and potential clinical follow-up. Reports presented (1) a pathogenic variant in *RYR1*, (2) a pathogenic variant in *PALB2*, (3) a pathogenic variant in *CHEK2*, (4) a variant of uncertain significance in *BRCA1* in a patient with strong family history of disease, and (5) carrier status in *HFE, MUTYH,* and *SPG7*. Analyses compared responses from genetic specialists (clinical geneticists and genetic counselors) and physicians without specialized genetics training, and examined how responses varied by type of secondary finding. **Results** Genetic specialists scored higher than non-genetic specialists on 4-point scales assessing understandings of reports (difference=0.50, p<0.001), and lower on scales assessing obligations about reporting (difference =-0.80, p<0.001) and burdens of responding (difference =-0.93, p<0.001). Nearly all attitudes differed according to what finding was reported, although genetic specialists were more sensitive to finding types than non-genetic specialists about laboratories' obligations to report (p<0.001 in interaction tests). The type of secondary finding also influenced whether participants rated reviewing patients' personal and family histories, documenting findings in the medical record, and recommending familial discussions as extremely important (all p<0.001). **Conclusion** Attitudes towards secondary findings and potential clinical responses of genetic specialists and non-genetic specialists differ greatly. Findings highlight the need to be cautious about generalizing findings from studies of genetic specialists to settings where providers are not genetic specialists, and add urgency to calls for specialty-specific approaches to supporting genomic medicine.

508W

Experiences and next steps in utilizing The Jackson Laboratory Clinical Knowledgebase (JAX-CKB), a relational database, for clinical and educational purposes. *K. Sanghavi, S.E. Patterson, C.M Statz, T. Yin, S. Mockus.* The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Statement of purpose: Access to clinically relevant, updated, and readily available synthesized data is critical to precision medicine in cancer genomics. The Jackson Laboratory Clinical Knowledgebase (JAX-CKB) is a relational database for interpreting complex genomic profiles for clinical and educational utilization. It includes genes and variants with descriptions, relevant therapeutic and prognostic information, clinical trials, efficacy and resistance evidences supporting treatment approaches, and an application that allows a methodical search for links between data elements in the database. **Methods/Approaches**: Build of the JAX-CKB database was initiated in early 2014 and the user interface for curation was launched in mid-2014. It was validated for clinical reporting in mid-2015 and launched for open access in mid-2016 (ckb.jax.org). The JAX-CKB continues to be used for clinical applications such as the 358 gene Cancer Treatment Profile assay and 212 gene ActionSeq™ assay, and molecular/genomic tumor boards. This database was used for provider education in an online continued medical education activity. The gene variants, efficacy evidence, and targeted therapies information are updated daily by Ph.D. staff. The clinical trials information is updated weekly. An internal survey was administered at a user-interface workshop to specifically test for user interface design elements such as the ease of navigation, search capabilities and understanding of associated content. **Summary of results**:10/12 respondents to the internal survey found everything they were looking for in the JAX-CKB database including molecular profile, treatment approach, and connections between different areas of the database, which they found very easy to understand (11/12). As of 6/6/17, the total number of JAX-CKB website hits include 46,473 with 12, 476 unique users. Average time spent on the website is 5.39 minutes. 73% of the first-time users returned to the website. The top gene searches include *TP53*, *EGFR*, *PIK3CA*, and *BRAF*. Survey results indicated the need for additional help documents related to transparency of curation methodologies. **Conclusion:** The JAX-CKB provides access to such data that empowers the global biomedical community in discovering and utilizing genomic solutions for somatic cancer conditions. Next steps include implementing a research protocol to cyclically evaluate the JAX-CKB, which would allow continuous offering of an innovative database.

509T

Hereditary colorectal cancer screening: A 10 year longitudinal study following an educational intervention. *J.C. Carroll[1,2], K. Semotiuk[3], J. Permaul[2], E.M. Yung[4], S. Blaine[1,5], E. Dicks[6], E. Warner[7], H. Rothenmund[8], M.J. Esplen[9], J. McLaughlin[10,11].* 1) Family & Community Medicine, University of Toronto, Toronto, ON, Canada; 2) Family Medicine, Mount Sinai Hospital, Sinai Health System, Toronto, ON, Canada; 3) Zane Cohen Centre for Digestive Diseases Familial Gastrointestinal Cancer Registry, Mount Sinai Hospital, Toronto, ON, Canada; 4) Faculty of Medicine, University of Toronto, Toronto, ON, Canada; 5) STAR Family Health Team, O'Loane Medical Clinic, Stratford, ON, Canada; 6) Memorial University of Newfoundland and Labrador, St. John's, NL, Canada; 7) Division of Medical Oncology, Sunnybrook Odette Cancer Centre,Toronto, ON, Canada; 8) WRHA Genetics and Metabolism Program, Health Science Centre, WPG, MB, Canada; 9) University Health Network, University of Toronto, Toronto, ON, Canada; 10) Public Health Ontario, Toronto, ON, Canada; 11) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

Context: Screening reduces mortality in individuals with a family history (FH) of colorectal cancer (CRC) yet rates are suboptimal. To address this, the team developed an educational intervention consisting of a mailed CRC Risk Triage/Management Tool for family physicians (FP) and CRC information booklet for patients. *Objective:* To describe CRC screening 10 years following the intervention. *Design:* Mailed self-complete questionnaires to FPs and patients: baseline, 5, and 10 years post-intervention. *Setting*: Family practices in Ontario and Newfoundland, Canada. *Participants:* Random sample of FPs and their adult patients with a FH of CRC. Results include FPs and patients who completed all questionnaires over 10 years. *Intervention:* CRC Risk Assessment/Management Tool indicating FH criteria and appropriate screening test and timing for patients at population, low, moderate, and high risk for CRC. Patient booklet: patient stories indicating CRC risk due to FH and appropriate screening. *Outcomes:* Appropriate patient-reported CRC screening based on FH risk. Logistic regression was used to determine predictors of appropriate screening. *Results:* Of 121 FPs at baseline, 41 (34%) completed all 4 questionnaires (54% female, mean age 45). Of 208 patients, 98 (47%) completed 5 and 10 year follow up (76% female, mean age 50, 66% low/34% moderate or high risk for hereditary CRC). At the 10 year follow up, 93% of FPs reported they routinely asked patients about FH of cancer (agree/strongly agree), 71% of patients reported they discussed FH of CRC with their FP and 60% of patients reported their FP recommended CRC screening. Appropriate patient-reported CRC screening (test + timing) declined over 10 years particularly in the moderate/high risk group (baseline: 76%, 5-year: 52%, 10-year: 54%; p=.036). Most moderate/high risk patients (~80%) underestimated their risk consistently over 10 years. At 10 years, patients reporting that their FP recommended screening were more likely to have had the appropriate test and timing (OR 5.4 95% CI 1.7, 16.7). Moderate/high risk patients were less likely to have had the appropriate test and timing (OR 0.20, 95% CI .07, 0.61). *Conclusion:* Only 65% of patients with a FH of CRC (patients most likely to benefit from screening) screened appropriately. Recommendation from the FP is an important predictor of screening.

510F

The Tell Me More Study: Returning medically actionable genomic variants in the setting of a neonatal genome sequencing study. *T. DeMarco[1], K. Hurley[2], S. Hull[3], B.E. Berkman[3], E. Klein[1], A. Fuller[1], K. Huddleston[1], B. Solomon[1], F.M. Facio[3].* 1) Division of Medical Genomics , Inova Translational Medicine Institute, Fairfax, VA; 2) Cleveland Clinic, Cleveland, Ohio; 3) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

The Tell Me More Study (TMM) investigated parental decision-making about enrolling newborns in genome sequencing (GS) research. This qualitative pilot study is part of the larger Inova Childhood Longitudinal Study at Inova Translational Medicine Institute, which follows offspring through the course of gestation to age 18 and provides GS to all participants. Through TMM, GS was performed concurrently for mother-father-child trios. Parents could learn about GS findings for themselves and/or for their child, deemed medically actionable under the ACMG guidelines. We report a case of a 45-year-old Asian-American male who was found to carry a deleterious *BRCA2* truncating mutation previously reported in individuals of Filipino descent. GS analysis also showed that his 3-year-old daughter carries the *BRCA2* mutation; however, the clinical team counseled the parents to defer clinical confirmatory testing to an age when she can make the decision autonomously. Thus, only the father's result was confirmed in a CLIA-certified lab. Family history collected after uncovering the *BRCA2* mutation revealed a paternal 1st cousin with breast cancer in her mid-40s. The father's unaffected parents and brother were provided genetic counseling and segregation testing, which showed paternal transmission, and revealed that his brother is a carrier. An interview at 1-month post-disclosure indicated that while the father was initially surprised by the results and was concerned about his daughter's risk, he felt that the information was empowering, allowing him and his at-risk relatives to participate in risk-reducing strategies. He reported sharing the information with relatives and a few colleagues, and intention to share it with his primary care provider. He envisioned sharing the results with his daughter when she is older and able to understand scientific principles. Despite learning this unexpected result, he did not regret his participation in the TMM study; his scientific background and interest in GS may have facilitated his positive adjustment. The father's result demonstrates that GS of infants may open another route to actionable results in adults who are not the original focus of testing. The case illustrates that people can adjust to and make use of unexpected GS results over time. However, it foreshadows new challenges in adhering to guidelines for deferral of *BRCA1/2* testing until adulthood if GS in newborns is adopted more widely. *TMM was supported by the 2014 JEMF.

511W

Genetic counseling assistants in a cancer genetics clinic: Genetic counselor time utilization and impact on patient volume. *M.L.G. Hallquist, E.P. Tricou, J.M. Savatt, H. Rocha, N. Deckard, A. Evans, A. Fan, W.A. Faucett, J. Pervola, A.K. Rahm, M. Rashkin, T. Schmidlen, M.L.B. Schwartz, J.L. Williams, M. Williams, A.H. Buchanan.* Genomic Medicine Institute, Geisinger Health System, Danville, PA.

Given steadily increasing numbers of patients seeking genetics services and the current workforce shortage of certified genetic counselors, it has become imperative that we adapt our service delivery models. One model that may increase genetic counselor efficiency while maintaining quality care is the use of a genetic counseling assistant (GCA) – a BS/BA trained assistant who has an interest in pursuing a graduate degree in genetic counseling. At Geisinger's Cancer Genetics Risk Assessment Clinic, we have delegated tasks such as triaging referrals, obtaining pedigrees, and confirming appointments to GCAs. By implementing a GCA, the goal was to maximize the time genetic counselors spent on tasks for which they are uniquely trained, such as direct patient care. This study explores the impact of GCAs on genetic counselor time utilization and patient volume. Using an online time tracking tool, genetic counselors and GCAs in the cancer genetics clinic recorded the amount of time spent on clinic-related tasks during three distinct time periods (Nov-Dec 2015, Mar-Apr 2016, Apr-May 2017). Tasks were binned into pre-appointment, direct patient care, post-appointment, or ancillary responsibilities. The genetic counselor FTE remained stable across all three time periods at 1.5 FTE, while the GCA FTE increased from baseline of no GCA, to 0.6 FTE at the second time point, to 1.3 FTE of GCA support at the third time period. Electronic medical record data were used to evaluate total numbers of patients seen (by phone or in person new appointments and results return appointments) during each tracked period. Total patients seen increased by 38% from baseline to the third time point (n=82 patients at baseline, 99 at time two, and 133 at time three). This corresponds to an increase in the number of new patients seen per genetic counselor FTE from 6.3 new patients per week per FTE at baseline to 10.2 new patients per week per FTE during the most recent time period. Moreover, genetic counselor time spent on direct patient care increased three-fold from baseline to the third time period (32 hours to 63 hours to 109 hours). These data suggest that use of a GCA may allow genetic counselors to focus their expertise on direct patient care, while simultaneously increasing patient volume.

512T

A reporting of secondary findings in whole exome sequencing in Japan: The Project of HOPE. *Y. Horiuchi[1,2], Y. Kiyozumi[2,4], H. Matsubayashi[1], S. Nishimura[1], K. Urakami[3], S. Oonami[3], M. Arai[1], M. Kusuhara[3], K. Yamaguchi[3].* 1) Psychiatry and Behavioral Sciences Dept, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Genetic Counseling Dept, Shizuoka Cancer Center Hospital, Shizuoka, Japan; 3) Research institute of Shizuoka Cancer Center, Shizuoka, Japan; 4) Graduate school of Ochanomizu University, Tokyo, Japan.

Recent advantage of whole genome sequencing (WGS) and whole exome sequencing (WES) techniques are increasingly used in both clinical and research contexts. On the other hand, new ethical issues regarding the disclosure of incidental or secondary findings, i.e., results that are unrelated to the original aims of the study may be expected to arise in the genome studies. In 2013, the American College of Medical Genetics (ACMG) announced recommendations for handling incidental findings in genome sequencing and published a revised version in 2016. However, the guidelines for the secondary findings and a system of the return of results for patients have not been well established yet in Japan. In this study, we systematically constructed the genetic counseling system from clinical genome study "Shizuoka Cancer Center (SCC) the Project HOPE (High-tech Omics-based Patient Evaluation)", as a basis for initial trials in Japan. The aim of Project HOPE is to evaluate biological characteristics of cancer and diathesis of each patient by multi-omics based analyses, which integrate genomics, transcriptomics, proteomics and metabolomics. We here report the prevalence of secondary findings and the system of the return of results to patient from the HOPE study. The disclosure of secondary findings for the patients were conducted carefully and systematically in SCC. To date, 3000 cases have been analyzed and approximately 1% of the total cases were reported to be associated with hereditary cancer syndromes and approximately 1% of the total cases were reported as based on the guidelines of the ACMG. Our systematic approach will contribute to a cancer genomic medicine and serve as a useful basis for genetic counseling system in Japan. This study was approved by the Ethics Committees of Shizuoka Cancer center and Tokyo Metropolitan Institute of Medical Science.

513F

Factors influencing sharing of genetic information: An in-depth interview study of patient who underwent cancer genetic testing in Singapore.

S.T. Li[1], S. Sun[2], D. Lie[3], M. Met-Domestici[1], J. Ngeow[1,4,5]. 1) Cancer Genetics Service, Division of Medical Oncology, National Cancer Centre Singapore, Singapore; 2) Division of Sociology, School of Humanities and Social Sciences, Nanyang Technology University, Singapore; 3) Office of Clinical Sciences, Duke-NUS Medical School, Singapore; 4) Oncology Academic Clinical Program, Duke-NUS Medical School, Singapore; 5) Division of Medical Oncology, National Cancer Centre Singapore, Singapore.

PURPOSE The uptake rate of predictive genetic testing of hereditary breast and ovarian cancer syndrome in at-risk relatives has been low in Asia, one known reason was proband's reluctance to share genetic results. This qualitative study aims to explore the factors that influence patient decision-making regarding whether or not to share their genetic test results with their at-risk relatives. **METHODS** In-depth interviews were conducted with 24 female patients with personal or family history of breast and/or ovarian cancer and received a positive, negative, or variant of uncertain significance genetic test result. Grounded theory with an inductive approach was used to identify major themes and develop a theoretical framework. **RESULTS** There were three overarching themes among the three result groups influencing the likelihood of patients sharing their genetic results: (i) family closeness; (ii) involvement of families in the testing process; and (iii) perception of low emotional impact. Presence of actionable genetic results and participant's perception of family members' acceptance increases likelihood of sharing in the positive results group. For the uncertain and negative results group, participants were less likely to disclose their genetic results if results were ambiguous or the participants perceived that there was no genetic or medical implication for their family. **CONCLUSION:** Our findings and theoretical framework provides a model of the influencing factors that involve in women's decision of sharing genetic result and highlights the important timing for the implementation of interventions to encourage or facilitate family communication of genetic results.

514W

Uptake of polygenic risk information among women at potentially high breast cancer risk.

B. Meiser[1], T. Yanes[1,2], R. Kaur[1], M. Scheepers-Joynt[3], M. Young[4], K. Barlow-Stewart[5], G. Lindeman[6], T. John[7], M. Harris[8], Y. Antill[9], J. Burke[10], T. Roscioli[11], J. Halliday[12,13]. 1) Prince of Wales Clinical School, UNSW Sydney; 2) School of Psychiatry, UNSW Sydney; 3) Familial Cancer Service, Peter MacCallum Cancer Centre, Melbourne; 4) Genome.One, Darlinghurst, Sydney,; 5) Sydney Medical School Northern, University of Sydney; 6) Familial Cancer Centre, The Royal Melbourne Hospital, Melbourne; 7) Clinical Genetics Service, Austin Hospital, Melbourne; 8) Family Cancer Clinic, Monash Medical Centre, Melbourne; 9) Family Cancer Clinic, Cabrini Health, Melbourne; 10) Tasmanian Clinical Genetics Service, Royal Hobart Hospital, Hobart; 11) Department of Medical Genetics, Sydney Children's Hospital, Randwick, Sydney; 12) Public Health Genetics, Murdoch Children's Research Institute, Melbourne; 13) Department of Paediatrics, University of Melbourne, Parkville, Melbourne.

Background: Despite increasing evidence for the utility of polygenic risk in families at high risk of breast cancer, research findings are yet to be integrated into clinical practice. This reflects the status of polygenic risk as an emerging technology and the limited evidence base on the psychosocial and behavioural outcomes of offering such testing. **Aims:** To ascertain the important psychosocial and behavioural implications of testing for polygenic breast cancer risk in the existing cohort the "Variants in Practice" (ViP) study. **Methods:** 400 women enrolled in VIP, who have either a high or low polygenic risk score (PRS), and whose personal and/or family history breast cancer remained unexplained after genetic testing for known cancer predisposition genes are being invited to participate in this study. Participants complete a baseline questionnaire assessing their knowledge of hereditary breast cancer, current breast cancer screening behaviours, psychological well-being, and intention to receive personal PRS results. **Results:** As of April 2017, 69/74 (93%) participants reported interest in receiving their PRS result, with 20/74 (27%) having received their results. The mean knowledge score among participants was 6 (out of 10), and 20% and 8 % of participants scored over the cut-offs for cancer-specific distress and general depression respectively. **Conclusion:** While, there is strong interest in receiving personal PRS result among women at high risk of breast cancer, the psychosocial and behavioural implications should be carefully considered. Data collection is ongoing, with additional data regarding uptake of results and short-term impact of receiving results to be presented.

515T

Assessing the possibility of *RUNX1*, *ETV6* and *GATA2* related germline predisposition in myeloid neoplasms in a somatic cancer setting. *K. Barber, J. Bissonnette, R. Klein, J. Weisberger, J. Zhao.* BioReference/GenPath, Elmwood Park, NJ.

Recently the World Health Organization (WHO) released an update to the classification of hematopoietic neoplasms with germline predisposition based on the presence of germline pathogenic variants in RUNX1, ETVS and GATA2 (WHO, 2016). Individuals who are heterozygous for a germline RUNX1 variant are at a 35-44% risk of myelodysplastic syndrome (MDS [614286]) or acute myeloid leukemia (AML [601626]) (Owen et al, 2008, Godley, 2014). Germline ETV6 variants are associated with Thrombocytopenia 5 [616216], an autosomal dominant thrombocytopenia. Clinical presentation for those with germline GATA2 variants is variable, however lifetime risk for MDS or AML is estimated to be 70%. We performed a retrospective analysis of all *RUNX1*, *ETV6* and *GATA2* variants identified on a Next Generation Sequencing panel for a myeloid neoplasm. Our goal was to determine how the new classification released by WHO would affect our strategy in reporting these variants. *RUNX1*, *ETV6* and *GATA2* variants were detected in 637 cases in our lab out of the total 7,092 myeloid cases analyzed over the last 17 months (637/7,092 = 9%). *RUNX1* variants are most commonly detected among these three genes (499/637= 78%). The median age of all patients in this study was 71. Of the cases where a diagnosis was known, *RUNX1/GATA2/ETV6* variants were most commonly seen in AML. Of a total of 756 unique *RUNX1*, *ETV6* or *GATA2* variants reported, 67% were known disease associated variants while 33% were novel or unclear variants. Of the 507 disease associated variants, 155 were found to have an allele frequency suggestive of a possible germline variant (31%). Of the ten youngest patients (ages 6-26) identified with one of these variants, seven (70%) were detected at an allele frequency suggestive of a germline mutation (40-60% allele frequency). Based on the relatively high proportion of cases with an allele frequency suggestive of a germline variant and the significance of a germline variant, we conclude that including a statement on the final report regarding the possibility of a germline mutation is reasonable in all cases where one of these variants is detected. Germline confirmation testing should be pursued when indicated. This will allow for appropriate management of those with germline predisposition to myeloid malignancies, and would also promote testing of family members that could be at risk. In addition, this study shows the value of having genetic counseling expertise within a somatic testing lab.

516F

Quality Improvement Project: Shared decision making in genetic cancer risk assessment and patient barriers to risk assessment and reduction in primary care. *D.M. Person, M.B Massart, I. Campwala.* University of Pittsburgh, Pittsburgh, PA.

Introduction Shared decision making and personalized medicine is becoming more common in primary care. It is especially valuable when considering family history for genetic cancer risk assessment as there are several risk reduction opportunities. Despite this, Genetic Cancer Risk Assessment (GCRA) has been underused in primary care (Trepanier at el 2016). Additionally, the barriers to genetic services amongst ethnic minorities have not been well characterized. We sought to evaluate the routine utilization of initial GCRA in our office and to survey patients regarding perceptions and barriers to genetic services. **Methods** GCRA was administered to 94 adult patients at our FQHC UPMC-Matilda Theiss. Patients were also asked to complete a survey about their perceptions and barriers to GCRA. Questions measured barriers based on religion, education about genetics/genetic services, family and community stigma, and medical mistrust. Survey answers were analyzed to assess the degree to which each of these barriers affects members of various patient populations. **Results** 31.91% screened positive and would be considered for additional genetic services based on GCRA. Without adjusting for socioeconomic factors, African American and Caucasian patients' answers reflected a similar education barrier to genetic services: 37.93% and 37.14%, respectively. African Americans had a greater religious barrier (19.28%) versus Caucasian patients (15.79%). 12.32% of African American patients' answers showed a possible barrier based on family/community stigma versus only 8.77% of Caucasian patients answers. Both African American and Caucasian patients selected Medical Mistrust as a significant barrier to genetic services, 41.42% and 40.81%, respectively. A t-test found that there is a significant difference in the religious barriers reported by African Americans versus Caucasian patients (p-value=0.027). Once we adjusted for socioeconomic factors we found that Medical Mistrust became a significant barrier for African American patients with a p-value of 0.047. Religion was almost significant with a p-value of 0.056. **Conclusion** Our African American patients reported a greater religious barrier to genetic services than their Caucasian counterparts. Once we adjusted for socioeconomic factors Medical Mistrust became a significant barrier for our African American patients. These findings are crucial to the continuing improvement of GCRA cultural competence and effectiveness in primary care.   .

517W

All in the Family: How family history affects diagnostic yield of hypertrophic aardiomyopathy multigene panel testing. *S.J. Martin, J. Dolinsky, L. Hoang, C. Antolik, T. Johnson, B. Dougall.* Ambry Genetics, Aliso Viejo, CA.

Family history is often a predictor of disease. For hypertrophic cardiomyopathy (HCM), a family history (fhx) is seen in approximately 72% of cases in which a mutation is identified (1). To further assess the relationship between fhx of HCM and the probability of identifying a pathogenic mutation in a HCM-associated gene, we examined the family history of all probands with HCM who underwent multigene panel (MGP) testing at our diagnostic laboratory from April 2015 to June 2016. Particular emphasis was placed on analysis of *MYBPC3* and *MYH7*. Clinical history was collected from test requisitions, available pedigrees and clinical notes. Among 237 patients referred for HCM MGP testing, 32% (n=77) were positive for a mutation, the majority of which were in *MYBPC3* (n=44; 57%) and *MYH7* (n=13; 17%). Among the 57 HCM probands with positive MGP testing results who reported fhx information to our laboratory, 84.2% (n=48) reported 'some cardiac fhx.' This included 36.8% (n=21) reporting a fhx of HCM, 28.0% (n=16) reporting a fhx of sudden cardiac arrest (SCA). In addition, 15.8% (n=9) reported no cardiac fhx. Probands with mutations in *MYBPC3* had a higher frequency of reporting a fhx of HCM (14/34; 41.2%) or SCA (10/34; 29.4%), compared to probands with *MYH7* mutations, in which fhx of HCM was reported in 22.2% (2/9) of probands and fhx of SCA was reported in 22.2% (2/9) of probands. Approximately half of the *MYBPC3* mutations were either protein truncating, haploinsufficient or splice site mutations (52.2%), while the remainder (47.7%) were missense mutations, whereas *MYH7* mutations were exclusively missense alterations, consistent with the mechanism of disease for this gene.   The prevalence of fhx of HCM and SCA in our cohort is lower than previously published rates of 72% and 89% (1), respectively. This may reflect ascertainment bias in previous cohorts, or an uptake of genetic testing among singleton cases of HCM in recent years. Nevertheless, these data demonstrate a role for genetic testing among sporadic or familial cases of HCM. Cardiologists and other healthcare providers should still take a thorough fhx among patients newly diagnosed with HCM. Further exploration of genotype-phenotype relationships will be valuable for better understanding the predictive value of genetic testing for HCM and may be useful in family risk counseling.   1-Genet Med. 2013 Dec;15(12):972-7. doi: 10.1038/gim.2013.44. Epub 2013 Apr 18. 2-Circ Cardiovasc Genet. 2013 Feb; 6(1): 118–131. .

518T

Potential impact of statin pharmacogenetic testing in an integrated healthcare system: The Integrating Pharmacogenetics in Clinical Care (I-PICC) Study. *S.J. Miller[1], C. Hau[1], N. Mahajalme[1], A.J. Zimolzak[1], L. MacMullen[1], J.L. Vassy[1,2,3].* 1) VA Boston Healthcare System, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Brigham and Women's Hospital, Boston, MA.

Background: The pharmacogenetic (PGx) association between *SLCO1B1* rs4149056 and simvastatin myopathy is well validated, but the impact of its integration into patient care is unknown. A key determinant of that impact is the pre-PGx landscape of statin use in a healthcare system. We aimed to describe the statin use in the primary care and women's health clinics of the VA Boston Healthcare System (VABHS) before the I-PICC Study, an ongoing randomized controlled trial of preemptive *SLCO1B1* genotyping. **Methods:** Eligible patients for the I-PICC Study are statin-naïve but have elevated cardiovascular disease (CVD) risk by American College of Cardiology/American Heart Association guidelines (ACC/AHA). Specifically, they are aged 40-75 and have ≥1 of the following criteria: 1) preexisting CVD, 2) diabetes, or 3) LDL cholesterol ≥190mg/dL. We used electronic health record data to quantify the patients seen in VABHS primary care and women's health in 2013-2016 who would have been eligible for the I-PICC Study. We defined diabetes and CVD from ICD codes. We determined prior and incident statin use from VA and non-VA medication orders. For each eligible patient, we defined a baseline date as the later of 1/1/2013 or the date the patient met I-PICC Study eligibility criteria. We calculated time to first statin prescription through 12/31/2016. **Results:** Of 34,254 unique patients seen at VABHS 2013-2016, 3,691 (10.8%) would have been eligible for the I-PICC Study at some time (171 [4.6%] women, 508 [13.4%] non-white). Of these, 1,827 (49.5%) had a first statin prescription during the observation period (incidence 280/1000 person-years). Of these, 911 (49.9%) were for atorvastatin (140/1000 person-years), and 689 (37.7%) were for simvastatin (106/1000 person-years). If 20% of the patients started on simvastatin had at least one copy of the C allele at rs4149056, then Clinical Pharmacogenetics Implementation Consortium guidelines for simvastatin prescribing might have guided alternative treatment for 138 (7.5%) of new statin prescriptions. **Discussion:** Preemptive *SLCO1B1* testing to guide statin initiation and avoid serious myopathy might change clinical management for 1 in 13 patients with elevated CVD risk. The I-PICC Study will examine the intended and unintended outcomes of integrating *SLCO1B1* testing, including rates of statin initiation, choice of statin, concordance with ACC/AHA guidelines, medication adherence, and documented adverse drug effects.

519F

Diabetes incidence after a polygenic risk intervention: Five-year follow-up in the Genetic Counseling/Lifestyle Change for Diabetes Prevention (GC/LC) Study. *J.L. Vassy[1,2], J.B. Meigs[2,3], W. He[3], J.C. Florez[2,3,4], R.W. Grant[5].* 1) VA Boston Healthcare System, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Department of Medicine, Massachusetts General Hospital, Boston, MA; 4) Broad Institute, Cambridge, MA; 5) Division of Research, Kaiser Permanente, Oakland, CA.

   Background   Although polygenic risk scores can predict type 2 diabetes (T2D), we have shown that learning one's polygenic risk did not impact short-term weight loss or motivation for behavior change among high-risk partici-pants in a T2D prevention program. The long-term impact of learning genetic risk on T2D incidence is unknown.   **Methods**   The GC/LC Study (2010-2011) was a randomized controlled trial of polygenic T2D risk counseling among adult primary care patients at one academic medical center. Participants were overweight and met at least one other metabolic syndrome criterion without a T2D diagnosis. They were randomized in a 4:1 ratio to undergo genotyping for a weighted 36-locus T2D risk score (Tested) vs. no genotyping (Control, *n*=34). Tested participants in the highest (High, *n*=42) and lowest (Low, *n*=32) quartiles of polygenic risk (together, Intervention) were retained in the study and received their results from a genetic counselor. All participants were then enrolled in a 12-week group weight loss program. Although 33 (31%) participants lost ≥5% body weight, weight loss and motivation did not differ between study arms. In this 5-year follow-up study, we identified cases of incident T2D among study participants, using electronic health records (EHR) and a validated algorithm of problem list terms and billing diagnoses from ≥2 outpatient visits. We used log-rank tests to compare T2D incidence rates (with exact 95% confidence intervals) among study arms through the date T2D criteria were met or the last EHR visit through 2016.   **Results**   Mean baseline age was 59 (11) years; 40% were women. Follow-up data (median 6.0 years) was available on all participants. Incident T2D occurred in 5 (15%), 13 (31%), and 9 (28%) of Control, High, and Low participants, respectively. Incidence rates per 1000 person-years were: Control 27 (12-64), Low 57 (30-107), High 70 (41-120). Between-group differences in time-to-T2D were not statistically significant (Intervention vs. Control, *p*=0.08; High vs. Control *p*=0.07; Low vs. Control *p*=0.18; High vs. Low *p*=0.63).   **Discussion**   Disclosing and coun-seling about polygenic risk did not reduce long-term T2D incidence among patients at high phenotypic risk for T2D. The possibility that T2D incidence was actually higher among intervention participants (both High and Low genetic risk) merits replication and investigation of whether receipt of genetic results resulted in poorer health behaviors vs. increased T2D screening and diagnosis.

520W

Acid sphingomyelinase deficiency (ASMD): Disease impact on families and caregivers. *R. Avetisyan[1], A. Hareendran[2], B.J. Sanson[1], S. Tan[1].* 1) Sanofi Genzyme, Cambridge, MA, USA; 2) Evidera, London, UK.

   Statement of Purpose: Acid sphingomyelinase deficiency (ASMD) is a rare, genetic, progressive, and potentially life-threatening disorder associated with significant disease burden. While the impact of ASMD on patients is known, it is also important to understand the extent of impact of the disease on the quality of life of the patients' families and caregivers. **Methods:** Qualitative research approaches were used to obtain information from the family caregiv-ers about the disease impact on patients, as well as to explore the impact of the disease on the caregivers themselves. This research was part of a larger patient-centered study; here we report the preliminary results of the self-re-ported impacts of ASMD on caregivers. **Summary of Results:** The interviews included a total of 10 primary caregivers/ family members of patients with ASMD ages 2-19 years. The recruitment was conducted in the US and UK. The average age of the caregivers was 38.7 (range 33-47) years; 70% of participants were mothers of the patients. Caregivers reported a wide range of major impacts on their lives. Significant distress and emotional impact, report-ed at the time of diagnosis ("crushed our heart", "felt really guilty", "cried every day", "it's really stressful"), evolved throughout care for the patients and includ-ed anxiety and worries about the child's life and future. Caregivers reported impacts on physical and mental health ("back hurts" from having to move/carry the child, "a huge level of fatigue", "interrupted sleep and it's constant", "can't exercise", postponing surgery, "went into depression", "needing a consultation with a mental health practitioner"). Social impacts and impact on relationships ("affected marriage", "cause of divorce") were also reported. The caregivers also experienced financial impacts, and a number of participants reported their employment being impacted by their child's disease ("lost job" because of fre-quent doctor's visits, "gave up career to be more flexible", stopped working as "felt guilty didn't have time" to take care of the child). A frequently mentioned source of support for caregivers was family members and friends. **Conclu-sions:** These findings demonstrate substantial quality of life impacts of ASMD on the family caregivers. This research provides further evidence about the broader disease burden of ASMD, and demonstrates the extended impact that rare diseases have, not only on patients, but also on their families. Supported by Sanofi Genzyme.

521T

Screening African Americans for *APOL1*-associated kidney disease risk: Stakeholder views. *S.M. Fullerton[1], E. Umeukeje[2], E. Blacksher[1], C. Spigner[3], J. Wilson[4], K. Cavanaugh[2], B. Young[5], W. Burke[1].* 1) Dept Bioethics & Humanities, University of Washington, Seattle, WA; 2) Vanderbilt Center for Kidney Disease, Vanderbilt University Medical Center, Nashville, TN; 3) Dept Health Services, University of Washington, Seattle, WA; 4) Dept Physiology & Biophysics and Medicine, University of Mississippi, Jackson, MS; 5) Dept Medicine (Nephrology), VA Puget Sound Health Care System and University of Washington, Seattle, WA.

   Background: The appropriate use of genetic information in addressing racial/ethnic health and healthcare disparities is hotly debated. For example, variants of the apolipoprotein L1 (*APOL1*) gene, associated with increased risk of end-stage renal disease (ESRD), are more common in individuals with recent West African ancestry, partially explaining disproportionate rates of ESRD among African Americans compared to whites. However, the biological mechanisms linking *APOL1* variation to ESRD risk remain poorly understood, factors affecting penetrance are uncertain, and appropriate preventive measures unclear. Routine clinical screening of *APOL1* variation in patients at increased risk of kidney disease may therefore pose risks as well as benefits. It is important to consider the views of major stakeholders before such testing is widely pursued. **Methods**: As part of a larger project focused on engagement with the African-American community about *APOL1* genetic testing in research, clinical care, and kidney transplantation [R01HG007879] we conducted and analyzed semi-structured interviews with 10 kidney disease researchers, 25 clinicians (including both primary care providers and nephrologists), and 25 community members (including patients, family members, and others not directly affected by kidney disease) in Seattle, WA, Jackson, MS, and Nashville, TN. Interviews included exploration of attitudes toward routine *APOL1* screening among African Americans. **Results**: Many community members responded positively to the prospect of routine clinical screening of *APOL1*, typically based on expectations of benefit; few expressed misgivings about misuses of test information. Researchers and clinicians, in contrast, were more skeptical about the value of routine clinical screening given current uncertainties surrounding clinical utility, favoring additional research before widespread clinical implementation. Expert key informants were also more likely to identify potential risks associated with routine screening, including psychosocial impacts, costs, insurance discrimination, and other forms of stigmatization. **Conclusions**: Community member views in this study reflect a positive societal discourse about risk identification that many researchers and clinicians do not share. There is a need for both better evidence and substantive discussion between experts and the public before wide-spread screening moves forward.

522F

Association between lower folic acid level and unbalanced neuron excitability in Chinese pediatric patients. *Y. Liu, X. Zhang, Y. Yu, X. Nie, X. Wang.* Pharmacy department, Beijing Children's Hospital, Beijing, China.

   Introduction: Folic acid (FA or VB_9) is an essential factor for many biochemical reactions. However, human cannot synthesis it on their own. The main source for folic acid is from one's diet. The deficiency of foliate intake would result in multiple neurological and hematological diseases. To confirm its association with different neuro-etiology, we were conducted an analysis of nine types of vitamin monitoring data from 4367 Chinese pediatric patients that visited our hospital. **Method and patient**: We collected data on nine types of vitamins (VA, VC, VD, VE, VB_1, VB_2, VB_6, VB_9, VB_{12}) in our monitoring data from 4367 pediatric patients that visited our hospital since August 2016 to April 2017, and extracted a diagnosis related to neuropsychiatry, including 'seizure', 'epilepsy', 'convulsion', 'mental/language deterioration', 'gait disturbances', etc. to evaluate if there is any association between phenotype and one's vitamin level. The monitoring method was an electrode detection. **Result**: Patients (n=475) with unbalanced neuron excitability (seizure, epilepsy, convulsion) symptoms are significantly ($P<0.001$) associated with a lower FA level [95%CI: -8.54 to -4.09, nmol/L] when compared to healthy controls (n=882) or the remaining patients (n=2903), but no difference was found in mental/language deterioration, gait disturbance, ASD or autism symptoms (n=107). **Conclusion**: The association between neuropsychiatry and FA deficiency was established several decades ago. However, our study reflects a slightly different result via a clinical outcome, which divides these neuropsychiatry into two groups: association with all cause unbalanced neuron excitability and no association with mental developmental disorder. The intrinsic cause of this result is still under investigation.

523W

Providing genomic medicine to the Hispanic population at the Stanford Center for Undiagnosed Diseases. *L. Fernandez[1,2], J. Kohler[1,2], A. Dries[1,2], D. Zastrow[1,2], J. Davidson[1,2], C. McCormack[1,2], M. Majcherska[1,2], D. Bonner[1,2], D. Waggott[1,2], S. Marwaha[1,2], N. Friedman[1,2], UDN. Members[3], P. Fisher[1,4,5], E. Ashley[1,2,6], JA. Bernstein[1,4,7], M. Wheeler[1,2].* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, California 94305-5406; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, California 94305; 3) NIH Undiagnosed Diseases Network, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892; 4) Department of Pediatrics, Stanford School of Medicine, Stanford, California 94305-5208; 5) Department of Neurology, Stanford School of Medicine, Stanford, California 94305; 6) Department of Genetics, Stanford School of Medicine, Stanford, California 94305-5208; 7) Lucille Packard Children's Hospital Stanford, Palo Alto, California 94304.

Minority groups, including Hispanic populations, have been historically underrepresented in genomic medicine studies. The Stanford Center for Undiagnosed Diseases (CUD), a clinical site of the Undiagnosed Diseases Network (UDN) has developed tools and workflows to overcome barriers and challenges present when working with self-identified Hispanic patients, such as culture, language, limited access to internet, difficulty to request medical records to multiple health providers, and unfamiliarity with genetics and genetic research. We are addressing these challenges, facilitating enrollment and active participation of Spanish-speaking participants in our study. One of the actions we've taken to overcome these challenges, is having a fluent team coordinator who guides these families throughout all their participation in our study. Our goal is to find answers for individuals whom exhaustive clinical work-up remains unrevealing. Eighteen of the CUD's first 70 participants are of Hispanic ethnicity. We have identified a new diagnosis in six of these cases to date. We present two of these cases. Case 1 is a 10 year-old with profound developmental delay, frontal lobe intractable epilepsy, neurogenic bladder, and choreathetoid cerebral palsy. He had previous extensive imaging, metabolic and genetic testing. Trio whole exome sequencing performed through the UDN identified a mosaic pathogenic variant in the *EEF1A2* gene: heterozygous, de novo c.796C>T (p.R226W). *EEF1A2* plays an essential role in protein synthesis, and mutations have been shown to cause epileptic encephalopathy and mental retardation. Interestingly, some of his clinic findings suggest mosaicism, including asymmetry on brain MRI. Case 2 is a 7 year-old with a history of epilepsy and speech delay. Trio whole exome sequencing performed through the UDN identified a pathogenic variant in the *FGF12* gene: heterozygous, de novo c.334G>A (p.G112S). *FGF12* is involved in the positive regulation of voltage-gated sodium channel activity; mutations cause epileptic encephalopathy. Interestingly, our patient has a variant, not described in the literature so far, with the exception of one unpublished case with apparently a more severe phenotype. These two interesting cases demonstrate the importance of having a workflow to approach this population and offer them these resources. By ensuring broad access to the UDN, we can increase knowledge of the underlying mechanisms and pathways of unknown and very rare diseases.

524T

Physicians attitudes, knowledge and barriers towards the inborn errors of the metabolism in a university hospital. *F. Suarez[1,4], M. Pedraza[2,5], S. Moreno[3].* 1) Instituto de Genetica Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Bogota Colombia; 2) Servicio de Pediatría, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana, Bogota Colombia; 3) Departamento de Epidemiologia y Bioestadistica, Facultad de Medicina, Pontificia Universidad Javeriana, Bogota Colombia; 4) Servicio de Genetica, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana, Bogota Colombia; 5) Clinica de Errores Innatos del Metabolismo, Hospital Universitario San Ignacio, Pontificia Universidad Javeriana, Bogota Colombia.

IEMs (Inborn errors of metabolism) are considered low-frequency diseases. Around every 1/100,000 births. However, considered together, their incidence can be close to 1/800 - 1/2500. Current knowledge on their biochemical basis allows pharmacological and nutritional treatments that change the natural history of some of these diseases. With the objective of describing knowledge of IEMs in medical specialists and residents in a University Hospital, we implemented a survey for assessing beliefs, attitudes, and barriers towards IEMs. A total of 94 physicians from Pediatrics, Internal Medicine, and Emergency room completed the survey. 75% of them had previous contact with IEMs patients, although, 64% consider that IEMs are not a common group of diseases. 90% consider that IEMs are chronic conditions, 61% consider them as degenerative diseases. 87% believes that diagnostic of IEMs can be suspect only with clinical signs. However, achieving a final diagnostic is a challenging task for 60% of physicians. 70% consider that the confirmatory laboratory results for IEMs take too long to arrive at their services. 72% believe that the HMOs impose barriers for authorizing specific test. 74% of the physicians are aware that the treatments are costly, 72% agree that current managements are not available for the majority of patients. The survey presented 4 clinical scenarios; the physicians chose if they were able to recognize the diseases and clinical signs of decompensation, 50% of the doctors in each clinical case, failed to recognize normal values of commonly used metabolites like ammonia. More than 90% of the physicians were not able to identify the clinical signs of disease decompensation and recognize no to know when to send a patient to urgent treatment. These results show that there are many misconceptions around IEMS. Improving education and training in these topics is imperative. The improvement of prognosis depends on the prompt and accurate diagnosis. Therefore physicians training must highlight the importance of recognition of important clinical signs and adequate interpretation of laboratory test. It is relevant that the physicians know that many of these diseases are not chronic but acute and mortal without treatment, wich is of particular importance is countries that lack newborn screening programs. We proposed the structure of a basic program for training physician in IEMs, and some other solutions for overcoming obstacles originated from the HMOs.

525F

A 25-year experience of Fragile X syndrome molecular diagnosis from a laboratory in Thailand. *P. Limprasert[1], D. Tangviriyapaiboon[2], O. Plong-On[1], T. Sripo[1], C. Charalsawadi[1].* 1) Department of Pathology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand; 2) Rajanagarindra Institute of Child Development, Chiangmai, Thailand.

Fragile X syndrome (FXS) is the most common form of X-linked intellectual disability (ID). To report the frequency of FXS in patients younger than 16 years referred for FXS DNA testing at a laboratory in Thailand, we retrospectively reviewed the results from previous research studies and unpublished data (293 males; 1991-1999; https://www.ncbi.nlm.nih.gov/pubmed/11400746) and 817 unrelated cases (771 males and 46 females) referred for FXS DNA testing at Songklanagarind Hospital during 2000-2015, to our knowledge the largest series of cases of molecular genetic FXS testing from Thailand. Overall, of the 1,064 males and 46 females tested in our laboratory, 80 males and 2 females with full mutation were identified (80/1,064 = 7.5% for males and 2/46 = 4.3% for females). The normal CGG repeats ranged from 5-47 with the two most common alleles of 29 (48.3%) and 30 (21.3%) CGG repeats. The frequencies of FXS from different cohorts and periods were all high (~7%), suggesting that patients with ID of unknown cause warrant molecular genetic testing for FXS (Table). In addition, female FXS carrier screening in the Thai population should be considered since our 25-year experience study found a high frequency of FXS. However, further studies of FXS carrier screening need to be done before implementing a population-based FXS screening policy.

Frequencies of FXS in 4 cohort studies in Thailand			
Cohort studies / Region of Thailand	Periods of samples collection	Sources	Positive FXS/ total screened cases (FXS Frequencies)
Males with ID Pediatric clinic/ Southern	1991-1999	Limprasert et al 1999 and unpublished data	9/132 (6.8%) males
Males with ID Pediatric clinic/Central	1997-1999	Limprasert et al 1999 and unpublished data	12/161 (7.5%) males
Males with ID Regional child psychiatric clinic/ Northern	2008-2009	This study	8/109 (7.3%) males
Children referred for FXS testing Human Genetics laboratory, University hospital/Mixed southern and central	2000-2015	This study	51/662 (7.7%) males 2/46 (4.3%) females

526W

Detection of genomic causes of developmental brain disorders among research participants in large-scale sequencing initiatives: Results disclosure, cascade testing, and psychosocial implications. *B. Finucane[1], K. Wain[1], E. Palen[1], L. Kasparson[1], J. Overton[2], L. Habegger[2], O. Gottesman[2], E. Maxwell[2], J. Reid[2], A. Hare Harris[1], C.L. Martin[1], D.H. Ledbetter[1] on behalf of the DiscovEHR collaboration.* 1) Autism & Developmental Medicine Inst, Geisinger Health System, Lewisburg, PA; 2) Regeneron Genetics Center, Tarrytown, NY.

Developmental brain disorders (DBD), including autism, intellectual disability, and schizophrenia, are clinically distinct conditions that share overlapping genomic causes, including discrete pathogenic copy number and sequence level variants. Many individuals live with symptom-based developmental and psychiatric diagnoses without ever knowing the underlying etiologies that explain their disabilities. As part of the Geisinger Health System–Regeneron Genetics Center DiscovEHR genomic sequencing initiative, we defined a set of 33 pathogenic segmental duplication-mediated copy number variants (CNVs), including deletions of 1q21.1, 22q11.2, and 15q13.3. Within a DiscovEHR cohort of 47,589 participants undergoing whole exome sequencing, we identified 386 (0.80%) with a pathogenic DBD-related CNV. The majority were adults of Caucasian background, reflective of the demographics of the overall DiscovEHR group. Most individuals had a DBD documented in their EHR but had never been genetically diagnosed. We developed a return of results (ROR) process for DBD-related genomic results, based on a broader ROR protocol already in place at Geisinger. We carefully considered the potential for negative psychological reactions to learning about a genetic DBD cause; strategies for clinical cascade testing for participants' family members; and the impact of cognitive and psychiatric symptoms on the ROR process. These challenges were balanced against the potential positive impact on medical management, as well as the perceived value to participants of having an etiological explanation for their lived experience with DBD. We completed CLIA confirmation for pathogenic DBD-related CNVs in 40 participants, and we initiated a disclosure process that included genetic counseling and family cascade testing. Participant responses were overwhelmingly positive, including poignant reactions to learning that there was a "medical reason" for cognitive and psychiatric disabilities. For some participants, identifying a genetic diagnosis resulted in important changes to medical care, including pharmacological approaches to epilepsy and psychiatric management. We conclude that identification and disclosure of causative variants is a clinically and psychologically important outcome for adults with DBD. Research and strategies for disclosure of incidentally-detected pathogenic DBD variants are needed as genomic sequencing becomes an increasingly common part of clinical care and population research.

527T

Maternal origin of familial 22q11.2 deletions negatively impacts FSIQ scores. *D.E. McGinn[1,2], T.B. Crowley[1], M. Unolt[1,3], B.S. Emanuel[1,4], E.H. Zackai[1,4], E. Moss[1], B. Morrow[5], J. Vermeesch[6], A. Swillen[6], D.M. McDonald-McGinn[1,4].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) Davidson College, Davidson, NC; 3) La Sapienza University, Rome, Italy; 4) Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA; 5) Albert Einstein Medical College, Bronx, NY; 6) KU Leuven, Leuven, Belgium.

   Background: Familial 22q11.2 deletions reportedly have a negative impact on full scale IQ (FSIQ) in affected offspring thought to be the result of cognitive deficits in the affected parent/assortative mating but no study has investigated differences based on parent of origin (POO) which would be useful for recurrence risk counseling. **Methods:** We retrospectively reviewed our database of 211 familial subjects with 22q11.2 deletion syndrome (22q) from 89 families followed at the Children's Hospital of Philadelphia (CHOP) for those affected offspring with a FSIQ score comparing the mean FSIQ (MFSIQ) of those with familial v *de novo* deletions and maternally inherited (MI) v paternally inherited (PI) familial deletions. We then compared our findings to those from Leuven, Belgium and to *de novo* cases where POO studies (POOS) were performed in NY. **Results:** CHOP cohort: N=26 children with 22q from 23 families where 65% were inherited from an affected mother. As expected the MFSIQ (M=71.3) was statistically lower (p=.02) than for *de novo* cases (N=342, M=76.6) but notably the MFSIQ for MI familial deletions (M=68.3) was statistically lower (p=.03) than for PI deletions (M=76.3). Likewise in Leuven: N=26 children from 22 families where 73% were MI and the MFSIQ (M=61.4) for familial cases was lower than *de novo* (M=74.5) cases and the MFSIQ for MI deletions (M=59.8) was lower than PI deletions (M=65.9). Combining the cohorts: 67% of familial deletions were MI and the MFSIQ (M=66.4) was much lower than *de novo* cases (M=76.2). Moreover the MFSIQ for children with MI deletions (M=63.7) remained statistically lower (p=.03) than PI deletions (M=72.0). Lastly comparing the familial cohorts to the *de novo* POOS cohort: N=57 children where the MFSIQ for maternal deletions (N=37, M=73.41) was no different (p=0.68) than for paternal deletions (N=20, M=73.41). **Conclusion:** Here we confirm the association of lower FSIQ in familial versus *de novo* cases of 22q and report the novel finding that maternal origin of the familial deletion negatively impacts FSIQ in our American and Belgian cohorts. Confounders such as maternal comorbidities (repaired CHD, epilepsy, psychiatric illness, etc.), mitochondrial effects, socioeconomics, and level of engagement of the unaffected parent are important considerations. Regardless, a maternally inherited familial 22q11.2 deletion should now be considered a significant risk factor for poorer cognitive outcome requiring inclusion in genetic counseling.

528F

The significance of *KIAA2022* gene in development of early epileptic encephalopathy and intellectual disability in a girl. *S. Zhilina[1,2], T. Kozhanova[1,2], T. Mescheryakova[1], N. Prokop`eva[1], K. Osipova[1], S. Aivazyan[1], I. Kanivets[3], F. Konovalov[3], E. Tolmacheva[3], F. Koschkin[3], A. Prityko[1,2].* 1) St.Luka Scientific and Practical Center of Specialized Medical Care for Children, Moscow, Russian Federation; 2) Pirogov Russian National Research Medical University, Moscow, Russian Federation; 3) Genetic Center «Genomed», Russian Federation.

   KIAA2022 is gene of X-linked intellectual disability (MIM 300260) and pathogenic variants in gene are responsible for severe intellectual disability in boys, as well as epilepsy, postnatal growth retardation, autistic behavior, strabismus and dysmorphic facial features. The patient (girl 5 years) with diagnosis of cryptogenic epilepsy with myoclonic seizures, psycho-speech and intellectual development delay was observed in the psychoneurology department. A child was from 1 pregnancy, after a long period of infertility (the two previous marriages were infertile, 3 marriage - infertility for 4 years) and after extracorporeal fertilization; first delivery; weight - 3200 g., length - 50 cm, Apgar 8/9. The onset of the disease was in 1,5 years after the vaccination - myoclonic seizures, regress in psychomotor and speech development. EEG revealed epileptiform activity. Therapy - levothiracetam (250 mg per day). The drug was replaced with topiromate (75 mg per day). The seizures persisted in therapy. Karyotype - 46, XX. Weight 20 kg, height 108 cm. The child is hyperactive, impulsive, aggressive, poorly falling asleep, hardly speaking, acting on his own motivation, the teams are performed after multiple repetitions. There are microbrachycephaly, high line of hair growth, thin hair, thin elongated, lowered eyebrows, straight section of the eyes, short eye cracks, short eyelashes, wide base of an elongated nose with thickened nostrils; short filter, pronounced lip groove, slightly open mouth with thickened lower lip, small carious teeth, pointed chin, high-placed large, back-rotated auricles with hypoplasia of the antics. The marriage is unrelated. Parents are phenotypically healthy. The father has a healthy girl from first marriage. The previously unknown heterozygous mutation in the 3 exon of *KIAA2022* gene (chrX: 73963039A> AT, p.Asp451fs, NM_001008537 .2) was detected by exome sequencing. The mutation was not registered in "1000 genomes", ESP6500 and ExAC. The mutation was validated in the proband by Sanger sequencing. No same mutation was in parents. The X-chromosome inactivation skewing was not determined in the study of the *AR* gene (CAG repeat, 1 exon). This clinical observation supports study of de Lange et al 2016 that mutations in *KIAA2022* gene can cause epileptic encephalopathy and intellectual disability both in boys and in girls. This is important for determining the tactics of genetic testing, medical maintenance and genetic counseling.

529W

Earlier answers are better answers: Family-centered utility of ge-nome-scale sequencing for children with intellectual disability. *K.B. Brothers[1], C.A. Rich[1], K.M. East[2], W.V. Kelley[2], J. Huff[4], K.M. Bowling[2], M.D. Amaral[2], C.R. Finnila[2], S.M. Hiatt[2], M.L. Thompson[2], D. Gray[2], E.J. Lose[3], S. Simmons[3], J. Whittle[2], R.M. Myers[2], N. Lamb[2], G.S. Barsh[2], E.M. Bebin[3], G.M. Cooper[2].* 1) Department of Pediatrics, University of Louisville, Louisville, KY; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) University of Alabama at Birmingham, Birmingham, AL; 4) Norton Healthcare, Louisville, KY.

The HudsonAlpha CSER Study is a translational study of the utility of whole genome sequencing for identifying a genetic cause in children or adults with intellectual disability. To date we have enrolled 463 families. We conducted 59 parent interviews, including a total of 81 individuals, most of whom were biological mothers (69%) and biological fathers (28%). Semi-structured interviews were conducted 1 to 12 months following the return of results visit. Using a purposive sampling strategy, participants were selected to maximize diversity across four factors: race, ethnicity, whether a primary result for their child was returned, and whether a secondary result was returned to a parent. Thematic analysis was conducted using Atlas.ti software that started with a deductive theme framework which was expanded inductively based on themes encountered in interviews. Families with younger children tended to report that receiving a diagnosis for their child was very important. These families cited a range of factors that motivated their search for answers, including alleviating guilt, planning for future medical and personal care, and communicating with family, friends, teachers, therapists, and medical providers. The perceived importance of obtaining a diagnosis was comparatively decreased in families with older children, especially those with adult-aged children. Families with adult-age children tended to express acceptance about living without a diagnosis, and viewed diagnostic results primarily as informational and interesting, rather than pragmatically useful. Although many families across the age spectrum admitted to hope for a cure, parents of older children tended to report that they accepted their child as he/she is, and denied hope for a cure. Parent expectations tended to be in line with experienced utility. Families with older children identified relatively few concrete benefits from sequencing results, although families of all ages reported emotional benefits and the perception that information is valuable. These findings carry important implications for sequencing in children with rare disease, and indicate that parents may receive greater utility if sequencing is performed earlier in a child's life.

530T

Early diagnosis of Usher syndrome in non-syndromic young patients with hearing loss, the importance of pretest counseling. *M.F. van Dooren[1], H.H.W. de Gier[2], M.W. Wessels[1], H.T. Bruggenwirth[1], A.M.W. van de Ouweland[1], S.G. Kant[3], E.H. Hoefsloot[1].* 1) Department of Clinical Genetics, Erasmus MC, Room 2018, 3000 CA, Rotterdam, Zuid Holland, the Netherlands; 2) Department of Otorhinolaryngology, Erasmus MC Sophia, Room Sp-1455, Wytemaweg 80, 3015 CN Rotterdam, the Netherlands; 3) Department of Clinical genetics, Leiden University Medical Centre, Postzone K5-R, PO box 9600, 2300 RC Leiden, The Netherlands.

Since the introduction of Next Generation Sequencing panels for analysis of hearing loss, more genetic causes of hearing loss in children have been found at a young age. In our cohort of 120 patients 21 cases (20 %) were under age of five. In this group we found four cases with non-syndromic hearing loss in which the diagnosis Usher syndrome was made using exome sequencing with a filter for 102 genes associated with hearing loss. The first patient, an 11 months old girl with bilateral sensorineural hearing loss (60-80dB) was seen on our joint ENT/genetics consultation. The inclusion of syndromic forms of hearing loss with co-morbidity in this NGS test, including possible future visual impairment was discussed prior to genetic testing. A frameshift mutation c.920_923dup, p.His308fs and a pathogenic missense mutation c.1256G>T, p.Cys419Phe in the *USH2A* gene were found in compound heterozygosity, establishing the diagnosis Usher syndrome type 2. The second 5-years old male patient was diagnosed with congenital bilateral sensorineural hearing loss of 60dB. An unclassified variant c.12868C>T, p.Arg4290Cys and a pathogenic mutation c.17668_17669delAT, p.Met5890fs were found in the *ADGRV1*-gen in compound heterozygosity, indicating the most likely diagnosis of Usher syndrome type 2. The third 2 years old male patient was diagnosed with congenital bilateral hearing loss of 50-55-dB. NGS analysis revealed compound heterozygosity for the c.1227G>A, p.Trp409* and c.10525A>T, p.Lys3509* in the *USH2A*-gen, establishing the diagnosis of Usher syndrome type 2. The fourth patient was a one year old girl who failed neonatal screening and no mutations in GJB2/6-gene. Using NGS two pathogenic mutations in the USH2A gene, c.2299delG, p.Glu767fs and c.13010C>T,p.Thr4337Met were identified. In Usher syndrome loss of vision/retinitis pigmentosa, is expected not to occur before the beginning of puberty. All patients with Usher syndrome diagnosed in our cohort were presymptomatic for visual involvement. After explaining these results to the parents, the insecurity regarding progression and severity of visual loss in future was the most important issue for the parents. As knowledge on these future problems on a very young age could be a burden for parents, pre-test counseling is essential prior to NGS analysis. However, an early diagnosis of Usher syndrome is of benefit for both clinical management of the patient and for future reproductive choices of their parents.

531F

Cri du Chat syndrome: Characteristics of 69 Brazilian patients. *C.A. Kim[1], C.B. Mello[2], L.S.E. Pimenta[1], L.M. Benedetto[2], E.C.N. Vaca[1], R.B.F. Khoury[1], D.M. Befi-Lopes[3], R.S. Honjo[1].* 1) Genetics Unit, Instituto da Criança, São Paulo, SP, Brazil; 2) Departamento de Psicobiologia - EPM/UNIFESP, São Paulo, SP, Brazil; 3) Departamento de Fonoaudiologia – USP, São Paulo, SP, Brazil.

Cri du Chat syndrome (CdCS) is a genetic syndrome caused by deletions in the short arm of chromosome 5. Although the main clinical features are well known, neurocognitive and behavioral characteristics of the phenotype are rarely described in detail in the literature. In this study we analyzed the main phenotypic features of CdCS according to parental perspective. **Methods:** A questionnaire was sent to 700 Brazilian families registered in the Brazilian Association of CdCS. The questions involved specific domains such as pregnancy and birth conditions, recurrence of the disease in the family, current major health problems, and aspects of cognitive development. **Results:** 69 questionnaires were fulfilled, regarding 42 females and 27 males, with current ages ranging from 1-41 years old (mean: 14y). Most of the parents noticed the typical cat-like cry at birth (93%). The age at diagnosis ranged from birth to 14 years old (mean: 12y). In 65% of the cases, parents reported their karyotypes were done (parental) and 49% said they did not know the recurrence risk of CdCS. The main health problems reported were: swallowing (75%) and feeding problems (75%), congenital heart disease (32%), spine abnormalities (28%), and neurological symptoms (22%). The behavioral problems were: aggressive behavior, stereotypies, anxiety, phobias and genital manipulation/masturbation. Neurodevelopmental delay was reported in most cases (96%), even though some questionnaires did not inform the milestones. Independent walk was achieved in 60%. Approximately 40% never presented expressive language and most of the patients are dependent in their daily activities. **Conclusion:** It is a pioneer initiative in the CdCS support group and the answers in the questionnaire can improve the health care assistance to these patients because it focuses the attention to the demands in a parental perspective. Besides, nearly half of the families state that they did not remember information regarding recurrence risk, which reinforces the importance of genetic counseling follow-up and the need for the expansion of Genetic services in Brazil. (Support: CNPQ).

532W

Over FISHing: Findings in 1400 patients with 22q11.2 deletion syndrome. *D. McDonald-McGinn[1,2], T.B. Crowley[1], D.E. McGinn[1,3], E. Moss[1], L. Katz[4], O. Jackson[5], M. Mascarenhas[6], J.W. Gaynor[7], E. Goldmuntz[8], V. Deeney[9], S. Sanna-Cherchi[10], J. Moldenhauer[11], R. Gur[12], B. Morrow[13], B.S. Emanuel[1,2], K. Sullivan[14], E.H. Zackai[1,2].* 1) Dept Clinical Genetics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Davidson College, Davidson, NC; 4) .Dept Endocrinology, Children's Hosp Philadelphia, Philadelphia, PA; 5) Dept Plastic and Reconstructive Surgery, Children's Hosp Philadelphia, Philadelphia, PA; 6) Dept Gastroenterology, Children's Hosp Philadelphia, Philadelphia, PA; 7) Dept Cardiothoracic Surgery, Children's Hosp Philadelphia, Philadelphia, PA; 8) Dept Cardiology, Children's Hosp Philadelphia, Philadelphia, PA; 9) Dept Orthopaedic Surgery, Children's Hosp Philadelphia, Philadelphia, PA; 10) Dept Nephrology, Columbia University, New York, NY; 11) .Dept Surgery, Children's Hosp Philadelphia, Philadelphia, PA; 12) Dept Neuropsychiatry Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 13) Albert Einstein College of Medicine, NY, NY; 14) Dept Allergy and Immunology, Children's Hosp Philadelphia, Philadelphia, PA.

Background: It has been a quarter since FISH studies were developed to identify the 22q11.2 deletion syndrome (22q). Formerly considered rare 22q is the most common microdeletion (~1:2000 children; 1:1000 unselected fetuses) leading to significant morbidity from multi-organ system involvement including cognitive deficits and behavioral phenotypes. Lack of recognition of the condition/lack of familiarity with current testing methods (e.g. identifying nested deletions missed by FISH) combined with phenotypic variability delays diagnosis which could improve outcomes. Early diagnosis offers medical and emotional preparedness. Careful phenotyping provides a window into understanding the genomic basis of associated conditions as evidenced by the recent description of *CRKL* (a gene within the nested LCR22C-LCR22D region) as a genetic driver of kidney development. Here we report findings on 1400 patients with 22q followed at a single center, demonstrating the value of maintaining a well characterized cohort. **Methods:** Data was abstracted for ascertainment, demographics, deletion size, associated features, familial cases and mortality. **Results:** 65% diagnosed by Genetics; 44 ascertained prenatally and delivered and followed at CHOP; 51% female; 75% White, 9% Black; 58% <18; Mean age at diagnosis - 3.8 years. Deletion size (883/1305) A-D 86%, A-B 5%, A-C 2%, B-D 4%, C-D 1%, Other 2%. Major associated features: CHD (69%) 1/2 with severe defects, palatal anomalies (68%), GI problems (66%), hypocalcemia (55%), renal anomalies (16%). Other significant features: skeletal (vertebral anomalies, scoliosis, clubfoot); CNS (idiopathic seizures, behavioral diagnoses, NTDs, polymicrogyria); autoimmune disease; GU (hypospadias, absent uterus); ENT (microtia, laryngeal web, choanal atresia, TEF); Eye (scleracornea, ptosis); CDH; polyhydramnios; malignancies. Mean FSIQ (361/1305): 76.05 (SD 13.5). 7% familial, 88 probands and 121 relatives; 60% familial in B-D deletions; 1 somatic mosaic; 1 germline mosaic; 7 with 22q11.2DS and another disorder; 7 with autosomal recessive associated conditions (CEDNIK, Bernard Souilier, van den Edne- Gupta) unmasked by a mutation on the other allele. 4% died most often related to CHD by 5 months of age. No association was found between CHD, hypocalcemia and FSIQ. **Conclusion:** Increasing the denominator in our aging cohort better defined or confirmed previously reported figures and contributed to the observation of additional/rare associations.

533T

Educational approach for TSC families at LeBonheur Children's Hospital Tuberous Sclerosis Center of Excellence. *N. Urraca[1,2], R. Mostafavi[3], T. Jones[1,4], E. Pivnick[3], J. Wheless[5], J.J Bissler[6].* 1) CFRI, Le Bonheur Children's Hospital, Memphis, TN; 2) Le Bonheur Children's Hospital Tuberous Sclerosis Center of Excellence; 3) Pediatric Genetics Division, UTHSC; 4) Pediatrics and Preventive Medicine Departments, UTHSC; 5) Pediatric Neurology Division, UTHSC; 6) Pediatric Nephrology Division, UTHSC.

   Background: Tuberous sclerosis complex (TSC) is an autosomal dominant disorder characterized by hamartomas in several organs. Le Bonheur Children's Hospital Tuberous Sclerosis Center of Excellence (LBCH-TSCE) serves patients from all over the world. Most patients are diagnosed at outside hospitals and referred to LBCH-TSCE for specific questions. The goals of this project were to assess the patient and families level of knowledge about genetics of TSC and to test the effectiveness of our educational approach. **Methods**: Subjects with TSC or their relatives seen at LBCH-TSCE were invited to participate. Each participant answered a 10-item questionnaire (pre-test) assessing knowledge level, completed an educational session, and received a TSC informational brochure. Six months after the same questionnaire (post-test) was administered by phone. Wilcoxon Signed Rank Test was used to test paired differences. **Results:** 23 TSC affected subjects and 44 relatives opted to participate. Almost half of participants (47.8%) reported they received genetic counseling previously. The median score for the pre-test was 6 with no significant (p=0.135) difference between affected individuals and relatives. Median difference between the pre- and post-test was significant (p=0.001; median 1; range -2 to 9). The difference was also significant by group: affected (p=0.034) and relatives (p=0.009). **Conclusions:** Providing an educational session and a TSC brochure improved participants' knowledge of the genetics of TSC. We suggest systemic implementation of educational sessions in TSC clinics. .

534F

Genetic evaluation and counselling of couples who lost children affected by rare disorders. *C. Skrypnyk[1,2].* 1) Molecular Genetics, Al Jawhara Center for Molecular Medicine, Arabian Gulf University, College of Medicine, Manama, Bahrain; 2) University Medical Clinics, King Abdulla Medical City.

   Background and aims: Some lethal rare genetic disorders may lead to the death before the clinical diagnosis was clarified and confirmed by genetic testing. The aim of this study is to present the steps and the difficulties in managing, from the genetic approach, couples who lost children affected by rare genetic disorders. **Methods:** Six couples who lost children affected by a rare genetic disorder, seen by the author at the genetic evaluation and counselling sessions, were retrospectively analyzed, revising the reason for referral, family history of genetic disorders, medical reports available and existence of biological samples from the affected children, the genetic testing plan and its results. **Results:** All couples looked to understand the recurrence risk and to find the method to avoid having another affected child. Medical reports were partially available, a clinical diagnosis assessment was made for their children but genetic testing was not performed or the result did not confirmed it. Biological samples from affected children were available only in two cases and genetic testing clarified the diagnosis for one case. Parental whole exome sequencing (WES) testing was performed for the other 4 couples and the pathogenic mutations were identified allowing a genotype-phenotype correlation. A couple was identified as carier of two rare genetic disorders. **Conclusions**: A correct genetic evaluation of the parents and adequate genetic counselling are mandatory for choosing the right genetic testing with the highest chance to certify or clarify, postmortem, the clinical diagnosis. In the absence of biological samples from the affected children, parental WES testing is the best approach to identify the responsible genetic defect and allow adequate management. Patience, empathy and support are highly required to help them to understand complex genetic situation, make their informed decision, give their consent for testing and future investigation and handle the results, in the delicate context of the parental grief for their lost children.

535W

The burden and economic impact of pediatric rare and undiagnosed genetic disease. *N. Gonzaludo[1], D. Bentley[2], J.W. Belmont[1], R.J. Taft[1].* 1) Population and Medical Genomics, Illumina Inc., San Diego, CA, USA; 2) Population and Medical Genomics, Illumina Cambridge Ltd., Chesterford Research Park, Saffron Walden, Essex, UK.

Rare and undiagnosed genetic diseases (RUGD) affect approximately 1 in 10 people in the United States. Patients with RUGD are high utilizers of the healthcare system and often undergo years of testing before arriving at a definitive diagnosis. As the majority of rare diseases have a genetic basis and 50–75% of are thought to begin in childhood, early genetic testing in suspected individuals may represent a large national cost savings. The aim of this study was to identify the population of suspected pediatric RUGD patients most likely to benefit from genetic testing, as well as quantify their healthcare utilization and estimate their national economic burden. Using the 2012 Kids' Inpatient Database (KID), we identified discharges for pediatric patients less than 18 years old with diagnoses suggestive of RUGD, based on a curated set of ICD-9-CM codes. To capture variability in diagnoses and clinical presentations, we created minimum and maximum estimates of RUGD discharges. To establish the minimum, we selected discharges with a principal diagnosis in the set of ICD-9-CM codes that were clinically abnormal given the age at discharge. For the maximum, we considered RUGD discharges for which any diagnosis (up to 15 diagnoses per discharge) was in the set of ICD-9-CM codes. For these cohorts, we analyzed demographics, clinical characteristics, and measures of healthcare utilization, as represented by inpatient total charges and length of stay.   We estimate that of the 5.85M weighted total pediatric inpatient discharges in the 2012 KID, 150,169 to 818,384, or 2.6–14% of discharges, represented suspected RUGD patients. After propensity score adjustment, total charges per inpatient discharge for neonatal RUGD patients were $59K to $243K higher ($p < 0.001$) than those of non-RUGD, with an additional 6.6–18.6 days spent in inpatient care. Total charges per discharge for non-neonatal RUGD patients were $38K to $52K higher ($p < 0.001$) than those of non-RUGD, with an additional length of stay of 1.8 to 2.4 days. Overall, the aggregate total charges for RUGD discharges accounted for $14B to $57B (11–46%) of the "national bill" for pediatric patients in 2012.   Our analysis shows that pediatric RUGD patients are a significantly higher burden to the national healthcare system. While further studies are necessary, early genetic testing in this population may serve to reduce costs associated with an otherwise lengthy diagnostic odyssey.

536T

Lessons learned about expanded carrier screening using genome sequencing: Implications for research and practice. *K.A.B. Goddard[1], T.L. Kauffman[1], C.K. McMullen[1], J.A. Reiss[2], C.S. Richards[3], G.P. Jarvik[4,5], B.S. Wilfond[6,7] on behalf of the NextGen project team.* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Department of Medical Genetics, Kaiser Permanente Northwest, Portland, OR; 3) Department of Molecular and Medical Genetics, Knight Diagnostic Laboratories, Oregon Health & Science University, Portland, OR; 4) Genome Sciences, Division of Medical Genetics, University of Washington, Seattle, WA; 5) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 6) Department of Pediatrics, Division of Bioethics, University of Washington, Seattle, WA; 7) Truman Katz Center for Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA.

Background: This comprehensive study of preconception expanded autosomal recessive or X-linked carrier screening allowed us to assess clinical integration from multiple viewpoints. **Methods:** We performed genome sequencing and reported 728 genes/conditions for carrier status, and 148 actionable secondary findings. Women were eligible if their clinician ordered any preconception carrier screening, were not pregnant at the time of consent, and were planning a future pregnancy. **Results:** We evaluated implementation processes including reasons for declining participation, participant choice in the results to receive, and genetic counseling time to disclose results. The primary reasons for declining were lack of time or interest rather than opposition to the rationale for testing. Participants valued choice about which results to receive as a sign of respect, although most participants (95%) chose to receive everything that was offered. The results disclosure process was often lengthy, particularly visit preparation time, with a median of 64 minutes in total. Laboratory outcomes included a framework for deciding on genes/conditions to offer, an analytic pipeline for sequence interpretation, and the yield for carrier and secondary findings. Overall, 78% of participants (132 women and 54 male partners) received ≥1 carrier result (mean: 1.5/person). Three women were carriers for X-linked conditions, and all were pregnant with a male fetus at the time of results disclosure. In twelve couples, both partners were carriers for the same autosomal recessive condition. Overall, 3.4% of participants had a secondary finding (2.9% for the ACMG SF v2.0 list). Patient or health system outcomes included economic outcomes and other impacts of learning carrier status. Most participants were willing to pay substantially more than a typical copay, although this was related to income. There was no significant difference in downstream utilization between those who received genome sequencing and a randomized control group, either overall or among subcategories of utilization. There was little evidence of harm in those with a negative result (<25% risk of having an affected child), but a small subset declined recommended care (e.g., prenatal screening in a subsequent pregnancy), perhaps due to a misunderstanding that it was not needed for them. **Conclusion:** We will discuss the implications of these findings for future clinical approaches and for future research on expanded carrier screening.

537F

Data sharing to advance understanding of genetic disease: A model for data sharing agreements from the Newborn Screening Translational Research Network. *A. Brower, J. Adelberg, M. Watson.* American College of Medical Genetics and Genomics, Bethesda, MD.

Exciting advances in newborn screening (NBS) and diagnostic technologies, including the use of genomics, enable identification of affected newborns prior to the onset of clinical symptoms for an ever increasing array of genetic disease. In concert with these advances, information technology tools to collect prospective clinical information to inform treatment, determine health outcomes, and understand the natural history of disease are in use by state-based NBS programs and treating clinicians. This unparalleled opportunity to diagnose a large number of genetic diseases in the newborn period, and begin to simultaneously capture longitudinal clincial data represents an incredible opporutnity to accelerate the translation of findings to routine clinical care, while advancing knowledge. Realization of this vision requires the shairng of a wide variety of data across a large number of diverse institutions. To capitalize on the availability of data collected across fifty-one NBS programs and hundreds of clinical sites, the Newborn Screening Translational Research Network (NBSTRN) has designed and implemented data sharing agreements (DSA) that facilitate uploading of data to a central database. The data sharing agreement addresses data privacy and security concerns, requires responsible data stewardship for both data contributors and data users, defines data types, and outlines the process for data access and secondary use. The NBSTRN has successfully executed DSAs with a state-based NBS program, several research teams and institutions. The DSAs enable the data owner to upload phenotypic and genomic data to a central database, review and manage requests for secondary use of the data, analyze and aggregate data across data owners and projects, and disseminate findings to the research community. Currently, the number of data points available for sharing is over one million across several thousand cases affected with one of forty-six genetic conditions. We will review the DSA definitions, workflow, and present two case studies. Our successful approach to DSAs serves as a model for the sharing of clinical and genomic data across the lifespan to advance disease understanding and improve patient outcomes.

538W

Cases of patients and informal caregivers who decided to handle their own health condition. *V.P.F. Francisco[1], S.A. Oliveira[1], P. Oliveira[2].* 1) Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa; 2) Católica Lisbon School of Business and Economics.

Our research in the area of "user innovation" has shown that patients, and their informal or non-professional caregivers, often develop innovative solutions to help them cope with their health condition. That same stream of research also concluded that these innovative solutions often improve the quality of life of the "patient innovators" (or their caregivers) and, if shared with others living under similar conditions, can significantly impact the lives of others. Since we increasingly live and work on online collaborative environments where the user can share, search and access information in a much faster, cost effective and efficient way, we have established an award-winning platform, www.patient-innovation.com aimed at empowering patients/caregivers under medical supervision and the support of two Nobel Laureates and other reputable individuals. (e.g. Patient Innovation was named Non-profit Startup of the Year and recognized by UN Secretary General Ban Ki-moon). Over 700 innovations were submitted and "screened" from a community of over 50.000 patients/caregivers from the 5 continents. In this poster we present some cases of patients and caregivers who developed innovative new solutions to help them cope with their diseases and critically discuss their similarities and differences.

539T

Demystifying biomedical big data analysis through a massive open online course (MOOC). *B.R. Haddad[1], PB. McGarvey[2,3], A. Kruse[4], Y. Gusev[1,2].* 1) Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC; 2) Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington, DC; 3) Protein Information Resource, Georgetown University Medical Center, Washington, DC; 4) Center for New Designs in Learning & Scholarship, Georgetown University, Washington, DC.

   Advances in genomics have led to the generation of massive amounts of data. However, the usefulness of these data to the basic scientist or to the clinical researcher, to the physician or ultimately to the patient, is highly dependent on understanding its complexity and extracting relevant information about specific questions. The challenge is to facilitate the comprehension and analysis of big datasets and make them more "user friendly". Towards this end we have developed a cross-disciplinary, Massive Open Online Course (MOOC), that aims to facilitate the understanding, analysis, and interpretation of biomedical big data for basic and clinical scientists, researchers, and librarians, with limited or no significant experience in bioinformatics. The 8-week course was funded by an NIH- BD2k R25 grant and was released in February 2017 on the edX platform. The course covered biomedical big data as it relates to five main areas: 1) Genomics; 2) Transcriptomics; 3) Proteomics; 4) Systems Medicine; and 5) Big Data usage in research and the clinic. Content consisted of short video lectures, interviews, and online hands-on-training on the use of various databases and analysis tools for different "omics" platforms. Students were provided with online readings and resources, and were expected to complete pre- and post-course surveys, formative assessments, and a final exam. Participation in discussion boards and other exercises was also encouraged. Over 2,000 students from over 150 countries have enrolled. We will share our experience in developing the course and discuss the pros and cons of this online approach of hands-on training in big data analysis. Quantitative data will be presented about students' demographics, motivation, and learning achievements. Themes from the discussion board and other exercises will be discussed. We plan to maintain this course as a "living resource," updating it regularly, and keeping it freely accessible. We believe that this will allow us to provide an educational opportunity to a large audience worldwide, particularly to individuals with limited access to traditional educational resources in this cutting-edge field.

540F

Section of Genetic Counseling at CHOP: Organizing for advocacy and professional development. *M.H. Harr[1], S. Woyciechowski[2], H. Dubbs[3], S. Kasperski[4], S. Biswas[5], A. Wilkins[5], A. Melchiorre[6], D.M. McDonald-McGinn[6], The Section of Genetic Counseling at CHOP.* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Center for Fetal Diagnosis and Treatment, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pathology, Children's Hospital of Philadelphia, Philadelphia, PA; 6) Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA.

   The Section of Genetic Counseling at Children's Hospital of Philadelphia (CHOP) was created in 2014 to provide a unified professional voice for all genetic counselors across the institution, aid in professional development and promote collaboration amongst genetic counselors. The purpose of this presentation is to provide insight into the process of formalizing the Section of Genetic Counseling at CHOP, outline our internal leadership and committee structure, and highlight accomplishments from the initial three years. Key deliverables of the Section included an institute-wide salary review that resulted in salary adjustments for new hires and employed genetic counselors, development of an enhanced career ladder for professional advancement, creation of a centralized system for student placement for observations and clinical rotations, institution of incident-to billing for genetic counselors, organization of educational and promotional activities for genetic counselors, and provision of onsite CEU opportunities. The methods and structural organization employed to achieve these deliverables will be reviewed.   The Section of Genetic Counseling at CHOP currently has 44 members practicing in a wide array of disciplines including pediatric genetics, metabolic medicine, prenatal diagnosis, hereditary cancer, cardiology, neurology, internal medicine, research, and pathology. The strategies utilized by the Section have not only resulted in increased camaraderie and collaboration amongst genetic counselors, but also in increased visibility as well as recognition of the profession at the institutional level across different diciplines. Here we will offer strategies for genetic counselors, and other professionals, to identify opportunities for advocacy and professional growth, recognize potential challenges to such formal organization, and present examples of forthcoming projects within the Section which aim to advance professional development of genetic counselors across the country.

541W

Genome Gateway: An online platform to increase communication between patients, providers and researchers. *A. Hott[1], K. East[2], W. Kelley[2], M. Cochran[2], D. Bick[3], B. Korf[4]. 1) Educational Outreach, HudsonAlpha, Huntsville, AL; 2) Genomic Counseling, HudsonAlpha, Huntsville, AL; 3) Genomic Medicine, HudsonAlpha, Huntsville, AL; 4) Department of Genetics, University of Alabama-Birmingham School of Medicine, Birmingham, AL.*

Genome Gateway is an online application developed to facilitate communication, education and the sharing of information between patients, the healthcare team, and clinical research teams. The purpose of the development of such a tool is three-fold: 1- to allow face-to-face patient/provider interaction to be as individualized as possible, 2- to provide a foundation for patients to understand genomic medicine evaluations and results, and 3- to create a well-organized, thorough set of educational materials that can "travel" with Genome Gateway and be utilized and customized. The development of Genome Gateway over the past 18 months has provided a clinical tool that alleviates the burden of long family history intakes, patient knowledge gaps, and pre/post-clinical visit communication. Customizable learning modules that include both written text and media, as well as, fully customizable questionnaires are delivered to individual participants tailored specifically for the participant's needs. Additionally, secure messaging to and from participants within the application is built in, as is, the ability to create a comprehensive family history pedigree. Informed consent and other on-boarding documents are also customizable and delivered through the platform. Over 150 patients utilize Genome Gateway currently through its implementation at the Smith Family Clinic for Genomic Medicine. Evaluation of patient satisfaction with the platform, overwhelmingly rate Genome Gateway as a useful and easy to use tool that aids clinical care. Many patients report the learning material and family history as the most useful aspects of the software. Over half of diagnostic genome patients self-reported reading all assigned learning topics. Providers indicate an average time savings during the first clinical visit to be 30-45 minutes per patient compared to time spent with patients that had chosen not to use Genome Gateway. The flexibility and ease of use inherent in this software makes Genome Gateway an impactful tool in clinical genomics research studies for patients and providers. Genome Gateway will be used in the newly launched Alabama Genomic Health Initiative (AGHI) to positively impact patient satisfaction, increase access to critical learning topics, and make clinical visits more individualized. The use of Genome Gateway in a population-based genomic health study, will provide additional insight into end-user decisions on use, completion, and clinical care time savings.

542T

Lumping and splitting: An age old dilemma with new age implications for disease classification. *C. Thaxton[1], J. Goldstein[1], K. Wallace[1], M. DiStefano[2], R. Ghosh[4], P. Witmer[5], E. Rooney Riggs[6], M. Haendal[7], A. Hamosh[5], H. Rehm[2,3], J. Berg[1], ClinGen. 1) Department of Genetics, The University of North Carolina, Chapel Hill, NC, USA; 2) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA; 3) Department of Pathology, Brigham & Women's Hospital, Harvard Medical School, Boston, MA USA; 4) Baylor College of Medicine, Houston, TX, USA; 5) McKusick-Nathans Institite of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA; 6) Autism and Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA; 7) Monarch Initiative, OHSU, Portland, Oregon, USA.*

The dilemma of categorizing and classifying scientific entities has been debated for centuries, and throughout this time several biological entities have cycled through lumping and splitting as scientific knowledge and technologies have advanced. In medical genetics, nosology — or the classification of disease — has historically approached classifying disorders based on sets of phenotypes, either lumping multiple phenotypes into a broader disease entity (i.e. syndrome) or splitting out isolated phenotypes. However, with progressive advancements in medical genetics and genomics, nosology has been turned on end. With NGS and WES sequencing technologies, we are now able to identify the the underlying genetic etiologies of diseases that may have previously been "lumped" based on phenotypic presentation; these etiologies are often distinct, and systematic reclassification and re-categorization of disease entities may be necessary. The Clinical Genome Resource (ClinGen) is a resource that defines the clinical relevance of genes and variants for use in precision medicine, and has developed a system to classify the strength of a gene:disease relationship through biocuration. One challenge faced by biocurators is how to curate genes associated with several phenotypic entities: when to lump or when to split a gene:disease entity given the evidence present in the literature? To assist biocurators and establish consistency within the consortium, ClinGen has formed a task group to assemble criteria to address the lumping and splitting predicament. Our preliminary criteria include: (1) defining disease entities; (2) establishing the molecular mechanism(s) underlying the associated entities; (3) discerning the phenotype and variable expressivity; (4) and assessing clinical management. We also formulated a pre-curation process that includes assessment of disease entities and binning of genes based on one of three possible phenotypic etiologies: isolated phenotype, complex phenotype, or syndromic phenotype. This pre-curation will ultimately facilitate and streamline lumping and splitting based on our criteria. Here, we will outline the pre-curation process, our guidelines, as well as specific examples to illustrate the process of determining when to "lump" and when to "split." Finally, we describe the implications for clinical diagnostic tools, informatics approaches, and care management to this more nuanced classification of disease entities.

543F

Diversity and inclusion in genomic research: Why the uneven progress?
S. Callier[1,2], A. Bentley[1], C. Rotimi[1]. 1) Center for Research on Genomics and Global Health, NHGRI, NIH, Washington, DC , US; 2) George Washington University School of Medicine and Health Sciences, Washington, DC, US.

   Conducting genomic research in diverse populations has led to numerous advances in our understanding of human history, biology, and health disparities, in addition to discoveries of vital clinical significance. Conducting genomic research in diverse populations is also important in ensuring that the genomics revolution does not exacerbate health disparities by facilitating discoveries that will disproportionately benefit well-represented populations. Despite the general agreement on the need for genomic research in diverse populations in terms of equity and scientific progress, genomic research remains largely focused on populations of European descent. In this article, we describe the rationale for conducting genomic research in diverse populations by reviewing examples of advances facilitated by their inclusion. We also explore some of the factors that perpetuate the disproportionate attention on well-represented populations. Finally, we discuss ongoing efforts to ameliorate this continuing bias. Collaborative and intensive efforts at all levels of research, from the funding of studies to the publication of their findings, will be necessary to ensure that genomic research does not conserve historical inequalities or curtail the contribution that genomics could make to the health of *all* humanity.

544W

Genetic testing in the criminal justice system: Human rights perspectives. *A. de Paor.* Law, Dublin City University , Dublin , Ireland.

   Genetic science and technology is advancing at a significant pace with discoveries as to the genetic basis of disease and health conditions emerging. Scientific advances are also revealing genetic details of behaviour and personality traits, including anti-social behaviour and violence. It is anticipated that behavioural genetics will develop further and will increasingly be used in various (medical and non- medical) settings, including in the criminal justice context. While much attention has been dedicated to the regulation of genetic information in commercial contexts such as employment and insurance, less focus has been dedicated to regulating genetics in the criminal justice context. This paper urges caution against the use of behavioural genetics in criminal proceedings, and argues that safeguards are needed to control the use and application of this emerging knowledge and any new technologies, and ensure appropriate awareness and education. There is a strong argument for regulating the use of genetic testing for predisposition to violence and criminality given the serious consequences for defendants and offenders. The UN Convention on the Rights of Persons with Disabilities has been used to bolster arguments for prohibiting genetic testing of persons in insurance and employment. This paper considers the potential of this Convention to inform the debate and provide a platform for discussion of the issues arising, from a disability and human rights perspective. The disability and human rights perspective is important in this debate as the absence of an appropriate regulatory framework means that it is likely that more individuals will be stigmatised and mistreated on the grounds of their genetic information in the criminal justice setting.

545T

Fine-scale demography and behavior of male and female human geneticists. *E. Glassberg[1], N. Telis[2], C. Gunter[3,4].* 1) Department of Biology, Stanford University, Stanford, CA; 2) Department of Biomedical Informatics, Stanford University, Stanford, CA; 3) Departments of Pediatrics and Human Genetics, Emory University School of Medicine, Atlanta, GA; 4) Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, GA.

   Improving the diversity of researchers in human genetics is likely to improve the diversity of problems and populations studied, subsequently decreasing health disparities of under-represented groups. One group that is under-represented in STEM is women; therefore, we wanted to quantify differences between male and female researchers in the American Society of Human Genetics (ASHG).   Over the past three Annual Meetings, we collected data regarding the question-asking behavior of men and women following platform presentations. We attended live sessions and, following the meeting, used videos on the web portal to gather data about the questions posed at talks in an array of fields. For each talk, we recorded the session, the speaker, and the gender of each attendee who asked a question of the speaker.   We found that, in platform presentations, female ASHG attendees tend to ask fewer questions than their male counterparts. However, analysis of abstract submissions demonstrated that the representation of women varies widely across sub-fields within ASHG (see abstract by Telis et al). In particular, we find that women are under-represented in statistical genetics and bioinformatics and over-represented in genetic counseling, genetics/genomics education, and ELSI.   One might hypothesize that the relatively small number of questions from women reflects underrepresentation of women in the audiences of the talks in our dataset. We therefore used the gender ratio of poster presenters in related sub-fields to estimate the proportion of women attending each talk. We show that women are under-represented in question-askers relative to their representation in each sub-field. This is true in data collected live (bioinformatics, statistical genetics) and via webcast (cancer genetics and clinical genetics). Of note, the poster-presenter data suggest that cancer and clinical genetics are female-biased sub-fields. Together, these findings suggest that the relative lack of questions from women is robust to the overall representation of women within the sub-field of the talk.   Finally, we show that women are more likely to ask questions of female speakers than they are of male speakers, though they remain less likely to ask questions of female speakers than their male counterparts.   These observations from within our own professional society suggest that increasing the representation of women may be insufficient to erase gendered behavioral differences in STEM.

546F

Re-writing regulation? A comparative ethics and policy study of "natural" vs. "synthetic" cells, organoids and human genomes. *R. Isasi, S. Zuchner.* Dpt. Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

   The Genome Project-write aims, among other goals, to tackle current health challenges by enabling human genome-scale synthesis and editing through the development of tools, methods and foundational technologies. GP-write activities will integrate human genetics, stem cells, engineering, among other areas that had been, and continue to be, subject to particular public scrutiny. GP-write seeks to be disruptive, defying existing conceptual and technical limits. Similarly, it will significantly confront the boundaries of governing national 'policy frameworks' or 'legal systems' (i.e. norms, principles, institutions, etc.). Influenced by historical contexts, these frameworks and systems reflect ethical or moral values as well as societal priorities. Thus, an important step to support initiatives such as GP-write is to determine if the policy *status quo*, as developed and implemented, is fit-for-purpose. As GP-write is also interdisciplinary and global, the latter requires identifying and understanding differences and similarities in national approaches, as well as the ethical principles and societal values underpinning them. Consequently, central questions to be addressed are: How these policy frameworks vary across scientific disciplines and their applications? Which legal (e.g. laws, regulations) and self-regulatory measures (e.g. professional guidelines, best practices, codes of conduct, etc.) are necessary to promote responsible innovation? This presentation will identify and analyze policy and ethical approaches to the regulation of 'natural' versus engineered/synthetic human cells and organoids. By mapping the policy/ethical landscapes, we will explore their strengths and challenges, as well as their ability to adapt to scientific advances, evolving technological uptake as well as social interests. It will further assist in determining whether the issues arising in projects such as GP-write fall within, or outside, the remit of existing policy dealing with interconnected scientific fields. Our work will help elucidate the consistency of the ethical principles, social values and scientific rationale underlying the policy choices.

547W

Actions and reactions to negative results from genome sequencing in a healthy preconception population. *T. Kauffman[1], S. Kraft[3,4], J. Dickerson[1], J. Schneider[1], A. Rosales[1], E. Shuster[1], J. Davis[1], C. McMullen[1], F. Lynch[1], B. Wilfond[2,3], K.A.B. Goddard[1] On behalf of the NextGen study team.* 1) Center for Health Research, Kaiser Permanente, Portland, OR; 2) Department of Pediatrics, Division of Bioethics, University of Washington, Seattle, WA; 3) Truman Katz Center for Pediatric Bioethics, Seattle Children's Hospital, Seattle, WA; 4) Division of Bioethics, Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington.

   Background: As expanded carrier screening becomes increasingly widespread, it is important to understand how its clinical implementation affects the patients receiving this testing, especially for the vast majority with a negative result who are at <25% risk. Methods: 180 women and 71 male partners in the NextGen study received information on carrier status for 728 gene/condition pairs and 114 medically actionable secondary findings. Participants completed surveys before and after results were disclosed and a subset were interviewed 12 months after result disclosure. Surveys included questions to measure the psychosocial effect of receiving results. Interviews further explored utility and actions taken following result disclosure. Electronic medical record data was reviewed to record genetic testing (or refusal) during pregnancy both prior to and after result disclosure. Participants with a negative result have < 25% risk of having a child that could be affected by one of the tested conditions. This included individuals or couples for whom (1) the woman was not a carrier of any autosomal recessive or X-linked conditions, (2) both partners were tested but were not carriers of the same autosomal recessive condition, and (3) no individual who was tested had an autosomal dominant secondary finding. Results: For nearly all survey respondents, a negative result with < 25% risk had no or minimal negative emotional impact. Most reported feeling glad, or at least neutral, that they had decided to get genome sequencing, not worried about the accuracy of the test results, and unconcerned about potential implications of the results for their families or their own privacy. Interview data revealed similar findings. Utilization of primary care and mental health services did not differ between participants with a negative result at < 25% risk and usual care (no additional testing) (26% vs 29%, p=.72). A small number of women declined recommended prenatal testing, at least one of whom declined it due to a misunderstanding that it was not needed. In the interviews, many reported feeling more confident in their family plans and relieved about a lower risk of having a child with a genetic condition. Discussion: There is little evidence of increased harms from genome sequencing in a healthy population for people with a negative result at < 25% risk. A small number of women may decline recommended prenatal care due to a misunderstanding that it is not needed for them.

548T

All, some or none: How wanting to learn genomic results differs within a national survey sample. *J.B. McCormick[1], Y. Zhou[2], K.D. Graves[2], P.S. Sinicrope[3], S.T. Vadaparampil[4], N.M. Lindor[5].* 1) Penn State University College Medicine, Hershey, PA; 2) Georgetown University, Washington, DC; 3) Mayo Clinic, Rochester, MN; 4) Moffitt Cancer Center, Tampa, FL; 5) Mayo Clinic, Scottsdale, AZ.

   Introduction. The volume and complexity of information derived from clinical genome sequencing presents major challenges to established genetic counseling approaches. Current efforts to inform patient choices for returning results have focused on genetics experts' efforts to bin the data into categories of actionability and severity. An alternative approach could be to determine if subgroups of patients could be "binned" according to their individual preferences for return of genomic results. Methods: We conducted a cross-sectional online survey with a representative sample of 895 US adults to assess perceptions and attitudes toward genomic testing for disease actionability and severity. After reading and ranking 18 hypothetical disease scenarios for disease actionability and severity, respondents indicated whether they would want to learn the results. We explored correlations with hypothetical interest in learning test results and grouped respondents according to whether they wanted to learn results for ALL, SOME or NONE of the scenarios. Results: The majority (70.8%) wanted all results returned and 25.7% wanted SOME results; only 3.5% wanted NONE. The profile across the ALL/SOME/NONE groups did not vary significantly by age, gender, marital status, education, having children, self-rated genetic knowledge, perceived risk for heart disease, or self-reported health status. The NONE group was more likely to be non-Hispanic White, inexperienced with genetic testing, report an average risk of cancer and have greater confidence in their decisions about genomics. Individuals in the NONE group also placed less value on healthy lifestyle and the potential importance of genetic results for self and family members. The NONE group also did not score severity or actionability as important in decision making as the SOME and ALL groups. Conclusion: Identifiable characteristics are evident across individuals who project wanting ALL, SOME, or NONE of hypothetically available genomic test results per this survey of a diverse US population. Results could be used to better understand the salience of genomic testing results for the lay population and help genetics care providers organize results in a way that is meaningful for their patients.

549F

Usability of family health history tools among underserved patients.
C. Wang[1], T. Bickmore[2], M. Paasche-Orlow[1], T. Norkunas Cunningham[1], H. Cabral[1], M. Winter[1], M. Campion[3], D. Bowen[4]. 1) Boston University, Boston, MA; 2) Northeastern University, Boston, MA; 3) Stanford University, Stanford, CA; 4) University of Washington, Seattle, WA.

Efforts to improve the accessibility of electronic, computer-based family health history tools necessitate the monitoring of navigational factors that may impede their use. As part of an ongoing trial examining the validity and accuracy of computer-based family health history tools, our team sought to identify specific navigational challenges to effectively using one of two tools: VICKY: Virtual Counselor (Wang et al., 2015) and My Family Health Portrait (MFHP). Participants in the study are recruited from a single, safety net hospital in the Boston, MA area. Study participants are primarily between the ages of 45-64 (83%), African American (74%), and have less than a college degree (72%). Navigation issues were tracked by study research assistants. Navigation data is available for the initial 58 patients enrolled in the trial (29 per arm). For these initial patients, approximately 7% (2/29) did not complete VICKY and 52% (15/29) did not complete MFHP, to the point where a pedigree could be generated. In the case of VICKY, one instance was due to a technical malfunction and another instance was because the participant chose not to continue, due to fatigue from using the tool. In the case of MFHP, 7/15 (45%) patients chose not to continue because they were frustrated with the tool; 10/15 (66%) were reported to have navigation issues. Navigation issues (not mutually exclusive) included inability to enter the tool from the home screen, enter date of birth for self, and adding and saving conditions for self and relatives. Ongoing efforts are needed to monitor the usability of electronic family health history tools to ensure quality data capture by these tools, particularly among underserved patient populations who may have more limited computer literacy skills.

550W

Why patients decline genomic sequencing studies: Experiences from the CSER consortium.
L.M. Amendola[1], J.O. Robinson[2], R. Hart[1], S. Biswas[3,4], K. Lee[2], B.A. Bernhardt[4], K. East[5], M.J. Gilmore[6], T.L. Kauffman[7], K.L. Lewis[8], M. Roche[9], S. Scollon[10], J. Wynn[11], C. Blout[12]. 1) Medical Genetics, University of Washington, Seattle, WA; 2) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030; 3) Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; 4) Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA 19104; 5) HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806; 6) Department of Medical Genetics, Kaiser Permanente Northwest, Portland OR 97227; 7) Center for Health Research, Kaiser Permanente Northwest, Portland, OR 97227; 8) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, Bethesda, MD 20892; 9) Departments of Pediatrics and Genetics, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599; 10) Baylor College of Medicine, Department of Pediatrics, Houston, TX 77030; 11) Department of Pediatrics, Columbia University, New York, NY 10032; 12) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115.

Background: Clinical and research settings are increasingly incorporating genomic sequencing (GS) technology. A better understanding of how often and why potential participants decline participation in GS research, and if there are reasons specific to GS, is essential to promoting informed decision making and identifying potential barriers to participation. The heterogeneity of populations and study designs in the National Institutes of Health funded Clinical Sequencing Exploratory Research (CSER) consortium provide a unique opportunity to explore this question in a diverse set of GS studies. **Methods:** We surveyed representatives from seven sites across the CSER consortium ascertaining the project's rates of decline and reasons potential participants declined to participate in GS research. Data were analyzed using descriptive statistics. **Results:** The decline rate for enrollment at the seven CSER sites ranged from 12-64% (median 28%), with the highest rate of decline reported by the project recruiting healthy participants for GS for reproductive carrier screening. Decline rates varied slightly across potential participants of differing ages and phenotypes. Reasons for declining GS research were reported for 1088 participants. Commonly cited reasons were similar to those reported for clinical single gene testing and non GS genetic research. Reasons included study logistics (31%), disinterest in research (16%), and privacy and discrimination concerns (13%). The potential psychological burden of pursuing and receiving results from GS has been cited as a concern in the literature; however, such impact was not cited frequently (4%) as a reason for decline. Not wanting to receive additional findings, a concern specific to GS, was also cited by a minority of potential participants (2%). **Conclusion:** Rates of decline at CSER GS study sites varied; further research is necessary to explore the impact, if any, of different participant groups and/or study protocols. Not wanting additional findings and psychological concerns related to GS did not appear to be a barrier to enrollment. Future consortia exploring GS implementation should consider using standardized collection methods to examine reasons for decline in larger populations and more diverse healthcare settings.

551T

Prostate cancer risk follow-up among *BRCA1/2* mutation carriers in Finland. *O. Kajula[1,2,3], O. Kuismin[2,3,4], H. Kyngäs[1,2,5].* 1) Research Unit of Nursing Science and Health Management, University of Oulu, Oulu, Finland; 2) Medical Research Center, Oulu University Hospital and University of Oulu, Oulu, Finland; 3) Department of Clinical Genetics, Oulu University Hospital, Oulu, Finland; 4) PEDEGO Research Unit, University of Oulu, Oulu, Finland; 5) Northern Ostrobothnia Hospital District, Finland.

Background: Mutations of *BRCA1/2* hereditary breast and ovarian cancer genes increase male carriers' lifetime risks for developing prostate cancer by 8.6-20 %. Thus, recommendations in genetic counseling for carriers include cancer risk follow-up procedures such as annual prostate specific antigen (PSA) test and digital rectal examination after age of 40 years old. However, despite increased risk for prostate cancer, there is no formal follow-up program for male *BRCA1/2* mutation carriers in Finland. Therefore, we have examined Finnish male *BRCA1/2* mutation carriers' the prostate cancer risk follow-ups based on their own experiences. **Methods:** Data were collected by theme interviews from Finnish male *BRCA1/2* mutation carriers (*n*= 31) registered in Departments of Clinical Genetics at five university hospitals in Finland. The themes were based on findings of previous studies and the data were analyzed using descriptive statistics and content analysis. **Results:** The mean age of the male *BRCA1/2* mutation carriers was 59 (range 33-82) years. Two men had been diagnosed with prostate cancer. Most of the male carriers (71 %; 22/31) had received recommendations for prostate cancer risk follow-up, and for 45 % (10/22) of those men the implementation was according to international recommendations. However, the men reported that they received inadequate information about follow-up plans in genetic counseling sessions. The men were also unsatisfied with the follow-up sessions, particularly because they felt that health-care professionals involved lacked knowledge of *BRCA1/2* and they doubted the importance of annual follow-ups. Moreover, the men worried about the continuity of prostate cancer risk follow-ups and expressed wishes that follow-ups should be organize automatically and regularity, and there should be opportunities to discuss possible problems necessary. **Conclusion:** This study shows that registered Finnish male *BRCA1/2* mutation carriers had received recommendations for prostate cancer risk follow-up. However, more attention to implementation of the surveillance is required. Departments of Clinical Genetics have an important role to play in organizing appropriate follow-ups, and findings of the study indicate needs to improve practices to provide cancer risks follow-ups in a comprehensive manner and adequate information about follow-ups in genetic counselling sessions for these carriers. Further studies are needed to explore benefits of prostate cancer follow-ups. .

552F

Sociodemographic influences on attitudes and beliefs about genetic research and services. *P.L. Bussies[1,2], K. Celis[1], S. Guo[1], K. Czape[1,2], L. Adams[1], J. Lee[1], S. Hahn[1], M. Cuccaro[1], M.A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, Miami, FL; 2) University of Miami Miller School of Medicine, Miami, FL.

Personalized medicine (PM) is individualized healthcare using multi-source information that is unique to a given person. A major source of PM is genomic information derived from ongoing clinical research. Hispanics (HI) and African Americans (AA) have been underrepresented in this research, decreasing the likelihood that PM will be clinically useful for them. Understanding the attitudes and beliefs of HI and AA about genetic research and genetic services can inform efforts to increase awareness of and participation in these activities. Our objective was to assess sociodemographic influences on attitudes and beliefs about genetic testing and genetic research among non-Hispanic whites (NHW), HI, and AA. We analyzed survey data collected from 426 individuals in South Florida (M=134/F=292; NHW=65, HI=258, AA=103). The 23-item survey collected sociodemographics (age, education, race-ethnicity, sex) and measured *genetic testing attitudes* (GTA; 4 items), *genetic research attitudes* (GRA; 4 items), *willingness to participate in genetic research* (WTP; 3 items), and *genetic knowledge* (GK; 12 items). We used univariate and multivariate methods to investigate group differences (sex, age, education, and race-ethnicity) on our dependent measures. Univariate ANOVAs (Bonferroni adjusted) showed (a) a **sex effect** in GRA (p=0.003); (b) an **education effect** in GK (p<0.001), as well as GRA (p<0.001) and GTA (p<0.001; and (c) a **race-ethnicity effect** in GTA (p<0.001) and GRA (p<0.001). There was no significant age effect. To model significant contributors to GRA and GTA, we performed logistic regression with sociodemographic variables as inputs. For our primary outcomes of interest, we found that education (p=0.005) and race-ethnicity (p<0.0001) were significant predictors of GTA while only education was a significant predictor of GRA (p<0.0001). Using these same predictors, we did not find a significant effect in our overall model for WTP. Our results suggest that both education and race-ethnicity play important roles in determining attitudes about genetic testing and genetic research. As previously reported by other investigators, our results point to less favorable attitudes among individuals with lower education as well as among AA and HI groups. These results support the importance of studying the complex interplay of sociodemographics and attitudes about genetics and genomics in order to enhance participation in understudied populations.

553W

Views of experts and the public on genome editing and its issues: A literature review. *I. Taguchi[1], T. Kona[1], Y. Sato[1], K. Matsuura[1], T. Wada[1,2], S. Kosugi[1,2].* 1) Department of Medical Ethics and Medical Genetics, Kyoto University Graduate School, Kyoto, Kyoto, Japan; 2) Clinical Genetics Unit, Kyoto University Hospital.

[Introduction] In 2012, the collaborative investigation team led by Professor Jennifer Doudna reported that a free genome editing using CRISPR-Cas9 was possible. They announced that National Academy of Sciences accepted the genome editing treatment for fertilized eggs only for serious genetic diseases in February, 2017. However, this ethical agreement was reached during the arguments for the basic and clinical researches. As there was a rapid shift towards the genome editing in practical use, we conducted a literature review to clarify future problems as well as grasping views of experts and the general public.   [Method] We used online databases (PubMed, ICHUSHI web) to find articles using the search query: ("genome editing" OR "gene editing") AND human. The inclusion criteria for this review are below; 1) Original articles or comments with authors' opinions 2) Obtained from the published in peer-reviewed journals 3) Written in English or Japanese 4) Published during the last five years (2012.1-2017.5) [Result] There is no unified agreement on the utility of genome editing of the human germline information yet. Some experts are totally against although others agree upon its use for fertilized eggs under certain conditions such as prevention of serious genetic disorders. In the public, its use for serious illnesses was generally accepted. However, the majority of opinions were against its use in the modification of expressions unrelated to the illnesses. In addition, there was no attitude survey conducted on the public in Japan. [Discussion] Detailed investigations on uneasiness and expectation on genome editing should be carried out in Japan. It is suggested that development of the legislation and education systems corresponding to genome editing is urgently needed. In addition, it is necessary for certified genetic counselors (CGCs) to predict outlook of the public's ethical views about genome editing. Depending on its use in the future generation, CGCs need to be able to handle possible impacts of the editing flexibly.

554T

The Rare Genomes Project: Improving our ability to diagnose rare genetic conditions through a nationwide partnership with families. *H. Brooks, A. O'Donnell-Luria, J. Hendrey, A. Zimmer, N. Lopez Zalba, L. Greenfield, M. Wojcik, M. Udler, S. Baxter, B. Weisburd, T. Mullen, H. Arachchi, H. Rehm, D. MacArthur.* Broad Institute of MIT and Harvard, Cambridge, MA.

More than half of all patients in the US with a suspected rare genetic disease do not currently have a formal genetic diagnosis for their disorder. Identifying all of the genes underlying rare diseases will require large-scale outreach to the rare and undiagnosed community to participate in genomic research; however, the low prevalence of rare diseases means that patients are geographically dispersed, making recruitment challenging. The Rare Genomes Project (RGP; raregenomes.org) is a direct-to-patient research study leveraging social media and online participation to dramatically increase the reach of rare disease genomic research. Patients and families are recruited through advocacy groups and social media, and directed to a simple online form to tell us about their condition. Families that meet inclusion criteria sign an electronic consent and are mailed blood or saliva collection kits. If available, archival tissue samples are collected from pathology departments. Samples are mailed overnight to our genomics platform for exome, whole genome and/or transcriptome sequencing. If we identify the genetic cause of the patient's condition, we will perform clinical validation and provide a report to the patient's family and their doctor. Families are engaged as stakeholders at each step. The first pilot study of the RGP launched in June 2017. We will describe the lessons learned from the first several hundred recruited families on the journey from social media contact through to return of clinically validated diagnoses, and details of the diagnoses made by the project to date. We will also outline key unsolved challenges in direct-to-patient genomic research, including the current regulatory barriers to the return of raw genetic data to participants, and the uncertainty around international expansion of the direct-to-patient model. We anticipate that this project will provide proof of principle of the value of a patient-engaged approach to diagnosis and gene discovery in rare disease.

555F

Impact of personal microbiome information on research volunteers. *C. Bloss[1], E. Hyde[1], R. Knight[1], L. Goetz[2]. 1) University of California San Diego, La Jolla, CA; 2) Scripps Health, San Diego, CA.*

   Introduction. Human microbiome research has advanced rapidly, but aside from fecal transplantation for Clostridium difficile infection, there are currently no proven health interventions that involve manipulation of an individual's microbiome. Against this background, personal microbiome analysis has become available to individuals for a fee through both biotechnology start-up companies, as well as non-profit research programs. There has been fairly rapid uptake of the service among early adopters, with the phenomena not unlike the emergence of direct-to-consumer genetic testing companies in the late 2000s. **Methods.** A retrospective survey study of participants in the American Gut Project, a non-profit, crowdfunded and crowdsourced human microbiome research study was conducted. Respondents underwent personal microbiome analysis with return of results six months later. Survey items assessed motivations for participating and impacts of results on behavior. **Results.** A total of N=210 (64% response rate) individuals participated. General curiosity most commonly motivated individuals to seek personal microbiome analysis (80%), however, desire to improve one's health was the second most commonly cited motivation (50%). Other health-related motivations included gastrointestinal problems (GI; 35%) and "other microbiome-relevant health problems" (15%). Collectively, 133 individuals (63% of the sample) cited at least one health-related motivation. A total of 37% of respondents reported that they either made or planned to make behavioral or lifestyle changes based on their results, with change in diet (88%) or use of nutritional supplements (24%) most commonly cited among these individuals. Finally, 29% of those surveyed reported sharing or intending to share their results with their physician or health care provider. **Conclusions.** Although research has shown links between the human microbiome and a range of health outcomes, and personal microbiome analysis has become widely available through companies and non-profit research programs, the interpretation of an individual's microbiome data is still uncertain to specialized scientists, let alone practicing physicians. Despite this, findings suggest that health-related factors motivate individuals to seek personal microbiome analysis, and in some cases results may lead to changes in behavior. Further research is needed to understand the implications of these findings for the field of human microbiome science.

556W

Establishing variant frequencies for pharmacogenomic data in a community health system based pharmacogenomics program. *P.J. Hulick, D.L. Helseth, K. Kaul, J. Khandekar, K. Gulukota, A.A. Lemke, K.A. Mangold, N. Miller, A. Sereika, L.M. Sabatini, N.M. Parsad, D.T. Wake, H.M. Dunnenberger. NorthShore University HealthSystem, Evanston, IL.*

   Purpose: NorthShore's Pharmacogenomics (PGx) Clinic, established in March of 2015, is one of a handful of clinics nationwide which preemptively provides testing to help predict how patients will respond to a variety of widely prescribed medications. Over 500 clinic visits have occurred which has saturated capacity. The patient-centered clinical implementation served as the foundation for our broader initiative to provide seamless integration of pharmacogenomics across our health system. In November of 2016, NorthShore piloted the ability for primary care physicians to order pharmacogenomics testing directly in the clinic through a process that integrates ordering, patient consenting, resulting, and clinical decision support driven by PGx data within the electronic medical record (EMR). At the same time, a transition from several external PGx laboratories to the NorthShore Laboratory using internal developed PGx panel occurred. As part of PGX integration, an in-house variant repository (Flype), was developed to meet the needs of our growing PGX program. Crucial to monitoring the quality control of our PGX program is capturing the variant data at a population level within our hospital system. Here we report on our initial variant data for three common drug metabolizing enzymes *CYP2D6*, *CYP2C19*, and *CYP2C9* as part of our quality control initiative. These three enzymes were chosen because each laboratory tested them and there are Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for each. The CPIC reported frequencies for a Caucasian population was chosen as the comparator given the ethnic makeup of the NorthShore population. **Summary:**Testing for 288 patients occurred at external laboratories compared to 234 patients at the internal laboratory**.** There was similar distribution between allele and phenotype frequencies between all populations. CYP2D6 had the greatest number of allele variants observed in the NorthShore population, 18. CYP2C19 is the enzyme with the most phenotypic variation with only 37% of our population being normal metabolizers. **Conclusion:** From a quality control standpoint, these were the expected findings. It is important to note that even though the variants being tested from each laboratory differ this did not translate to a major difference at the phenotype level. It is important for any large system implementing pharmacogenomics to carry out similar evaluations to identify any expected or unexpected findings.

557T

Primary care physician views on direct access pharmacogenomic testing in a community health system. *A.A. Lemke, P.J. Hulick, C.G. Hutten Selkirk, A.W. Sereika, N.S. Glaser, D.T. Wake, H.M. Dunnenberger.* Center for Personalized Medicine, NorthShore University HealthSystem, Evanston, IL.

Clinical pharmacogenomics (PGx) is a relatively new field with variable delivery models. In fall of 2016, a pilot program of provider-ordered, direct access PGx testing was launched that involves mailing a testing kit to the patient, obtaining a buccal sample, and returning results by mail. Through this process, clinical PGx testing can be completed at home by the patient, while still receiving supervision from the ordering provider. Little is known about how health care providers view the benefits and challenges of this type of testing. To explore issues salient to physicians, this study utilized semi-structured interviews and a brief survey to assess perspectives of 15 community primary care providers who had undergone PGx kit testing themselves, and who also had received complimentary kits for five of their patients. Participants were drawn from six community practices and all spent 51% or more of their time in patient care. The majority specialized in internal medicine and their years in practice varied. Three main themes emerged from the qualitative interviews: 1) perceived value and utility of PGx testing; 2) specific challenges to implementation in practice; and 3) provider needs going forward. Reported utility of PGx testing included avoiding medication trial and error odysseys, reducing side effects, improving medication adherence, facilitating shared decision-making, and providing patients with validation of their past medication experiences. Participants described challenges in terms of patient data privacy concerns, cost of testing, insurance coverage, and understanding complex PGx test results. Provider needs reported were more PGx education, guidance on how to address cost and insurance issues with patients, and improved clarity in the PGx results report. The survey findings indicated that most (93%) participants believed that there is "clinical utility" in PGx testing, and all strongly or somewhat agreed that undergoing self-testing improved their own understanding of PGx testing. The majority (93%) also felt that the operational logistics of implementing direct access PGx testing in their workflow seems reasonable. However, one-third (33%) did not feel confident in their ability to explain PGx results to their patients. Our study findings reveal a broadened view of the concept of clinical utility of PGx testing and the types of initiatives primary care physicians feel are needed in order to reduce barriers in PGx service delivery.

558F

Pharmacogenomic counseling: Exploring participant questions about PGx test results. *T. Schmidlen[1,2], A.C. Sturm[2], L. Scheinfeldt[1,3], M.F. Christman[1].* 1) Coriell Personalized Medicine Collaborative, Coriell Institute for Medical Research, Camden, NJ; 2) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 3) Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA.

As the use of pharmacogenomic (PGx) information in healthcare increases, a better understanding of the needs of individuals receiving this information will be necessary to help guide PGx result delivery. The Coriell Personalized Medicine Collaborative (CPMC) is a prospective study investigating the utility of personalized medicine. Participants receive online genetic risk reports for 27 actionable complex diseases and 8 drug-gene pairs and can request free telephone-based genetic counseling (GC). To explore the needs of individuals receiving PGx test results, we conducted a retrospective qualitative review of inquiries from participants who requested counseling to discuss CPMC results from March 2009-February 2017. Eighty out of 690 (12%) total GC inquiries were focused on discussion of PGx results. All 80 PGx inquiries were coded and six salient themes emerged: "general help", "issues with drugs", "relevant disease experience", "what do I do now?", "sharing results" and "other drugs". Forty-three (54%) participants had general questions about their PGx results while only 7 (9%) were looking for specific guidance on dosing or drug selection. Past issues with drug side effects (n=14, 18%), dosing (n=2, 3%), and efficacy (n=3, 4%) were also mentioned. Some alluded to a personal (n=12, 15%) or family history (n=6, 8%) of diseases treated by drugs reported on in the CPMC study. Seventeen participants (21%) were interested in availability of PGx results for other non-study related drugs, while 28 (35%) asked about the impact of currently available PGx study results on other drugs. Some indicated sharing their PGx results with a doctor (n=5, 6%), and some questioned the impact of their PGx results on family members (n=5, 6%). No demographic factors predicted PGx genetic counseling requests (P>0.05). Total number of medications reported by a participant was also not predictive (P>0.05). However, number of reported medications with a corresponding PGx result was significantly associated with PGx GC requests (P<0.05). Highly engaged participants (viewed more study reports) were significantly more likely to request PGx GC (P<0.001). Our work illustrates a range of questions raised by study participants receiving PGx test results, most of which were addressed by a genetic counselor with few requiring input from prescribing providers or pharmacists. These results further support a role for genetic counselors in the team-based approach to optimal PGx result delivery.

559W

Implementation of precision medicine intiatives: Special considerations for underserved communities. *C.W. Brown[1,2], J. Ward[1,2], E. Pivnick[1,2], H.J. Mroczkowski[1,2], R. Mostafavi[2], R. Rooney[2,3], R. Joyner[2], A. Ryder[2,3], V. Park[1], N. Urraca-Gutierrez[2], LeBonheur Children's Foundation Research Institute.* 1) Pediatrics/Genetics Division, University of Tennessee Health Science Center, Memphis, TN; 2) LeBonheur Children's Hospital, Memphis, TN; 3) Department of Pediatrics, University of Tennessee Health Science Center.

  'Precision Medicine' promises to improve the quality of health care based in part on increasingly detailed patient information at the molecular level, thereby enhancing the precision and accuracy of diagnosis and affecting clinical management decisions. Once a precise diagnosis is established, more targeted and effective prevention, surveillance and treatment strategies can be implemented, and ineffective treatments that waste resources or can be associated with adverse outcomes can be avoided, including unnecessary procedures and medications. In addition to direct effects on medical care, genetic information can provide a diagnostic endpoint and greater sense of control, strengthen patient and family support networks that can favorably affect patient outcomes, and provide important health information that extends to family members, characteristics that challenge the conventional definition of clinical utility. At the societal level, effective genetic testing should have a favorable impact on health and allow effective health interventions throughout an entire population. However, substantial challenges occur in communities where limited resources preclude access to the 'high end' technologies that are associated with precision medicine initiatives, running the risk of creating even greater health care disparities in those communities. We describe several ongoing initiatives to address this challenge in underserved communities of Memphis, Tennessee, using a multi-faceted strategy that includes community outreach and education, partnerships with local foundations and industry, streamlining of processes to facilitate clinical DNA testing, and encouraging and facilitating participation in genomic research studies. Included in these efforts are a hospital-wide initiative to collect DNA from all patients to build a general population cohort to support genetic and genomic research; an industrial partnership, *Teaching the Genome Generation,* an educational program to implement laboratory based genomics exercises in local high schools; and other community-focused activities for genomics education. These measures are designed to help 'level the playing field' of accessibility and to minimize precision medicine disparities.

560T

Integration of a tool for electronic education and consenting within primary care to enable precision prevention. *M. Moore[1], M. Mboob[1], K. Rageth[1], M. Robinson[2], K. Hood[4], S. Jurgens[1], M. Haendel[3], C. Boerkoel[1], M. Sincan[1], C. Hajek[1], R. Pyatt[1], Monarch Initiative.* 1) Imagenetics, Sanford Health, Sioux Falls, SD; 2) Micromattie Consulting Inc, Orlando, FL; 3) Oregon Health and Science University, Portland, OR; 4) HAART (HIV/AIDS Alliance for Region Two), Baton Rouge, LA.

  Implementation of large scale genotyping within primary care for purposes of precision preventive medicine remains problematic; in part due to the lack of clinical resources for enrollment of patients. We hypothesized that electronic educational modules and electronic consenting can address the lack of resources and enable enrollment of patients into such genotyping protocols. To test this hypothesis, we developed an interactive, integrated consent and educational tools for enrolling individuals in pharmacogenetic and predisposition testing at the time of a visit to their primary care physician. Following refinement from feedback of the medical system, customer relations groups, and analysis of focus group responses; we launched this tool as part of a precision prevention initiative within a rural primary care setting. We present the logic underlying this process, as well as our initial results. These results include feedback from patient interaction with the electronic tools, educational initiatives to encourage patient uptake, and clinician responses.

561F

The missing and the vulnerable: Developing responsible science policy for applying DNA to cross-border humanitarian causes. *S.H. Katsanis[1], L. Snyder[1], A.Z. Mundorff[2], S.A. Faith[3].* 1) Duke Initiative for Science and Society, Duke University, Durham, NC, USA; 2) Anthropology, University of Tennessee, Knoxville, TN, USA; 3) North Carolina State University, College of Veterinary Medicine and Forensic Sciences Institute, Raleigh, NC, USA.

If scientists, humanitarians and forensic investigators were to convene today to design an international missing persons system, it likely would look very different from the systems currently in place. Within the U.S. and around the world, the current procedures for missing persons' identification using DNA are inefficient and fragmented, reliant upon law enforcement prioritization and policies borne from criminal investigations. Across borders, identifications of the deceased or displaced are mired in bureaucracy, jurisdictional regulations, and limited resources that prevent the science from being applied expeditiously to aid investigations. In addition, as concerns for border security increase, policymakers are turning to genomics as a biometric for tracing individuals entering the country, processing refugee claims, and screening for human trafficking. We have examined the science policy as applied to cross-border cases, conducted a stakeholder analysis on DNA applications for human rights, assessed cross-border DNA database tools, and developed model language for family reference sample (FRS) collection. We have found that genetic verification of claimed relationships is required in the U.S. for some refugees and proposed legislation would expand this practice. Hundreds of human remains of unidentified migrants are found along the southern U.S. border each year and thousands of migrants are dying as they cross the Mediterranean Sea. Such high-risk migrants often lack identification documents, so DNA is an important tool for death investigators to identify human remains. Cross-border sharing of FRS DNA profiles is essential and currently ineffective, requiring the resolution of many technical, ethical, political, and administrative challenges, particularly to protect vulnerable populations from abuses of power. While guidelines and best practices have developed for handling disaster victim identifications and for identification of the missing, standard consent practices for FRS have not developed. Continuing global migration in the coming decades necessitates rethinking of the existing infrastructure to enable cross-border DNA sharing. Networking the multiple efforts around the world is a first step towards improved communication for humanitarian identifications. Careful design of a DNA database system to protect sensitive data is vital.

562W

Tipping the scales: Participants make healthy dietary changes in response to direct-to-consumer genetic test results. *S.B. Laskey, J.F. Shelton, M. Johnson, C. Wilson, R. Smith, B.L. Koelsch, The 23andMe Research Team.* 23andMe, Mountain View, CA.

The growing direct-to-consumer (DTC) genetic testing industry serves as a promising opportunity to provide personalized recommendations and empower consumers to make healthy lifestyle decisions. However, motivating behavior change is a notorious public health challenge, and the impact of DTC genetic test results on short-term and long-term behavior is still an active area of investigation. In this study, we measured self-reported dietary behavior in a subset of consented 23andMe customers who received the personalized Saturated Fat and Weight Report (SFWR), which describes the impact on BMI of an interaction between rs5082 genotype and dietary saturated fat intake. One month after publication of the SFWR, 47% of the survey participants who reported viewing their results (n = 17,253, 62% female) also reported that they were considering, planning, or had already started making dietary changes in response to the report. These participants were more likely to be female (p < 10^{-12}), have GG genotype at rs5082 (p < 10^{-149}), and have higher BMI (p < 10^{-147}). In a follow-up survey six months later (n = 10,387), 26% of participants reported having made dietary changes in response to the SWFR, including 55%, 45%, and 29% of participants who were considering, planning, or had started changes at the one-month time point, respectively. The dietary changes reported by participants were overwhelmingly healthy and aligned with the recommendations published in the SFWR. For example, of participants who reported changing dietary saturated fat intake, 94% reported eating less saturated fat in response to the SFWR. At the six-month follow-up time point, 91% and 97% of participants who reported making changes to dietary saturated fat and unsaturated fat intake, respectively, reported that they had maintained those changes. These results suggest that personalized, DTC genetic results can be a powerful motivator for positive lifestyle changes and a trusted resource for informing health and wellness decisions. Of participants who reported making dietary changes since the release of the SFWR, 66% say that the report impacted their decision. Importantly, because survey respondents may take a more proactive role in their health than non-respondents, these findings may not be generalizable to users who did not answer our surveys.

563T

Genotypes associated with phenotypes: A human genetics laboratory exercise. *D. Caporale, A. Pritchard.* Biology, University of Wisconsin-Stevens Point, Stevens Point, WI.

Within the human genome, certain genes seem to be associated with particular phenotypes. As part of a new laboratory exercise in a Human Genetics course, single nucleotide polymorphisms (SNPs) of potential interest to students were selected for assay development. Some SNPs discovered within previous studies included: 1) an APOA2 gene variant associated with weight gain when consuming saturated fatty food, 2) variants of the FGF21 and FTO genes associated with a person's preference towards sweet foods, 3) an OXTR heightened empathy variant, and 4) CYP1A2 variants associated with caffeine metabolism. We designed allele-specific primers to amplify the wild type alleles and variant alleles of these genes and others, in order for students to identify their genotypes. Computer websites were used in the following order: UCSC Genome browser to obtain gene sequences containing each SNP, NCBI PRIMER BLAST to design primer pairs, Fisher Scientific Primer Design Tool to determine the Tm compatibilities among the four primers, and BLAT to assess primer specificity within the human genome. Primers were then tested with three DNA control samples to optimize temperature profiles. After optimization, these assays were performed by students enrolled in Human Genetics. Here we report the results of the class to demonstrate the success of the assays and illustrate the genetic diversity of these genes within the class. Finally, associations between students' genotypes and phenotypes were assessed.

564F

f-treeGC: Questionnaire-based pedigree chart creation software in compliance with recommendations for standardized human pedigree nomenclature. *T. Tokutomi[1,2], K. Yamamoto[1,2], A. Shimizu[2], M. Sasaki[2], A. Fukushima[1,2].* 1) Department of Clinical Genetics, School of Medicine, Iwate Medical University, Morioka, Iwate 020-8505, Japan; 2) Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Shiwa, Iwate 028-3694, Japan.

Objective The Tohoku Medical Megabank project aims to create a next-generation personalized healthcare system by conducting large-scale genome-cohort studies involving three generations of local residents in the areas affected by the Great East Japan Earthquake. We collected medical and genomic information for developing a biobank to be used for this healthcare system. We designed a questionnaire-based pedigree-creation software program named "f-treeGC," which enables even less experienced medical practitioners to accurately and rapidly collect family health history and create pedigree charts according to recommendations. **Methods** f-treeGC may be run on Adobe AIR. Pedigree charts are created in the following manner: 1) At system startup, the client is prompted to provide required information on the presence or absence of children; f-treeGC is capable of creating a pedigree for up to five generations. 2) An interviewer fills out a multiple-choice questionnaire on genealogical information. 3) The information requested includes name, age, gender, general status, fertility status, pregnancy status, fetal status, and physical features or health conditions of individuals over five generations. In addition, information regarding the client and the proband, and birth order information, including multiple gestation, custody, multiple individuals, donor or surrogate, adoption, and consanguinity may be included. 4) f-treeGC shows only marriages between first cousins via the overlay function. 5) f-treeGC automatically creates a pedigree chart, and the chart-creation process is visible for inspection on the screen in real time. 6) The genealogical data may be saved as a file in the password-protected original format and/or in read-only format. The created/modified date and time may be changed as required. By default, the file name automatically contains the terms typed into the entry fields, including physical features or health conditions, to enable sorting or searching from the database. **Summary** f-treeGC is fully compliant with international recommendations for standardized human pedigree nomenclature, and has a variety of uses, from genome cohort studies or primary care to genetic counseling. The present software is available for use, at no monetary cost, at the Iwate Medical University Hospital website (http://www.iwate-med.ac.jp/hospital/clinics/medical/m26/).

565W

Everything is actionable: Patient values and perceived utility of incidental genome sequencing results. *Y. Bombard[1,2], M. Clausen[2], C. Mighton[2], S. Shickh[2], E. Glogowski[3], K. Schrader[4], M. Robson[5], A. Scheer[1,2], C. Elser[6], A. Eisen[7], M. Evans[1,2], J.C. Carroll[1,8], J. Hamilton[5], K. Offit[5], K. Thorpe[1], J. Lerner-Ellis[1,8], A. Laupacis[1,2].* 1) University of Toronto, Toronto, ON; 2) St. Michael's Hospital, Toronto, ON; 3) GeneDx, Gaithersburg, MD; 4) BC Cancer Agency, Vancouver, BC; 5) Memorial Sloan Kettering Cancer Center, New York, NY; 6) University Health Network, Toronto, ON; 7) Sunnybrook Health Sciences Centre, Toronto, ON; 8) Mount Sinai Hospital, Sinai Health System, Toronto, ON.

   Introduction: Genome sequencing (GS) is a driver of personalized health, facilitating diagnosis and optimized care, especially in precision oncology. One complex feature of this technology is its capacity to generate incidental results (IR), which has created much controversy. Guidelines suggest that 'medically actionable' IR should be offered to patients undergoing GS, with calls to offer additional IR based on patient preferences. Research explores patient preferences for IR in broad, often abstract terms, yet little is known about patient informed and deliberated values towards IR. Our objective was to describe patient considered values and perceived utility of receiving their IR following a deliberative, educational intervention. **Methods:** Semi-structured interviews were conducted with 15 cancer patients participating in usability testing of a decision aid (DA) designed to facilitate informed, value-based selection of incidental findings. We created an interactive, online DA (www.genomicsadvisor.com) to guide patients' selection of IR by "binning" IR into 5 categories: (1) Medically actionable & pharmacogenetic variants, (2) Common disease SNPs, (3) Mendelian disease variants, (4) Early-onset neurological variants, (5) Carrier results. Content analysis was used to analyse the data. **Results**: Participants were enthusiastic towards GS itself, and expressed a perceived inherent value for its use in their own healthcare. All participants chose to receive some incidental findings; 9 participants selected all incidental findings. Many participants saw knowledge as empowering, and stated that the more information they had, the more actions they could take. Participants perceived all categories of IR as being "actionable," even ones defined as non-medically actionable (2, 3, 4, 5). Participants said that learning any IR could inform actions to delay disease onset, reduce symptoms, help with planning and to inform their families' future financial or reproductive planning. **Conclusion:** Despite this small sample size, participants' enthusiasm for IR suggests that uptake and use of results will be high. Many participants perceived all categories of IR to be "actionable," which creates a divergence in how patients and experts conceive IR as well as their utility and obligations for return. This divergence may cause gatekeeping challenges for providers if they are positioned to offer only medically actionable IR, and variations in access to information among patients.

566T

Repatriation of out-of-country molecular testing for disorders targeted by the provincial newborn screening program: Improving access to diagnostic testing in a cost and time efficient manner in Ontario, Canada - Our first year experience. *B. Bélanger[1], D.E. Bulman[1,2].* 1) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Research Institute and University of Ottawa, Ottawa, Ontario, Canada.

   Newborn Screening Ontario (NSO) screens about 140 000 newborns every year in the province of Ontario, Canada. Following a positive newborn screen (NBS) result, newborns at risk for one of the multiple conditions screened have to go through diagnostic testing including molecular testing. The majority of the molecular diagnostic tests were available in laboratories outside the province, leading to additional costs and delays associated with sample transportation and prior approval by our provincial evaluation process. NSO developed a Next generation sequencing (NGS) targeted gene panel including genes associated to primary and secondary diseases targeted by the provincial NBS program. The genes targeted for interpretation can be chosen by positive screen results (i.e. elevated C5OH, PA/MMA, etc.), disease targeted, or it can be gene specific. Samples from newborns are all treated as STAT sample with a maximal turnaround time of 6 weeks (12 weeks for non-STAT samples). Within the first year of operation, our lab received 123 samples to test for the metabolic targets of newborn screening including 79 STAT samples for newborns or carrier testing for pregnant couples. All results were communicated to the ordering physician with an average turn-around time of 24.3 days between receiving a requisition and sending the report. Results were positives (2 variants) in 42% of the reports, carriers (1 variant) in 21% of the reports and negative in 38% of the reports. For at least 2 patients, the option to select interpretation by positive screen results helped us identify variants in secondary targets that would likely have been missed if only genes associated with primary targets were requested. We also received samples from older children and adults with a clinical diagnosis of a metabolic disorder, but for which genetic testing was not initially provided. The Ontario Government repatriation initiative helped us to offer a cost-effective, targeted and manageable approach for molecular diagnostic testing in newborns and families in Ontario. This additional testing by the provincial NBS laboratory helped to repatriate testing within the province and to facilitate molecular testing for health care provider, increasing accessibility for patients and family.

567F

Developing a genomics ready clinical workforce: From raising awareness to the establishment of a faculty of genomic medicine. *M. Bishop, D. Callaghan, E. Miller, S. Simpson, A. McPherson, A. Pope, M. Foster, A. Seller.* Genomics Education Programme, Health Education England, Birmingham, West Midlands, United Kingdom.

By April 2018 whole genome sequencing (WGS) is expected to be commissioned for use in England's National Health Service (NHS). Health Education England's Genomics Education Programme (GEP) was established to ensure relevant staff groups are equipped in knowledge and skills to safely deliver their role in the patient pathway. While ensuring capacity and capability of the highly specialised genomic workforce is key, the greater challenge is to provide the right level of education to the wider workforce. The GEP has taken a top-down and bottom-up approach to address this issue. In 2014 the GEP established the flagship Master's programme in Genomic Medicine, delivered by 10 universities across England. The aim of this programme is to expand and extend the knowledge of trained health professionals by providing an academic grounding in genomics so they can act as clinical champions within their own specialism. To date 750 NHS professionals from all sectors of the workforce have completed at least 1 module of the Master's programme. Alumni will join the new NHS Faculty of Genomic Medicine which will serve as the centre of knowledge and expertise for genomic medicine. Concurrently, the GEP administered a cross-sectional survey of the NHS workforce (n=2578) to identify education and training needs. Results show the majority of staff (80%) want further education in genomics, primarily on how genomics will impact their clinical role. In response to this the GEP has developed educational activities using innovative methods to raise awareness of the use of genomics in healthcare. In 2016, a 3 week Massive Open Online Course (MOOC) on WGS and its application in healthcare was launched. Over 3 iterations, the course registered >14,000 learners representing a range of professions within the NHS. Evaluation of course comments showed learners understood the use of genomic technologies and how genomics will influence their own sphere of practice. The GEP has also used social media as a tool to educate and inform different NHS staff groups through the facilitation of twitter chats. These 1 hour chats enable awareness raising and education at scale, with each chat having a total reach of >5 million. Analyses of tweet conversations show an evolution of participants' understanding of the relevance of genomics. These two approaches have ensured the GEP are responding to the learning needs of NHS staff, whilst also establishing a network of clinical champions within the NHS.

568W

Use of problem-based team learning to improve success of underrepresented students in an undergraduate genetics course. *B. Bowling, E. Strome, D. Robertson.* Department of Biological Sciences, Northern Kentucky University, Highland Height, KY.

Active learning techniques in various science, engineering, technology, and mathematics courses have been shown to increase student success, particularly students coming from underrepresented backgrounds (Freeman 2014, Wieman 2014). Recent studies have begun to focus on specific ways to optimize active learning (e.g. Adams, Garcia, and Traustadottir 2016). We have implemented a model combining active learning techniques we refer to as problem-based team learning (PBTL) in which students are engaged in their learning inside and outside of class. This sophomore/junior level genetics course is required for all biology majors and a prerequisite for upper-level courses. The course includes introductory concepts in genetics and molecular biology such as DNA structure and function, gene expression and regulation, and population genetics. PBTL requires students complete readings and some online lectures prior to class and students are held accountable through individual and team readiness assessment quizzes (Michaelsen *et al.* 2004). In-class, students work collaboratively in permanent teams on application problems. We analyzed student success in the course based on risk factors including low income (as determined by Pell grant eligibility), underrepresented minority (self-reported), and first generation college student (self-reported). The average class size was approximately 50 students and those with risk factors comprised around half of the class. Four years of data with the same instructor indicate students with one or more of the risk factors were ~5 times more likely to succeed in the course (earning a C- or higher) when it was taught using PBTL than using less consistent active learning techniques. Student were also more likely to persist rather than abandon the course before and after the withdrawal deadline. Future work involves testing the efficacy of the PBTL model when adopted by other instructors in the genetics course as well as other courses within the curriculum, and further supporting underrepresented students in biology.

569T

Phenotate: Crowdsourcing phenotype annotations of genetic disorders through student exercises. *W.H. Chang[1,3], A.X. Lozano[2,3], M. Brudno[1,3], G. Baynam[4,5].* 1) Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; 2) Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; 3) Centre for Computational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Western Australia; 5) Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, Perth, Western Australia.

Creating and cataloging accurate, up-to-date, and standardized phenotype annotations for genetic diseases has tremendous clinical and research applications. In addition to providing clinicians with clear references to assist with diagnoses, especially for rare diseases, structured and computable annotations are used by algorithms to suggest a diagnosis based on a given set of phenotypes from a standardized terminology. The Human Phenotype Ontology (HPO) project provides a standardized nomenclature of phenotypes. However, the linkages between HPO and disease terminologies such as Online Mendelian Inheritance in Man (OMIM) and Orphanet Rare Disease Ontology (ORDO) are incomplete, while methods for differential diagnoses are heavily reliant on the quality of the annotations. Our project harnesses crowdsourcing, the use of large numbers of non-experts to collect high-accuracy knowledge, to generate and improve phenotype annotations between HPO and OMIM/ORDO. Phenotate is a web-based platform for undergraduate or medical students, as well as medical residents, to annotate OMIM/ORDO diseases with HPO phenotypes by completing classroom exercises. Students' responses are graded automatically against an expert-validated set of annotations for a given disease, while the ability to assign diseases without existing expert annotations expands the knowledgebase to new disorders. Annotations must reference the literature. Through these classroom exercises, new insights for a disease can be gained by observing which annotations are the most frequent, weighted by the students' grades for the exercise (a proxy of quality and reliability). The first class to use Phenotate was a second year undergraduate molecular genetics class (MGY200) at the University of Toronto in which 78 students annotated Marfan syndrome (MFS), congenital myasthenic syndrome (CMS), and Friedreich's ataxia (FA) with over 500 unique phenotypes across 72, 68, and 69 annotations, respectively. Overall, students provided more comprehensive annotations than clinicians who also submitted annotations: for MFS and FA, the most commonly entered phenotypes (ectopia lentis and dysarthria, respectively) were not present in the expert annotations, despite their relevance. Phenotate is an open platform, available for use by anyone teaching clinical genetics. Crowdsourcing annotations will enable a better understanding of genetic disorders, while offering students an educational tool that supplements their coursework.

570F

The Genomics Education Partnership: Authentic big data course-based research projects for undergraduates. *R.L. Glaser[1], W. Leung[2], D. Lopatto[3], C.D. Shaffer[2], S.C.R. Elgin[2].* 1) Stevenson University, Owings Mills, MD; 2) Washington University in St. Louis, St. Louis, MO; 3) Grinnell College, Grinnell, IA.

Undergraduate research is one of the ten high impact practices (HIPs) in undergraduate education, as identified by the Association of American Colleges and Universities, that has been shown to positively impact retention and engagement. Since 2006, the Genomics Education Partnership (GEP), a collaboration of more than 100 diverse undergraduate colleges and universities and The McDonnell Genome Institute at Washington University in St. Louis (WUSTL), has provided authentic genomics research projects to undergraduate students. The GEP incorporates genomics research into the undergraduate biology curriculum to help faculty keep up with the rapid advances in the fields of genetics and genomics, and to introduce students to big data analyses. Current GEP research projects utilize comparative genomics to understand the evolution of the heterochromatic domains of the *Drosophila* Muller F element. GEP courses can be developed for any level and for multiple course settings. GEP courses can focus on the gene annotation of a recently sequenced *Drosophila* species, and/or sequence improvement, the latter involving a wet lab component. Annotation projects are designed to teach students about eukaryotic gene structure and genome organization, while introducing them to bioinformatics tools such as the UCSC genome browser. All GEP faculty have attended a training workshop at WUSTL to gain familiarity with the GEP protocols and curriculum materials. Completed student projects are submitted to WUSTL for quality control checks, after which they are pooled and the assembled data used in publications. Students who review, critique and approve the draft manuscript are eligible for co-authorship on the publication. Students enrolled in GEP courses report gains in knowledge about eukaryotic genes and genomes, and the scientific process, irrespective of institution type or course structure, with the learning gains in GEP courses being similar to those in summer research experiences. The GEP has recently developed curriculum for beginning college students, which focuses on involving freshman as well as community college students in the projects. In addition to supporting the *Drosophila* genome projects, the GEP is collaborating with the Galaxy Project to develop G-OnRamp, a platform that enables faculty researchers working with other species to develop genome browsers to aid in engaging students in genome annotation of their species of interest.

571W

Utility of video-based education in the genetics clinic and beyond: Developing alternative service delivery models. *P. Magoulas, D. Riconda, A. Lewis, H. Streff, S. Lalani, C. Shaw, A. Wallis, I. Machol, S. Chargois, L. Rosales, B. Lee.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

   Education about genetic testing in a clinic setting has historically been provided by genetics specialists or other health care providers during face-to-face encounters. While this can be effective, it may not be the most efficient way to provide education in an increasingly busy and time-constrained genetics clinic environment. In addition, specialists with limited genetics training may order genetic tests for their patients without the assistance of genetics specialists. Thus, there is a greater need to educate an increasing number of patients about the nuances of genetic tests. We surmised that creating a series of educational videos to aid in the education and consenting process for certain genetic tests, such as whole exome sequencing (WES), could improve efficiency and increase patient understanding and satisfaction. The videos are meant to supplement the genetic counseling process by explaining basics of genetics, complexity of specific genetic tests, and types of results that may be returned. By providing patients standardized information, more clinician time can be spent providing answers to patient-specific concerns. Patient understanding may also be improved by viewing the video more than once. We were able to successfully integrate educational videos into the clinic workflow both during the clinic visit and outside of clinic by creating multiple internet-based platforms in which to view the videos. We also developed surveys that are completed before and after viewing the video to capture familiarity of the specific test and effectiveness of the video as an educational tool. For example, the average familiarity with WES before viewing the video was 2.05 on a Likert scale of 1 to 5 (1 = not familiar; 5= very familiar) (N=20). All of the respondents agreed that they had a better understanding of WES after viewing the video with an average score of 4.29 out of 5 (1=strongly disagree; 5=strongly agree) and that it provided a good explanation of WES (average score of 4.23). These preliminary data suggest that the video is an effective tool that increases knowledge and understanding. The utility of educational videos in both the genetics clinic and other subspecialty clinics can be an efficient mechanism to provide general education about genetic tests so that patient-specific questions and concerns can be addressed. Alternative delivery models, such as video education, will be necessary as complex testing in the genetics clinic and beyond increases.

572T

An interactive video vignette successfully teaches pedigree analysis to undergraduates. *D.L. Newman[1], L.K. Wright[1], J.A. Cardinale[2].* 1) Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY; 2) Dept. of Biology, Alfred University, Alfred, NY.

   Interactive Video Vignettes (IVVs) are an innovative online medium that engage users by incorporating prediction questions, data analysis and comprehension questions in an engaging and accessible way. A list of available biology-content IVVs can be found at https://www.rit.edu/cos/interactive/. We developed an IVV to teach students about Mendelian genetics in the context of a human disease. The storyline involves a college student working with his mother to construct their family's pedigree and decide whether or not they are at risk for Marfan Syndrome. This IVV was a little longer than most other IVVs (about 20 minutes) and incorporated many complex ideas that are known to be difficult for undergraduate students, including 1) *The relationship between genes, alleles and traits,* 2) *The genetic meaning of dominance,* 3) *Pedigree construction,* 4) *Pedigree analysis of a Mendelian trait,* and 5) *Calculating probability of inheritance.* Assessment questions that aligned with each of these concepts were developed and administered before and after students completed the IVV activity. Validation interviews were conducted to ensure assessment question clarity and to gain deeper insight into student thinking (N=14). Pre/post data and IVV usage metrics were gathered from eight classes (22-38 students per class) from three different institutions to investigate student understanding of the concepts linked to the main ideas addressed in the IVVs. All student populations studied showed improvement on the pre/post questions related to *pedigree construction*, *pedigree analysis of a Mendelian trait*, and *calculating probability of inheritance*. Our analysis also revealed that students enjoyed the activity, found the story engaging, and agreed that it enhanced their learning.

573F

RGEOde: Mining big data in the high school or undergraduate biology classroom. *K. Pirc-Hoffman[1,2], J. Pevsner[3,4], A.L. Norris[1,3,4].* 1) Krieger School of Arts and Sciences, Johns Hopkins University, Rockville, MD; 2) Project Lead the Way (PLTW) Biomedical Science Program, Gaithersburg High School, 101 Education Blvd, Gaithersburg, MD 20877; 3) Department of Neurology, Kennedy Krieger Institute, 707 N. Broadway, Baltimore, MD 21205; 4) Department of Neuroscience, Johns Hopkins School of Medicine, 725 N. Wolfe St, Baltimore, MD 21205.

Despite the emerging role of big data in science today, bioinformatics is absent in most high school biology classrooms. Some science-focused curricula have integrated bioinformatics, but to a very limited extent. Given that data and tools are freely available online, an internet-enabled computer is all that is required to learn and perform bioinformatics analyses. Here we present R GEO Data Explorer (RGEOde), a gene expression module to facilitate integration of bioinformatics into the high school or introductory college biology classroom. To assist teacher-implementation, RGEOde is a complete module that includes resources for both the teacher and students. These resources include tutorial PowerPoint presentations, pre- and post-quizzes, student worksheets, and an example RGEOde analysis. Through the lesson sequence, students: (1) gain content knowledge of how eukaryotic genes are regulated and expressed, (2) learn basic R programming, and (3) implement R programming for gene expression analysis of a GEO microarray experiment. Field testing of the RGEOde module was conducted with a group of Project Lead the Way (PLTW) Biomedical Science students (n = 33) at Gaithersburg High School (Montgomery County, Maryland). The student cohort included both juniors (88%) and seniors (12%), with a slight sex-bias (64%) for self-identified females. Student knowledge content scores on gene expression increased an average of 10% from pre-quiz to post-quiz. The majority of students self-reported an increased interest in bioinformatics after the activity (80%). The challenges to implementation were technical in nature, notably restricted permissions and limited hard drive space on the school-provided laptops. Students who used their own laptops did not face these challenges. As a result, the module has been adapted to circumvent such challenges, through use of alternative implementation methods, including online web apps and pre-loaded USBs for students. RGEOde is a complete bioinformatics activity module appropriate for high school and introductory undergraduate biology classes. RGEOde leverages freely available gene expression data from NCBI GEO (www.ncbi.nlm.nih.gov/geo/), analysis tools in R, and best practices from field testing. The RGEOde module complements existing high school curricula, including PLTW Biomedical Science (MI 3.1.4-5, BI capstone), NGSS (HS-LS1-1), and AP Biology Framework (3.A.1.e).

574W

No one's genome is more interesting than your own: Understand Your Genome® through experiential education. *E. Ramos, M.A. McGinniss, J. Esposito, E. Thorpe.* Illumina, Inc., San Diego, CA.

Background Clinical whole genome sequencing (cWGS) continues to gain acceptance as the cost of sequencing decreases and studies demonstrating clinical utility are highlighted. Despite the accelerating pace of adoption, there are limited education options about cWGS available to a broad population. To address growing interest, the Understand Your Genome® (UYG) program was created to improve understanding, awareness, and engagement around cWGS via an experiential educational forum. **Methods** Full participants received physician-mediated cWGS through the CLIA-certified, CAP-accredited Illumina Clinical Services Laboratory, which included clinical interpretation of variants found in ~1690 genes associated with more than 1200 genetic disorders, access to their genome in an online visualization tool, and attendance at an educational symposium. A "watch and learn" option was introduced starting in November 2015 for symposium attendance only. At registration, attendees were asked questions about their professional backgrounds, understanding of genomics, and their motivations for attending the UYG program. Beginning in 2016, participants were surveyed immediately following the conference. Participants of 10 symposia completed identical surveys, and responses were tallied and analyzed. **Results** As of May 2017, 1932 individuals participated in one of 35 programs and 1437 had cWGS. 264 individuals completed the survey for a completion rate of 40% (range of 13-63%). Across these 10 symposia, 64% of participants responded that their interest in genomics had increased significantly. For the 3 programs that had the most diverse attendee background with lower genomic and medical literacy, 78%, 88% and 89% reporting a significant increase in interest. >82% reported an increase in their understanding of genomics, biology and common diseases, how genomics affects them, the clinical utility of genomics, and how genomics is being applied to improve patient care. 89% rated their experience as excellent or very satisfied. **Conclusion** A review of survey data from individuals participating in an experiential education program report increased interest in genetics, understanding of multiple facets of clinical genomics, and intent to continue engagement with their genomes. As the price of cWGS continues to drop, these data suggest that experiential education may support engagement and genomics education for a diverse population, including community genomics initiatives like the All of Us Study.

575T

Breakthroughs and barriers in secondary/post-secondary partnerships designed to promote research in molecular genetics, human genetics and GMOs. *T.N. Turley-Stoulig[1], A. Anderson[2], A. Bertaut[1], A. Call[2], C. Carolina[2], A. Hartley[2], C. Hudspeth[1], K. Humble[1], I. Juarez[1], T. Leonard[2], N. Lorena[1], M. McCormick[2], D. Newton[2], K. Phan[1], J. Pursell[2], R. Rieger[1], C. Taylor[1], H. Williams[1], A. Ulloa[1], H. Ulloa[1], T. Tinney[2].* 1) Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA; 2) Northshore Technical Community College, Lacombe, LA.

   The field of genetics and associated technologies are advancing at a rapid pace, creating a great need to interest students in genetics for future study and better educate the general population to foster a better understanding of advancements being made and genetic services available. The purpose of this study was to increase genetics-based knowledge and comprehension, increase awareness of genetics technologies and research and encourage study in genetics and STEM disciplines in general for a range of students at progressive levels. Four high schools participated during the 2016-17 year following a pilot program the previous year with one participating high school. We used an innovative, multi-dimensional approach involving a university/community college partnership coupled with outreach that spans the fields of genetically modified organisms (GMOs) and human genetics by bringing laboratory techniques of both disciplines into classrooms of high school students, exposing them to genetics-based materials they would otherwise not encounter. A series of primer modules were used to educate students in basic genetics and train them in DNA isolation, PCR and GMO analysis through a series of practical modules. Instructional modalities included animations and hands-on techniques. The study spanned the academic year and involved training of undergraduate students from Southeastern Louisiana University who then taught Northshore Technical Community College students. The undergraduate cohorts then team-taught high school students in their high school classrooms and on the undergraduate campuses. Primer modules covered basic genetics concepts, human genetic disease, GMOs, and modern laboratory techniques. Practical modules provided laboratory extensions for conducting GMO experiments using techniques discussed in primer modules. Challenges were largely associated with attrition of community college students. We plan to modify the program in the future using feedback from community college students and greater involvement of community college faculty in an effort to increase retention. Results at all student levels indicated an increase in student understanding of genetics concepts and stronger interest in the study of cells and/or genes in the future. We conclude that this 'ladder of learning" approach enhances the educational experience of community college, university and high school students in a unique and invaluable way and may be applied to other programs.

576F

Effectiveness of a dedicated rotation in genomics, genetic medicine, and undiagnosed diseases for internal medicine residents. *M. Wheeler[1,2], L. Geng[2,3], J. Kohler[1,2], P. Levonian[4], R. Witteles[1,3], J. Hom[2,6], J. Ford[4,5].* 1) Division of Cardiovascular Medicine, Stanford University, Palo Alto, CA; 2) Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 3) Internal Medicine Training Program, Department of Medicine, Stanford University, Stanford, CA; 4) Department of Genetics, Stanford University, Stanford, CA; 5) Division of Medical Oncology, Stanford University, Stanford, CA; 6) Division of General Internal Medicine, Stanford University, Stanford, CA.

   Though genetics and genetically driven therapies are gaining importance in the care of adult patients, formal graduate medical education in genetics and genomics has not been a routine part of the internal medicine residency curriculum. To address this educational gap, we developed a rotation to provide targeted genetics and genomics education for internal medicine residents at Stanford University. The two-week rotation included subspecialty inherited diseases clinics with practitioners in medical oncology, cardiovascular disease, neurology, and medical genetics. In addition, residents participated in sessions with advanced practice members of the clinics including genetic counselors, physician-scientists, and nurse specialists. To complement the clinical educational component, residents attended clinical conferences focused on genetic diseases and were introduced to the Center for Undiagnosed Diseases (CUD), a clinical site of the Undiagnosed Diseases Network. Under the supervision of physician members of the CUD, residents could participate in the clinical evaluation of patients referred to the CUD. We surveyed participants to assess their experience in the rotation, their genetic literacy prior to and after the rotation, and to compare knowledge gained during the rotation with the educational experiences in other internal medicine residency rotations. Survey respondents identified increases in genetic literacy, new appreciation of genetic counseling, importance of family history taking, availability and affordability of genetic testing, and utility of genetics for clarifying diagnosis. Residents started with a wide range of genetic literacy, from a single course in medical school to doctorate in computational genomics; despite this diversity, all participants gained additional knowledge and described the rotation as a valuable addition to their graduate medical education. Exposure to the CUD was cited as a valuable addition, as was the opportunity to focus on diagnosis and to delve deeply into a challenging case. As graduate medical education continues to evolve, our experience demonstrates that integration of genetics and genomics into the training of internal medicine residents is feasible, effective, and desired by trainees. Similar rotations more broadly implemented may be a viable intermediate to existing joint training programs in internal medicine and medical genetics.

577W

Perspectives and barriers to adoption of infobutton-related technologies for genomic medicine. *L.V. Rasmussen[1], S.J. Aronson[2], J.J. Connolly[3], G. Del Fiol[4], R.R. Freimuth[5], J.J. Murray[6], D.B. Pet[7], J.F. Peterson[7], B.H. Shirts[8], M.S. Williams[9], C.L. Overby[9,10].* 1) Department of Preventive Medicine, Northwestern University, Chicago, IL; 2) Partners HealthCare, Boston, MA; 3) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT; 5) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 6) Department of Internal Medicine, Meharry Medical College, Nashville, TN; 7) Vanderbilt University School of Medicine, Nashville, TN; 8) Department of Laboratory Medicine, University of Washington, Seattle, WA; 9) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 10) Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD.

Background Infobuttons – context-sensitive links embedded in electronic health record (EHR) systems or patient portals – can connect providers and patients to more relevant information resources for complex medical topics, including genomic medicine. Little is known about individual and institutional perspectives on infobutton implementation. **Methods** A survey was developed by the electronic Medical Records and Genomics (eMERGE) consortium's Infobutton sub-group to assess perceptions about infobuttons, the OpenInfobutton software (OIB), ClinGen EHR toolkit (CG), and two eMERGE-developed systems (DocUBuild [DUB] and MyResults.org [MR]). The survey was distributed to several major consortium-driven initiatives that are focused on implementing genomic medicine: ClinGen, CPIC, CSER, DIGITizE, eMERGE, GA4GH, IGNITE, and ISCC. **Results** In total, 25 unique responses from 19 institutions were received. The majority of respondents did not have infobuttons enabled for genomic content at their institution (n=16, 64%), and did not have or were not aware of plans to expand their availability (n=20, 80%). Respondents generally agreed they felt knowledgeable about infobuttons (n=17, 68%), but less so for specific software platforms (OIB: n=11, 44%; CG: n=5, 20%; DUB: n=8, 32%; MR: n=9, 36%). Regarding experiences and opinions about specific software systems, responses in the survey were optional, and 24 respondents completed at least one question about their experiences with and opinions about OIB, 20 about CG, 19 about DUB and 19 about MR. For two software systems (CG, DUB), respondents were mixed neutral (NT) and in agreement (AG) on the impact the systems would have on improving infobutton adoption (CG: NT=8/20 [40%], AG=9/20 [45%]; DUB: NT=9/19 [47.4%], AG=7/19 [36.8%]). For these systems, the largest perceived adoption barrier was lack of adaptability to local needs (CG=8, DUB=6, MR=6), with other perceived barriers including system complexity, and lack of ability to conduct a small-scale trial. **Conclusion** Institutions involved in genomic medicine initiatives are not widely using infobuttons for genetic results, but do anticipate benefits from recent projects focused on infobutton implementation and adoption. More importantly, perceived barriers are noted toward adoption, which system developers may proactively address.

578T

Evaluating nonresponse bias in a longitudinal study of healthy adults receiving genome sequencing. *S.D. Crawford[1], M.D. Linderman[2,9], M. Ball[3], M.A. McGinniss[4], J. Esposito[4], E. Ramos[4], R.H. Young[1], R.C. Green[5,6,7,8], PeopleSeq Consortium.* 1) SoundRocket, Ann Arbor, MI; 2) Department of Computer Science, Middlebury College, Middlebury, VT; 3) Open Humans Foundation, Boston, MA; 4) Illumina, Inc., San Diego, CA; 5) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 6) Broad Institute of MIT and Harvard, Cambridge, MA; 7) Harvard Medical School, Boston, MA; 8) Partners HealthCare Personalized Medicine, Cambridge, MA; 9) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

The PeopleSeq Consortium is a multi-site longitudinal study designed to measure medical, behavioral and economic outcomes after elective sequencing of ostensibly healthy individuals. The study design consists of a pre-test survey (fielded prior to sequencing report disclosure), a post-test survey (fielded approximately three months following report disclosure), and annual follow-up surveys thereafter. Participants are recruited via email invitations, and nonresponders are reminded to participate via a series of follow-up emails. In an ongoing recruitment, approximately 1,437 individuals have been invited to participate. A[proximately 513 have participated, giving us a response rate of ~36%. Nonresponse is only problematic if those who do not respond are different than those who do respond, resulting in a nonresponse induced bias. Current best practices in survey research suggest that response rates are not a good proxy for data quality, and that other information is needed to evaluate the potential for nonresponse bias at the case or item level. Individuals who undergo personal genome sequencing at one of the collaborating sites are eligible for inclusion in the PeopleSeq Consortium so the study is in a unique position to know a little bit about all eligible participants. As such, aggregate and individual data, including basic demographics and other measures of interest in evaluating possible nonresponse bias, can be obtained about both nonresponders and responders. In our study, these data include gender/sex, race/ethnicity, age, education, adoption status, and employment field. By analyzing the differences between responders and nonresponders on these measures, we can build an understanding for how nonresponse may or may not be introducing a bias in our study. These data will be the first data on nonresponse in a study of outcomes from personal genome sequencing of ostensibly healthy individuals who have had healthy sequencing. Understanding the potential for bias in this way also enables us to assess the overall quality of the data that is collected. We will discuss the likely impact response rate increasing efforts may have on any possible bias we detect.

579F

MyCode participants' research priorities for precision health. *E. Huang[1], A.K. Rahm[2], C. Flansburg[2], J.K. Wagner[2].* 1) Lafayette College, Easton, PA; 2) Geisinger Health System, Danville, PA.

   Introduction: Participant engagement has previously been identified as a key component to long-term success of precision health initiatives, whether such initiatives are at a national scale such as the All of Us Research Program or a regional scale such as the MyCode Community Health Initiative. Building communitas, including participants in research design and implementation decisions, providing access to information, and reporting findings from such research and innovation projects foster durable, trusted relationships and shared investment in our biomedical future. **Methods:** An online survey was administered to MyCode participants to elicit their input on environmental data to be integrated with genomic and electronic health record data in precision health projects conducted at Geisinger Health System. The survey collector was open from June 7 to July 31, 2016, and a total of 77,629 participants were invited to the survey (62,629 using the MyGeisinger patient portal were recruited online and a random sample of 15,000 not using MyGeisinger were recruited via postcard). Content analysis was performed on free-text responses submitted. **Results:** A total of 6,091 participants responded (7.8% response rate) and submitted 9,815 free-text responses to a question eliciting suggestions for health outcomes, medical conditions, or topics for prioritized study. Respondents emphasized the study of causes of disease and health conditions and not just their cures; treatments and diagnostics; or prevention. While cancers were most prominent among specific conditions, the most popular medical specialties covering respondents' suggestions were, in ranked order, neurology, gastroenterology, and immunology. Environmental aspects suggested focused on water quality, fracking impacts, air quality, genetically modified organisms, and exposures to asbestos, radon, and lead. Respondents suggested a few interesting unexpected topics for study, including the health effects of technology usage. **Conclusions:** Participant engagement requires considerable institutional commitment, dedicated resources, and recognition of the value in shared decision-making. Empowering participants to shape research priorities is one way to remind participants of their enrollment in longitudinal projects, build mutual respect, manage expectations for progress, and promote shared scientific vision. Next steps are to integrate these results into future plans for MyCode.

580W

A conceptual framework for genome sequencing: Adaptation of the Theory of Planned Behavior. *L. Hull[1], J. Vassy[1,2,3].* 1) VA Boston Healthcare System, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Brigham and Women's Hospital, Boston, MA.

   The outcomes of whole-genome or whole-exome sequencing depend on the quality of the decision-making process; these outcomes may include patient understanding of sequencing results, psychological impact of results, changes in behavior in response to sequencing, healthcare utilization and cost, and decisional satisfaction and regret. The decision to pursue sequencing is complicated given that the results may be inconclusive or contain additional information unrelated to the original reason for sequencing. Therefore, true informed consent for sequencing depends on a patient's understanding of its risks and benefits. Many studies have explored the motivations and concerns of patients undergoing sequencing; however, a framework to approach how patients assess the risks and benefits of sequencing, and ultimately decide whether and how they pursue sequencing, has not been formally described. We performed a narrative review in which we synthesized the themes elicited in previous work regarding the decision to pursue or decline sequencing, mapping these to the Theory of Planned Behavior (TPB) model. We found several descriptive, qualitative, mixed-methods, and survey-based studies that have evaluated the motivations and concerns of individuals pursuing sequencing; however, only one study specifically focused on the concerns of sequencing study decliners. Participants' motivations and concerns have stemmed from personal, altruistic, and family-related considerations. Trust in providers and research institutions were high among individuals participating in studies, but may not be representative of the larger population. Intention to pursue sequencing was high and did not change after informed consent among well-educated study enrollees. We conclude that further study is needed of more diverse cohorts of individuals both pursuing and declining sequencing, to better understand their decision-making process. Additional areas of study could include studying trust and barriers to participating in sequencing, especially among sequencing decliners. The effects of informed consent as a possible modifier of initial intention to pursue sequencing should be evaluated in populations with lower baseline enthusiasm or lower education levels. Given, these findings, we propose a theoretical framework, derived from the TPB, to conceptualize how patients approach the decision to pursue or decline sequencing; this framework helps to highlight areas for future investigation.

581T

Special considerations for genomic sequencing research in the United States Air Force: Development of the MilSeq Project. *M.D. Maxwell[1], J. Vassy[1], C. Gardner[4], C. Blout[1], S. Pereira[5], M. Mehlman[9,10], E. Parasidis[11], K. Christensen[1], J.O. Robinson[5], A. McGuire[5], M. Lebo[6,7,8], M. De Castro[2], R. Brenner[3], B. Morgan[4], R. Green[1,6,8,12].* 1) Department of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Air Force Medical Genetics Center and Molecular Diagnostic Laboratory, Keesler AFB, MS; 3) Air Force Medical Support Agency, Falls Church, VA; 4) 59th Medical Wing, Joint Base San Antonio – Lackland AFB, TX; 5) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 6) Harvard Medical School, Boston, MA; 7) Department of Pathology, Brigham and Women's Hospital, Boston MA; 8) Partners Personalized Medicine, Boston, MA; 9) The Law-Medicine Center, Case Western Reserve University School of Law, Cleveland, OH; 10) Biomedical Ethics, Case Western Reserve University School of Medicine, Cleveland, OH; 11) Moritz College of Law and College of Public Health, Ohio State University, Columbus, OH; 12) The Broad Institute of MIT and Harvard, Cambridge, MA.

Purpose The Genetic Information Nondiscrimination Act (GINA) of 2008 protects against health insurance and employment discrimination among civilians who undergo genomic sequencing (GS), but does not apply to active-duty military personnel. Although the military routinely performs select carrier screenings and incorporates protections afforded by the National Defense Authorization Act (NDAA) of 2008, we aim to explore a range of considerations unique to GS in the military. **Methods** In collaboration with United States Air Force (USAF), we have developed a pilot research protocol, the MilSeq Project, in which GS will be performed on 75 ostensibly healthy, active-duty Airmen. Pathogenic and likely pathogenic variants as well as autosomal recessive carrier status will be reported to Airmen participants by USAF healthcare providers who have received brief genomics training and are supported by a genetic counselor. Relevant pharmacogenomic variants and disease-associated risk alleles will also be reported. Reports will be permanently incorporated into the Airmen's military medical record after disclosing results. To maximize the safety of the research protocol, we examined the applicability of GINA, NDAA, other federal regulations, and basic human subject research rights to the military population. **Results** Special considerations for genomic research in the military include: 1) institutional considerations, as the military operates within a hierarchical rank structure which can influence a service member's voluntary participation in research; 2) the potential for adverse career outcomes, as medical records can be reviewed when evaluating assignments and promotions, with injury prevention and duty disruption as primary concerns. Therefore, Airmen in certain military occupational specialties may be deemed unfit for duty or discharged if determined to be at risk or symptomatic of certain genetic conditions. However, based on current literature, a variant that could impact an Airman's career is expected to be found in less than 1% of individuals who receive GS. These considerations are clearly delineated in the MilSeq Project informed consent process. **Conclusion** The implementation of GS in the military includes considerations that are distinct from a civilian cohort and therefore data are not directly comparable. We have developed a protocol for the MilSeq Project that is sensitively designed to begin collecting reliable outcome data regarding the use of GS in the military.

582F

"It's not an easy language for me to understand": Australians' expectations of support from health professionals for interpretation of personal genomic test results. *S. Metcalfe[1,2], C. Hickerton[1], R. Tytherleigh[1], J. Savard[3], E. Tutty[1], C. Gaff[4], K. Gray[5], A. Middleton[6], A. Newson[4], A. Nisselle[1,2], E. Stackpoole[1], B. Terrill[7], E. Turbitt[1], B. Wilson[8].* 1) Genetics Education and Health Research and Dept Paediatrics, Murdoch Childrens Research Institute and The University of Melbourne, Parkville, Melbourne, Victoria, Australia; 2) Australian Genomics Health Alliance; 3) The University of Sydney, Sydney, NSW, Australia; 4) The Walter & Eliza Hall Institute and Melbourne Genomics Health Alliance, Parkville, Melbourne, Victoria, Australia; 5) The University of Melbourne, Parkville, Melbourne, Victoria, Australia; 6) Wellcome Trust Sanger Institute, Hinxton, UK; 7) Garvan Institute of Medical Research, Sydney, NSW, Australia; 8) University of Ottawa, Ottawa, Ontario, Canada.

Background. Australians can order personal genomic tests (PGT) directly from companies online, via medical professionals or alternative health practitioners. PGT companies claim to provide information about an individual's health (including nutrition/'genomic wellness'/fitness) and/or ancestry. Surveys of Australian genetic health professionals (GHPs) have shown an increase from 11.3% (2011) to 51.1% (2017) seeing clients about PGT, raising concerns about the impact on publicly-funded health services. Many are not comfortable discussing PGT results and some services now refuse to do this. **Aim.** To explore Australians' interest in PGT and their expectations of support for result interpretation. **Methods.** The online GeNIOz survey was available to the public from April 2016 to May 2017. Quantitative data were analysed descriptively in STATA. Qualitative interviews were also conducted and coded independently using thematic analysis. **Results.** 2,806 Australians attempted and 2,393 fully completed the survey. 22.5% (572/2,538) had genetic testing, with 15.1% having had a medical/health-related PGT, while another 32.0% considered such a test, and 10.7% had ancestry testing. All respondents were asked about whom they would approach to help them understand their health-related test results: 77.7% selected general practitioners (GPs), 57.4% selected GHPs, and 7.5% selected alternative health practitioners. Furthermore, 55.8% selected GPs and 44.2% selected GHPs, even for non-health related tests. Some interviewees who had ancestry testing discussed downloading their raw data to conduct further analysis using online programs or companies, including to obtain health information, occasionally taking these results to a health practitioner for interpretation. Other interviewees had *MTHFR* and other 'genomic wellness' tests and some turned to alternative health practitioners for support. **Conclusion.** In our sample, 47.1% Australians have either had or are interested in having a personal genomic test for health-related reasons, including testing for 'genomic wellness', as marketed in Australia. Their views regarding from whom they would seek support to interpret these tests have workforce implications, not only for genetic health services which are traditionally publicly funded with a medical focus, but also for GPs and alternative health practitioners (naturopaths/nutritionists) who often have limited understanding of these types of tests and their clinical significance.

583W

The duty to warn at-risk relatives: The experience of genetic counselors and medical geneticists. *T.J. Perry, M.B Farmer, C.B. Hurst, G. McGwin, N.H. Robin.* University of Alabama at Birmingham, Birmingham, AL.

Most genetic counselors and medical geneticists will experience patient refusal to warn an at-risk relative. When faced with such a situation, conflicting laws, professional guidelines, and ethical obligations may lead to confusion regarding whether a clinician has a duty to warn their patient's relatives of genetic risk. Dugan, et al. (2003) and Falk, et al. (2003) surveyed genetic counselors and medical geneticists, respectively, to examine their clinical experiences with the conflict of duty to warn versus patient confidentiality. This issue has become increasingly relevant as the actionability and breadth of genetic test results has significantly increased since 2003, and pertinent federal laws have been implemented which did not exist at the time of those past studies. Using a modified and merged version of the Dugan and Falk surveys, we sought to understand if interim changes in the field of medical genetics have led to shifts in how clinicians are experiencing, and therefore reacting to, patient refusal to inform at-risk relatives. Compared to the majority of clinicians surveyed by Dugan and Falk, only 30% (77/258) of medical geneticists and genetic counselors believed they had a duty to warn their patients' relatives of genetic risk. However, neither profession was less likely to seriously consider notifying at-risk relatives without patient consent than in the previous studies. Of those who considered warning an at-risk relative, two genetic counselors and no medical geneticists actually proceeded with notification of an at-risk relative without patient consent. Only 8% (21/259) of participants believed professional guidelines effectively address the issue of duty to warn. Of those who felt current guidelines do not address the issue, 60% were not aware of relevant American Society of Human Genetics guidelines. Similar to the results of Dugan and Falk, the majority of medical geneticists (59% 30/51) and almost half of genetic counselors (46%, 95/206) experienced patient refusal. Yet awareness of federal and state laws which regulate the disclosure of genetic information to at-risk relatives remains surprisingly low. These results suggest that because the conflict of duty to warn remains a common experience among genetics professionals, educational resources are needed to facilitate informed decision making. Furthermore, participants' opinions of current policies and clinical decisions may guide future policies and professional positions.

584T

A legal analysis of the loss of chance doctrine in the context of genomics and precision medicine. *J. Wagner.* Geisinger Health System, Danville, PA.

Genetic professionals have long been concerned about legal liability, including liability risks that might attach when decisions are made to return or alternatively withhold data or results from individuals to whom those pertain. Professional organizations have statements and recommendations regarding what to return and when. Labs do not routinely reanalyze data or recontact individuals with updated information as genomic science and technology advances, and there is general consensus that researchers do not have duties to look for or report all findings. Nevertheless, when broad genomic research is conducted where participants are also patients and when institutional policy or research protocol stipulates data or results will be withheld (regardless of rationale), there are concerns about when and how individuals might ultimately learn this information and whether the discovery will be too late to avoid (1) the progression of a condition for which prevention or treatment was available or (2) unnecessary harms such as ineffective treatments for which substitutes were available. The loss of chance doctrine is a legal theory available in some states that, in the context of medical malpractice but not research, enables a plaintiff to recover from a defendant whose breach of duty reduced the chance of a favorable outcome (such as a delayed diagnosis or treatment resulting in diminished chance of recovery or survival). There is a paucity of critical analysis of this legal doctrine, despite acknowledgement by some of its relevance to the genetics field. Even the American Law Institute has not yet taken a position despite the doctrine's first court appearance 50 years ago. A systematic search of cases invoking the loss of chance doctrine was performed using Thomson Reuters Westlaw in Spring 2017. Cases identified were briefed using standard legal research methodologies (i.e., analysis of parties, facts, history, holding, principles, and rationales). 862 cases were identified, including 681 state and 181 federal and including 70 in PA, NJ, & DE (states in US 3rd Cir). This legal doctrine is poorly understood and manifested differently across jurisdictions within the medical malpractice context. In light of the legal uncertainty and given the perceived burdens making ongoing reanalysis and recontact of all participants infeasible, adoption of policies to provide prompt and broad data access to patient-participants as a risk mitigation strategy should be considered.

585F

Medical relevant findings from whole exome sequencing for apparently healthy individuals in physical examination of Chinese people. *Z. Xia[1], S. Han[1], D. Chen[1], W. Yuan[1], N. Li[1], S. Wang[1], M. Qi[1,2,3].* 1) School of Medicine, Zhejiang University, Hangzhou, Zhejiang province, China; 2) James Watson Institute of Genome Sciences; 3) Department of pathologyand laboratory Medicine, University of Rochester, NY, USA.

   Background: Whole exome sequencing (WES) has been proved to be a very powerful approach to identify new disease genes and molecular diagnosis of unknown clinical cases. However, limited studies reported on if this approach could provide useful information and should be offered to the apparently healthy people. **Methods**: WES was performed on 2 patients with clinical findings and 29 self-claimed 'healthy' individuals at an Illumina HiSeq platform. 1493 genes were analyzed for 303 monogenic disorders including inherited cancers, cardiac arrhythmias, high prevalent genetic diseases, inborn errors of metabolism, age-related diseases, neuromuscular diseases, skin disorders, and eye diseases. Variants of pathogenic, likely pathogenic as well as variants of unknown significance (VUS) were reported according to the current ACMG guidelines. Genetic counseling was provided. **Results**: One pathogenic mutation consistent with the clinical diagnosis, a VUS matched to family history, and 112 susceptible variants were identified and reported. Seventy-five pathogenic or likely pathogenic (P/LP) variants in 56 genes occurred 102 times in 30 individuals who might or would be affected. Thirty-six P/LP variants appeared 45 times (34 gene, 22 individuals) in autosomal recessive carriers. One X-linked recessive pathogenic variant appeared in 2 unrelated female carriers. **Conclusions**: This study suggests that WES for apparently healthy people, will enable effectively physicians to take appropriate measures for disease diagnosis, prevention and treatment in their patients. It will also help these individuals to prevent the genetic disease passed down to their offspring.

586W

How Filipino parents cope with having a child with Maple Syrup Urine Disease. *M.R. Tumulak[1], C.D. Padilla[1], M.Y. Laurino[1], E.S. Regalado[2], A.V. Legaspi[3], E.R. Ventura[4].* 1) University of the Philippines-Manila, Manila, Philippines; 2) University of Texas Health Science Center, USA; 3) Bow Valley College, Canada; 4) University of the Philippines-Diliman, Quezon City, Philippines.

   Introduction: In the Philippines, maple syrup urine disease (MSUD) is the most common inborn error of metabolism (IEM). A family may experience stress, anxiety, sorrow, or feelings of helplessness when a child is diagnosed to have a genetic disorder that may lead to chronic care and disability. An illness in a child affects not only the child, but the whole family as well. Methods: The phenomenological approach was used in this study, in order to capture the coping experience of the participants. A total of 12 parents were interviewed using a semi-structured set of questions.   Result: Knowledge of MSUD, the coping strategies used and their advice for other parents were considered to be part of their coping mechanism in the diagnosis of the disease in their child. The families used different coping strategies, such as information seeking, acceptance of their fate, and strengthening their belief in God, for them to be able to adapt to their situation.   Conclusion: The findings of this study describe the different coping processes used by the families. It will help form policies and guidelines for genetic counseling of MSUD patients and their families in the Philippines. This aims to assist families in coping with the diagnosis and improve outcomes of affected children.

587T

Experience and specialty impact processes for interpretation and clinical application of genetic test results. *C. Berrios[1], E. Hurley[2], I. Thiffault[1], L. Zellmer[1], C. Saunders[1], S. Soden[1], K. Goggin[2], L. Willig[1], E. Farrow[1]. 1) Center for Pediatric Genomic Medicine, Children's Mercy, Kansas City, MO; 2) Division of Health Services and Outcomes Research, Children's Mercy, Kansas City, MO.*

Significant efforts have been made in recent years to standardize laboratory classification of genetic variants, including the publication of guidelines by professional societies. However, the impact of genetic testing on patient care ultimately depends on the clinician's interpretation of results. Clinician interpretation and application of results can be highly variable. To date, little is known about clinicians' cognitive processes and behaviors that may explain this variability. This study explores how clinicians across diverse specialties review and interpret genetic test results. We surveyed clinicians who return test results to patients/families regarding frequency of returning results, investigation of differing variant classifications, and resources used. A subset of participants completed in-depth interviews about their approach to reviewing lab reports, interpreting variants, applying findings, and communicating results to patients/families. Of the 93 survey respondents, 33% returned genetic results > 3x/month and 46% did so < 1x/month. Behaviors in investigating results varied, but for variants of uncertain significance (VUS) a majority "sometimes/frequently/always" sought laboratory or genetic specialist consultation (80.7%) and conducted additional research (74.1%). Clinicians who less frequently returned results were more likely to seek consultation ($\chi^2(2)=6.87$, $p=0.032$) or conduct additional research ($\chi^2(2)=10.64$, $p=0.005$) on pathogenic variants. All participants were equally likely to use PubMed and OMIM, with a minority accessing other resources. Those who frequently returned results were more likely to use HGMD ($\chi^2(2)= 9.18$, $p=0.010$) and ClinVar ($\chi^2(2)=25.15$, $p<0.001$). Follow-up interviews (n=24) corroborate survey findings regarding frequent investigation of VUS and differences in resources used. Providers with specialty or significant experience in genetics referenced more detailed discussions of VUS with patients and, in some cases, reinterpretation of a laboratory variant classification for clinical care. Reinterpretation was influenced by representative and availability heuristics, trust in the laboratory and institutional resources, and confidence in clinical expertise. Our results reveal divergent processes in clinician interpretation of genetics results based on specialty and experience. Further study of factors influencing clinical variant interpretation are needed to improve standardization of genomic medicine.

588F

A comparison of international policies on CRISPR and gene modification technologies and the risk of lag in Canadian science. *Z. Master[1], P. Bedford[2]. 1) Alden March Bioethics Institute, Albany Medical College, Albany, NY; 2) Centre for Commercialization of Regenerative Medicine, Toronto, ON.*

CRISPR/Cas gene editing technology has been touted to have widespread applications to human health because of its accuracy in genome modification and ease of use. While many scientific societies and national and international organizations have drafted statements and policies in an attempt to promote research, the situation remains unclear on the permissibility of germline gene modification research using CRISPR under Canadian legislation. Compared with the U.S. and U.K., Canada's law on genome modification may be considered restrictive. Under appropriate ethical oversight, research involving germ cell gene editing is authorized by the European Academies Science Advisory Council and the National Academies of Sciences, Engineering, and Medicine. A group of Canadian scholars contend that the Assisted Human Reproduction Act (AHRA) is ambiguous, and have interpreted the legislation to prohibit the use of CRISPR technology on human embryos. We interpret the policy to permit the use of CRISPR under specific conditions, but agree that clarification from Health Canada is required before moving forward. Based on our analysis of the AHRA, Parliamentary debates and previous policy reports, we argue that CRISPR research might be permissible using surplus embryos where appropriate evidence demonstrates negligible likelihood of transmission to descendants or when researchers do not intend to transfer embryos to induce pregnancy. Canadian legislation contains a germline prohibition focused on the transmission of genes to future descendants. An examination of the legislative history shows that the intent behind the prohibition was to prevent genetically modified children and concerns at the time did not focus on research using surplus embryos. We also contend that CRISPR technology may be allowed in cases where gametes used to create embryos focus on studies intended to improve assisted reproductive technologies as Canadian legislation permits the creation of embryos for such purposes. Aligned with some national policies, Canada could permit germline editing under appropriate ethical safeguards while prohibiting germline modification to future generations. Clear direction is required from Health Canada on interpreting the legislation as international trends will compel Canada to make more permissible legislative amendments or risk a lag in Canadian science and innovation regarding the use of gene modifying technologies.

589W

What determines value for genomics-informed healthcare? Evidence from a discrete choice experiment. *D. Regier[1,2], D. Veenstra[3], A. Basu[3], S. Jagori[3], J. Carlson[3].* 1) ARCC-CCR, BC Cancer Agency, Vancouver, British Columbia, Canada; 2) School of Population and Public Health, University of British Columbia; 3) Pharmaceutical Outcomes Research and Policy Program, Department of Pharmacy, University of Washington.

   Introduction: The uptake of precision medicine has been varied. Evidence that enumerates individuals' preferences is needed to inform the worth of precision medicine interventions and future research agendas. **Objective:** To determine the impact of genomic information, scientific uncertainties, cost, and health outcomes on preference-based utility and willingness to pay (WTP) for genomics-informed healthcare. **Methods:** We used a sequential mixed methods approach. Qualitative research with patients identified the following attributes as important: type of genetic test, penetrance, medical expert agreement on changing medical care based on genomic information, quality of life, life expectancy with statistical uncertainty on outcomes, and cost. The discrete choice experiment method was used to enumerate preference-based utility and value. The target population was a sample of the American lay public. Analyses allowed for preference heterogeneity to characterize individual-level utility. **Results:** The lay public (n=101) stated strong positive utility for increasing levels of medical expert agreement and moderate disutility for quantity of life gains with a high degree of statistical uncertainty. Average WTP to receive genomics-guided care where identified disorders had high-penetrance, a high level of medical expert agreement, improved quality of life, and improved quantity of life by 3 months with a high degree of statistical certainty was $1700 (95% CI, $1029,2371). WTP decreased to $1231 (95% CI 532,1929) when the uncertainty around quantity of life gains significantly increased. **Discussion**: The value of precision medicine technology to the public is significantly influenced by medical expert agreement and the degree of scientific uncertainty surrounding health outcomes. These findings can be used to project the economic value of and demand for genomics-guided health care, as well as inform a translational genomics research agenda.

590T

Measuring health outcomes in telegenetics. *J. Stock[1], D. Maeise[2], S. Mann[3], M. Lyon[2], A. Keehn[2], D. Flannery[2], M. Watson[2].* 1) Seattle Children's , Seattle, WA.USA; 2) American College of Medical Genetics and Genomics, Bethesda, MD, USA; 3) Hawaii Department of Health, Honolulu, HI, USA.

   Purpose To improve access to genetic services, the National Coordinating Center (NCC) for the seven Regional Genetic Networks (RGNs) supports the use of telegenetics – genetic services delivered through interactive video conferencing. In 2017, work began on a menu of health outcome measures for genetic services. The goal is to standardize the collection of data about genetic services; to describe the value of genetic services to insurers, other funding sources, and healthcare decision-makers, as well as families; and to assess the similarities or differences in patient health when genetic services are provided via interactive videoconferencing or as in-person visits. This performance measurement work complements the current NCC/RGN set of common evaluation data elements to assess telegenetic visit characteristics; provider and family satisfaction; provider and family time and travel costs; and family experience. **Methods** The NCC Telegenetics Workgroup developed a *Framework for Health Outcomes Measurement in TeleGenetics* that includes descriptions of telemedicine models, genetic services interventions, and outcome types (individual patient, family, health system). A literature review was conducted by Seattle Children's medical librarian using Medline and PsychInfo from 2006 forward. The search found 78 abstracts that measured outcomes after the provision of interactive video conferencing for health purposes between a provider and patient. Abstracts which measured only satisfaction were excluded. Two surveys conducted by the NCC found low rates of interactive video conferencing utilization by genetic professionals (18%-35.7%). **Results** With too few clinical geneticists and genetic counselors; geographic disparities in the distribution in genetic professionals; and patient distance to genetic services, telegenetics is a potential solution to current access barriers. Performance measures that encompass prenatal counseling, cancer genetics, newborn screening follow-up and disease management are needed to evaluate changes in access to services for patients with genetic conditions, especially the medically underserved. Through technical assistance to promote the use of telemedicine, the NCC seeks input from genetic specialists and affected individuals and their families to make for meaningful outcomes measures. The NCC at the American College of Medical Genetics and Genomics is funded by the Health Resources and Services Administration *UH9MC30770-01-00.*.

591F

Reproductive decision making process, medical concerns and special needs of preimplantation genetic diagnosis (PGD) users: Lessons from qualitative and quantitative analysis. *S. Zuckerman[1,2], S. Gooldin[3], B. Savitsky[4], G. Altarescu[1,2].* 1) Shaare Zedek Medical Center, Jerusalem, Israel; 2) Hebrew University - Hadassah Medical School, Jerusalem, Israel; 3) Hebrew University - The Department of Sociology and Anthropology, Jerusalem, Israel; 4) Hebrew University- The Braun School of Public Health, Hadassah Medical Center, Jerusalem, Israel.

   PGD is an accepted procedure for prenatal diagnosis (PND), worldwide. It may eliminate some of the obstacles related to conservative options of PND, but can raise personal, social and moral questions. The psychosocial aspects of the technology have been discussed among experts and were evaluated among various population groups. Yet, only scant empirical data focused on PGD users' experiences. A combined methodology was used to investigate PGD's implications on its users: qualitative analysis of semi-structured in-depth face-to-face interviews with 43 PGD users, representing variety of population's sub-groups. On the basis of the interviews, a detailed closed web-based questionnaire was developed. Univariate and multivariate adjustment was performed on data obtained from 155 PGD users in different stages of the procedure**.** PGD is considered a preferable diagnostic procedure for 139 (95%) subjects. Nevertheless, 71 (47 %) reported a complex decision-making process. Perceived advantages are: assurance of the embryo's unaffected status from the beginning of the pregnancy (48%), thus avoiding the need for pregnancy termination (33%) and invasive prenatal tests (20%). Perceived disadvantages focused on the medical actions involved (63%) and the delay in time between the first counseling and the PGD procedure itself (52%). Other future needs included improving the communication with medical staff (47%) and implementing emotional support (29%). The study indicates special needs of respondents from minority groups and distinct genetic, as well as socioeconomic, backgrounds. Subjects introduce permissive moral stand regarding different PGD uses but made a clear distinction between medically justified applications to unjustified social and aesthetic uses. This interdisciplinary qualitative and quantitative analysis of a large sample of PGD users will deepen the insights of the special needs of this particular group and can form a basis for development of counseling and guidance programs of future PGD users. Furthermore, it will also help foster a public debate concerning medical, ethical, sociological and economic aspects of the technology.

592W

Next-generation sequencing experience: Impact of early diagnosis of Usher syndrome. *C. Wright[1], A. Brown[1], M. Albert, Jr.[3], A. Mukherjee[4], G. McGwin[4], C. Hurst[1], J. Smith[2], N.H. Robin[1].* 1) Department of Genetics, University of Alabama at Birmingham, Birmingham , AL; 2) Hearing and Speech Dept, Children's of Alabama Birmingham, AL; 3) Retina Consultants of Alabama Birmingham, AL; 4) Department of Biostatistics, University of Alabama at Birmingham, AL.

   Purpose: The aim of this study was to assess the parental psychosocial implications, such as emotions and coping, of earlier diagnosis of USH via genetic testing compared to parents of children who were diagnosed later via ophthalmologic findings. **Method**: Thirty-six participants were recruited through an online posting on the Usher Syndrome Coalition website. Two comparison groups were formed based on the method of diagnosis (i.e. genetic diagnosis vs. ophthalmologic diagnosis). Semi-structured interviews were recorded and transcribed. Comparison, using thematic and statistical analysis, of psychosocial impact on parents of children diagnosed early (genetic testing) and later (ophthalmologic findings) was completed. **Results**: There were no statistically significant differences in emotions between the two groups of participants, suggesting that earlier diagnosis via genetic testing does not lead to increased anxiety or psychosocial issues for parents. Additional themes identified from parent interviews and their application to patient care are described. **Conclusions**: Earlier diagnosis of Usher syndrome via genetic testing does not cause a more harmful emotional impact than later diagnosis via ophthalmologic findings. In fact, there are multiple benefits to earlier diagnosis via genetic testing .Earlier diagnosis allows parents to emotionally process and prepare the child for independence throughout life.

593T

Targeted genomic screening in unselected adults. *K.R. Muessig[1], K. Kuczynski[2], P. Himes[1], M. Gilmore[1], S. Westaway[1], B.C. Powell[2], J.S. Berg[2], J. Cadigan[2], K. Weck[2], J. O'Daniel[2], K. Foreman[2], K.AB. Goddard[1], J.P. Evans[2], G.E. Henderson[2].* 1) Center for Health Research, Kaiser Permanente NW Portland, OR; 2) University of North Carolina at Chapel Hill, Chapel Hill, NC.

Background: Targeted genomic screening in unselected adults has not been widely studied. 1-2% of adults in the US unknowingly carry pathogenic, highly penetrant variants that confer high risk for severe but preventable disease. Before we implement screening for adults, we must answer questions about the true prevalence of genetic conditions in unselected populations, acceptance of genetic counseling, penetrance of conditions ascertained by screening, and adherence to recommended clinical follow up. Methods: We recruited individuals at a research biobank at Kaiser Permanente Northwest [KPNW] (N=650) and a primary care practice at UNC-Chapel Hill (N=436). The study used online consent, survey, and interview procedures. Joiners agreed to genomic screening for 17 genes related to 11 medically actionable conditions. 262 of those approached were enrolled. Joiners who screened positive were offered their choice of an in-person or phone appointment with a genetic counselor to disclose results. During results disclosure, the genetic counselor recommended clinical follow up actions, and placed the result in the electronic medical record. Results: The GeneScreen test identified 15 positive screening results; 14 were CLIA-confirmed; 1 was not confirmed per participant request. Confirmed results included pathogenic or likely pathogenic variants for 6 conditions including variants in *HFE* (N=7, 2 C282Y homozygotes, 5 C282Y/H63D); *KCNQ1* (N=2), *BRCA2* (N=1), *LDLR* (N=2), *RET* (N=1), and *RYR1* (N=1). The 14 participants with CLIA-confirmed positive results (10 KPNW+ 4 UNC) were offered genetic counseling to learn their test result. Four attended a visit in-person and 10 requested a phone call. Nine participants declined referrals for clinical follow up with Genetics; however, of the 9, three wanted to follow up with their PCP, and one accepted a referral to cardiology. Among the 10 KPNW participants, several had prior knowledge about their condition (N=5); most preferred results over the phone (N=4). This was also true for participants who had signs or symptoms prior to joining (N=2). Conclusion: Although participants invited to join the study were not selected due to phenotypic characteristics, some self-selected on the basis of prior personal or family history of the targeted conditions. The results disclosure process and willingness for clinical follow up is influenced by pre-existing signs or symptoms of the condition or prior knowledge of the molecular diagnosis.

594F

Burkitt lymphoma and skeletal dysplasia. *S.L. Campbell[1,2], J. Flores[1,2,3], R. Antony[1,2,4], B.A. Speckhart[2], E. Bugaieski[1,2], K.S. Fernandez[1,2,5].* 1) Department of Pediatrics, University of Illinois College of Medicine at Peoria, Peoria, IL; 2) Children's Hospital of Illinois at OSF-Medical Center, Peoria, IL; 3) University of Utah School of Medicine, Pediatric Genetics Division, Salt Lake City, UT; 4) University of California Davis, Sacramento, CA; 5) Valley Children's Hospital, Madera, CA.

BACKGROUND: Cartilage-hair hypoplasia (CHH) is an autosomal recessive chondrodysplasia with short stature, sparse hair and defective cell-mediated immunity. It is caused by mutations in the ribonuclease mitochondrial RNA processing (*RMRP)* gene. Cancer incidence is 7-fold higher in patients with CHH than in the general population, especially non-Hodgkin lymphoma (NHL) and basal cell carcinoma in the third and fourth decade of life. **OBJECTIVE:** To report the association of skeletal dysplasia and Burkitt lymphoma in a young adult female. **DESIGN/METHOD:** Case Report. **RESULTS:** An 18-year old Amish female with disproportionate short stature presented to our center for management of Burkitt lymphoma. She had presented a week earlier with lymphadenopathy, fevers, night sweats, and weight loss for 2-4 weeks. A tonsillar biopsy was consistent with mature B-cell lymphoma. On exam, her height was three feet associated with spine involvement and brachydactyly, mild bowing of the legs, normal size head without frontal bossing, fine and sparse hair. She had normal intelligence. Her pattern of dysmorphisms was suggestive of CHH (parents declined genetic testing). She had extensive cervical, occipital, and submandibular lymphadenopathy and splenomegaly felt across the midline. Repeat biopsy confirmed Burkitt lymphoma with *Myc* rearrangement. PET-CT scan showed widespread disease involving cervical lymph nodes, spleen, iliac bone and bone marrow. Treatment with standard-intensity FAB/LMB therapy (group C) with the addition of Rituximab was initiated. She had an incomplete response to COP (~80% reduction of tumoral masses) but achieved complete remission after COPADAM1. She is 16-months off therapy without evidence of disease. **CONCLUSIONS:** While the association between CCH and NHL has been established we are unclear as to the specific mutation in our patient. Immunodeficiency related to *RMRP* mutation could have been the driver for tumorigenesis. On the other hand, fibroblast growth factor receptor 3 (*FGFR3*), a glycoprotein belonging to the tyrosine kinase receptor family, is a regulator of bone growth and gain of function germline mutations of the *FGFR3* gene result in congenital skeletal dysplasias such as hypochondroplasia/achondroplasia. *FGFR3* overexpression can be oncogenic and cooperate with *Myc* in B-cell lymphomas development. We hope to increase awareness of skeletal dysplasias and cancer predisposition in the pediatric population.

595W

RB1 **gene mutations in retinoblastoma patients from Central America.** *M. Dean[1], E. Torres Gonzalez[1], L. Garland[2], G. Bendfelt[1,3], V. Giron[4], C. Garrido[4], H. Lou[2], P. Valverde[4], M. Barnoya[2], M. Castellanos[4], G. Mejia[5], R. Ortiz[5], S. Luna-Fineman[4,6,7].* 1) Laboratory of Translational Genomics, NCI, Gaithersburg, MD; 2) Leidos Biomedical Research, Inc., National Laboratory for Cancer Research, Gaithersburg, MD; 3) Universidad de San Carlos Medical School, Guatemala City, Guatemala; 4) Unidad Nacional de Oncología Pediátrica, Guatemala City, Guatemala; 5) Hospital La Mascota, Managua, Nicaragua; 6) Stanford University, Stanford, CA; 7) Pediatric Hematology/Oncology/SCT/Cancer Bio, Stanford University School of Medicine, Stanford, CA.

Previous analysis of 327 consecutive cases at a pediatric referral hospital in Guatemala revealed that retinoblastoma accounts for 9.4% of all cancers and the estimated incidence is 7.0 cases/million children, higher than the United States or Europe. The number of familial cases is low, and there is a striking disparity in indigenous children due to late diagnosis, advanced disease, rapid progression and elevated mortality. Complete *RB1* sequencing of germline DNA from 73 retinoblastoma patients from Guatemala (n=66), Nicaragua (n=6), and Honduras (n=1) identified 21 patients with pathogenic mutations. The mutations were nearly all predicted null alleles, 43% nonsense, 29% frameshift and 24% splice site mutations. Twelve of the mutations have not been previously reported. Pathogenic mutations were identified in 75% (9/12) of the bilateral cases, versus 13% of the unilateral cases (X2=17, P=10E-5). One unilateral case was found to be mosaic (19%) for an R552X mutation, and another unilateral case with a c.55_56delGA mutation has a sibling with retinoblastoma. There was no significant difference in gender between cases with and without mutations, and 78% of the children with mutations in Guatemala were Mestizos versus 60% of WT subjects (not significant). The higher incidence of retinoblastoma in Guatemala is associated with a lower rate of bilateral disease (24%) than all US cases (31%) or US Hispanic cases (39%, X2=10, P=0.0016) in the SEER database. However, this could be related to a higher mortality (82%) in bilateral indigenous cases. There is no difference in the percentage of cases diagnosed before 1 year of age (51% in Guatemala vs. 50% in the US). An early diagnosis program could identify cases at an earlier age and improve the outcome of retinoblastoma in this diverse population, and a pilot educational poster program resulted in early diagnoses.

596T

Male breast cancer susceptibility due to FANCM mutation: A case report. *R. Janavicius[1,2].* 1) Vilnius University Hospital Santara Klinikos, Vilnius, Lithuania; 2) State Research Institute, Innovative Medicine Center.

Male breast cancer (MBC) is a rare, poorly understood disease. Germline susceptibility to MBC is substantial and is mostly confined to *BRCA2/BRCA1* genes mutations, however majority familial MBC has no known genetic basis. Recently, with the broader application of NGS, newer female breast cancer (FBC) susceptibility genes were established. Interestingly, a vast majority of FBC susceptibility genes are related to Fanconi anemia pathway. FANCM was recently identified as a moderate FBC predisposition gene, with the preponderance to triple negative BC. However, the contribution of these genes to MBC is still understudied. Case report. 66 yrs old male patient was referred for VULSK Oncogenetic Unit due to MBC diagnosed at the age of 65 (pT1cN0M0, ductal carcinoma G2, ER-98%, PR-90%, HER2(FISH)(–)). Family history showed prostate cancer in paternal uncle (dx 87). As part of routine comprehensive genetic testing, the patient underwent 96 cancer genes TruSight Cancer NGS testing panel on MiSeq (Illumina). Truncating FANCM (NM_020937) gene mutation c.5101C>T (p.Gln1701*) was identified; no other pathogenic/likely pathogenic genes mutations were revealed. During predictive testing, healthy daughter carrier (age 29) was identified. Discussion: From the newly identified FBC genes to date, only PALB2 was associated with possible MBC risk. No clear FANCM involvement in MBC susceptibility was demonstrated in published studies; only one MBC case (untested) was noted in a family with FANCM c.5791C>T mutation from Spain. Conclusion: FANCM c.5101C>T in MBC patient is reported for the first time, highlighting potential contribution of FANCM to MBC susceptibility. Funding: Lithuanian Research Council, SEN18/2015.

597F

Beyond the NMD boundary: Characterizing the phenotypes of C-terminal *CDH1* mutations. *K. Krempely, C. Espenschied, I. Lu, A. Chamberlin, R. Karam.* Ambry Genetics, Aliso Viejo, CA.

The nonsense-mediated mRNA decay (NMD) pathway is an mRNA surveillance system that degrades transcripts containing premature termination codons (PTC) located approximately 50 nucleotides upstream of the last exon-exon junction. Transcripts containing C-terminal PTCs located downstream the 50 nucleotide "NMD boundary" are not targeted for decay, potentially coding for truncated proteins retaining partial function. Therefore, one must be cautious when interpreting C-terminal truncating variants. *CDH1* truncating mutations cause Hereditary Diffuse Gastric & Lobular Breast Cancer (OMIM 137215). *CDH1* transcripts containing C-terminal PTCs downstream the NMD boundary have been shown experimentally to escape NMD; consequently, breast and gastric cancer risks associated with these truncations should be carefully evaluated. The aim of this study was to characterize and report *CDH1* C-terminal truncating alterations identified downstream of the NMD boundary in a large diagnostic laboratory cohort of ~400,000 *CDH1* alleles. Using molecular evidence, structural information, and clinical data, we characterized the most C-terminal *CDH1* pathogenic and likely pathogenic alterations in our cohort. The most C-terminal truncation we identified was p.E836* (c.2506G>T); it was confirmed *de novo* in an individual diagnosed with lobular breast cancer at 54 years of age who did not meet the International Gastric Cancer Linkage Consortium diagnostic criteria (IGCLC negative). Other truncations identified in this region include c.2398delC (p.R800Afs*16), c.2430delT (p.F810Lfs*6), and c.2490dupG (p.L831Afs*4), all seen in families meeting IGCLC criteria. We also identified, in an IGCLC negative family, a nonstop variant of unknown significance predicted to extend the protein by 29 amino acids (c.2647T>C (p.*883Qext*29)). Due to the uncertainty surrounding the functional impact of these alterations, clinical and molecular evidence proved essential in determining their pathogenicity. Characterization of the most C-terminal *CDH1* pathogenic truncation increases our understanding of the sequences located beyond the NMD boundary, resulting in improved classification of alterations in this clinically actionable gene.

598W

Prevalence of the *UGT1A1*28* polymorphism in a population-based African American breast cancer cohort: A pilot study. *A. Starlard-Davenport[1], A. Smith[1], G. Vidal[2,3], J. Cordero[1], R. Cordero[1], R. Blue[4], C. Simspon[1], A. Stanfill[4], B. Lyn-Cook[5].* 1) Department of Genetics, Genomics and Informatics, College of Medicine, University of Tennessee Health Sciences Center, Memphis, TN 38163; 2) Division of Hematology/Oncology, College of Medicine, University of Tennessee Health Science Center, Memphis, TN 38163; 3) The University of Tennessee West Cancer Center, Memphis, TN 38163; 4) Department of Acute and Tertiary Care, College of Nursing, University of Tennessee Health Science Center, Memphis, TN 38163; 5) Division of Biochemical Toxicology, FDA National Center for Toxicological Research, Jefferson, AR 72079.

   Background and Objectives: Inherited variations in UDP-glucuronosyltransferase (*UGT*) genes involved in estrogen metabolism are associated with an increased breast cancer risk in primarily non-Hispanic white and Chinese populations. UGT1A1 is a major enzyme that conjugates estrogens and their metabolites, thus preventing estrogen-mediated mitosis and mutagenesis. UGT1A1 is polymorphic. The *UGT1A1*28* promoter polymorphism is characterized by the presence of an additional TA repeat in the TATA sequence of the *UGT1A1* promoter, ((TA)7TAA, instead of (TA)6TAA). Homozygous ((TA)7/(TA)7) and heterozygous ((TA)6/(TA)7) *UGT1A1*28* genotypes are associated with reduced *UGT1A1* expression and transcriptional activity. The objective of this study was to assess the prevalence of *UGT1A1*28* polymorphisms in a population-based African American cohort of women with and without breast cancer. **Methods:** Saliva specimens were collected from African American women with a history of breast cancer (cases (n=70)) and controls (n=120) recruited through community-based participatory events and West Cancer Center in Memphis, TN between May 2016 to April 2017. Women age 18 and above were included in the study. Saliva was collected using Oragene®•OG-500 DNA Self-Collection kits. DNA was isolated using DNA Genotek's Prep-It L2P DNA isolation kit. The *UGT1A1*28* polymorphism was identified from PCR amplified and purified DNA. Purified PCR products (351 bp) were bidirectional sequenced using forward or reverse primers. *UGT1A1* genotypes were assigned based on the number of TA repeats for each allele (i.e., 6/6, 6/7, 7/7, or 7/8). This study was approved by the UTHSC Institutional Review Board. **Results:** Homozygous *UGT1A1*28* ((TA)7/(TA)7) genotypes were detected in 24% of women. Most women (41%) were heterozygous for the ((TA)6/(TA)7) genotypes while 20% of women had the wildtype *UGT1A1*28* ((TA)6/(TA)6) genotype (where no polymorphism was detected). Of women with breast cancer, 53% were heterozygous for the ((TA)6/(TA)7) genotypes whereas only 19% carried the wildtype genotype. **Conclusions:** The results suggest that homozygous and heterozygous *UGT1A1*28* promoter polymorphisms are prevalent among African American women. Recruitment of African American women to participate in our research study is ongoing to determine whether screening for the *UGT1A1*28* promoter polymorphism may be useful to identify women with reduced estrogen metabolism rates and greater susceptibility to breast cancer.

599T

Muir-Torre syndrome presenting with juvenile astrocytoma. *G.E. Tiller, S. Jeddi, J. Goff, E. Parkhurst.* Dept Genetics, Kaiser Permanente, Los Angeles, CA.

   Astrocytomas are primary brain tumors which may arise due to somatic or germline mutations in a number of genes, including p53, NF1, BRCA1, and the MMR genes associated with Lynch syndrome. We describe a 24 year old Hispanic male who was diagnosed with a frontal lobe astrocytoma (anaplastic ganglioglioma, WHO grade III) at 17 years of age, followed by the development of acute lymphoblastic leukemia at age 18 years. Family history was negative for central nervous system, endometrial and colorectal cancers, although his mother was diagnosed with non-Hodgkins lymphoma at age 59 years. A 20-gene cancer panel implemented on the patient revealed a germline MSH6 mutation (c.1168-1170delGATinsAA); immunohistochemical analysis of the brain tumor with anti-mismatch repair protein antibodies revealed absent staining for MSH6. Colonoscopy revealed no polyps. Parental testing revealed that the 74 year old father carried the mutation as well. He recently had 2 sebaceous adenomas removed, in which subsequent IHC testing revealed absent staining for MSH6. He has been referred for screening colonoscopy, and it has been recommended that his older son and daughter receive genetic counseling and testing. Among 184 Kaiser Permanente patients with personal and/or family history of Lynch syndrome-associated malignancies as well as MMR gene mutations, only three were identified with primary brain tumors: astrocytoma at 17 years (MSH6), glioblastoma at age 57 years (MSH2), and glioblastoma at age 75 years (MLH1). The present case illustrates two key points: the utility of cancer gene panels in patients with juvenile brain tumors, and the value of cascade testing despite negative family cancer history.

600F

In-frame germline deletion of exons 23 and 24 results in DICER1 syndrome. *M. Apellaniz-Ruiz[1,2], L. de Kock[1,2], N. Sabbaghian[1], M. Wu[1,2], F. Guaraldi[3], L. Ghizzoni[3], G. Beccuti[3], W.D. Foulkes[1,2,4,5].* 1) Lady Davis Institute, Segal Cancer Centre, Jewish General Hospital, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 3) Division of Endocrinology, Diabetes and Metabolism, Department of Medical Sciences, University of Turin, Turin, Italy; 4) Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montréal, Québec, Canada; 5) Department of Medical Genetics, Research Institute of the McGill University Health Centre, Montréal, Québec, Canada.

   Introduction: *DICER1* syndrome is a rare human disease (OMIM #601200) associated with a predisposition to several rare phenotypes, including pleuro-pulmonary blastoma, cystic nephroma, Sertoli-Leydig cell tumor (SLCT) and multinodular goiter (MNG), among other manifestations. Patients with this syndrome have been shown to harbour an unusual combination of a wide-ly-distributed, generally-truncating germline *DICER1* mutation, coupled with a somatic missense mutation located in exons 24 or 25 which partially encode the RNase IIIb domain. Here we describe a large family with remarkable history of MNG and SLCT, carrying an in-frame germline *DICER1* deletion. **Objective:** We aimed to study the genetic background of this large pedigree and to further explore the impact of the identified germline mutation. **Methods:** DNA from blood and tumor samples was analyzed for *DICER1* muta-tions. Multiplex ligation-dependent probe amplification (MLPA) assay was used to screen for germline *DICER1* duplications or deletions. cDNA was synthe-sized from RNA from LCLs to characterize the expression of *DICER1* alleles. In addition, *DICER1* targeted capture and/or Sanger sequencing of RNase IIIa and IIIb domains was performed on tumor DNA. **Results:** MLPA showed that 8 of the 15 tested family members carried a germline *DICER1* in-frame deletion (c.4207-41_5096+1034del). mRNA transcripts with the deletion were not eliminated by nonsense-mediated mRNA decay and were expressed in the cDNA of the patients. Thus, individuals carrying the deletion would express a protein devoid of exons 23 and 24. Additional second somatic hotspot muta-tions were found in the tumors from 4 of the *DICER1* germline deletion carriers from whom tissue was available. **Conclusion** Most patients with DICER1 syndrome have a germline loss of function *DICER1* mutation on one allele, and usually lose both the RNase IIIa and IIIb catalytic domains. Here, for the first time, we have shown that the expression of an almost full-length DICER1 protein in which the RNase IIIa domain and two (1810, 1813) of four metal-ion binding amino acids in the RNase IIIb domain remain intact, is not sufficient to prevent tumorigenesis.

601W

Improving variant classification by incorporating pre-curated gene-spe-cific knowledge into hereditary cancer multi-gene panel testing. *H. Kang, C. Pardo, J. Thusberg, K. Erhard, R. Quintana, S-H. Wang, I. Wilson, K. Ouyang, M. Anderson.* Invitae Corporation, San Francisco, CA.

   Multi-gene panel testing in the diagnosis of inherited disease is routinely recommended in the clinic and broadly provided by genetic testing laborato-ries. As panels and tested genes grow in number, novel variant observations also increase. To guide the interpretation of a variant during multi-gene panel testing, our laboratory developed Sherloc, an objective evidence- and point-based variant classification framework. Although truncating variants, splice-site variants, and large copy number variations are often de facto pathogenic, variants truncating in the last exon of a gene (i.e., expected to escape nonsense-mediated decay), causing in-frame gross deletions or duplications, and occurring in the initiator codon are not necessarily pathogenic. When clinical or functional data are unavailable, our framework is designed to query whether a variant impacts critical functional domains and structural residues, pre-defined pathogenic variants, or the molecular mechanism for a gene. To improve variant classification for approximately 130 hereditary cancer genes, we established a database for this key information. Herein, we demonstrate our approach with a series of novel or rare variants identified during the clinical diagnostic testing of hereditary cancer genes. For *FH* and *VHL*, which utilize two alternative initiating ATGs that generate different protein isoforms, the interpretation of novel truncating variants between the two methionines can be confusing without gene-based rules. For the functionally and structurally well-characterized *BRCA1*, *BRCA2*, and *CHEK2*, our approach is important in pre-determining whether in-frame gross deletions and duplications occur in functionally critical domains or in exons in which naturally occurring alternative transcripts have been reported. For *POLE* and *POLD1*, in which loss-of-func-tion missense variants in specific domains have been established as the dis-ease-causing mechanism, our approach is useful in excluding novel truncating variants from de facto pathogenic classification. For *CDKN1C*, with its highly repetitive sequence, we have pre-defined rules to exclude commonly detected variants within the PAPA domain. Finally, we present a nonsense variant in the last exon of *TMEM127*, in which the classification was established with clinical data because studies of the protein functional domains are limited. Our approach demonstrates that pre-curated, gene-specific knowledge is essential for variant classification.

602T

Evaluation of QIAGEN Clinical Insight as a content resource for variant curation in a CLIA laboratory. *K. Karimi[1], S. Cox[1], K. Moyer[1], P. Kang[1], E. Evans[1], S. Tanenbaum[2], D. Richards[2].* 1) Counsyl, South San Francisco, CA; 2) QIAGEN, Redwood City, CA.

Counsyl performs an expanded carrier screen for >175 recessive diseases and a panel of up to 36 genes for hereditary cancer risk assessment. Our CLIA laboratory utilizes NGS to analyze all exons for panel genes, excluding some regions with high homology, and a comprehensive curation workflow is used to interpret the clinical impact of detected variants. Automated proprietary software initially classifies variants with high population frequency or which have not been reported. Remaining variants undergo curation by PhD scientists and genetic counselors before approval by board-certified laboratory directors. Pathogenicity is assessed using ACMG guidelines and is based on published case and functional studies, variant databases, population frequency, conservation, and in silico predictors. Identification of articles mentioning a variant is crucial for accurate appraisal of pathogenicity, and Counsyl searches variant databases, an in-house library, and online. We investigated QIAGEN QCI as a solution for reducing manual searches, with quantitative and qualitative assessment of variant coverage. QIAGEN Clinical Insight (QCI) is a clinical decision support platform that provides manually curated clinical case evidence with computed ACMG classifications and a comprehensive bibliography of articles. Articles describing variants in the relevant genes are identified through natural language processing of abstracts and PubMed annotations. Full texts are reviewed by scientists that have undergone training, and data is entered into a web-based curation tool. QIAGEN uses third party User Acceptance Testing to validate high-level coverage and accuracy to comply with quality targets. For 1,919 variants curated by Counsyl, we evaluated reference overlap with QCI and whether additional QIAGEN references would alter classifications. We found QCI coverage to be comprehensive, containing 89.3% of article-variant pairs identified by Counsyl, and an additional 13,938 article-variant pairs for the evaluated variants. The vast majority of our classifications (98.8%) were unaltered, thus demonstrating the benefits of adopting QCI for reference selection, while validating the efficacy of our previous workflow. With initial integration, QCI has significantly increased curation efficiency, as evidenced by ~75% time savings in a reference search process that can take up to 45 minutes. More time-intensive methods can now be focused on variants nearer evidence thresholds.

603F

Genotoxic chemotherapies and radiotherapy contribute to the development of multiple primary tumors in patients with Li-Fraumeni syndrome. *E. Kasper[1], E. Angot[2], E. Colasse[2], L. Nicol[3], J-C. Sabourin[2], S. Adriouch[4], Y. Lacoume[5], C. Le Clezio[1], S. Raad[1], Y. Zerdoumi[1], T. Frebourg[1], J-M. Flaman[1], G. Bougeard[1].* 1) Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; 2) Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Pathology, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; 3) PICTUR – Small Animal Imaging, Rouen University, Rouen, France; 4) Normandie Univ, UNIROUEN, Inserm U1234, Rouen University, Rouen, France; 5) Animal Facility, Faculty of Medicine and Pharmacy, Rouen University, Rouen, France.

Li-Fraumeni syndrome (LFS), due to *TP53* germline mutations, is one of the most severe predispositions to cancer, characterized by an early age of onset and a broad tumor spectrum. We recently reported in LFS patients a very high incidence of multiple primary cancers (MPC), above 40%, including secondary tumors in radiation fields. Considering these clinical observations and the mechanism of anti-cancer treatments we speculated a link between these treatments and the development of MPC in LFS patients. To address this question, we first adapted to mouse cells the p53 genotoxicity assay previously developed in our laboratory and based on the p53-transcriptional response to DNA damage in human lymphocytes. We showed that most of classical anti-cancer treatments, except spindle poisons, are genotoxic in mouse like in human cells. We then evaluated the *in vivo* genotoxicity of treatments using a LFS mouse model. *Trp53* KO-/-, *wt/-* and *wt/wt* mice were exposed to X-rays or to various anti-cancer drugs: a mitotic spindle poison (Docetaxel), and a topoisomerase inhibitor (Etoposide). Mice treated with NaCl were used as control group. Tumor development was monthly monitored using whole-body MRI, and histopathological analysis was performed to confirm the presence of tumors. We analyzed a total of 208 mice. X-rays and Etoposide were shown to accelerate tumor development in *Trp53-/-* and *wt/-* mice (OR=6.46, p-value=$1.24.10^{-3}$ and OR=3.91, p-value=$5.76.10^{-4}$, respectively), unlike Docetaxel which was devoid of effect on tumor development kinetics. This study provides evidence that genotoxic chemotherapies and radiotherapy contribute to MPC development in LFS patients. Therefore, in germline *TP53* mutation carriers, radiotherapy should be avoided whenever possible and surgical treatment prioritized. Diagnosis laboratories should be able to perform fast analysis of *TP53* before radiotherapy in clinical situations strongly suggestive of LFS such as very early-breast cancer < 31 years of age, and non-genotoxic treatments should be considered in the future to reduce the risk of MPC.

604W

FLCN gene pathogenic variants in individuals presenting with lung cysts without pneumothorax, skin features or renal tumours typical of Birt-Hogg-Dube syndrome: A case series and review of the literature. *K. Kohut[1], I. Vlahos[1], R. Aul[1], K. Snape[1,2], H. Hanson[1,2].* 1) Clinical Genetics, St George's University Hospital NHS Foundation Trust, London, United Kingdom; 2) St George's University of London, London, United Kingdom.

Background: Birt-Hogg-Dube syndrome (BHDS) is an autosomal dominant, highly penetrant condition caused by a pathogenic variant in the *FLCN* gene. Clinical features include skin fibrofolliculomas, pulmonary cysts often resulting in spontaneous pneumothorax and certain renal tumours, often bilateral and multifocal. Genetic testing is typically initiated in probands with ≥5 fibrofolliculomas, facial angiofibroma histologically confirmed and not linked to another syndromic cause, multiple/bilateral chromophobe, oncocytic, and/or hybrid renal tumors or renal tumours with a family history of renal tumour, or personal and family history of pneumothorax. **Case series:** We describe two cases in which an *FLCN* pathogenic variant was found in individuals under investigation by the respiratory team whose radiological work-up identified lung cysts suggestive of BHDS, in the absence of any other features of the condition. One was a 37 year-old man who presented with chest pain. CT chest identified thin-walled pulmonary cysts involving all lung zones. He had no obvious renal cysts and no family history. He did not have a formal diagnosis of BHDS-related skin features, but on review in Genetics he was referred to dermatology to determine if some lumps on his face, neck and upper back could constitute fibrofolliculomas. The second case was a 46 year-old woman who was under review after developing left lower lobe bullae post pneumonia. CT demonstrated cystic lung disease with a "stacked" appearance, which has been described in BHDS. She had no renal cysts, family history or skin features of BHDS. Both cases had some intrafissural cysts, which are unusual in other cases of cystic lung disease. **Results:** These two cases exemplify the possibility to identify BHDS based on isolated presentation of lung cysts. **Conclusions:** Health professionals in genetics, general practice and respiratory/thoracic specialities including radiology should be aware of the possibility of BHDS in individuals with lung cysts, and refer to genetics for assessment and genetic testing of the *FLCN* gene. Careful questioning about personal and family history as well as a detailed examination could help to distinguish between BHDS and other conditions presenting with lung cysts. A diagnosis of BHDS will identify the need for further dermatological and renal follow-up, as well as at-risk family members who should be screened. This could lead to earlier diagnosis and improved prognosis.

605T

Screening of over 1000 Indian patients with breast and/or ovarian cancer with a multi-gene panel: Prevalence of BRCA1/2 and non-BRCA mutations. *A.U. Mannan[1], J. Singh[1], N. Thota[1], S. Singh[1], S. Padhi[1], P. Mohan[1], S. Deshwal[1], S. Sur[1], M. Ghosh[2], A. Agarwal[3], R. Sarin[4], S. Gupta[5], R. Ahmed[6], S. Almel[7], B. Chakraborti[6], V. Raina[8], P. DadiReddy[9], B.K. Smruti[10], S. Rajappa[11], C. Dodagoud[3], S. Aggarwal[12], M. Singhal[4], A. Joshi[13], R. Kumar[14], A. Kumar[15], D. Mishra[6], S. Sankaran[1], S. Katragadda[1], V. Veeramachaneni[1], R. Hariharan[1,16].* 1) Strand Center for Genomics and Personalized Medicine, Strand Life Sciences, Bangalore, India; 2) Triesta Sciences, Health Care Global Enterprises Limited, Bangalore, India; 3) BLK Super Specialty Hospital, Delhi, India; 4) Indraprastha Apollo Hospitals, Delhi, India; 5) Tata Memorial Centre, Mumbai, India; 6) Tata Medical Center, Kolkata, India; 7) P. D. Hinduja Hospital and Medical Research Centre, Mumbai, India; 8) Fortis Healthcare, Gurgaon, India; 9) Continental Hospitals, Hyderabad, India; 10) Bombay Hospital and Medical Research Centre, Mumbai, India; 11) Basavatarakam Indo American Cancer Hospital and Research Institute, Hyderabad, India; 12) Sir Ganga Ram Hospital, Delhi, India; 13) Nanavati Super Specialty Hospital, Mumbai, India; 14) Rajiv Gandhi Cancer Institute and Research Centre, Delhi; 15) Health Care Global Enterprises Limited, Bangalore, India; 16) Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India.

Breast and/or ovarian cancers (BOC) are among the most common cancers in women across the world. In the Indian population, healthcare burden due to BOC has been steadily rising, thus stressing the need for early detection, surveillance and disease management measures. However, the burden attributable to inherited mutations is not known. We sequenced 1010 unrelated patients and families from India with an indication of BOC, using the TruSight Cancer panel which includes 14 genes strongly associated with risk of hereditary BOC (HBOC). Genetic variations were identified using the Strand NGS software and interpreted using the StrandOmics platform. We were able to detect mutations in 304 (30.1%) cases, of which, 55 mutations were novel. Majority *(*84.5%) of the mutations were detected in the *BRCA1/2* genes as compared to non-BRCA genes (15.5%). When the cases were stratified on the basis of age at diagnosis (AAD) and family history of cancer, the highest detection rate of 72% was observed in patients with AAD below 40 years and with first degree family member(s) affected with BOC. Our findings indicate that in the Indian population, there is a high-prevalence of mutations in the high-risk breast cancer genes: *BRCA1*, *BRCA2*, *TP53* and *PALB2*. In India, socioeconomic inequality limiting access to treatment is a major factor towards increased cancer burden; therefore incorporation of a cost-effective multi-gene test will be helpful in ensuring widespread implementation of genetic screening in the clinical practice for HBOC.

606F

Cancer following radiotherapy for primary cancers in Li-Fraumeni syndrome patients. *L. Oba[1], P.P. Khincha[2], J.T. Loud[1], L. Morton[1], S.A. Savage[1], M.I. Achatz[1].* 1) Clinical Genetics Branch, DCEG, National Cancer Institute, NIH, Bethesda, MD; 2) Radiation Epidemiology Branch, DCEG, National Institution, NIH, Bethesda.

Li-Fraumeni syndrome (LFS) is a rare autosomal dominant hereditary cancer predisposition syndrome associated with early age of onset, and a high lifetime risk of multiple primary malignancies. Approximately 70% of individuals with LFS have a germline *TP53* mutation. *In vivo* studies have shown that Trp53 heterozygous mice are radiation-sensitive and likely to develop a second cancer after exposure to ionizing radiation. These findings, along with reports of second malignancies occurring in the radiation field of LFS patients treated with ionizing radiation, led to the recommendation to minimize radiation exposure in germline *TP53* mutation carriers. The Li-Fraumeni Syndrome Study, a longitudinal cohort study conducted by the Clinical Genetics Branch, National Cancer Institute, evaluated the association of previous radiotherapy for cancer treatment with the onset of a second cancer in the area exposed to ionizing radiation. Medical charts were reviewed of 68 LFS patients (14 males and 54 females) who carried a germline *TP53* mutation. Of these 68 patients, 27 developed one primary cancer and 41 had two or more primary cancers. Twenty-four patients of 68 were treated with therapeutic radiation, 15 of whom developed a subsequent primary malignancy. Of the 15, seven patients (five breast cancers and two sarcomas) developed a new primary malignancy within the field of radiation. The subsequent malignancies developed were spindle cell sarcoma of the chest in three patients, breast cancer in two patients, adenocarcinoma of the lung in one patient and glioblastoma in one patient. After radiation exposure, the median time to the appearance of a subsequent cancer in those seven patients was four years (range from 1-42 years). In summary, 29% of LFS patients treated with radiation therapy developed a cancer within the radiation field of therapeutic radiation. This analysis demonstrates the importance of larger studies on radiation treatment exposure in LFS patients to determine the risk of radiation-induced cancers, and the importance of minimizing radiation exposure in *TP53* germline mutation carriers.

607W

Germline mutations in DNA repair genes are overrepresented in children and young adults with rhabdomyosarcoma: A discovery and validation cohort study. *T. Wegman-Ostrosky[1], N. Light[2], E.L. Young[3], H.C. Chen[9], Y.Y. Chi[4], D. Hall[5], D.S. Hawkins[6], J. Kim[1], P.J. Lupo[7], R. Patidar[9], A. Shlien[2], S. Skapek[8], M. Tyagi[9], J. Wei[9], X. Wen[9], J.D. Schiffman[3], J. Khan[9], D. Malkin[2,10], D.R. Stewart[1].* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, Natio, Rockville, MD; 2) Genetics and Genome Biology Program, The Hospital for Sick Children, University of Toronto, Toronto, Canada; 3) Dept. of Pediatrics and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 4) University of Florida, Gainesville, FL, USA; 5) Children's Oncology Group, Monrovia, CA, USA; 6) Seattle Children's Hospital, Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA; 7) Dept. of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 8) Dept. of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA; 9) Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA; 10) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

Intro. Rhabdomyosarcoma (RMS) accounts for 3% of all pediatric cancer and is the most prevalent soft tissue tumor in childhood and adolescents. The etiology of RMS remains largely unknown; it is reported that hereditary factors account for 7-33% of RMS. We sought to determine the prevalence of pathogenic germline genetic variation in RMS patients unselected for family history. *Methods.* All samples were ascertained from the D9902 correlative biology study of the Children's Oncology Group. Germline whole-exome sequencing (WES) was conducted at the National Cancer Institute (NCI) on the discovery cohort (n=121); whole-genome sequencing (WGS) was conducted at SickKids on the validation cohort (n=122). WGS is underway on another 150 samples. WES was performed on 598 cancer-free adult controls from NCI studies. Analysis on 137 cancer-predisposition genes from the combined COSMIC germline and Rahman lists focused on those with previously annotated dominant inheritance. We reviewed variants with a MAF <0.05%. Truncating and canonical splice-site variants were considered pathogenic (P); missense variants with CADD >20 and evidence of pathogenicity in ClinVar or HGMD-DM (after literature review) were considered likely pathogenic (LP). P/LP frequency in each cohort was compared with controls using the Fisher's exact test. *Results.* The NCI and Toronto cohorts were comparable in age, gender and histologic type. In the NCI cohort, 12 P/LP variants were identified (10%; P=0.009) in *ATM, DICER1, FLCN, MSH6, NF1, PTPN11, RB1, RET, TP53* and *SDHA*. In the Toronto cohort, 13 P/LP variants were identified (11%; P=0.004) in *ATM, BRCA2, CHEK2, EXT2, HRAS, NF1, RB1, RET,* and *TP53*. Genes with P/LP variation common to both cohorts were *ATM, NF1, RB1* and *TP53*. In the controls, there were 23 P/LP variants (4%) in *ATM, BRCA2, CHEK2, EXT2, HRAS, MSH6, RET,* and *TP53*. *Conclusion.* We observed a significant excess of pathologic germline variation in RMS in multiple established cancer predisposition genes. WES and WGS explained only a minority of risk in sporadic RMS. Some genes (*NF1, RB1, PTPN11, TP53, DICER1*) underlie syndromes known to be associated with RMS, whereas others (*BRCA2, CHEK2, MSH6, SHDB*) merit additional validation in larger cohorts. Our results suggest that germline genetic testing may be of clinical utility for patients with sporadic RMS, and that pathogenetic variants in DNA repair pathway genes may be broadly relevant to the etiology of these cancers.

608T

Unexpected cancer-predisposition gene mutations in Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome patients without underlying germline *PTEN* mutations. *L. Yehia[1,4], Y. Ni[1,2], K. Sesock[1], F. Niazi[1], B. Fletcher[1], J. Chen[1], T. LaFramboise[1,3,6], C. Eng[1,3,5,6].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Center for Clinical Genomics, Cleveland Clinic, Cleveland, OH; 3) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 4) Department of Pathology, Case Western Reserve University, Cleveland, OH; 5) Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH; 6) Germline High Risk Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

Patients with *PTEN* hamartoma tumor syndrome (PHTS), which encompasses Cowden/Cowden-like syndrome (CS/CS-like) and Bannayan-Riley-Ruvalcaba syndrome (BRRS) characterized by germline *PTEN* mutations, benefit from *PTEN*-enabled cancer risk assessment and clinical management. *PTEN* wildtype patients (~50%) remain at an increased risk of developing certain cancers. Existence of germline mutations in other known cancer susceptibility genes has not been previously explored in these patients, with implications for different medical management. We conducted a 4-year multicenter prospective study (2011-2015) of incident patients with features of CS/CS-like and BRRS without *PTEN* mutations. Whole-exome sequencing and targeted analysis were performed including 59 clinically actionable genes from the American College of Medical Genetics and Genomics (ACMG) and 24 additional genes associated with inherited cancer syndromes. Germline mutation spectrum and frequencies were compared to The Cancer Genome Atlas (TCGA), including 6 apparently sporadic cancer types associated with PHTS. Pathogenic or likely pathogenic cancer susceptibility gene alterations were found in 13 of 87 (14.9%) CS/CS-like and BRRS patients and included *MET* (n=6), *MUTYH* (n=1), *RET* (n=1), *TSC2* (n=1), *BRCA1* (n=1), *BRCA2* (n=1), and *HRAS* (n=1). We found concordant genotype-phenotype associations in 5 (38.5%) patients. Neither proband age at consent, nor personal/family history of cancer predicted the presence of pathogenic germline mutations. Mutations were only found in 119 of 3152 (3.7%) patients from TCGA apparently sporadic PHTS component cancers (OR=4.08, 95% CI 2.07-7.52, p=1.87x10^{-4}). Our findings demonstrate that broad multigene testing in the context of genetic counseling should be considered in *PTEN* wildtype CS/CS-like and BRRS patients due to the enrichment and clinical actionability of the identified and unexpected mutated cancer genes. Finding non-canonical gene mutations in CS/CS-like and BRRS patients might alter medical management.

609F

Clinical and genetic analysis of patients with hamartomatous polyposis syndromes. *J. Oh, H.T. Kim, J.W. Kim.* Department of Laboratory Medicine & Genetics, Samsung Medical Center, Sungkyunkwan University School of medicine, Seoul, South Korea.

Background: Peutz-Jeghers syndrome (PJS) is an autosomal dominant inherited disorder characterized by hamartomatous polyps in the gastrointestinal tract and germline mutation of the serine threonine kinase 11 (*STK11*) gene are known to be the leading cause of PJS. Also, multiple hamartomatous polyps can be found in the patients with mutation of the *SMAD4* gene and patients with mutations of *STK11* and *SMAD4* are known to have a tendency of younger. In this study, we examined the largest sample size in the *STK11* gene on Asian patients to date. **Methods:** Patients referred to molecular genetic center of Samsung medical center for genetic study of the *STK11* and *SMAD4* between January 2006 and May 2017 were examined in this study. An analysis of the *STK11* and *SMAD4* gene was performed in 70 patients and 12 patients, respectively. We sequenced the coding regions and intron-exon boundaries of the *STK11* and *SMAD4* gene and obtained clinical data from electronic medical records. We also conducted a meta-analysis in the STK11 and SMAD4 gene on East Asian patients based on literature search results. **Results:** Genetic abnormalities were observed in 30% of patients (21/70 patients) in the *STK11* gene and 17% (2/12 patients) in the *SMAD4* gene. Twenty different variants including eight frameshift, five nonsense, five missense and three aberrant splicing were found in 21 patients in the *STK11* gene. Pathogenic variants were found most frequently in exon 1 (8/21) and other pathogenic variants were scattered throughout the *STK11* gene. Seven novel variants that were not recorded in the *STK11* databases were identified: c.232A>G, c.233A>C, c.239_240del, c.443del, c.598-2A>G, c.652_667del, c.884_891dup. **Conclusions:** In previous studies, the type of the *STK11* gene variants reported in Korean patients was mostly missense, but our study showed high rate of truncating variant similar to other studies. Since mutation of the *STK11* and *SMAD4* gene is associated with the disease of childhood and increases risk of cancer at multiple sites, screening of the *STK11* and *SMAD4* gene variants on patients with hamartomatous polyposis syndrome is very important to identify individuals with suspected patients. Key Words: Peutz-Jeghers syndrome, Polyposis, *STK11*, *SMAD4*, Korean.

610W

WGS in pediatric neuroncology patients shows a preponderance of germline Mendelian disease gene mutations. *M. Bainbridge[1], S. Nahas[1], L. Farnaes[1], D. Dimmock[1], D. Malicki[2], M. Bondy[3], S. Chowdhury[1], S. Kingsmore[1], J. Crawford[1], R. Wechsler-Reya[4].* 1) Institute for Genomic Medicine, Rady Children's Hospital, San Diego, CA; 2) Rady Children's Hospital, San Diego, CA; 3) Baylor College of Medicine, Houston, TX; 4) Sanford-Burnham Medical Research Institute in La Jolla, CA.

   Pediatric central nervous system (CNS) cancers are rare but devastating tumors with poorly understood, likely genetic, etiology. Here we present results from a neuro-oncology pilot project at Rady Children's Institute for Genomic Medicine combined with a previous study of 200 familial glioma cases. We employed high coverage whole genome tumor (120x) normal (40x) sequencing with somatic variant detection at a minimum of 5% allele frequency combined with CNV and methylation arrays to rapidly interrogate pediatric CNS tumors. At data cutoff, we identified germline tumor predisposition mutations in *NF1, DICER, FANCA,* and *VCP*. In the last case the patient had a metastatic lesion in the CNS from an apparent unidentified primary osteosarcoma. This patient also had a known pathogenic mutation in the autosomal dominant gene causative for Paget's disease [MIM: 167320], *VCP*, a bone over-growth disorder associated with increased risk of osteosarcoma. The patient with *DICER1* mutation also had a splicing mutation in *PTCH1*, the cause of Basal cell nevus syndrome [MIM: 109400] and Holoprosencephaly 7[MIM:610828]. We examined a cohort of ~200 familial glioma cases with WGS and observed a single highly deleterious (CADD > 30) missense mutation in PTCH1, whereas no similar variants in *PTCH1* was found in a control cohort of >1000 samples. In another patient with no known causative germline mutations we observed syndromic findings including encephalopathy, speech delay and growth hormone deficiency. In this patient, we discovered a likely pathogenic missense mutation in *SMARCA4,* the cause of Coffin-Siris Syndrome 4 [MIM:614609] and rhabdoid tumor predisposition syndrome 2 [MIM:613325]. Disruption of SMARCA4 function in this patient may, therefore, explain syndromic findings or account for predisposition to tumor formation. In addition, molecular findings in one case completely altered clinical treatment: Reclassifying the tumor from a pineoblastoma to high grade glioblastoma and, owing to a mutation in *H3F3A* (p.Lys27Met), permitting the patient to enroll in a clinical trial. These results demonstrate the value of WGS sequencing in identifying critically important molecular findings in pediatric patients with CNS tumors. The identification of rare, deleterious mutations in known Mendelian genes underscores the hypothesis that genes which are critical for embryological brain development may contribute to predisposition to cancer of the CNS.

611T

Identification of three novel loss of function mutations within *APC*, the causal gene for classical familial adenomatous polyposis. *A. Gupta[1], S. Macklin[3], A. Sigafoos[2], K. Clark[2], F. Vairo[1], P. Atwal[3], E. Klee[1].* 1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 2) Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN; 3) Clinical Genomics, Mayo Clinic, Jacksonville, Florida.

   Statement of purpose: Familial adenomatous polyposis (FAP) is characterized by multiple adenomatous colorectal polyps developing in the second decade of life, which if left untreated lead to colorectal cancer by the third or fourth decade of their life, and contribute to 1% of colorectal cancer cases. Classical FAP is caused by germline mutations in the Adenomatous polyposis coli (*APC*) gene which are inherited in an autosomal dominant manner. To date, a total of 1795 mutations have been identified within *APC* as listed on Human Gene Mutation Database (HGMD), of which there are 332 nonsense, 97 missense, 112 splice site substitutions, 8 regulatory, 731 small deletions, 315 small insertions/duplications, 46 small indels, 129 gross deletions, 12 gross insertions/duplications and 13 complex rearrangements. Notably a substantial number of these are loss of function (LOF) mutations. This study reports three novel LOFs identified within *APC*. **Methods:** Genomic DNA was extracted from whole blood of the patients. DNA of these patients was tested using Invitae colorectal cancer gene panel consisting of 12 genes including *APC*. Sequence changes as well as deletions/duplications were evaluated. **Results:** A likely pathogenic duplication of exons 3-11 was identified in one patient. Reports of large gross duplications spanning several exons within *APC* are relatively rare, and only one study (Kerr et al., 2012) has reported two such mutations, of which one was a large duplication of the 5'UTR and exons 2-15 and the other was a large duplication spanning exons 4-11, and both were identified in only one patient each. A potential cryptic splice site mutation, classified as a Variant of Unknown Significance (VUS) (c.221-5T>G) was identified in another patient. The most likely consequence of this mutation is the shifting of acceptor splice site upstream by 5 bps according to various splice prediction programs. This may lead to an insertion of 4 bps, causing a frameshift and subsequent truncation (p.E74Vfs*2) after two amino acid residues. Another patient was found to have a pathogenic 2 bp deletion (c.421_422delAG), also leading to a frameshift and subsequent truncation (p.R141Vfs*6). As mentioned above, a large number of splice site mutations and small deletions within this gene have been previously implicated with FAP. **Discussion:** The three novel mutations identified within *APC* in this study contribute to the existing vast database of germline mutations already known for FAP.

612F

Functional characterization of germline *TMEM127* mutations in familial pheochromocytoma. *S.K. Flores, Y. Deng, Y. Qin, P.L. Dahia.* Dept. of Medicine, The University of Texas Health Science Center at San Antonio, San Antonio, TX.

Background: Germline mutations in the transmembrane (TM) protein-encoding gene, *TMEM127*, have been associated with pheochromocytomas (PHEOs), highly hereditary tumors of neural crest origin. TMEM127 is a tumor suppressor that localizes to endo-lysosomal membranes and negatively regulates mTOR signaling. However, the functional relevance and clinical significance of germline *TMEM127* mutations have not been well defined. Investigating the *TMEM127* mutation spectrum and the related mutant phenotypes can provide insights into TMEM127 function and allow for better interpretation of genetic screening results and surveillance of affected and at-risk individuals. **Purpose:** The aim of this study was to characterize the functional consequences of germline *TMEM127* mutations detected in patients with PHEOs by determining the subcellular localization and protein stability of the resulting mutant TMEM127 proteins. **Methods:** Nine *TMEM127* missense mutations and two frameshift mutations were selected for characterization. GFP-tagged wild-type (WT) and mutant TMEM127 plasmids were generated and transiently transfected into HEK293FT cells. Confocal images of GFP signals were taken 48 hours post-transfection. Cell pellets were collected at 24, 48 and 72 hours post-transfection for Western blot analysis and probed for TMEM127, GFP and B-actin (as loading control). Band intensity was calculated using ImageJ. **Results:** Missense mutations (N=5) and frameshift mutations (N=2) that occurred within or disrupted putative TM domains resulted in mutant proteins with a diffuse, cytoplasmic pattern indicative of loss of membrane localization. Missense mutations (N=4) located outside of putative TM domains resulted in mutant proteins with a punctate pattern similar to that of WT TMEM127 indicative of retention of membrane localization. All mutant proteins with a diffuse, cytoplasmic pattern showed a markedly decreased abundance when compared with the WT TMEM127. **Conclusion:** Membrane localization of TMEM127 is necessary for its stability and, likely, for its tumor suppressor function. Loss of membrane localization is associated with mutations disrupting putative TM domains. Thus, TM-associated mutations are likely to lead to loss of function. It is conceivable that mutations, where the mutant protein retains membrane localization, may impair other function(s) of TMEM127, such as binding to an associated protein, and will require further investigation.

613W

Systematic characterization of germline variants in endometrial carcinoma from the DiscovEHR study. *J.E. Miller[1], R.P. Metpally[2], T.N. Pearson[1], S.B. Krishnamurthy[2], V.R. Dasari[2], D.R. Lavage[1], A.M. Cook[2], A.L. Smith[2], J. Nash[2], D.J. Carey[2], M.D. Ritchie[1,3], D. Kim[1,3], R. Gogoi[2] on behalf of the DiscovEHR collaboration.* 1) Biomedical & Translational Informatics Institute, Geisinger Health System, Danville, PA; 2) Weis Center for Research, Geisinger Medical Center, Danville, PA; 3) Huck Institute of the Life Sciences, Pennsylvania State University, University Park, PA.

Endometrial cancer (EMCA) is one of the most common cancers among women in the world. Currently, there is a need to identify which genes contain potentially pathogenic germline variants among EMCA patients. The Geisinger Health System (GHS) has the unique opportunity to address this problem using the DiscovEHR cohort which includes whole-exome sequencing on over 50,000 consenting patients, 300 of whom were diagnosed with EMCA. Here, a bioinformatics pipeline was developed to annotate rare variants that are potentially pathogenic using ClinVar, HGMD (Human Gene Mutation Database) and VEP (Variant Effect Predictor). Potentially pathogenic variants in EMCA participants (N = 300) were compared to potentially pathogenic variants present in non-cancer controls (N = 2120, female participants older than 80 with no history of cancer) and other hormone related cancers (N = 1486, OHRC, e.g. ovarian and breast cancer) cohorts to identify which genes and variants are unique in EMCA. Twenty-six genes had a greater burden in EMCA relative to the non-cancer controls. Of the 26 genes with potentially pathogenic variants, only 2 also had an increased burden in the OHRC cohort, suggesting the 24 genes identified may help distinguish participants with EMCA from those with non-EMCA cancer or no cancer. Additionally, comparing GHS to the cancer genome atlas (TCGA) EMCA germline data illustrated there were 91 genes with potentially pathogenic variation and 55 unique potentially pathogenic variants that were present in both studies. Thus, similar germline variation can be observed in unique EMCA cohorts and could help prioritize genes and variants to investigate for future work. In summary, this systematic characterization of potentially pathogenic germline variants provides new insights for distinguishing those diagnosed with EMCA from those with non-EMCA cancer or no cancer.

614T

The HBV receptor gene *SLC10A1* is frequently down-regulated in hepatocellular carcinoma and is associated with poor survival. *P. An, C.A. Winkler.* Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD.

Sodium taurocholate cotransporting polypeptide (NTCP, SLC10A1) was recently identified as the cell-entry receptor for Hepatitis B virus (HBV). SLC10A1, exclusively expressed in the basolateral membranes of hepatocytes, is physiologically responsible for hepatic uptake of sodium-dependent bile salts. *In vitro* and animal experiments show that down-regulation of *SLC10A1* reduces HBV replication. Chronic HBV infection is a major risk factor to development of hepatocellular carcinoma (HCC). However, the clinicopathological role of SLC10A1 in HCC is largely unknown. In this study, we investigated the expression and prognostic value of *SLC10A1* in HCC by integrating and meta-analyzing eight gene expression datasets (n=1200) derived from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) data sets. We assessed *SLC10A1* gene expression difference between tumor and normal tissues, and its correlation with overall patient survival with Kaplan-Meier method. The expression level of *SLC10A1* was markedly decreased in HCC tumor tissues compared with corresponding normal tissues in 6 out of 8 datasets (meta-p-value <0.0001, fold change 1.03-9.6). Interestingly, differential expression was observed in both HBV-positive and HBV-negative HCCs, suggesting that other mechanisms than HBV receptor function of SLC10A1 underlie the observed expression differences. Furthermore, in the TCGA HCC dataset, low expression in tumor tissue was associated with poor survival (Log-rank p-value<0.001). We postulate that decreased SLC10A1 may lead to over-accumulation of bile acids, which is potentially cytotoxic to hepatocytes, causing liver inflammation and regeneration. In summary, the *SLC10A1* gene is down-regulated in liver cancers and is associated with poor prognosis. The intriguing role of SLC10A1 in HCC tumorigenesis warrants further investigation. (Funded by the National Cancer Institute under Contract HHSN261200800001E).

615F

Germline variants in mismatch repair genes are associated with microsatellite instability in sporadic tumors. *A.R. Buckley[1,2], H. Carter[3,4,5], O. Harismendy[4,6], N.J. Schork[2].* 1) Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA, United States of America; 2) Human Biology Program, J. Craig Venter Institute, La Jolla, CA, United States of America; 3) Division of Medical Genetics, Department of Medicine, University of California San Diego, La Jolla, CA, United States of America; 4) Moores Cancer Center, University of California San Diego, La Jolla, CA, United States of America; 5) Cancer Cell Map Initiative (CCMI), University of California San Diego, La Jolla, CA, United States of America; 6) Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, CA, United States of America.

Cancer research to date has largely focused on somatic genetic aberrations, whereas the contribution of germline variation is controversial. While the role of germline variation in cancer susceptibility has been clearly demonstrated in rare familial cancer predisposition syndromes and in *BRCA*-associated cancers, it is arguable that inherited variation can additionally shape tumor growth and influence tumor phenotypes in sporadic cancers. Microsatellite instability (MSI) is a tumor phenotype characterized by expansion or contraction of repetitive microsatellite sequence elements due to somatically-acquired defects in the mismatch repair (MMR) machinery, either through somatic mutation or promoter hypermethylation. We hypothesize that deleterious germline variants in MMR pathway genes can similarly contribute to tumor MSI. To test this hypothesis, we called germline variants on 5,365 individuals spanning 17 cancer types from The Cancer Genome Atlas (TCGA) and studied their contribution to previously characterized somatic MSI. Using a rare variant burden approach, we tested the association between germline variants and somatic MSI accounting for somatic mutation and promoter hypermethylation of 225 genes involved in DNA repair. We found that germline loss of function (LOF) variants in the MMR pathway are significantly associated with an increased number of somatic MSI loci. This association was multiple-hypothesis corrected significant at both the pathway level and for *PMS2* and *PNKP* individually. Importantly, in some instances we observed that somatic LOF mutations, germline LOF variants, and somatic promoter hypermethylation have a similar effect size. Our results demonstrate that germline variation can influence somatic phenotypes to the same degree as somatic mutation, highlighting the need to integrate germline variants in cancer analyses.

616W

Genetic polymorphisms of *CD40* ligand gene and susceptibility to cervical cancer. *T. Chang[1], Y. Yang[2,4], Y. Lee[1,3,5], T. Chen[2], W. Lin[1], C. Lin[1].* 1) Med Res Dept, Mackay Memorial Hosp, New Taipei, Taiwan; 2) Gynecology and Obstetrics Dept, Mackay Memorial Hosp, Taipei, Taiwan; 3) Pediatrics Dept, Mackay Memorial Hospital, Taipei, Taiwan; 4) Gynecology and Obstetrics Dept, Taipei Medical University, Taipei, Taiwan; 5) Pediatrics Dept, Taipei Medical University, Taipei, Taiwan.

Although human papillomavirus (HPV) infection is the main cause of cervical cancer, the host immunogenetic background also plays an important role in the development of cervical cancer. CD40 ligand (CD40L, also known as CD154 or gp39), a type II transmembrane protein, is primarily expressed on activated T cells and critical in T cell immunity. This study aims to investigate if variants of the *CD40L* gene are associated with cervical cancer in Taiwanese women. The *−3459 A/G* and *CA* dinucleotide repeat polymorphisms were genotyped in 518 cervical squamous cell carcinoma (CSCC) patients and 508 age/sex matched healthy controls. The presence and genotypes of HPV in CSCC patients were also determined. We found no significant association between any polymorphisms or haplotypes examined and CSCC risk. In addition, no significant association was observed between HPV-16 positive CSCC patients and controls. Our findings provide no support for the hypothesis that specific *CD40L* polymorphisms are associated with CSCC susceptibility in Taiwanese women.

617T

The progression of global gene expression in melanoma: From normal skin to metastatic disease. *L. Cordeiro, O. Lupi.* UNIRIO, Rio de Janeiro, RJ, Brazil.

Cutaneous melanoma is the most deadly form of skin cancer and its worldwide incidence has been increasing at a rate higher than of any other tumor over the last several decades. In 2017, over 87000 new cases of melanoma will be diagnosed in the United States, making it the fifth cancer in incidence, and more than 10000 deaths will be attributed to the disease. While five-year survival rates have been improving, it can be lower than 15% in advanced cases, despite major advances in the treatment of metastatic disease in the last decade. High-throughput molecular techniques have been key to the development of such novel systemic therapies and will most likely play a major role in melanoma research. We systematically mined public gene expression data and built a unique dataset of over 700 samples, with expression measures of over 12000 genes, of normal skin, primary and metastatic melanoma. We analyzed the progression of changes in global gene expression from normal skin (N) to primary melanoma (P) and from primary to metastatic melanoma (M) and found 1429 differentially expressed genes in the N-P transition and 405 differentially expressed genes in the P-M transition. Moreover, 111 and 132 genes are consistently up and downregulated, respectively, in both N-P and P-M transitions, indicating that certain molecular changes take place at the outset and are maintained throughout advanced stages of the disease. Several genes were upregulated more than ten fold from normal skin to metastatic melanoma – *SPP1*, *GDF15*, *BCL2A* and *SERPINE2* – while some were downregulated by more than 95% – *SERPINB5*, *CST6*, *PERP*, *SLURP1* and *HOPX*. Pathway enrichment analysis uncovered distinctive patterns of changes in global expression in each transition. In the N-P transition, several immune-related pathways are activated – antigen processing and presentation, immune-related signaling and inflammatory response – as well as the p53 signaling pathway. In the P-M transition, the immune-related pathways and the p53 signaling pathway remain strongly activated and though cell cycle pathways seemed somewhat perturbed in the N-P state change, it is clear that these changes became statistscally stronger in the P-M transition. Also, the deregulation of epithelial–mesenchymal transition pathways observed in the P-M transition is a known hallmark of metastatic disease.

618F

CDKN2A germline polymorphisms demonstrate parallel associations of disease risk and clinical outcome in melanoma patients. *S. Fang[1], Y. Wang[1], M. Ross[1], J. Gershenwald[1], J. Cormier[1], R. Royal[1], J. Reveille[2], Q. Wei[3], C. Amos[4], J. Lee[1].* 1) University of Texas MD Anderson Cancer Center, Houston, TX; 2) The University of Texas Health Science Center at Houston; 3) Duke University School of Medicine; 4) Dartmouth College.

Germline mutations of the *CDKN2A* gene account for 1-2% of sporadic cases of melanoma and up to 40% of familial melanoma; furthermore, the *CDKN2A* region represents the strongest melanoma risk locus in genome-wide association studies (GWAS). A recent investigation has suggested that melanoma patients with high-penetrance *CDKN2A* germline mutations have worse melanoma-specific survival compared to patients with *CDKN2A* wild-type status. We therefore sought to evaluate associations between *CDKN2A* germline polymorphisms and both melanoma risk and patient outcomes. Previous studies have reported significant associations between melanoma susceptibility and multiple SNPs in the 9p21 region adjacent to *MTAP* and flanking *CDKN2A*, of which 9 SNPs were genotyped in the MD Anderson GWAS cohort of 1804 melanoma patients and 1026 normal controls, with a total of 818,237 genotyped SNPs and a median follow-up for the melanoma patients of 8.3 years. We evaluated these 9 SNPs for their associations with melanoma risk and melanoma-specific survival. The results indicated that 6 of 9 SNPs were significantly associated with melanoma risk ($P<0.05$). The A allele of the SNP rs4636294 in *CDKN2A* was associated with increased melanoma risk (odds ratio [OR]=1.23, $P=0.0002$) and increased risk of melanoma-related death (hazard ratio [HR]=1.22, $P=0.0135$). After adjustment for age, sex and disease stage at presentation, this SNP remained associated with increased risk of melanoma-related death (HR=1.20, $P=0.0281$). Similarly, the C allele of rs2218220 was associated with increased melanoma risk (OR=1.23, $P=0.0002$) and increased risk of melanoma-related death in both univariate (HR=1.22, $P=0.0141$) and multivariable analysis (HR=1.20, $P=0.0285$). Finally, the G allele of rs935053 was associated with increased melanoma risk (OR=1.25, $P=0.000049$) and increased risk of melanoma-related death in univariate (HR=1.22, $P=0.0144$) and multivariable analysis (HR=1.22, $P=0.0151$). These associations suggest that common variants in the tumor suppressor gene *CDKN2A* may influence not just melanoma risk but also, in parallel fashion, the risk of melanoma-related death. Thus, common *CDKN2A* polymorphisms may act, similarly to rare high-penetrance mutations, in controlling not just melanoma risk but also disease severity. Additional investigation is indicated to confirm these findings in an independent cohort and to evaluate alterations in *CDKN2A* expression and function associated with these polymorphisms. .

619W

Association of cytokine polymorphisms with gastric cancer prognosis in Santiago, Chile. *P. Gonzalez-Hormazabal[1], M. Musleh[2], S. Romero[1], VG. Castro[1], M. Bustamante[3], J. Stambuk[4], F. Izquierdo[4], R. Pisano[4].* 1) Human Genetics Program, Institute of Biomedical Sciences, University of Chile; 2) Department of Surgery, University of Chile Clinical Hospital; 3) Department of Surgery, School of Medicine at Eastern Campus, University of Chile; 4) San Juan de Dios Hospital.

Chile has one of the world's highest mortality rate from gastric cancer. Polymorphisms in cytokine gene has been associated with gastric cancer, but little is known about their role in prognosis for this disease. We aimed to investigate the association of 12 polymorphisms in eight cytokine genes (IL-1B, IL-8, IL-17A, IL-17F, IL-32, TNF, IL1-RN, and IL-10) with overall survival in 165 gastric cancer patients from Santiago, Chile, who received gastrectomy between December 2010 and August 2014. Mean survival time was 35.5 months (minimum 0.3 months, maximum 75.9 months). TaqMan assay was used to genotype the polymorphism. We used Stepwise Cox regression to select the polymorphisms and clinicopathological variables to be included in the Cox proportional hazard model. Our results suggest that rs1143627 (*IL1B* -31T>C) and rs4073 (*IL8* -251A>T) are associated with poor survival (p=0.037 and p=0.010, respectively). Hazard Risk (HR) for rs1143627 TC or CC genotypes and rs4073 AT or TT genotypes was 1.33 [1.02 – 1.74] and 1.31 [1.07 – 1.63], respectively, both adjusted for age, sex, TNM score and tumor size. In conclusion, rs1143627 and rs4073 polymorphism are associated with poor overall survival of gastric cancer. Further larger studies are needed to validate our findings.   Grant support: Fondo Nacional de Desarrollo Científico y Tecnológico, Chile. FONDECYT #1151015.

620T

Multi-gene hereditary cancer panel testing for *BAP1*. *S. Hiraki, L. Susswein, J. Bissonnette, Y. Wang, RT. Klein, KS. Hruska.* GeneDx, Gaithersburg, MD.

   Statement of Purpose: *BAP1* is a recently identified cancer predisposition gene associated with an increased risk of uveal melanoma, cutaneous melanoma, mesothelioma, clear cell renal cell carcinoma (RCC), and other cancers. However, the full phenotypic spectrum and penetrance of the gene has yet to be delineated. Our aim was to describe the spectrum of *BAP1* variants identified in our clinical testing cohort, and to characterize the clinical phenotype of individuals with *BAP1* pathogenic and likely pathogenic variants (collectively denoted PV). **Methods:** Results of *BAP1* testing were reviewed for individuals undergoing next-generation sequencing and deletion/duplication analysis of *BAP1* as a single-gene test, targeted variant test, or as part of a custom multi-gene hereditary cancer panel of up to 61 genes between January 2016 and March 2017. Personal and family histories, as well as the spectrum of variants identified, were assessed for those who carried a *BAP1* PV. **Results**: Of 835 individuals whose testing included *BAP1*, 7 unique *BAP1* PV were identified in 13 individuals. All of the variants identified were loss-of-function and included 3 nonsense, 2 frameshift, and 2 canonical splice variants. Six individuals had panel testing for 2-44 genes, 2 individuals had single-gene *BAP1* testing, and 5 individuals from 2 unrelated families had targeted testing for a known familial PV. One individual who underwent panel testing also had a known PV in the family. Ten of the 13 PV carriers had phenotypes consistent with *BAP1*, including clear cell RCC (5), melanoma (3), and dysplastic nevi (2). The remaining 3 were unaffected, 2 of whom were under the age of 25 years and had targeted variant testing. Family histories included melanoma in 3 cases, one of which was uveal, and renal cell carcinoma in 3 cases. **Conclusions:** All *BAP1*-positive individuals had a clinical presentation consistent with *BAP1* tumor predisposition in terms of personal and/or family histories. For individuals with characteristic *BAP1*-associated tumors, adding this gene to a testing panel may identify variants that could influence future screening.

621F

Role of *HOXB13* in breast and ovarian cancer: Preliminary data from a laboratory-based multigene panel testing cohort. *C. Horton, H. LaDuca, J. Grzybowski, C. Gao.* Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA.

   Background: The p.G84E alteration in *HOXB13* has been associated with an increased risk for prostate cancer in male carriers. While the phenotypic spectrum in carriers unselected for personal or family history of prostate cancer has not been well studied, predisposition to breast or ovarian cancers in females has been suggested. Here we aim to evaluate the role of p.G84E in predisposition to breast and ovarian cancer using a clinical laboratory cohort. Methods: De-identified molecular results and clinical histories were retrospectively reviewed for 33146 individuals who had multigene panel testing (MGPT) that included analysis of the *HOXB13* p.G84E locus. The frequency of *HOXB13* p.G84E was compared between non-Finnish European (NFE) MGPT cases and NFE controls from the Exome Aggregation Consortium (ExAC). Results: Eighty-nine individuals were *HOXB13* p.G84E carriers (0.27% of 33146), including nine with a mutation in a second gene. The most common cancers reported in p.G84E carriers were female breast (42 of 70; 60%), ovarian (3 of 70; 4.3%), and prostate (2 of 10; 20%). The frequency of p.G84E was significantly higher in NFE prostate probands than the ExAC NFE controls (1.35%; OR 4.4 p=0.03), but not in NFE female breast (0.35%; OR 1.1 p=0.5) or ovarian (0.16%; OR 0.5 p=0.4) probands. Individuals with family history of prostate cancer were more likely to carry p.G84E than those without (OR 2.43 p=5.06e-05) and the effect was strengthened in individuals with two or more family members with prostate cancer compared to those with no family history (OR 3.2 p=0.002). In contrast, individuals with family history of breast or ovarian cancer were not more likely to carry p.G84E than those without family history of these cancers. Conclusions: Although breast and ovarian cancer were among the most common cancers in p.G84E carriers, the mutation rate in these probands was not increased compared to the general population. Therefore, the preponderance of breast and ovarian cancer likely reflects the nature of this laboratory cohort rather than a risk association. The impact of prostate cancer family history on *HOXB13* mutation frequency highlights the importance of including prostate cancer when collecting family history regardless of the primary cancer indication for testing. The results of this study reinforce the association between *HOXB13* and prostate cancer; however, more data is needed to determine if female carriers are at increased risk to develop cancer. .

622W

Genome-wide analysis of shared and distinct germline genetic risk variants in colorectal cancer stratified by primary tumor site. *J.R. Huyghe[1], S.L. Schmit[2], S. Chen[3], H.M. Kang[3], C. Qu[1], T.A. Harrison[1], S.I. Berndt[4], S. Küry[5], H. Brenner[6], G. Casey[7], A.T. Chan[8], J. Chang-Claude[6], S. Gallinger[9], M. Hoffmeister[6], T. Hudson[10], M. Jenkins[11], L. Le Marchand[12], P.A. Newcomb[1], J.D. Potter[1], M.L. Slattery[13], J. Smith[14], E. White[1], G.R. Abecasis[3], D.A. Nickerson[14], L. Hsu[1], S.B. Gruber[15], U. Peters[1] on behalf of CCFR, CORECT and GECCO.* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Moffitt Cancer Center, Tampa, FL; 3) University of Michigan, Ann Arbor, MI; 4) National Cancer Institute, Bethesda, MD; 5) CHU Nantes, Nantes, France; 6) German Cancer Research Center, Heidelberg, Germany; 7) University of Virginia, Charlottesville, VA; 8) Massachusetts General Hospital, Boston, MA; 9) Mount Sinai Hospital, Toronto, ON, Canada; 10) Ontario Institute for Cancer Research, Toronto, ON, Canada; 11) The University of Melbourne, Melbourne, Australia; 12) University of Hawaii Cancer Center, Honolulu, HI; 13) University of Utah Health Sciences Center, Salt Lake City, UT; 14) University of Washington, Seattle, WA; 15) USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA.

Colorectal cancer (CRC) is a heterogeneous disease, consisting of biologically and clinically distinct subgroups that can affect prognosis and drug response. Yet, to date, CRC GWAS have assumed a genetically homogeneous case group, mainly because comprehensive molecular profiles are not readily available for well-powered samples. Recently, it has been established that primary tumor site is correlated with molecular subtype, and prognosis. The proximal, right-sided colon has a different embryonal origin than the distal, left-sided colon and the rectum, and right-sided tumors more frequently have microsatellite instability, the CpG island methylator phenotype, and *BRAF*^{V600E} mutations. By contrast, left-sided tumors have higher frequencies of *APC*, *KRAS*, and *TP53* mutations. In this study, we explored whether genetic risk factors differ between CRC case subgroups defined by primary tumor site. We performed whole-genome sequencing of 1,961 CRC cases and 981 controls, and subsequently imputed these haplotypes into 11,895 CRC cases and 14,659 controls that are part of the Colorectal Cancer Family Registry (CCFR) and the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO). Stratified GWAS revealed a genome-wide significant common SNP signal associated specifically with proximal colon cancer risk (rs1800734; P=1.3e-9; 4,349 cases) at the *MLH1* gene, which frequently harbors causative mutations in Lynch syndrome (OMIM #609310). When comparing effects at known CRC risk loci between tumor sites, we found strong differences at several loci, e.g., for rs3087967 at *COLCA1*/*COLCA2*, odds ratios were 1.18 (95% CI: 1.11-1.24; P=4.0e-9) based on 6,344 left-sided tumor cases, and 1.02 (95% CI: 0.96-1.08; P=0.49) based on 4,344 right-sided tumor cases. To quantify the shared genetic basis, we performed bivariate GREML analysis as implemented in GCTA. We estimated the genetic correlation r_g to be 0.74 (95% CI: 0.45-1.00). A likelihood ratio test for rejecting the hypothesis that r_g = 1 approached significance (P=0.055). On the underlying liability scale, the heritability was estimated to be 0.068 (95% CI: 0.046-0.091) and 0.080 (95% CI: 0.050-0.102) for left- and right-sided tumors, respectively. We are currently expanding this analysis to >60,000 CRC cases and controls. Additionally, we will use the BUHMBOX statistic to test whether the shared genetic basis can be explained by subgroup heterogeneity such that risk alleles are shared by only a subset of all cases.

623T

Birth order affects risk of multiple lymphoid cancers in lymphoid cancer families. *S.J. Jones[1,2], O.S. Uvarov[3], D. Salema[2], J. Connors[4], A. Brooks-Wilson[2,3].* 1) Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada; 3) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada; 4) Centre for Lymphoid Cancers, BC Cancer Agency, Vancouver, British Columbia, Canada.

The hygiene hypothesis proposes that an early life environment that has a relative lack of early exposure to microorganisms and infectious disease inhibits a child's immune system from maturing optimally. Consequently, such individuals are more susceptible to atopic disorders, including allergies and autoimmune diseases, as well as some lymphoid cancers including Hodgkin lymphoma (HL), acute lymphocytic leukemia, and possibly non-Hodgkin lymphoma (NHL). Birth order and family size relate to the hygiene hypothesis, as they are likely to affect exposure to infectious diseases, with low birth order and smaller families correlating with higher risk of immune-related disorders. We examined sibships within 182 families with a history of lymphoid cancer. Birth order and other information was collected from telephone interviews and questionnaires, from multiple family members whenever possible. 380 sibships with 481 lymphoid cancer cases were examined. We hypothesized that early life factors in these families would support the hygiene hypothesis. An inverse relationship between birth order and lymphoid cancer diagnosis was observed for all types of lymphoid cancers together (p-value <0.0001), as well as for NHL (p-value <0.0001), chronic lymphocytic leukemia (p-value = 0.0069), and multiple myeloma (p-value = 0.0011) separately, but was not significant for HL (p-value = 0.2265), possibly due to small sample size. The observed inverse relationship between birth order and risk of lymphoma is supportive of the hygiene hypothesis, and that childhood exposures to infectious disease may play a role in risk of multiple types of lymphoid cancers. The familial nature of the cancers studied here implies a role of shared genetic and/or environmental factors. The current observation implies that such effect(s) may be modified by lifestyle factors that correlate with birth order, and could lead to identification of lifestyle factors that protect against lymphoid cancers even in the familial context.

624F

Rare DNA repair gene mutations predispose to young onset and lethal prostate cancer in the UK. *Z. Kote-Jarai[1], DA. Leongamornlert[1], EJ. Saunders[1], S. Wakerell[1], I. Whitmore[1], T. Dadaev[1], C. Cieza-Borrella[1], K. Govindasami[1], MN. Brook[1], DV. Conti[2], RA. Eeles[1].* 1) Oncogenetics, The Institute of Cancer Research, London, London, United Kingdom; 2) Keck School of Medicine, USC, Los Angeles, CA, US.

Prostate cancer (PrCa) is the most common solid tumour in men in the Western world. There is substantial evidence that PrCa predisposition is due both to common and rare germline variation. We screened 167 genes from DNA damage response and repair pathways, within a UK based cohort of young onset cases (diagnosed at <65 years) and controls. Samples were sequenced using a custom Agilent SureSelectXT bait library and Illumina HiSeq technology and processed using a BWA/GATK 2.8 pipeline. Following sample QC, data was analysed from 1,285 PrCa cases and 1,163 controls, with mean target coverage of 76x. We identified 5,086 single nucleotide variants (SNVs) and 175 indels. 233 unique protein truncating variants (PTVs) with MAF <0.5% in controls were found in 97 genes of the screening panel. The total proportion of PTV carriers in cases was higher than in controls (14.5% vs. 11.6%, $P = 0.036$). This enrichment was greater within the previously reported BROCA gene set of 22 tumour suppressor genes (4.5% vs 2.2%, $P = 2.5 \times 10^{-3}$) and subset of *ATM, BRCA1, BRCA2* and *CHEK2* (3.7% vs. 1.4%, $P = 3.1 \times 10^{-4}$).To identify genes which distinguish PrCa cases from controls within our sample set, we calculated the OR for each gene with at least 2 carriers of PTVs. This analysis selected 21 genes with OR ≥1.5; men with PTVs in these were 3.6-fold more likely to have PrCa ($P = 4.29 \times 10^{-10}$). We subsequently compared lethal PrCa cases (cause of death PrCa, $n = 183$) to indolent cases (Gleason score ≤6, $n = 563$) to evaluate genes associated with poor clinical prognosis. 8 genes had OR ≥1.5 for lethal PrCa, 7 of which overlapped the predisposition gene set (*ATM1, CHEK2, BRCA2, GEN1, MSH2, MSH5* & *POLD1*). Within our cohort, carriers of PTVs in the lethal PrCa gene set were 3.9-fold more likely to die from PrCa (11.5% vs. 3.2%, $P = 6.20 \times 10^{-5}$). The union of these gene sets provides a 22 gene panel of which only 6 genes overlap with the BROCA panel. 7.8% of men diagnosed <65yrs with PrCa carried a PTV in this gene set compared to 2.3% of controls. These findings will help facilitate the development of a PrCa specific sequencing panel with both predictive and prognostic potential.

625W

Genome-wide scan of single-nucleotide polymorphisms associated with prostate cancer specific death. *W. Li[1], D. Sjoberg[2], E. Vertosick[2], A. Dahlin[3], O. Melander[3], D. Ulmert[4], H. Lilja[5,6,7], R. Klein[1].* 1) Icahn Institute for Genomics and Multiscale Biology and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY USA; 2) Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY USA; 3) Lund University, Lund, Sweden; 4) Molecular pharmacology program, Sloan Kettering Institute, New York, NY USA; 5) Departments of Laboratory Medicine, Surgery, and Medicine, Memorial Sloan Kettering Cancer Center, New York, NY USA; 6) Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK; 7) Department of Translational Medicine, Lund University, MalmÖ, Sweden.

While over 100 SNPs are associated with risk of prostate cancer diagnosis, SNPs associated with prostate cancer specific mortality remain to be identified. We used GWAS to find single nucleotide polymorphisms (SNPs) predictive of prostate cancer specific death. This study used a time-to-event survival analysis based on 11,506 men aged 45-73 in Malmö, Sweden accepting participation during 1991-96. In the Malmö Diet and Cancer (MDC) study, 1053 men were diagnosed with prostate cancer of whom 245 died from the disease by December 31, 2014. Men diagnosed with prostate cancer were genotyped genome-wide and all common variants were imputed using 1000 Genomes data. Stage and grade at diagnosis and outcome information was obtained from the Swedish Cancer Registry, the National Prostate Cancer Registry of the Southern Region, and cause of death obtained from the Causes of Death Registry at the National Board of Health and Welfare in Sweden, a reliable source on the cause of death from prostate cancer. We conducted a genome-wide association study using Cox proportional hazards model, adjusted for age at diagnosis and stage and grade at diagnosis. Top candidate SNPs were further studied using eQTL analysis and explored in other prostate cancer cohorts. We tested each of 6,126,633 SNPs for association with prostate cancer specific survival. Two of SNPs tested in this analysis reached genome-wide significance ($p < 5 \times 10^{-8}$): rs73055188 ($p = 5.27 \times 10^{-9}$, per-allele HR = 2.27, 95% CI 1.72–2.98) located in the *AOX1* gene and rs35105661 ($p = 1.10 \times 10^{-8}$, per-allele HR = 2.03, 95% CI 1.59–2.59) located in the *HERC4* gene. Additionally, 10 candidate SNPs with $P < 1 \times 10^{-6}$ were found. eQTL study shows that rs73055188 is correlated with *AOX1* gene expression level, which is associated with biochemical recurrence of prostate cancer. Our GWAS identified several SNPs associated with prostate cancer specific death. Multiple lines of evidence showed that *AOX1* gene is related to prostate cancer progression. These putative associations suggest that GWAS may be useful in identifying SNPs and genes associated with clinical outcome in prostate cancer.

626T

Analysis of circulating tumor cells in multiple myeloma patients reveals mutations in proto-oncogenes and tumor supressor genes of NF-kB, Ras/MAPK, and PI3K/Akt pathways. *D.S. Manjegowda[1], A.M. Veerappa[2].* 1) Department of Biomedical Science, Nitte University Centre for Science Education and Research, Nitte University, Paneer Campus, Deralakatte, Mangalore, Karnataka, India; 2) Department of Genetics and Genomics, University of Mysore, Manasagangotri, Mysore,Karnataka, India.

Multiple myeloma (MM), a malignancy of plasma cells, is characterized by widespread genomic heterogeneity and, consequently, differences in disease progression and drug response. The characterization of cirulating tumor cells (CTCs) may respresent a non-invasive method to capture relevant mutations present in plasma cell clones. Here, we conducted a study, to analyze the genomic landscape of CTCs in five MM patients. We performed whole-exome sequencing in sorted CTCs and compared their mutational profile. The disease pathway analysis was performed to identify driver and passenger gene mutations. Present investigation revealed several deleterious and damaging driver gene mutations in *MLL3*, *BRCA2*, and *LTB* contributing towards early initiating mutagenic events leading to plasma cell immortalization and disease initiation. This loss of genomic integrity led to accumulation of additional translocation mutations involving *RAG1*, *IGH* and *IGK*. This transposition occured near to an enhancer present 149 kb upstream enhancing the production of IGK. The molecular events driving and sustaining the cancer identified crucial genomic alterations in genes such as *KRAS*, *DIS3*, *NANOG*, *FLT3*, *IL6*, *JAG1* and *MAP3K1*. Secondary mutations in *KRAS*, a proto-oncogene and member of MAPK signaling pathway, was found causing continous proliferation and differentiation of cells leading to MM. These genes belong to NOTCH, wingless (WNT), and nuclear factor-kappa B (NF-kB), Ras/MAPK, phosphatidyl/inositol3-kinase/Akt (PI3K/Akt) pathways and mutations in genes of these pathways contribute towards activation of oncogenic pathways. These muations were found to precede, succeed or jointly initiate several stages of activations in oncogenic pathways contributing towards sustained activation in MM cells.

627F

BRA-STRAP: BRCA Refined Analysis of Sequence Tests: Risk And Penetrance. *T. Nguyen-Dumont[1], F. Hammet[1], J.J. Burke[1], H. Tsimiklis[1], K. Tucker[2], J. Kirk[3], P. James[4], A. Trainer[4], I. Winship[5], N. Pachter[6], N. Poplawski[7], S. Grist[8], D.J. Park[9], E. Thompson[4], I. Campbell[4], J. Weitzel[10], F.J. Couch[11], J. Leary[12], J.L. Hopper[13], D.E. Goldgar[14], A. Morrow[2], M.C. Southey[1].* 1) Department of Pathology, The University of Melbourne, Parkville, VIC, Australia; 2) Prince of Wales Hospital, Sydney, NSW, Australia; 3) Familial Cancer Service, Westmead Institute for Cancer Research, University of Sydney, NSW, Australia; 4) Peter MacCallum Cancer Centre, VIC, Australia; 5) Royal Melbourne Hospital, VIC, Australia; 6) King Edward Memorial Hospital, WA Australia; 7) Adult Genetics Unit, South Australian Clinical Genetics Service, SA 5006, Australia; 8) South Australia Pathology, Flinders Medical Centre, SA 5042, Australia; 9) Melbourne Bioinformatics, The University of Melbourne, VIC 3052, Australia; 10) City of Hope, Duarte CA, USA; 11) Mayo Clinic Cancer Center, Rochester, MN, USA; 12) Westmead Institute for Medical Research, NSW, Australia; 13) Center for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, VIC 3010, Australia; 14) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA.

The rapid introduction of massive parallel sequencing into clinical genetics services is enabling the screening of multiple breast cancer susceptibility genes in one assay at reduced cost for women who are at increased risk of breast (and other) cancer. These gene panels now typically include a large number of genes but only a few of these genes have established clinical validity. These tests therefore pose considerable challenge to clinical genetic services, as very little is known about the breast cancer risk associated with any of the observed genetic variation. Accumulated research on *BRCA1* and *BRCA2* mutation carriers now means that these women can be offered personalized risk assessment, targeted treatment regimens and informed decision making about the use of chemo-preventive agents, bilateral salpingo-oophorectomy, mammography, risk reducing mastectomy, magnetic resonance imaging (MRI) and other screening modalities. This information is lacking for high-risk women (as defined on family history) who are not found to carry an identifiable mutation in *BRCA1* or *BRCA2*. These women represent about 80% of those tested. Some evidence has been accumulated to enable improved management for women found to carry mutations in *PALB2*, *ATM* and *CHEK2*, and their families. The community now needs similar information about the larger number of genes that are being tested routinely on gene panels to provide the evidence-base from which clinical management and genetic counseling can be provided. We are conducting a large nation-wide study of 10,000 Australian women at high-risk of breast cancer who have tested negative for mutations in *BRCA1* and *BRCA2*. We are performing targeted sequencing of *BRCA1, BRCA2, PALB2, TP53, ATM, MLH1, MSH2, MSH6, PMS2, BARD1, BRIP1, CDH1, CHEK2, MRE11A, RAD50, NBN/NBS1, MUTYH, NF1, PTEN, RAD51C, STK11, FANCM, RECQL, RAD51D* and will present the spectrum of genetic variation observed in these genes. The BRA-STRAP study will enable i) Australian population-based specific estimates of mutation prevalence and penetrance and ii) pooling with similar data from other international efforts including BRIDGES (European Commission) and CARRIERS (NIH). This work will enable the evidence-based translation of new genetic information and a new genetic testing model into clinical practice.

628W

Genome-wide association study to identify the novel biomarker for response to tamoxifen. *H. Ohnishi[1,2], I. Endo[1], S. Nakamura[3], T. Ishikawa[4], M. Kubo[5], C. Udagawa[1], G. Kutomi[6], Y. Sagara[7], Y. Hasegawa[8], H. Zaha[9], H. Takei[10], H. Zembutsu[1].* 1) Cancer Precision Medicine Center, Cancer Institute, Tokyo, Japan; 2) Department of Gastrointestinal Surgery, Graduate School of Medicine, Yokohama City University, Yokohama, Japan; 3) Division of Breast Surgical Oncology, Department of Surgery, Showa University School of Medicine, Tokyo, Japan; 4) Department of Breast Surgery, Tokyo Medical University, Tokyo, Japan; 5) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) 1st Department of Surgery, Sapporo Medical University, Sapporo, Japan; 7) Department of Breast Surgery, Sagara Hospital, Kagoshima, Japan; 8) Department of Breast Surgery, Hirosaki Municipal Hospital, Hirosaki, Japan; 9) Department of Breast Surgery, Nakagami Hospital, Okinawa, Japan; 10) Department of Breast Surgery, Nippon Medical School, Tokyo, Japan.

Purpose: Although many association studies of genetic variants with the clinical outcomes of breast cancer patients receiving adjuvant tamoxifen therapy have been reported, genetic factors determining individual response to tamoxifen are not fully understood. To identify novel genetic markers for response to tamoxifen, we conducted a genome-wide association study (GWAS). Material and Methods: We enrolled 347 pathologically diagnosed with estrogen receptor (ER)-positive (> 10 %), human epidermal growth factor receptor 2 (HER2)-negative, invasive breast cancer patients receiving preoperative tamoxifen monotherapy for 14-28 days. Ki-67 response in breast cancer tissues after tamoxifen therapy was used as a surrogate marker for response to tamoxifen. We performed a GWAS using 275 patients with the Illumina HumanOmniExpressExome-8-v1-2 (Illumina), and 72 patients were used for a replication study. Results: In the GWAS, the top 100 single nucleotide polymorphisms (SNPs) which revealed the smallest P values showed possible associations with Ki-67 response (P value = 3.55×10^{-4} to 8.44×10^{-6}). We selected 74 SNPs of the top 100 SNPs, after considering linkage disequilibrium (LD; $r^2 < 0.8$), for a replication study. Furthermore, a combined result of the GWAS and replication study revealed stronger association of three SNPs with Ki-67 response than the result of GWAS (SNPA; $P_{combined} = 8.34 \times 10^{-6}$, SNPB; $P_{combined} = 1.64 \times 10^{-5}$, and SNPC; $P_{combined} = 2.03 \times 10^{-5}$). To further characterize the three SNPs, we imputed genotypes of SNPs that were not genotyped in the GWAS but are located within 500 kb upstream or downstream of the three SNPs. In this study, we finally identified three SNPs (SNPD on chromosome 4; $P_{combined} = 5.69 \times 10^{-6}$, SNPB on chromosome 6; $P_{combined} = 1.64 \times 10^{-5}$, and SNPE on chromosome 10; $P_{combined} = 9.77 \times 10^{-6}$) possibly associated with Ki-67 response. We examined the combined effects of the three SNPs by classifying patients into three groups on the basis of the total number of risk alleles of these three SNPs, and found that the number of risk alleles of these three genes had cumulative effects on Ki-67 response after tamoxifen therapy ($P_{combined} = 1.86 \times 10^{-15}$). Conclusions: Our results could contribute useful information for identifying biomarkers for predicting response to tamoxifen therapy in patients with breast cancer.

629T

Gene environment interactions in the context of lobular breast cancer. *C. Petridis[1,2], M.A. Simpson[1], R. Roylance[3], E.J. Sawyer[2].* 1) Medical and Molecular Genetics, King's College London, London, United Kingdom; 2) Research Oncology, King's College London, London, United Kingdom; 3) UCLH NHS Foundation Trust, University College London, United Kingdom.

Invasive lobular breast cancer (ILC) is the second most common histological subtype of breast cancer accounting for approximately 15% of all invasive cases. A pre-invasive lesion, lobular carcinoma *in situ* (LCIS) is also present in approximately 60% of these cases. We have previously shown that there is a significant overlap in the genetic predisposition of ILC and the most common invasive ductal carcinoma (IDC) but there are also distinct associations. It has also been shown that the majority of the environmental risk factors have similar associations between the two subtypes but that the use of hormone replacement therapy is more strongly associated with ILC compared to IDC. The aim of this study was to assess whether there is any of an interaction between breast cancer susceptibility loci and use of HRT in ILC. 12 single nucleotide polymorphisms (SNPs) previously found to be associated with ILC were genotyped using the iCOGS platform (cases) and KASP genotyping technology (controls), in postmenopausal women (406 cases and 481 controls) from the GLACIER study. Seven SNPs reached a corrected level of significance (P<0.0042) and were included in the analysis. Three loci show some evidence of interaction with HRT: rs704010, *ZMIZ1* (*P*=0.0065); rs865686, 9q31.2 (*P*=0.019) and rs2981582, *FGFR2* (*P*=0.03). For rs704010, 10:79081391, *ZMIZ1* and rs865686, 9:108126198, 9q31.2 the risk increased for homozygous carriers of risk allele in the presence of HRT (rs704010: OR of HRT for carriers of 2 risk alleles = 1.82 (1.06, 3.11)). In contrast for rs2981582, 10:121592803, *FGFR2* the risk conferred by the risk allele was observed only in the group of individuals who have not used HRT. None of these interactions were deemed significant after correcting for multiple testing, however these interactions warrant further investigation with a larger sample set.

630F

Oral findings of cancer predisposition conditions: "Red flags" that dental, oral and craniofacial providers should recognize for early diagnosis, referral and management. *A.M. Pham[1,2,3], P.A. Sanchez-Lara[1,2,3], L.T. Odono[1,2,4], D.D Yamashita[1,2,4].* 1) Children's Hospital Los Angeles, Los Angeles, CA; 2) Center for Craniofacial Molecular Biology, Los Angeles, CA; 3) Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) Ostrow School of Dentistry, University of Southern California, Los Angeles, CA.

In this paper, we highlight the presentation of a recent 7-year-old Hispanic female patient who presented to the oral surgery clinic for evaluation of her multiple distal tongue papillary nodules. She had a history of chronic constipation, mild intellectual disability, seizures and a negative family history. On exam she had ten 3mm papillary nodules on the distal 1/3 of her tongue and two pea-sized buccal prominences at the inner edge of the oral commissure. She had a narrow face, prominent lips, mild periorbital and conjunctival erythema. Biopsy of the nodules confirmed that they were papillary neuromas, which supported the concern for MEN2B. Subsequent screening test for calcitonin and genetic analysis for the *RET* gene confirmed our diagnosis. Multiple endocrine neoplasia type 2B (MEN2B) is a rare congenital disorder that involves high risk for development of medullary carcinoma of the thyroid (MTC), pheochromocytoma, and parathyroid adenoma or hyperplasia. We also highlight several other recent patients with typical and atypical oral presentations of other cancer predisposition syndromes. Although the cases highlighted in this series are not an exhaustive list, several features identified should be considered "red flags" that trigger prompt referral for confirmation, counseling and management.

631W

Exome sequencing of individuals with testicular germ gell tumor and family history reveals novel germline variants. *L.C. Pyle[1,2], B. Wubbenhorst[1], R.R. Godse[1], K. D'Andrea[1], B. Weathers[1], P.A. Kanetsky[3], NCI. DCEG Cancer Gen Res Lab[4], D.J. Vaughn[1], J.T. Loud[5], M.H. Greene[5], K.L. Nathanson[1], D.R. Stewart[4,5].* 1) CHOP, Philadelphia, PA.Translational Medicine and Human Genetics, Department of Medicine, University of Pennsylvania; 2) Division of Human Genetics and Metabolic Disease, Department of Pediatrics, The Children's Hospital of Philadelphia; 3) Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center & Research Institute; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 5) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute.

Testicular Germ Cell Tumor (TGCT) is the most common cancer affecting men aged 15-45. More than 230,000 US men are living with the disease. No highly penetrant genetic has been identified in TGCT, despite high estimated heritability. Sons and brothers of TGCT cases have 4-6-fold and 8-10-fold risk increases, respectively. Linkage and candidate gene studies have been uninformative, with the exception of a modest association with *PDE11A* inactivating mutations and the Y-deletion "*gr/gr.*" GWAS studies, however, have revealed almost 40 independently associated loci, involving pathways unique to germ cell tumors (germ cell development, sex determination), in addition to pathways associated with other cancer types (DNA damage response and telomere length). This study aimed to identify candidate germline variants with a role in TGCT predisposition. We performed exome sequencing on 221 probands with TGCT who also had either a first- or second-degree family history of TGCT, or bilateral TGCT. Variants were considered of interest if they were: rare (MAF<0.01 ExAC European non-Finnish population); missense, frameshift, or stop mutation; assessed as deleterious in MetaLR and MetaSVM; and heterozygous. Several genes presented three or more pathogenic mutations throughout the cohort. These genes included *TYR* and *OCA2*, important for tyrosine metabolism, and of particular interest given that TGCT almost exclusively affects white men. Variants were found in several genes related to hypogonadism and fertility, including *WDR11* and *PROK2*. Pathway analysis showed significant over-representation of variants in three pathways: KREB zinc fingers (p=1.08E-18, FDR=1.23E-07), ABC transporters (p=2.20E-18, FDR=2.92E-15), proteins related to collagen deposition and remodeling (p=8.21E-09, FDR= 1.09E-5). The latter are of particular interest, given that collagen remodeling drives descent, and maldescent (cryptorchidism) is one of the best-defined TGCT risk factors. Association was independent of whether the patients were clinically cryptorchid (Fisher's exact t-test p=0.637). These data suggest that heritability of TGCT is genetically highly heterogeneous, concordant with existing models based on GWAS findings. Our findings suggest novel associations of TGCT predisposition with ABC transporters, collagen remodeling genes, and zinc fingers.

632T

Exome sequencing provides evidence of novel genes associated with colorectal cancer and polyps. *E.A. Rosenthal[1], B.H. Shirts[2], L.M. Amendola[1], M. Horike-Pyne[1], P.D. Robertson[3], K.R. Patterson[3], J.D. Smith[3], C. Nefcy[4], D. Veenstra[5], F.M. Hisama[1,6], R. Bennett[1], M.O. Dorschner[3,7,8], D. Nickerson[3], D. Crosslin[3,9], I.B. Stanaway[9], R. Nassir[10], N. Zubair[11], T.A. Harrison[11], U. Peters[11,12], G.P. Jarvik[1,3,12].* 1) Department of Medical Genetocs, University of Washington School of Medicine, Seattle, WA; 2) Department of Laboratory Medicine, School of Medicine, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA; 5) Department of Pharmacy, School of Pharmacy, University of Washington, Seattle, WA; 6) Department of Neurology, School of Medicine, University of Washington, Seattle, WA; 7) Department of Pathology, School of Medicine, University of Washington, Seattle, WA; 8) Department of Psychiatry & Behavioral Sciences, School of Medicine, University of Washington, Seattle, WA; 9) Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA; 10) Department of Biochemistry and Molecular Medicine,University of California Davis,Davis,CA; 11) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; 12) Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington.

The lifetime risk to develop colorectal cancer (CRC) is ~4.5% and doubles with one affected 1st degree relative and increases with additional affected 1st degree relatives. There are ~16 major genes known to be associated with high penetrance CRC or polyposis (CRC/P). Although ~25% of all CRC/P cases appear familial, current genetic testing fails to identify a causal genotype in ~80% of those cases. There are likely other genes underlying heritable high penetrance CRC/P that have not been validated and are not on current genetic testing panels. To test for evidence of novel CRC/P associated loci we evaluated 178 candidate genes nominated by linkage, GWAS, or involvement in DNA repair. We compared the number of rare (minor allele frequency < 0.005), potentially disruptive variants (PDVs: stop gain, splice acceptor/donor change, frameshift) found in 83 unexplained CRC/P cases and 2440 controls. Cases included individuals from the University of Washington Clinical Sequencing Exploratory Research NEXT Medicine study (N=51) and Family Polyposis Study (N=10), ascertained from patients referred to medical genetics, and from the Women's Health Initiative (N=22). European ancestry controls were selected from the Exome Sequencing Project, and were not known to have any Lynch [MIM 120435] associated cancers. All participants were verified to have European ancestry and did not have identifiable pathogenic or likely pathogenic variants in any CRC/P associated gene. Multiple sequencing platforms were used as the data were ascertained from multiple sites and at different times. Sequences for each participant were aligned to the whole genome and jointly genotyped using Genome Analysis Tool Kit. As the target regions differed, we analyzed regions covered by all targets and with >90% genotype rate. We found a significant association between case status and rare PDV presence: 19% of cases versus 12% of controls with ≥ 1 PDV in the 178 genes considered (OR=1.7, p=0.04). PDVs were found in 15 genes for cases and 92 genes for the larger control group. As the number of burden tests is large, no gene met test-wide statistical significance (p<2.8e-4). However, logistic regression, adjusting for ancestry and sequencing center/target, indicated 3 genes had suggestive evidence for association with CRC/P: *BRIP1* (OR=36, p=0.003), *NTHL1* (OR=37, p=0.0016), and *RECQL* (OR=20, p=0.008). In summary, these data support the existence of additional genes that underlie high penetrance CRC/P.

633F

Identification of novel prostate cancer susceptibility loci in Finnish population. *C. Sipeky[1], T.L.J. Tammela[2], A. Auvinen[3], C. PRACTICAL[4], J. Schleutker[1,5].* 1) Institute of Biomedicine, University of Turku, Turku, Finland; 2) Faculty of Medicine and Life Sciences, University of Tampere, Tampere, FinlandTampere, Tampere, Finland; 3) Faculty of Social Sciences, University of Tampere, Tampere, Finland; 4) Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL); 5) Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Turku, Finland.

Prostate cancer (PrCa) is the second most prevalent cancer in men worldwide, with 420.000 new cases and with over 100.000 annual deaths (2012). PrCa has a significant heritable component, with genetic factors accounting for 58% of the risk. Accordingly, genome-wide association studies (GWAS) identified 104 low penetrance PrCa susceptibility loci to date, predominantly in populations of mixed European ancestry. However, effect allele frequencies (EAF) and strength of association (Odds ratio, OR) are often variable across populations. Our aim is to identify common PrCa susceptibility alleles specific for Finnish population. Therefore, we conducted an analyses of >200.000 single nucleotide polymorphisms (SNPs) in 2764 PrCa cases and 2401 controls of Finnish origin. Germline DNA samples were genotyped within the Collaborative Oncological Gene-Environment Study (COGS) using Illumina iSelect custom SNP genotyping platform. Statistical analyses of unconditional case-control logistic regression model was performed by using PLINK software. After FDR adjustment altogether 160 variants remain significant on GWAS level (p<5x10[-8]). The PrCa associated signals are concentrated on chromosomes (chr) 7, 8, 11, 17, 19, and in close proximity to one another on the same chr. The strength of association varies between 0.59-1.86. In our study we identified 101 risk variants on chr 8, 11, 17 (OR 1.24-1.86) and 59 protective variants on chr 7, 8, 17, 19 (OR 0.59-0.80). The OR 1.86 is the highest by now found in GWAS studies examining common SNPs in non-familiar cases. Our data suggests that chromosome 11 seem to be risk conferring, whereas chromosomes 7 and 19 possess protective variants. The results facilitate population risk stratification for screening, clinical studies, treatment choices and functional research of prostate cancer in Finnish population.

634W

Co-heritability between aggressive and non-aggressive prostate cancer: Are germline risk loci the same for all prostate cancers? *C.G. Tai[1], N.C. Emami[1,2], T.J. Hoffmann[1,3], L.C. Sakoda[4], E. Jorgenson[4], L.A. Habel[4], J. Shan[4], D.K. Rantunga[4], C.R. Chao[5], N.R. Ghai[5], D. Aaronson[6], J. Presti[6], C. Schaefer[4], N. Risch[1,3,4], S.K. Van Den Eeden[4,7], J.S. Witte[1,3,7].* 1) University of California, San Francisco (UCSF), San Francisco, CA; 2) Program in Biological and Medical Informatics, UCSF, San Francisco, CA; 3) Institute for Human Genetics, UCSF, San Francisco, CA; 4) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 5) Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena CA; 6) Department of Urology, Kaiser Oakland Medical Center, Oakland, CA; 7) Department of Urology, UCSF, San Francisco, CA.

   Introduction: There is considerable debate over whether germline genetic risk loci for overall prostate cancer confer susceptibility for aggressive prostate cancer phenotypes. Previous studies of aggressive prostate cancer have been unable to replicate many of the same genetic associations previously found for overall prostate cancer, of which the majority is non-aggressive. But often these studies have limited sample size and thus likely underpowered. In this analysis we aim to evaluate the degree to which risk loci are shared between aggressive and non-aggressive prostate cancer and to estimate the genetic correlation between the two phenotypes. **Methods:** We conducted our analysis using a subset of unrelated individuals of European ancestry from the Kaiser Permanente Genetic Epidemiology Research on Aging (GERA) cohort which included 3,501 prostate cancer cases (1,090 aggressive and 2,411 non-aggressive) and 24,664 controls. Aggressive phenotypes were prostate cancer cases with high Gleason score, early age at diagnosis (less than 55 yrs), or biochemical recurrence. Using the method implemented in GCTA v1.26.0 we conducted a bivariate restricted maximum likelihood estimation of the genetic covariance between aggressive and non-aggressive prostate cancer using controls without prostate cancer as the referent for both phenotypes. Genetic relationship matrices were built using a set of LD-pruned SNPs filtered based on minor allele frequency, call rate, missingness between cases and controls, and 1000 Genomes allele frequency. **Results:** SNP heritability for aggressive and non-aggressive prostate cancer was 0.19 (SE = 0.07) and 0.26 (SE = 0.06), respectively. The genetic correlation between these two phenotypes was estimated to be r_G= 1.0 (SE = 0.27). Using imputed SNP data, the heritability estimates increased to 0.25 (SE = 0.10) and 0.35 (SE = 0.09) for aggressive and non-aggressive prostate cancer, respectively, while the genetic correlation remained the same r_G= 1.0 (SE = 0.26). **Conclusions:** Although there may be some genetic risk loci specific to aggressive prostate cancer, it seems many of the loci previously found for overall prostate cancer are also risk loci for aggressive prostate cancer.

635T

A rare variant in GOLM1 predisposes to cutaneous malignant melanoma. *C. Teerlink[1], C. Huff[2], J. Stevens[1], S. Holmen[3,4], K. Trombetti[3,4], D. Grossman[3,5], J. Farnham[1], J. Wen[6], J. Facelli[7], L. Meyer[5,8], J. Zone[5], S. Leachman[9], L. Cannon-Albright[1,8].* 1) Dept Internal Medicine, Univ Utah Sch Med, Salt Lake City, UT; 2) Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX; 3) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 4) Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT; 5) Department of Dermatology, University of Utah School of Medicine, Salt Lake City, UT; 6) Utah Department of Health, Salt Lake City Utah, UT; 7) Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT; 8) George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT; 9) Department of Dermatology, Oregon Health and Science University, Portland, OR.

   Statement of purpose: Significant linkage of cutaneous malignant melanoma (CMM) to chromosome 9q21 was previously reported in a Danish pedigree resource and confirmed in Utah high-risk pedigrees. Sequence data was analyzed to identify the responsible predisposition gene. Methods: Whole exome sequencing of pairs of related CMM cases from two pedigrees with evidence of 9q21 linkage was performed to identify the responsible predisposition gene. Results: A rare non-synonymous variant in Golgi Membrane Protein 1 (*GOLM1*), (rs149739829, S307L, chromosome 9:88,650,378 base pair (hg19)) shared in both carriers in one linked pedigree was observed. Segregation of this variant in additional CMM-affected relatives of the index carriers was confirmed. A significant excess of carriers of the variant was observed among 491 unrelated Utah CMM cases compared to 207 population non-cancer controls (OR=9.7, p=0.0045). This significant excess was validated in an independent set of 1,534 CMM cases and 1,146 population controls from Texas, (OR=2.5, p=0.015). Multiple other pedigrees segregating this variant have been identified; protein prediction modeling suggests large changes in the cytoplasm region shown to bind EGFR and other RTK proteins. Conclusion: These findings suggest that *GOLM1* is a CMM predisposition gene.

636F

Association of TLL1 variant with hepatocellular carcinoma developed after eradication of hepatitis C virus. *K. Tokunaga[1], K. Matsuura[2], H. Sawai[1], Y. Kawai[3], K. Kojima[3], M. Nagasaki[3], Y. Tanaka[2], Japanese Genome-Wide Association Study Group for Viral Hepatitis.* 1) Dept Human Genetics, Univ Tokyo, Grad Sch Med, Tokyo, Japan; 2) Dept Virology & Liver Unit, Nagoya City University, Grad Sch Med Sci, Nagoya, Japan; 3) Dept Integrative Genomics, Tohoku Univ, Tohoku Med Megabank Org, Sendai, Japan.

The risk of developing hepatocellular carcinoma (HCC) does not completely disappear even after the eradication of hepatitis C virus (HCV) by anti-viral therapy. We aimed to identify host genomic variations associated with HCC development after achieving sustained virological response (SVR) in chronic hepatitis C (CHC) patients. We conducted a genome-wide association study (GWAS) in 456 Japanese patients who achieved SVR by interferon-based therapy, followed by a replication analysis for 79 candidate SNPs in an independent set of 486 patients. A SNP located within the intron of *TLL1* (rs17047200) on chromosome 4, showed a strong association with developing HCC at a genome-wide level of significance when the results of GWAS and the replication stage were combined (odds ratio = 2.4, P = 2.7 × 10^{-8}) in the allele frequency model. Multivariate analysis showed that the *TLL1* variant was an independent risk factor of developing HCC (hazard ratio = 1.9, P = 0.002) as well as male gender, elder age, lower platelet count, albumin level, and higher post-treatment α-fetoprotein level. In combination of the risk genotypes with distinct other factors, we could propose each prediction model for HCC development in patients with mild and advanced hepatic fibrosis. The *TLL1* expression analyses showed that mRNA levels in human stellate cells increased according to their activation; those in liver tissues of an animal model with hepatic fibrosis and CHC patients increased according to hepatic fibrosis progression. We suggest that the genetic testing of *TLL1* variant would be useful for implementing personalized surveillance of HCC after achieving SVR.

637W

Biology and clinical implications of the 19q13 aggressive prostate cancer susceptibility locus. *G.-H. Wei[1], P. Gao[1], J. Xia[1], C. Sipeky[2], Y. Yang[1], Q. Zhang[1], P. Zhang[3], K. Zhang[1], J. Zhu[3], H-M. Lee[1], S. Suleman[1], N. Giannareas[1], S. Liu[4], T. Whitington[5], J.F. Lindberg[5], X. Wang[4], Q. Huang[1,6], F. Wiklund[5], A. Manninen[1], M.H. Vaarala[7,8], L. Wang[3], J. Schleutker[2,9].* 1) Biocenter Oulu, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland; 2) Institute of Biomedicine, University of Turku, Turku, Finland; 3) Department of Pathology, MCW Cancer Center, Medical College of Wisconsin, Milwaukee, USA; 4) State Key Laboratory of Medical Molecular Biology, Center for Bioinformatics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 6) School of Life Science, Shandong University, Jinan, China; 7) Department of Pathology, Oulu University Hospital, Oulu, Finland; 8) Medical Research Center, University of Oulu, Oulu, Finland; 9) Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Turku, Finland.

Genome-wide association studies have identified the rs11672691 intronic variant in the long non-coding RNA gene *PCAT19* at 19q13 associated with aggressive prostate cancer (PCa). In the current study, we observe a significant association of rs11672691 with additional features indicative of clinically aggressive disease including high tumor stage, PSA progression and the development of castration-resistant PCa in a cohort of 2738 PCa patients. Yet, the underlying molecular mechanisms by which rs11672691 impacts PCa severity remain totally unknown. Here we find that rs11672691 resides in an enhancer element and alters the binding site of homeodomain transcription factors including HOXA2, a novel oncogenic transcriptional regulator with prognostic potential in PCa. HOXA2 displays preferential binding to the aggressive PCa-associated G allele of rs11672691, and regulates the expression of rs11672691-associated genes *PCAT19* and *CEACAM21* that are revealed by an eQTL analysis of four independent data sets with over 1000 prostate samples. Furthermore, *PCAT19* regulates *CEACAM21* expression, and rs11672691 enhancer forms a long-range chromatin interaction with *CEACAM21* promoter region. Suppression of *PCAT19* or *CEACAM21* inhibits PCa cell growth and metastasis. *CEACAM21* upregulation correlates with PCa progression and relapse. Strikingly, CRISPR-Cas9-mediated single nucleotide editing indicates that rs11672691 directly impacts the expression of *PCAT19* and *CEACAM21*, and PCa cellular properties. Importantly, clinical data indicate that the PCa patients carrying the GG genotype of rs11672691 have increased risk of biochemical relapse and decreased metastasis-free survival. In addition, *CEACAM21* expression has obvious prognostic value in the PCa patient group with GG genotype compared with those carrying GA or AA genotype of rs11672691. Together, these results demonstrate the biology and clinical implications of rs11672691 in association with aggressive PCa, and lay the ground work for future translation of this finding to the clinic.

638T

Cisplatin induced gene expression in lung and ovarian cancer cell lines reveal tissue specificity. *S.W. Williams, A.G. Thornburg, A.L. Stark.* University of Notre Dame, Notre Dame, IN 46556.

Lung and ovarian cancers are two of the most prevalent and deadly forms of cancer. Despite extensive research into the nature of these types of cancer individually, broader relationships between the cancers are not as well understood. Cisplatin, a platinum based drug, is used in the treatment of both lung and ovarian cancers. The drug works by attaching to and disrupting DNA in dividing cells, inducing apoptosis. We sought to gain a more complete understanding of the relationship between lung and ovarian cancer by examining the expression profiles of genes in cisplatin treated cell lines. We utilized the Expression Atlas, a continually developing database of gene expression information created by the European Bioinformatics Institute to identify genes with significant expression changes (fold change greater than 5) following cisplatin treatment in ovarian cancer cells. We then evaluated those genes in comparably treated lung cancer cells. Interestingly, MMP12 exhibits a fold change of 7.1, with an adjusted p-value of 2.056E-07, in ovarian cancer cells treated with cisplatin. Despite MMP12 being identified as expressed in lung tissue, we found that this gene exhibits no expression in cisplatin treated lung cancer cells suggesting a difference in how cisplatin-induced gene expression is different in these distinct cancer tissues. MMP12 has been implicated in numerous lung phenotypes including COPD and emphysema, further pointing to fundamental differences between ovarian and lung cancer. Further understanding chemotherapy induced gene expression in different cancer cells can lead to important therapy improvements for patients.

639F

No association between 135G>C polymorphism of *RAD51* and colorectal cancer in Iranian population. *N. Yazdanpanahi[1], R. Salehi[2], S. Kamali[2].* 1) Department of Genetics, Falavarjan Branch, Islamic Azad University, Isfahan, Iran; 2) Department of Genetics and Molecular Biology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran.

Introduction: Colorectal cancer (CRC) is of the most frequent cancers and accounts as the second leading cause of death from cancers in the world. The gene *RAD51* involves in double-strand breaks repair of DNA. Genetic variants within this gene could affect the potential of DNA repair and susceptibility to various tumors such as CRC. The current study evaluated the association of *RAD51* 135G>C polymorphism with sporadic CRC in a subgroup of Iranian population. **Materials and Methods:** One hundred patients with sporadic CRC and 100 controls were investigated from two referral centers in Isfahan. Genotyping for the *RAD51* 135G>C was carried out using polymerase chain reaction-restriction fragment length polymorphism assay. **Results:** No significant association between the *RAD51* 135G>C and sporadic CRC (odds ratio = 0.86, 95% confidence interval = 0.464–1.595) was revealed. The frequency of genotypes and also alleles of the mentioned polymorphism were not significantly different between case and control groups (P = 0.2 and 0.4, respectively). **Conclusion:** The results suggest that *RAD51* 135G>C probably has not a crucial role in Iranian CRC risk and is not a crucial potential risk factor in molecular diagnostics of mentioned disease among Iranian population.

640W

Functional evaluation of the correlation of gene expression changes in pathogenesis and platinum chemotherapy in lung and breast cancer cell lines. *M.L. Yoder, A.L. Stark.* University of Notre Dame, Notre Dame, IN 46556.

Gene expression can be altered by many factors including developmental changes and drug treatment. We sought to investigate whether induced gene expression changes between cancer pathogenesis and chemotherapy response could be correlated. We interrogated the Expression Atlas created by the European Bioinformatics Institute for genes with a significant (p<10-10) expression change in A549 and other lung cancer cell lines. We identified 171 genes that met that threshold. The gene with the largest fold change (8 fold induction) was ALDH1A1 (p = 4.9E-13). Interestingly, there has been a correlation established between ALDH1A1 and cancer pathogenesis in breast cancer so we decided to further follow-up on this gene and others from the list. Because cisplatin is a chemotherapeutic used to treat both lung and breast cancer, we wanted to assess the gene expression changes after cisplatin treatment. We exposed lung (A549) and breast (MCF7) cancer cells to 5 uM cisplatin and collected pellets at 2, 4, 6, 24, and 48 hours after exposure. We utilized quantitative RT-PCR to determine the log-fold change of ALDH1A1 and other genes from the list. Further experimentation is necessary to validate the significance of these findings, however, these results point to a promising new avenue to identify genes' whose expression modification could improve treatment response.

641T

A genetic analysis of the change in prostate-specific antigen concentrations over time. *S. Rashkin[1], T. Hoffmann[1,2], L. Sakoda[3], E. Jorgenson[3], L. Habel[3], J. Shan[3], D. Ranatunga[3], C. Chao[4], N. Ghai[4], D. Aaronson[5], J. Presti[6], C. Schaefer[3], N. Risch[1,2,3], S. Van Den Eeden[3,6], J. Witte[1,2].* 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 3) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 4) Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, CA; 5) Department of Urology, Kaiser Oakland Medical Center, Oakland, CA; 6) Department of Urology, University of California San Francisco, San Francisco, CA.

Prostate-specific antigen (PSA) measures have been used to screen for prostate cancer as increased PSA concentrations are indicative of prostate cancer. However, PSA concentration can vary between individuals due to various factors, leading to only moderate accuracy of PSA testing with a universal threshold. It has been proposed that PSA screening for prostate cancer can be improved by incorporating information from genetic markers influencing PSA concentrations independent of prostate cancer to determine individual-specific screening thresholds. Studies have been conducted examining the relationship between PSA concentrations and genetic variation, but most use PSA levels measured at a single point in time or simplify longitudinal measurements to a single summary statistic. However, knowledge of how genetics affects PSA concentrations over time may better aid in determining a personalized screening threshold for an individual by incorporating how PSA is likely to change for an individual over time independent of prostate cancer to determine if an increase in PSA over time falls within normal bounds for that individual. Here, we explore the genetic basis of changes in PSA over time in Kaiser Permanente health plan members. A previous GWAS in this dataset (28,503 non-Hispanic white Kaiser Permanente health plan members and 18,615 men from replication cohorts) compared each SNP to the median PSA level over time (an average nine measures per man) and found multiple SNPs associated with PSA levels, many of which were not associated with prostate cancer. We now extend this analysis by examining how genetics affects change in PSA over time. First, we model PSA using a linear mixed effects model, adjusting for age at each measurement and principle components, with a random intercept and slope term to model within-person variation in baseline PSA and change in PSA as age increases. We then use the random slopes as the outcome for a GWAS. Preliminary results indicate that, while most significant SNPs are replicating previous results, there may be novel findings associated with change in PSA but not median PSA, including one on 17p12. Additionally, there are some SNPs that were previously associated with median PSA that are not significant in our analysis, indicating that they may be related to baseline PSA rather than changes in PSA over time.

642F

Characterization of global molecular architecture and regulatory mechanisms underlying hepatocellular carcinoma. *H.M. Natri, M. Wilson Sayres, K. Buetow. Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ.*

   Cancer is a complex disease shaped by diverse gene-by-gene and gene-by-environment effects. While genome-wide association studies have generated numerous associations between variants and complex traits, the underlying mechanisms connecting these variants to phenotypes are largely unknown. The majority of known loci associated with complex diseases are located in noncoding regions and are thus likely to affect disease risk by altering gene regulation. Expression quantitative trait locus (eQTL) analysis has emerged as an important tool for translational medicine by enabling the identification of causal variants affecting complex diseases. Detection of eQTLs and epistatic effects between loci requires the calculation of a large number of associations. Multiple-testing creates a computational challenge which has limited the number of detected eQTLs and construction of genetic networks underlying complex diseases. Hepatocellular carcinoma (HCC) is the second leading cause of cancer death worldwide. HCC is influenced by genetic susceptibility, environmental factors such as infections, metabolic syndrome, and alcohol use, and shaped by numerous biological processes, resulting in a high degree of genetic and transcriptional heterogeneity. HCC incidence in the US has doubled in the last 3 decades, attributable to increased rates of obesity. HCC exhibits sex bias in occurrence with 2.1-fold higher rate in HBV+ males, 1.9-fold in HCV+ females, but equal rates in individuals with metabolic syndrome. While sex bias in cancer is partly attributed to sex-specific differences in risk behaviors and environmental exposure, the relationship of these differences to sex as a biological variable has not been systematically investigated. Moreover, most eQTL studies as yet have failed to explore the role of sex-specific impacts of regulatory variants. Here, we utilize an integrative systems genetic approach to investigate the mechanisms involved in HCC. By sex-specific analyses of genotypic, transcriptomic, and phenotypic data from TCGA, GTEx, and ICGC datasets, we produce a detailed outlook on the role of inherited and somatic genetic variation and gene regulation in HCC. We use a Hadoop Big Data framework and multiple mapping algorithms for efficient and robust identification of both cis and trans-eQTLs. We present a characterization of regulatory networks underlying HCC occurrence and susceptibility, highlighting the role of sex-specific impacts of genetic variants.

643W

Molecular characterization of Brazilian patients with hereditary breast and ovarian cancer syndrome: What can we find beyond *BRCA1* and *BRCA2* genes? *S.C.S. Carvalho[1], D.B. Brotto[1], R. Torrieri[2], L.A. Teixeira[1,2], L.F. Araujo[1], J.R. Plaça[3], K.C. Peronni[3], T.O. Anjos[3], W.A. Silva Jr[1,2,3]. 1) Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil; 2) Center for Medical Genomics at General Hospital of the Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil; 3) Regional Blood Center at General Hospital of the Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil.*

   The Hereditary Breast and Ovarian Cancer Syndrome (HBOC) is characterized by families with breast and/or ovarian cancer inherited in an autosomal dominant manner and it comprehends almost 10% and 15% of all cases of breast and ovarian cancer, respectively. The *BRCA1* and *BRCA2* genes were associated with high risk of hereditary breast and ovarian cancer with a prevalence of pathogenic mutations of 20-30% in HBOC cases, leading to a 20-fold increase in risk. Many studies identified other genes, although less penetrant, associated with an increased risk to HBOC. Genetic screening in other genes than *BRCA1*/*BRCA2* in the Brazilian population has shown mutations in *TP53* (p.R337H), *CHEK2* (c.1100delc) and *PTEN* (microdeletions) genes, accounting for 4,16% of the evaluated cases. Thus, this study aimed to characterize the mutational profile of HBOC patients based on the screening of 21 genes associated with DNA damage repair (homologous and mismatch repair), including *BRCA1* and *BRCA2*, in a Brazilian cohort. Using a targeted gene sequencing panel for mutation screening in 94 patients of Southeast Brazil, it was identified 726 point mutations and small insertions/deletions. Almost 28% were mutations located in coding regions and affecting splicing regions and 16,7% were probably pathogenic mutations found in 20 of the 21 analyzed genes. The five top mutated genes were *BRCA1*, *BRCA2*, *CHEK2*, *PALB2* and *PMS2*. The prevalence of pathogenic mutations in *BRCA1* and/or *BRCA2* was 57,44% with one case presenting CNVs in *BRCA2*. This group of patients was younger (the oldest one were diagnosed with 37 years old) than the non-BRCA mutated group, displayed higher tumor grades and comprised the major number of deaths (70%) and metastatic (61,5%) cases, with almost 68% being triple negative breast cancer cases. The prevalence of pathogenic mutations in other genes than *BRCA1* and *BRCA2* was 33%, in which 9,57% of the cases did not present any mutation in *BRCA1*/*BRCA2* genes. In these non-*BRCA* mutated cases it was found pathogenic mutations in *PMS2* and 3 novel mutations in *BRIP1*, *ATR* and *RAD50* genes. Only 2 of 94 samples were not mutated in any gene of the panel and both will proceed to exome analyses. Although all these mutations are in process of Sanger validation, we suggest in advance that other genes than *BRCA1*/*BRCA2* may contribute to HBOC phenotype in 1/3 of the Brazilian cohort. The next steps will focus on functional characterization of these mutations on the DNA repair kinetics.

644T

Fine-mapping analysis of 152 breast cancer risk loci from OncoArray and iCOGS data. *L. Fachal[1], H. Aschard[2], J. Allen[3], D. Barnes[3], J. Beesley[4], M. Ghoussaini[1], S. Khar[1], J.S. Carroll[5], V.N. Kristensen[6], G. Chenevix-Trench[4], A.C. Antoniou[3], J. Simard[7], P. Kraft[8], D.F. Easton[1,3], A. Dunning[1], the Consortium of Investigators of Modifiers of BRCA1/2, the Breast Cancer Association Consortium.* 1) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 2) Department of Genomes and Genetics, Center for Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France; 3) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Cancer Division, QIMR Berghofer Medical Research Institute, Brisbane Australia; 5) Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge, UK; 6) Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Oslo, Norway; 7) Genomics Centre, Centre Hospitalier Universitaire de Québec-Université Laval Research Centre, Quebec city, Canada; 8) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

One hundred and fifty-two breast cancer risk loci have been discovered through genome-wide association studies (GWAS). For most, the causal variants and the mechanisms underlying these associations remain unknown. To identify candidate causal variants and better understand how these loci affect carcinogenesis, we combined association data for all known susceptibility regions with *in-silico* genomic feature annotation. Genotypes for 639,121 variants across these regions were determined using the custom iCOGS and OncoArray platforms, or estimated by imputation, in 84,642 cases and 88,937 controls of European ancestry from the Breast Cancer Association Consortium (BCAC). Using stepwise multinomial logistic regression, we identified independent signals specifically associated with ER-positive (ER+) or ER- breast cancer risk, or associated with both subtypes. We also determined the credible causal variants within each signal, and whether these variants overlapped with transcription factor (TF) binding sites, histone marks, DNase hypersensitivity sites (DHS) (ENCODE, IHEC, GEO), or TF motifs (Factorbook, JASPAR), specifically in breast tissue. Across the 150 confirmed loci, we identified 196 independent association signals. Seventy-nine loci contained two or more independent signals. Sixty-seven signals were more strongly associated with ER+ disease and 28 with ER- disease. We also conducted a meta-analysis between the ER- specific BCAC GWAS and the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) GWAS of 9,439 unaffected and 9,369 affected *BRCA1* mutation carriers; based on this meta-analysis we identified 10 additional signals. For ER+ disease, we found significant enrichment of binding sites for 21 different TFs, including FOXA1, GATA3, and ESR1; and histone marks H3K27ac and H3K4me1 in ER+ breast cancer cell lines coincident with the positions of the candidate variants. For ER- disease, the most significantly enriched annotation is DHS in mammary epithelial tissue, indicative of open chromatin. Empirical Bayes fine-mapping analyses incorporating enriched functional annotations were able to refine association signals and identify individual SNPs with high posterior probability of causality (>95%) in 17 regions, including those near *FGFR2, MAP3K1, and TERT/CLPTM1L*. Our results suggest significant overlap of credible causal variants with active gene regulatory elements and binding sites for certain TFs in a subtype specific manner in breast cancer.

645F

Gene expression imputation identifies novel candidate genes and susceptibility loci associated with risk of cutaneous squamous cell carcinoma. *N.M. Ioannidis[1,2], W. Wang[1], E. Jorgenson[3], M.M. Asgari[3,4], A.S. Whittemore[1,5].* 1) Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 3) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 4) Department of Dermatology, Massachusetts General Hospital, Boston, MA; 5) Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA.

Cutaneous squamous cell carcinoma (cSCC) is a common and costly form of skin cancer, with risk factors including fair skin pigmentation and exposure to ultraviolet radiation. Recent genome-wide association studies (GWASs) have identified genetic susceptibility loci for cSCC, although the roles of some loci in disease pathogenesis remain unclear. Examining associations between cSCC risk and imputed gene expression levels in GWAS subjects has the potential to identify candidate causal genes at these previously associated loci, as well as to identify genes in novel cSCC associated loci for further study. We imputed expression levels of autosomal genes in blood and skin tissue of 6,891 cSCC cases and 54,566 controls participating in the Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort of the Kaiser Permanente Northern California healthcare system. We identified 19 loci containing 33 genes whose imputed expression levels differed between cSCC cases and controls in at least one tested tissue type at a false discovery rate < 10%. Many of these genes are located in novel cSCC susceptibility loci that were not identified in previous GWAS analyses. In addition, several associated genes located in previously associated loci had not been previously considered as candidate causal genes. Thus, our results suggest novel genetic mechanisms contributing to cSCC risk and illustrate the ability of gene expression imputation methods to identify previously undiscovered susceptibility loci from GWAS datasets. Our work also illustrates several limitations in interpreting results based on gene expression imputation that are relevant to future studies of other complex traits.

646W

Novel susceptibility loci associated with *BRCA*-negative *BRCA-like* breast cancer ("BRCAX") for Korean women. *J.Y. Lee[1], J. Kim[2], J.W. Lee[2], S.W. Kim[3], S.H. Ahn[2], J.L. Hopper[4], J. Sung[1].* 1) Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, South Korea; 2) Department of Surgery, Asan Medical Center, Seoul, South Korea; 3) Department of Surgery, Daerim St.Mary's Hospital, Seoul, South Korea; 4) Centre for Molecular, Environmental, Genetic and Analytical Epidemiology, University of Melbourne, Carlton, Victoria, Australia.

INTRODUCTION: *"BRCAX"* refers breast cancer types mimicking those with *BRCA1/2* gene mutation carriers, but genetic screening for *BRCA* genes found to be non-significant tests. In Korea, breast cancer cases with very early onset (age <40) or signs suggesting strong genetic background are allowed to conduct *BRCA* gene test covered by National Health Insurance. Burden from the *BRCAX* spectrum cases is higher for Asians than Westerners. Around 10% of examined were reported to have significant *BRCA1/2* mutations. Usual suspects include unknown rare genes with high penetrance, epigenetics, or interactions between already known loci. In this study, we report initial findings from a genome-wide search for new candidate genes explaining *BRCAX* spectrum cancer. **METHODS:** 1,505 of *BRCAX* cases from Korean Hereditary Breast Cancer (KOHBRA) study and 5,979 control women from Korean Genome and Epidemiology Study (KoGES) were analyzed using genome-wide SNP/SNV panels (>3.3 million markers after imputation). Asian women's cases from the Breast Cancer Association Consortium (BCAC) were used for replication. **RESULTS:** From initial analysis, we have identified 17 loci approaching suggestive significance level (10^{-5}) including 6 known and 11 novel regions. Three novel loci (*PDE7B, UBL3* and *SLC39A11*) were replicated for high-risk Asian women's breast cancer cases from the BCAC (1,482 cases and 3,612 controls). A candidate variant is located on intron of *PDE7B.* Evidences that this variant alters gene expression of *PDE7B* in several tissues and is involved in regulatory functions were observed. Moreover, effect size of known susceptibility genes seemed different from general breast cancer cases. To compare the impact of identified loci for *BRCAX*, we estimated a portion of familial relative risk (%FRR) for breast cancer explained by known and novel variants. Sum of %FRR for *BRCAX* cases was increased from 25.0% to 27.6% by addition of new loci. For known and novel variants which showed significant association results for both datasets, sum of %FRR for *BRCAX* cases (9.57%) was higher than that of breast cancer cases with unknown mutation status for *BRCA* genes (6.50%). **CONCLUSIONS:** We have identified several novel loci that might explain the *BRCAX* cases. Our findings suggest new candidate genes exist and might be used for screening test. **ACKNOWLEDGEMENT:** Data of breast cancer cases were provided by Korean Hereditary Breast Cancer (KOHBRA) and Breast Cancer Association Consortium (BCAC).

647T

Melanoma genetics: Larger sample size identifies novel loci and enables causal inference. *S. Macgregor[1], U. Liyanage[1], J.S. Ong[1], M. Iles[2], D.T. Bishop[2], M.T. Landi[3], P.A. Kanetsky[4], F. Demenais[5], A. Goldstein[3], A. Cust[6], M.H. Law[1], Melanoma genetics consortium.* 1) Statistical Genetics, QIMR Berghofer Medical Research Insitute, Brisbane, Australia; 2) University of Leeds, Leeds, UK; 3) National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA; 4) H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA; 5) Institut National de la Santé et de la Recherche Médicale (INSERM), UMR-946, Genetic Variation and Human Diseases Unit, Paris, France; 6) Sydney School of Public Health, The University of Sydney, Sydney, Australia.

We conducted a genome wide association study (GWAS) meta-analysis comprising 17569 cutaneous melanoma cases and 28006 controls from Australia, Europe and the United States. Haplotype Reference Consortium data was used to allow imputation of rarer variants, allowing us to identify, at genomewide levels of significance, the intermediate penetrance melanoma risk variant in *MITF* initially found via sequencing in families. Furthermore, due to increased sample size, we identified 5 loci not previously detected. The peak SNP at each new locus lies within or near *KIAA0390*, *CDH1*, *GPRC5A*, *DMTN* and *GPR37*. Confidence in determination of which is the key gene (or genes) varies by locus. The *GPR37* hit is a common variant (minor allele frequency~0.25) which is immediately adjacent to the gene *POT1*, a telomere biology gene previous found to harbor rare loss of function variants conferring melanoma risk. For both *KIAA0390* and *DMTN*, the lead SNP altered gene expression levels. We applied a functional enrichment approach using fgwas to test which tissues were most relevant to the peak SNPs. Enrichment was largest for DNAse-I hypersensitive sites in melanocytes, with enrichment also seen in other tissues including skin fibroblasts and thymus. We also used the GWAS data to conduct Mendelian Randomization to evaluate the role of 8 putative risk factors in melanoma risk. Our power to draw strong causal inferences was high for height, obesity, cholesterol and fatty acid levels. For height, we showed that a predicted 10cm increase in height led to a 14% increase in melanoma risk (P=0.02), consistent with observational study estimates. For obesity, cholesterol and fatty acid levels we could rule out all but very small effects of these risk factors on melanoma risk. Our power was lower for vitamin D levels, waist hip ratio, coffee and alcohol consumption; these traits were not significantly related to melanoma risk but at current sample sizes we can only rule out these traits having a large effect on melanoma risk. By October 2017 we anticipate increasing our sample size by several thousand cases; this will increase the number of novel loci as well as strengthening our causal inferences. In this analysis of over 45,000 samples, we identified several novel loci conferring melanoma risk and used Mendelian randomization to draw strong inferences regarding causality of some modifiable risk factors.

648F

A genome-wide association study of cisplatin-induced tinnitus in adult cancer survivors. *B.L. Mapes[1], O. El Charif[1], H.E. Wheeler[2], R.D. Frisina[3], S.D. Fossa[4], D.R. Feldman[5], R.J. Hamilton[6], D.J. Vaughn[7], C.J. Beard[8], C. Fung[9], C. Kollmannsberger[10], J. Kim[11], T. Mushiroda[12], M. Kubo[12], E. Gamazon[13], N.J. Cox[13], P. Dihn[14], S. Ardeshir-Rouhani-Fard[14], P. Monahan[14], L.H. Einhorn[14], L.B. Travis[14], M.E. Dolan[1].* 1) Department of Medicine, University of Chicago, Chicago, IL, USA; 2) Departments of Biology and Computer Science, Loyola University Chicago, Chicago, IL, USA; 3) Departments of Medical and Chemical & Biomolecular Engineering and Communication Sciences & Disorders, Global Center for Hearing and Speech Research, University of South Florida, Tampa, FL, USA; 4) Department of Oncology, Oslo University Hospital, Radiumhospital, Oslo, Norway; 5) Department of Medical Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 6) Department of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, ON, Canada; 7) Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 8) Department of Radiation Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; 9) J.P. Wilmot Cancer Institute, University of Rochester Medical Center, Rochester, NY, USA; 10) Division of Medical Oncology, University of British Columbia, Vancouver, BC, Canada; 11) Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 12) RIKEN Center for Integrative Medical Science, Yokohama, Japan; 13) Division of Genetic Medicine, Vanderbilt University, Nashville, TN, USA; 14) Department of Medical Oncology, Indiana University, Indianapolis, IN, USA.

Despite its efficacy as a curative agent in testicular cancer, cisplatin can result in severe, often permanent toxicities, including tinnitus. Treatment of tinnitus is limited to palliative behavioral modifications and attenuation of associated anxiety and stress. We performed a novel genome-wide association study of cisplatin-induced tinnitus in 835 well-characterized testicular cancer survivors (Median age at assessment: 38 ± 10 (SD) years; median cumulative dose of cisplatin: 400 mg/m^2). There are no external quantitative means to assess tinnitus clinically, therefore, we assessed it as a binary response to the question derived from the validated Scale for Chemotherapy-Induced Neurotoxicity: "Have you suffered in the last 4 weeks from: Ringing in your ears?", with "not at all" (662) respondents as controls, and "Quite a bit" or "Very much"(173) as cases, excluding "a little bit" respondents (467). We observed a significant association between tinnitus and cisplatin dose by 100mg/m^2 [OR: 1.36, p =0.01], age by decade [OR: 1.28, p = 0.003], noise exposure [OR: 1.7, p = 4x10^{-6}], vertigo [OR: 6.4, p = 1.4x10^{-7}], use of psychotropic medication (anxiolytics/antidepressants)[OR: 2.4, p = 0.003], and difficulty hearing in crowds [OR: 7.9, p < 10^{-16}]. Nearly 7.3 million SNPs passed quality control in 835 genetically European individuals. Imputed SNP dosages and covariates including age, noise exposure, cisplatin dose, and 10 genetic principal components were logistically regressed on tinnitus. Intronic SNPs with p < 1.0x10^{-5} were observed in several genes that are associated with hearing loss, neurodegenerative and anxiety disorders. The top signal (p = 1.5x10^{-6}) was the variant rs7606353, about 14kb upstream of *OTOS*; a gene involved in cochlear development and implicated in vestibular disease. Other implicated genes include *OR6B3, DPYSL2,* as well as *PPFIBP1*, a gene also associated with tinnitus in Vanderbilt BioVU (p = 0.002; 105 cases, 7899 controls). Gene set enrichment analysis revealed that the ten most enriched biological processes (FDR q < 0.15) are neurological in origin, including central nervous system axonogenesis (q = 0.04). Heritability was assessed using GCTA, revealing potential polygenic architecture (h^2 = 0.57±0.48, p = 0.10). Discovering new links between tinnitus, genetic predispositions and interactions with cisplatin therapies can pave the way for more effective preventative strategies or novel treatment options for cancer survivors.

649W

Genetic association analysis of advanced neoplasia in a colon cancer screening cohort. *X. Qin[1,2], E.R. Hauser[1,2], A. Stone[3], W.C. Carter[3], D. Humphries[4], M.C. O'Leary[1], T.S. Redding[1], Z.F. Gellad[1,2], D. Provenzale[1,2], D. Lieberman[5,6].* 1) VA Cooperative Studies Program Epidemiology Center, Durham VA Health Care System, Durham, NC; 2) Duke University, Durham, NC; 3) VA Pharmacogenomics Analysis Laboratory, Central Arkansas Veterans Healthcare System, Little Rock, AR; 4) Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare System, Boston, MA; 5) VA Portland Health Care System, Portland, OR; 6) Oregon Health & Science University, Portland, OR.

Numerous advances have been made in identifying genetic risk factors for colon cancer through the analysis of large case-control studies. Fewer studies have been performed to identify genetic factors for cancer precursors, i.e. adenomatous polyps, which are identified and removed during colonoscopy. Understanding which genetic variants increase risk in progression from cancerous precursors to cancer could result in better risk stratification of individuals at risk of colon cancer. We developed a biorepository for the Department of Veterans Affairs (VA) Cooperative Studies Program Study #380 (CSP 380) "Prospective Evaluation of Risk Factors for Large (≥1cm) Colonic Adenomas in Asymptomatic Subjects." CSP 380 is a longitudinal study of 3121 participants aged 50-75 enrolled for screening colonoscopy from 1994-1997. Follow-up using the VA electronic health records is continuing. The biorepository includes 815 individuals with blood (lymphocytes and/or serum), DNA, and/or normal rectal tissue samples. Approximately 170 of these individuals have stored polyp samples frozen at the time of the baseline colonoscopy. To date approximately 1700 archived polyp samples (formalin or Bouin's fixed) have been obtained, representing a longitudinal tissue resource from repeat colonoscopies. Our initial genetic association study has three goals: 1) evaluate association of previously identified colon cancer SNPs in genetic risk of advanced neoplasia; 2) evaluate association of variants in genes differentially expressed in colon cancer from The Cancer Genome Atlas (TCGA) project; and 3) perform a GWAS analysis using DNA from the baseline blood samples. Advanced neoplasia at baseline, defined as at least one tubular adenoma ≥10 mm, adenoma with villous histology, adenoma with high-grade dysplasia, or invasive cancer, was identified in approximately 26% of individuals in the biorepository. Initial power calculations give 80% power for odds ratios of 1.4 at an alpha of 0.05 (colon cancer candidate SNPs), 1.5 at an alpha of 0.0001 and 1.8 at genome-wide significance assuming a minor allele frequency of 0.10. Genotyping with the Illumina Infinium Omni2.5-8 GWAS chip is complete with excellent genotyping quality and will be available for collaborative research. Information on the CSP 380 data and biorepository, study forms and data dictionary is available at https://www.research.va.gov/programs/csp/cspec/initiatives.cfm.

650T

Genome-wide association study identifies novel loci for mammographic breast density. *W. Sieh[1], J.H. Rothstein[1], J.A. Lipson[2], S.E. Alexeeff[3], L. Sakoda[3], R. McBride[1], R. Graff[4], N. Achacoso[3], V. McGuire[2], D.L. Rubin[2], A.S. Whittemore[2], L.A. Habel[3].* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Stanford University School of Medicine, Stanford, CA; 3) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 4) University of California, San Francisco, CA.

Purpose: Mammographic density is a strong risk factor for breast cancer. Most of the variation in mammographic density is explained by hereditary factors that remain largely unknown. We conducted a genome-wide association study to identify novel loci for mammographic density. **Methods**: This study included 24,192 non-Hispanic white women genotyped on the Affymetrix Axiom European array as part the Research Program on Genes, Environment and Health (RPGEH). All women were 40 years or older and underwent screening with full-field digital mammography at 37 Kaiser Permanente clinics throughout Northern California. Density measurements were obtained by a single radiological technologist using Cumulus on the processed images. In Stage 1, 20,311 women imaged using Hologic machines were analyzed. In Stage 2, 3881 additional women imaged using GE machines were analyzed independently. Meta-analysis was performed using METAL. **Results**: Combined analysis of Stages 1 and 2 identified 37 loci associated with dense area (n = 24), non-dense area (n = 15) or percent dense area (n = 16) with p-values $< 5 \times 10^{-8}$ and concordant associations in Stages 1 and 2. Interestingly, alleles at 2 loci were associated with dense area and non-dense area in the same direction, and consequently were not significantly associated with percent dense area. Among the 37 loci, 10 overlapped with known susceptibility alleles for breast cancer and 8 were previously known to be associated with mammographic density. **Conclusions**: Mammographic density is a complex trait influenced by polygenic and environmental factors. We found evidence that some loci have pleiotropic effects on the amount of dense and fatty tissue in the breast, and many loci are also associated with breast cancer. Further analyses are needed to identify potentially functional variants, and biological pathways that may be involved in mammographic density.

651F

Phenome wide association study of breast cancer genetics reveals novel association with seborrheic keratosis. *J. Liu[1], Z. Ye[2], J. Mayer[2], B. Hoch[2], S. Hebbring[1].* 1) Department of Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 2) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI.

Background: The phenome-wide association study (PheWAS) is an innovative and complementary approach to genome-wide association study (GWAS) with particular advantages in identifying disease susceptibility loci using a disease independent approach. The genetic etiology of breast cancer is very complex including both common and rare genetic variants contributing significantly to disease risk. We hypothesized that the PheWAS approach may be utilized to comprehensively study breast cancer genetics and identify novel gene-disease relationships. **Methods**: We evaluated associations between 48,953 genetic variants in 65 breast cancer candidate genes with 7,457 phenotypes in a cohort of 11,000 subjects linked to extensive electronic health record data at Marshfield Clinic. The study population is homogenous with 77% of participants claiming German ancestry. Association testing was conducted by Firth analysis. **Results:** Multiple statistically significant (p<1.37E-10) SNP-disease pairs were identified including expected associations between *TCF7L2* and type II diabetes, along with multiple novel SNP-disease pairs. For example, a novel association between variant rs10069690 on gene *TERT* and Seborrheic Keratosis (p=9.29E-13) was identified. Furthermore, when restricting analysis to only 172 variants previously defined as GWAS significant SNPs for breast cancer, we further identified a statistically significant association between *FTO* and obesity (p=1.36E-08) and suggestive association between *FRY* and arthropathy (p=6.30E-06). **Conclusions:** The PheWAS approach has proven to be a powerful method to discover and rediscover SNP-disease associations. This PheWAS study confirmed the associations between breast cancer genes and obesity and type II diabetes, and further emphasizes the genetic variants associated with breast cancer have pleiotropic properties.

652W

Profiling of genomic alterations of mitochondrial DNA in gingivobuccal oral squamous cell carcinoma indicates that somatic mutations modulate prognosis in patients. *A. Palodhi, S. Ghosh, N.K. Biswas, A. Basu, P.P. Majumder, A. Maitra, India Project Team of the International Cancer Genome Consortium.* National Institute of Biomedical Genomics, Kalyani, West Bengal, India.

Statement of Purpose: Gingivobuccal Oral Squamous Cell Carcinoma (OSCC-GB) is of high prevalence in many lower and middle income countries, including India where this cancer type is primarily driven by the wide-spread use of chewing tobacco. We have undertaken high resolution genomic investigations of the mitochondria in patients of OSCC-GB from India in order to elucidate the role of mitochondrial genome in initiation and progress of this disease. We expect our findings to be useful for patient classification with respect to disease progress. Methods: Mitochondrial DNA sequence data was generated from paired tumor and blood from 89 OSCC-GB patients using Illumina HiSeq 2500 and Ion Torrent PGM at high depth (>2500X). BWA-MEM was used to align the sequence data to the mitochondrial reference sequence (rCRS-revised Cambridge Reference Sequence). VarScan2 was used to detect somatic and germline variants in the mtDNA. The fold change in mtDNA copy number in tumour DNA compared to DNA from paired adjacent normal tissue was measured by Real Time PCR. Information on mitochondrial transcripts were extracted from whole transcriptome RNA-Seq data generated from paired tumor and adjacent normal tissue RNA using the Tuxedo set of tools. Results: The major highlights of our findings are: (i) We found that a significantly higher *(Chi-sq test, P= 0.01)* proportion of OSCC-GB patients who harbor somatic non-synonymous mutations in mitochondrial genes coding for respiratory Complexes III, IV and V have lymph node metastasis compared to the other OSCC-GB patients. (ii) The somatic mutation signature 5'-TCN-3'>5'-TTN-3' was overrepresented *(Chi-sq test, P= 0.0018)* in the mitochondrial genomes, which is different from the tobacco related mutation signature in the autosomal genome previously reported in OSCC-GB. (iii) mtDNA copy number in tumor tissue was observed to be reduced than that in adjacent normal tissue in a significant *(Mann-Whitney test, P< 0.000001)* number of OSCC-GB patients. Additionally, patients with somatic mutations in mitochondrial coding genes, presented lower expression of mitochondrial transcripts *(Two sample unpaired t-test, P= 0.05)*. Overall, we conclude that somatic mutations in mitochondrial DNA might influence prognosis in gingivobuccal oral cancer and the results suggest that such information might be useful in gaining clinical insights in the progress of the disease.

653T

Prognostic inherited genetic variation in non-small cell lung cancer. *F. Abbas Aghababazadeh, X. Wang, B. L. Fridley.* Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL.

Background: Lung cancer is a common and severe disease which ranks the top among cancers worldwide in terms of mortality for both men and women. Environmental exposures, such as tobacco smoking, contribute to over 80% of lung cancer deaths. However, in a small percent of cases, germline single nucleotide polymorphisms (SNPs) have been shown to contribute to the risk of developing non-small cell lung cancer (NSCLC), particularly in families with strong family histories for NSCLC. Much research has been conducted to determine prognostic somatic mutations and drug-able mutations, thus leading to targeted therapies for NSCLC. However, little research has been reported on genetic variation and their impact on survival following diagnosis of NSCLC. Therefore, we set out to determine genetic variation related to overall survival (OS) in two major subtypes of NSCLC using data from the TCGA; lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD). **Methods:** Genome-wide genetic data from the TCGA for subjects with confirmed diagnosis of lung cancer were downloaded along with clinical features such as age, smoking status, race, stage of the disease and overall survival (OS) status. Under each lung cancer subtypes, quality control (QC) and population stratification estimation were applied, followed by the testing for association of each SNP marker; under an additive genetic model; with OS was modeled with Cox proportional hazard regression models, while adjusting for age of diagnosis, stage, and population stratification. Meta-analysis was then completed to combine the results across the two cancer subtypes. **Results:** 411 LUAD and 317 LUSC patients were included in the analyses on 676766 and 766758 SNPs, respectively. The median survival time (MST) for patients corresponding to the each subtype LUAD and LUSC were 20.6 and 19.1 months, respectively. For the analysis of LUAD, we observed two SNPs with p-value < 10-4. Meta-analysis is on-going to combine the results from the two independent genome-wide association studies. **Conclusion:** Future research is needed to understand the genetic variation that relates to the etiology and progression / outcome of this deadly cancer, and how environmental factors by modify these effects. Additional pharmacogenomics studies are also needed to determine genetic variations- either inherited (germline), acquired (somatic), or both - that contributed to toxicity or response to commonly used therapies for NSCLC.

654F

Replication study and functional analysis identify a novel gene associated with gemcitabine-induced leukopenia/neutropenia. *C. Udagawa[1,2,3], Y. Teratake[1], H. Ueno[4], H. Okamura[3], T. Okusaka[4], H. Yamano[5], K. Hirayama[5], N. Nakamura[5], H. Zembutsu[1,2].* 1) Cancer Precision Medicine Center, Cancer Institute, Tokyo, Japan; 2) Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan; 3) New Business Development Life Science Group, Toyo Kohan Co., Ltd., Tokyo, Japan; 4) Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital, Tokyo, Japan; 5) Research and Development Center Life Science Group, Toyo Kohan Co., Ltd., Yamaguchi, Japan.

Gemcitabine is used for the treatment of patients with various tumors. The use of gemcitabine is often limited by unpredictable dose-limiting hematological toxicities. Genetic factors are thought to be one of the causes of individual variability in the leukopenia/neutropenia observed in cancer patients who received gemcitabine therapy. In our previous studies, we identified four novel SNPs associated with gemcitabine-induced severe leukopenia/neutropenia (combined $P = 1.27 \times 10^{-6} - 5.98 \times 10^{-5}$, odds ratio (OR) = 2.60 – 34.00). To identify a genetic marker(s) for gemcitabine-induced leukopenia/neutropenia, which could be feasible for routine clinical use, we conducted a replication study by genotyping the four single nucleotide polymorphisms (SNPs) using 171 patients treated with gemcitabine monotherapy, and a combined study in a total of 345 patients, including 119 patients with at least grade 3 leukopenia/neutropenia and 226 controls. We identified two loci significantly associated with gemcitabine-induced leukopenia/neutropenia (SNPA in *AERG1*, $P_{fisher} = 7.24 \times 10^{-4}$, odds ratio (OR) = 3.11; SNPB in *AERG2*, $P_{fisher} = 1.22 \times 10^{-2}$, OR = 2.39). Combined analysis indicated a strong association of the SNPA in *AERG1* with gemcitabine-induced leukopenia/neutropenia (combined $P = 3.35 \times 10^{-8}$, OR = 3.68), and possible association of the SNP B in *AERG2* with the leukopenia/neutropenia (combined $P = 3.85 \times 10^{-6}$, OR = 3.01). Furthermore, we confirmed that cell death on dose-dependent manner was observed after treatment of gemcitabine and that mRNA expression level of *AERG1* was significantly increased by dose escalation of gemcitabine in cell lines (MCF7 and HS-5). To investigate the functional role of this gene, we examined the relationship between expression level of *AERG1* and gemcitabine-induced toxicity. We further investigated the effect of *AERG1* on cell viability and growth after gemcitabine treatment by shRNA-mediated knockdown system. We observed the resistance to the apoptosis in *AERG1* knocked down cells compared to the control cells. Although further investigation is required to clarify the effect of the SNP on gemcitabine-induced myelosuppression, our results suggested that *AERG1* might contribute to gemcitabine-induced apoptosis.

655W

Risks of melanoma in melanoma-prone families with and without *CDKN2A/CDK4* mutations over four decades. *A.M. Goldstein[1], D.E. Elder[2], M. Curry[3], M.C. Fraser[1], V. Pichler[4], D. Zametkin[4], X.R. Yang[1], M.A. Tucker[1].* 1) Human Genetics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 2) Department of Pathology, University of Pennsylvania, Philadelphia, PA; 3) Information Management Services, Inc., Silver Spring, MD; 4) Westat, Inc., Rockville, MD.

We assessed the risks of melanoma over four decades in American melanoma-prone families. From 1976-1989, eligibility criteria included documented invasive cutaneous melanoma (CM) in two living family members. After 1990, this criterion was changed to three living members with documented CM because of increased CM incidence. We clinically evaluated 1226 members of 56 families and followed them prospectively. All families were tested for mutations in *CDKN2A* and *CDK4*; 29 were mutation positive (Mut+) and 27 mutation negative (Mut-). We compared the rates of invasive CM, both first and second, by family mutation status to population rates in the Surveillance, Epidemiology, and End Results (SEER) program. For estimation of CM risk by calendar period, we divided the time into three intervals: prior to study initiation in 1976, 1976-1989, and 1990+. For estimating CM risk by study participation, we divided time into a retrospective period before the participant's first clinical exam (including initial exam) and a prospective period after initial exam. Comparing the three calendar periods, the risk of first CM decreased slightly. Chiefly, risks in the 1990+ period were lower than risks in 1976-1989 (13.0 vs 30.8 for all; 7.8 vs 15.3 for Mut-, 15.6 vs 38.6 for Mut+). Also, the risks of CM after first exam were approximately one third the risks prior to the first exam, in both Mut+ (12.7 vs 34.6) and Mut- (5.4 vs 16.2) families. Among CM patients, the risk of a second melanoma was increased 10-fold in all families; the risk was somewhat higher in Mut+ families versus Mut- families. Cumulative risks of first melanoma for both Mut+ and Mut- families were significantly different from those in SEER over the same period. All families had earlier onset and substantially higher risk than the general population with Mut+ families having earliest onset and highest risk. Similarly, risks of second CM were significantly elevated in all, Mut+, and Mut- families. Second CM risks in these families were, however, similar to general population risks in SEER from parallel periods suggesting that given susceptibility and exposure to develop one melanoma, individuals continue to be at risk and need surveillance, whether or not they carry a high-risk mutation. In this family-based study spanning four decades, the prospective risk of melanoma has decreased substantially in both Mut+ and Mut- families, during a period when melanoma risk in the general population has rapidly increased.

656T

Impact of interaction between single nucleotide polymorphisms (SNPs) in neural repair genes and patient characteristics on cognitive outcomes in patients with hematological malignancies treated with hematopoietic cell transplantation (HCT). *N. Sharafeldin[1], A. Bosworth[2], S.K. Patel[2], Y. Chen[2], P. Singh[1], X. Wang[3], L. Francisco[1], S.J. Forman[4], F.L. Wong[2], S. Bhatia[1].* 1) Institute for Cancer Outcomes and Survivorship, University of Alabama at Birmingham, Birmingham, AL; 2) Population Sciences, City of Hope, Durate, CA; 3) Department of Mathematics, University of North Texas, Denton, TX; 4) Hematology and Hematopoietic Cell Transplantation, City of Hope, Duarte, CA.

Neural repair genes have been previously linked to cognitive impairment in neurodegenerative diseases such as Alzheimer's disease and in association with chemotherapy-induced cognitive decline in cancer patients. The aim of this study was to explore the association between SNPs on neural repair genes *APOE*, *TOMM40*, and *BDNF* and cognitive impairments after hematopoietic cell transplantation (HCT). We also aimed to examine the impact of interactions between candidate SNPs and key demographic and clinical factors on cognitive impairment. In a prospective cohort of 277 hematologic cancer patients treated with autologous or allogeneic HCT (median age 52.2y; 58.5% males; 68.6% non-Hispanic whites [NHW]), cognitive function was assessed using standardized neuropsychological testing at pre-HCT, 6 months, 1, 2, and 3 years post-HCT. Most common primary diagnoses were non-Hodgkin lymphoma (30.1%) and acute myeloid leukemia (24.7%). We calculated a Global Deficit Score representing the overall performance across all tests in the neuropsychological test battery, and used it in generalized estimating equation models as a binary indicator of cognitive impairment. We identified significant associations between cognitive impairment post-HCT and age ≥50y at HCT (OR=3.0, 95%CI: 1.6-5.7, p=0.001), male sex (OR=2.6, 95%CI: 1.5-4.6, p=0.001), race other than NHW (OR=2.8, 95%CI: 1.5-5.2, p=0.001), and low cognitive reserve (OR=5.9, 95%CI: 3.2-10.7, p<0.0001). Analysis involved 14 candidate SNPs across the 3 genes, LD pruning was used to identify number of independent tests (6 tests) yielding a p-value threshold of 0.0083 using Bonferroni correction for multiple testing. There were no statistically significant SNP main effects, however, we identified significant interactions between rs1519479 on *BDNF* gene and age ≥50y at HCT (OR=0.2, 95%CI: 0.1-0.6, p=0.006); between rs10835210 on *BDNF* gene and race other than NHW (OR=5.5, 95%CI: 1.7-18.1, p=0.005); and between rs1160985 on *TOMM40* gene and low cognitive reserve pre-HCT (OR=10.4, 95%CI: 2.3-47.5, p=0.003). There were no significant interactions between sex and any of the candidate SNPs. These results suggest a significant effect of age, race, and cognitive reserve on cognitive impairment in HCT recipients and that the effect of SNPs on neural repair genes may be modified by these factors.

657F

Novel analysis incorporating multiple tumor characteristics provide evidence of highly heterogeneous associations for known breast cancer risk loci. *T.U. Ahearn[1], H. Zhang[2], J. Lecarpentier[3], K. Michailidou[3,4], R.L. Milne[5,6], F.J. Couch[7], J. Simard[8], P. Kraft[9,10], D. Easton[3,11], P.D.P. Pharoah[3,11], M. Schmidt[12,13], M. Garcia-Closas[1], N. Chatterjee[2] on behalf of the Breast Cancer Association Consortium.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute , Rockville, MD; 2) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 3) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Department of Electron Microscopy/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 5) Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Australia; 6) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global health, The University of Melbourne, Melbourne, Australia; 7) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 8) Genomics Center, Centre Hospitalier Universitaire de Québec Research Center, Laval University, Québec City, QC, Canada; 9) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA; 10) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 11) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 12) Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; 13) Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands.

Background: Breast cancer genome-wide studies have identified 181 single nucleotide polymorphisms (SNPs), many of which have shown differential associations by estrogen receptor (ER) status, or have been identified in ER-negative/triple negative (TN) specific scans. Because multiple clinico-pathological tumor features are correlated with each other, additional analyses are needed to identify specific tumor features driving the observed heterogeneity. **Methods**: Analyses included 106,571 cases and 95,762 controls of European descent participating in 81 studies of the Breast Cancer Association Consortium. Genotypes for 181 previously identified risk SNPs were determined using genome-wide arrays. Subtypes were defined based on ER, progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). We used a two-stage polychotomous logistic regression to perform a global test that simultaneously tests subtype heterogeneity defined by one or more tumor marker. The model was also used to test for heterogeneity by specific markers adjusting for the others, and to efficiently handle missing data on tumor markers via an Expectation-Maximization algorithm. **Results**: Of 181 known risk SNPs, 86 had a P value for the global test for heterogeneity <0.05, and 43 had P < 3x10^{-4} (corresponding to a Bonferoni corrected P<0.05). Of these 43 SNPs, 20 had a P_{het}<0.05 for marker-specific tests for heterogeneity only for ER:15 of which were primarily associated with ER-positive and 5 with ER-negative disease. Two SNPs located in 9q31.2 and 10q26.12 had P_{het}<0.05 only for PR and were associated with PR-positive but not PR-negative disease. One SNP located in 6q25/*ESR1* showed P_{het}<0.05 only for HER2 and was more strongly associated with HER2-positive than negative disease. The remaining 20 SNPs showed P_{het}<0.05 for more than one marker: five of them were primarily associated with TN disease and the rest of other marker combinations. Ongoing analyses incorporating information on tumor grade and histology while accounting for ER, PR and HER2 may provide further insights into the underlying heterogeneity of SNP associations. **Conclusion**: We have identified novel evidence of heterogeneity for breast cancer SNPs by tumor subtypes defined by ER, PR, and HER2. The two-stage regression method is an efficient and systematic approach for identifying heterogeneous associations between SNPs and disease subtypes defined by multiple, correlated disease characteristics.

658W

Development of breast cancer risk prediction for the UK population using the UK Biobank dataset. *K. Alajmi, A. Lophatananon, K. Muir.* University of Manchester , Manchester-UK , United Kingdom.

Several risk prediction models have been developed to estimate the likelihood of developing breast cancer based on specific risk factors in currently healthy individuals within a specific period of time. The majority of these models are however not user-friendly, do not focus on modifiable factors entirely and are not specifically designed for the general public. Our research group is developing an individualised risk prediction model for breast cancer focussing on the modifiable risk factors using the UK Biobank data. A nested case-control study within the 273,467 female participants is being used to develop the model. Two models have been built based on the menopausal status (Pre (64,116 subjects) and post menopause (196,948)). Backward stepwise logistic regression was used to fit the model using the significant risk factors from the univariate regression analyses. Moreover, calibration and discrimination were performed in both models. Additionally a cross validation approach of 10 folds was used to test the internal validation. For model external validation, we will further seek external validation in other Caucasian cohorts. The model provides risk scores derived from the presence or absence of specific risk factors and will be compared to the general public score. The model will allow people to modify their risk profile with appropriate prevention measures. The main goal of the model is to be used in cancer education and prevention. The preliminary results from pre and post-menopause models show moderately good calibration and discrimination (ROC=0.691) among pre-menopaused females. While in the post-menopausal model it was well calibrated and moderately discriminated with an (ROC 0.6882). Further risk factors (SNPs) will be identified and assessed for causality using a Mendelian Randomisation approach (SNPs as surrogate of exposures) and genetic predisposition will also be included in the model with the aim of improving the performance of the model. In conclusion, we are developing an individualised breast cancer risk prediction model for the UK population based on the modifiable risk factors. The model will enable us to educate and to design appropriate interventions tailored to the individual with the aim of assisting them to make appropriate changes to modify their cancer risk profile.

659T

Integrative genomic analyses revealed candidate susceptibility genes in GWAS identified loci for colorectal cancer risk. *J. Bao[1,2], X. Shu[1], J. Wang[3], W. Wen[1], J. Long[1], Q. Cai[1], Q. Liu[3], W. Zheng[1], X. Guo[1].* 1) Division of Epidemiology, Vanderbilt University Medicine Center, Nashville, TN, USA; 2) College of life Science, Fujian Agriculture and Forestry University, Fuzhou,Fujian, China; 3) Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, TN, USA.

Genome-wide association studies (GWAS) have identified approximately 50 loci associated with colorectal cancer (CRC) risk. However, the gene and biological mechanisms underlying these observed associations are largely unknown. We conducted a cis-expression quantitative trait loci (cis-eQTL) analysis for all GWAS-identified single nucleotide polymorphisms (index SNPs) and their flanking genes (± 1 MB region), in three publically available datasets including normal and/or colorectal cancer tissue samples from the Genotype-Tissue Expression (GTEx) (N = 169), The Cancer Genome Atlas (TCGA) (N = 242), and Colomics project (N = 144). Applied the Benjamini-Hochberg adjusted *P* < 0.05 in at least one of three datasets, eQTL analysis revealed that 19 index SNPs are significantly linked with expression levels of 32 genes, suggesting they may be candidate susceptibility genes for CRC. Among them, nine were previously reported as candidate genes linked to six index SNPs, and 23 genes linked to 15 index SNPs are newly identified. To further investigate the possible underlying biological mechanisms, we performed an integrative analysis of Chromatin Immunoprecipitation Sequencing (ChIP-seq), DNase I hypersensitive sites sequencing (DNase-seq), and chromatin-interaction data from various genomic resources that were generated from multiple cell types, including CRC cells, to examine the chromatin interactions between promoter/enhancer regions of all SNPs in strong linkage disequilibrium (LD, R[2] > 0.8) with index SNPs and promoter regions of the identified candidate susceptibility genes. Our result suggested that approximate 60% of these genes including five previously reported and 14 novel genes showed evidence of the chromatin interactions with the regions of functional SNPs (located in promoter/enhancer regions). A functional enrichment analysis using Ingenuity Pathway Analysis (IPA) revealed that the most enriched network for these target genes are related to lipid metabolism and cancer function. Our study revealed several novel candidate genes for CRC risk and shield lights on biological mechanisms for CRC carcinogenesis.

660F

Genome-wide association study of Waldenström macroglobulinemia identifies novel loci. *S.I. Berndt[1], M.L. McMaster[1], J. Zhang[2], S. Slager[3], J. Vijai[4], S. Li[1,5], G. Kleinstern[3], K. Ekström-Smedby[6], E.E. Brown[2], C.C. Chung[1], B. Zhu[1,5], B. Hicks[1,5], B.M. Birmann[7], B.K. Link[8], J. McKay[9], A. Monnereau[10], A. Nieters[11], A. Smith[12], L.R. Teras[13], C.M. Vajdic[14], R. Vermeulen[15], H. Hjalgrim[16], J. Park[17], N. Chatterjee[18], K. Offit[4], J.R. Cerhan[3], N. Rothman[1], L.R. Goldin[1], C.F. Skibola[19], N.E. Caporaso[1].* 1) Division of Cancer Epidemiology & Genetics, National Cancer Institute, Rockville, MD, USA; 2) University of Alabama at Birmingham, Birmingham, AL, USA; 3) Mayo Clinic, Rochester, MN, USA; 4) Memorial Sloan Kettering Cancer Center, New York, NY, USA; 5) Leidos Biomedical Research Inc., Frederick, MD, USA; 6) Karolinska Institutet, Stockholm, Sweden; 7) Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 8) The University of Iowa, Iowa City, IA, USA; 9) International Agency for Research on Cancer (IARC), Lyon, France; 10) Sorbonne Paris Cité, Paris, France; 11) University Medical Center Freiburg, Freiburg, Germany; 12) University of York, York, UK; 13) American Cancer Society, Atlanta, GA, USA; 14) University of New South Wales, Sydney, Australia; 15) Utrecht University, Utrecht, Netherlands; 16) Statens Serum Institut, Copenhagen, Denmark; 17) Dongguk University, Seoul, Korea; 18) Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA; 19) Emory University, Atlanta, GA, USA.

Waldenström macroglobulinemia (WM) is a unique subset of lymphoplasmacytic lymphoma (LPL) that is defined by the presence of an LPL infiltrate in the bone marrow together with a monoclonal IgM protein in the serum. Although WM/LPL has been estimated to have high heritability, the genetic basis for WM predisposition remains unknown. To identify susceptibility loci for WM, we conducted a two-stage genome-wide association study (GWAS) in over 500 WM/LPL cases and 4,300 controls of European ancestry. In stage 1, 217 WM cases and 3,798 controls were genotyped on the Illumina Omni Express or Illumina Omni2.5 platforms and passed quality control filters. The data were imputed using the Haplotype Reference Consortium panel and analyzed using logistic regression. Eleven promising loci ($P<5.0$x10^{-7}) were selected for replication in 312 WM/LPL cases and 564 controls. Two novel loci were convincingly replicated at stage 2 ($P<1.2$ x 10^{-5}) and were genome-wide significant in the combined analysis of both stages: 6p25.3 (*EXOC2*, OR = 21.14, P=1.36 x 10^{-54}) and 14q32.13 (intergenic near *TCL1*, OR=4.90, P=8.74 x 10^{-19}). A possible second signal with weak linkage disequilibrium to the first variant (r^2=0.002) was observed at 14q32.13, but did not reach genome-wide significance (P=3.16 x 10^{-6} after conditioning on the first signal). Both loci are located near genes of particular interest for WM/LPL. Dysregulated TCL1 expression in B cells enhances cell proliferation and survival, leading to cell transformation, and it is aberrantly expressed in WM. EXOC2 interacts with Ral proteins to mediate oncogenic RAS signaling, which is critical for cancer cell survival and proliferation. In laboratory experiments, we found that cells transduced with the *EXOC2* variant allele reporter pCS-EGFP-3'G showed significantly increased EGFP fluorescence compared to the wildtype, suggesting the variant is associated with increased EXOC2 protein levels. Cells harboring the *EXOC2* variant allele were also observed to have significantly increased cell proliferation compared to wildtype cells. We further discovered that the E*XOC2* variant abrogates a miRNA binding site, possibly contributing to gene expression changes for NF-κB pathway constituents. Together these two new, relatively high penetrant loci explain ~3.9% of the familial risk, significantly increasing our understanding of the genetic susceptibility of WM/LPL and pointing to possible biological pathways that may be important for risk.

661W

Evaluation of the impact of rare variants on glioblastoma susceptibility. *R. Bohlender, Y. Yu, H. Zhao, C. Huff.* Epidemiology, MD Anderson Cancer Center, Pearland, TX.

Glioblastoma (GBM) is a deadly disease, with a 5-year survival rate of only 5%. GBM has a strong familial component, with an an estimated heritability of up to 66%. However, only 6% of the excess familial risk of GBM is accounted for by known genetic risk factors. We conducted a whole-exome sequencing case-control study involving 324 GBM cases of European ancestry from The Cancer Genome Atlas (TCGA) and 3507 controls of European ancestry from the National Database for Autism Research (NDAR) to investigate sources of risk. We used the Variant Annotation, Analysis, and Search Tool (VAAST) to conduct a case-control association analysis. VAAST is a gene-based test, which leverages variant level information about phylogenetic conservation and amino acid substitution to prioritize variants likely to influence protein function. Due to differences in sequencing platforms, laboratory preparation, and capture technologies, we observed high levels of Type I error inflation. We used the Cross Platform Association Toolkit (XPAT), which is designed to control for cross-platform artifacts in association studies with heterogeneous sequencing datasets. Using XPAT, we observed consistency between empirical and observed p-values to alpha levels of 0.005. Two genes previously implicated in GBM as susceptibility and somatic driver genes, TP53 and EGFR, were nominally statistically significant (p=7.4x10^{-4} and 3x10^{-3}, respectively). To estimate the risk of GBM conferred by rare variants, we calculated gene-level odds ratios for the following variant categories: All rare variants (minor allele frequency < 0.005), rare likely gene-disrupting (LGD) variants (insertion of premature stop codon, frameshift, or splice-site variant), and rare missense variants predicted to be potentially damaging by both PolyPhen2 and VAAST. Effect size estimates for EGFR (OR=3.290 (1.312, 8.251)) and TP53 (OR=6.546 (1.557, 27.516)) are nominally significant. We observed an excess of rare LGD variants in RAD50 (OR: 5.439, (0.993, 29.812); p = 0.0153), suggestive of a role in GBM risk. RAD50 is a part of the MRN protein complex, involved in double stranded break repair, and has been previously associated with breast cancer risk. We report the full results from our case-control analyses, and provide an initial assessment of somatic-germline interaction in GBM at EGFR, TP53, and RAD50. The results from this work yield new insights into the impact of rare genetic variation on GBM susceptibility.   .

662T

Familial-aggregation of somatic mutations in lung cancers. *Y. Chang[1,2], M. Lee[3], H. Chen[4], S. Su[5], G. Chang[6,7], H. Chen[2].* 1) Department of Clinical Laboratory Sciences and Medical Biotechnology, College of Medicine, National Taiwan University, Taipei, Taiwan; 2) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; 3) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan; 4) Graduate Institute of Toxicology, College of Medicine, National Taiwan University, Taipei, Taiwan; 5) Center of Biotechnology, National Taiwan University, Taipei, Taiwan; 6) Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan; 7) Comprehensive Cancer Center, Taichung Veterans General Hospital, Taichung, Taiwan.

   Genetic variations plays import role in lung cancer, especially somatic mutations. Somatic mutations in epidermal growth factor receptor (*EGFR*) have great impact on tumor progression and treatment strategy. Patients with familial history of lung cancer are more susceptible to lung cancer. The familial-aggregation of somatic mutations is seldom investigated. In this study, we sequenced 143 genes for 27 lung cancer patients from 13 families. These 143 genes are well established oncogene or tumor suppressor, and relevant to clinical trials. The sequenced sites are about 2661 base pairs (bp) with average sequence depth 1655X (range 1005X-2208X). We identified 570 single nucleotide variants (SNVs) in 99 genes. These 570 SNVs could group 21 out of 27 patients into correct families (77.78%) in cluster analysis. The percentage of somatic mutations shared among a family (co-mutation) is 77.99% (range 60.02%-88.08%, p-value 2.2×10^{-16} estimated from Cochran–Mantel–Haenszel test). Additionally, among the identified 570 SNVs, 222 somatic SNVs (%) in 70 genes were not observed in Taiwanese populations. Co-mutation rates were observed significant different between family members both having EGFR-positive lung cancers and at least one members with EGFR-negative lung cancers (21.50% versus 42.10%, p-value of t-test is 0.0427). These results indicated that somatic mutations in lung cancer had the property of familial-aggregation and family members both having EGFR-positive lung cancers had lower co-mutation rate.

663F

Classification of lung adenocarcinoma using optimized support vector machines applied to gene expression data. *J.S. Diaz[1,2], A.C Gonzaga[2].* 1) Faculty of Medicine and Health, University of Leeds, Leeds, West Yorkshire, United Kingdom; 2) Department of Physical Science and Mathematics, University of the Philippines Manila, Manila, Philippines.

   Introduction: Lung cancer has been identified as a major health issue for both developed and developing countries. Non-small-cell lung carcinoma (NSCLC) is the most common type of lung cancer accounting to 85 to 95 % of lung cancer cases with adenocarcinoma as common form. Early diagnosis of lung cancer is important to provide the patients with more timely treatment preventing spread and worsening of the disease. To classify lung cancer patients on the basis of gene expression data, this study used two types of response variables namely 1. Normal and Cancer(Binary) and 2. C1, C2, C3, C4, CM and G1 (multiclass). **Methods:** The 12600 genes in the dataset were initially filtered based on the top ranking genes in terms highest variability. The cut-off is 50 standard deviation units leaving 3312 genes used in the next step of analysis. Support vector machines optimized by genetic algorithm were used to model patient class. **Results**: Results indicate that almost all lung cancer patients were correctly classified by the model at 99% accuracy rate while 94% accuracy rate was observed for normal patients. Overall, the accuracy rate observe for this classification analysis is 96.3 % with 2 genes in the model. Maximal overlap wavelet transformation was applied to the data set using filters of the Daubechies (*DB or db*) class and checked the classification performance versus the analysis without the transformation. Five wavelet filters namely db2, db4, db6, db8 and db10 at eight levels (1-8) were tested. It was found out that in general, models using maximal overlap wavelet transforms of the data were able to classify patients accurately at a rate ranging from 94.1% to 98.75 % composed of 13 to 50 features (functions of genes). For the multiclass problem, it was found out that selected wavelet transforms were able to classify adenocarcinoma patients accurately at a rate of 93.22 to 99.58% and outperformed the model without using the wavelet transform of the data. **Conclusion:** Maximal overlap wavelet transformation improves classification accuracy of the genetic algorithm optimized support vector machines but selected more number of genes as compared with analysis without the transformation. Therefore, if the primary consideration is simplicity in terms of number of genes to consider in decision making, method without the transformation is recommended. On the other hand, allowing some increase in the number of genes in the model leads to more accurate classification.

664W

Whole-genome DNA methylation profiling in breast cancer by the Illumina MethylationEPIC array and the TruSeq EPIC sequencing platforms.
C. He[1], G. Jiang[2], H. Lin[2], N. Lin[1], A. Shendre[1], J. Wan[2], Y. Liu[2]. 1) Cancer Prevention and Control Program, University of Kentucky Markey Cancer Center, Lexington, KY; 2) Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN.

The Illumina MethylationEPIC array (EPIC-array) has been widely used to profile genome-wide DNA methylation in human samples using hybridization techniques. The new Illumina TruSeq Methyl Capture EPIC sequencing (EPIC-seq) leverages the power of next-generation sequencing (NGS) technology and aims to providing much higher resolution data. We empirically compared whole-genome DNA methylation profiling from the two platforms (EPIC-array and EPIC-seq) in our breast cancer study and critically evaluated the performance of the two platforms in identifying differentially methylated CpG sites and islands that were associated with breast cancer risk. We showed that EPIC-Seq covered much more CpG sites within the same region, on average, 1,485,036 CpG sites compared to 155,862 CpG sites by array profiling (about 9.5-fold higher). When we compared the difference of CpG methylation (β) between breast tumor and normal samples, we found that in general Δβ showed high consistence between two approaches in terms of both CpG sites and CpG islands. However, EPIC-seq tended to have larger Δβ compared to EPIC-array in two extremes of the distribution. Despite the high consistency for the differentially methylated sites and regions, we observed that EPIC-Seq provided higher measures in the variation of the beta value (wider dynamic range) comparing to EPIC-Array technology, particularly around the extreme hypo- and hyper-methylation regions (beta value is close to 0 and 1). These results suggested that EPIC-seq had a higher sensitivity in distinguishing methylation difference and may lead to identification of more DNA methylation sites and regions associated with breast cancer risk.

665T

Fine mapping of the 6q25 breast cancer risk locus among Latinas reveals additional variants associated with risk. *J. Hoffman[1], L. Fejerman[2], D. Hu[2], P. Lott[3], S. Huntsman[2], E.M. John[4,5], G. Torres-Mejía[6], M. Echeverry[7], Y. Ding[8], L. Kushi[9], J. Weitzel[8], S. Neuhausen[8], L. Carvajal-Carmona[3], C. Haiman[10], E. Ziv[2], COLUMBUS Consortium.* 1) Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Department of Medicine, Institute for Human Genetics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA; 3) University of California, Davis, Davis, CA; 4) Cancer Prevention Institute of California, Fremont, CA; 5) Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA; 6) Instituto Nacional de Salud Pública, Cuernavaca, México; 7) Universidad del Tolima, Ibague, Colombia; 8) City of Hope Comprehensive Cancer Center, Duarte, CA; 9) Kaiser Permanente Northern California, Oakland, CA; 10) University of Southern California, Los Angeles, CA.

Introduction: Numerous loci have been identified by genome-wide association studies (GWAS) for breast cancer (BC), but the majority have been identified in populations of European and Asian ancestry. We have conducted a GWAS of BC in Latinas and have previously identified at least one locus at 6q25 associated with risk. Here we report on fine mapping at the 6q25 locus in this population. Methods: For discovery we used data from 2396 cases and 7468 controls. We included Latina BC cases and controls from the San Francisco Bay Area Cancer Study (SFBACS), Northern California BC Family Registry (NCBCFR) (942 cases and 699 controls), Kaiser Research Program on Genes Environment and Health cohort (225 cases and 3576 controls), Multiethnic cohort (MEC) (520 cases and 2491 controls), and the Cancer de Mama (CAMA) Study (709 cases and 702 controls). All datasets were imputed to the Haplotype Reference Consortium (HRC). We performed association analysis using logistic regression and adjusted for genetic ancestry by means of principal components analysis. For replication, we used a subset of the CAMA study without GWAS data (310 cases and 464 controls), Latina BC cases (1148) from the Clinical Cancer Genetics Research Network and matched controls (387), and samples from a Colombian study (954 cases and 769 controls). Replication samples were genotyped by Taqman probes except for the Colombian study, which was imputed to the HRC. Results: We identified 34 SNPs that were genome-wide significant (p<5e-8) at 6q25. The top SNP, rs140068132, previously identified in our GWAS as a Latina specific SNP, was associated with a strong protective effect for the minor allele, with an odds ratio (OR) of 0.58 per allele (p=1.6e-12). Of the remaining SNPs, 21 were in high to moderate LD (r2>0.5) with rs140068132 and 7 SNPs were in low LD (r2<0.2). Of these 7 SNPs, 4 were in near perfect LD, characterized by rs851984 (OR 1.26, p=1e-8), which had previously been reported in GWAS of Caucasians and Asians. Another 3 SNPs, characterized by rs3778609 (OR 0.75, p=2e-8), were in low LD with rs140068132 and other previously reported SNPs. This SNP replicated in a combined analysis of additional studies (OR 0.89, p=0.039). Summary: We have performed fine mapping at 6q25 in Latinas and have identified one novel genome-wide significant block of SNPs. This result supports the notion that multiple variants within 6q25 contribute to breast cancer risk, some of which are population specific.

666F

Prevalence of cancer predisposition gene mutations among unselected pancreatic cancer patients. *C. Hu[1], S.N. Hart[2], E. Polley[2], R. Gnanolivu[2], J. Na[2], K.Y. Lee[1], W.R. Bamlet[2], K.G. Chaffee[2], R.R. McWilliams[3], G.M. Petersen[2], F.J. Couch[1,2].* 1) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Oncology, Mayo Clinic, Rochester, MN.

Early detection and intervention offers the possibility of improved survival for individuals diagnosed with pancreatic cancer (PC). In the absence of bio-markers predicting the presence of PC, individuals with genetic predisposition to this disease have the greatest potential for improved survival. Several PC predisposition genes such as BRCA2, PALB2, and ATM have been identified through family-based studies. However, the relevance of all known cancer predisposition genes to PC, and the risks of PC associated with mutations in established PC predisposition genes are not known. Here we evaluated the largest cohort to date of unselected PC patients subjected to hereditary cancer panel testing for the prevalence and risks of mutations in cancer predisposition genes. Germline DNA from 3030 unselected PC patients enrolled in a prediagnostic rapid ascertainment cohort at Mayo Clinic from 2000 to 2016. A customized 37 gene (1733 primer pairs) Qiaseq multiplex PCR-based panel along with Hiseq4000 paired-end sequencing was used to identify germline variants. Pathogenic variants were combined by gene and compared to mutation frequency in ExAC reference controls to estimate associations with PC. A total of 171 pathogenic mutations and 8 CNVs in established PC predisposition genes were found in 6.04% of patients: ATM (2.46%), BRCA2 (2.05%), BRCA1 (0.64%), PALB2 (0.50%), CDKN2A (0.40%). Multiple pathogenic variants in the CHEK2 (1.08%), PMS2 (0.28%), MSH6 (0.23%), MLH1 (0.22%), and TP53 (0.20%) cancer predisposition genes were identified in another 2.01% of patients. Pathogenic mutations in ATM, PALB2, BRCA1, BRCA2, CDKN2A, and MLH1 were associated with moderate or high risks (odds ratio (OR): 2.0-6.0) of pancreatic cancer, whereas mutations in CHEK2 and MSH6 were associated with lower risks (OR<2.0). Overall, mutations in cancer predisposition genes were found in up to 8% of unselected pancreatic cancer cases. Importantly, not all patients with mutations met hereditary cancer testing criteria defined by the American College of Gastroenterology and National Comprehensive Cancer Network. Together these findings suggest that refinement of the guidelines for clinical genetic testing may be needed.

667W

Identification of hereditary mutations of breast cancer susceptibility and candidate genes in high-risk patients. *M. Janatova[1], J. Soukupova[1], K. Zdarilova[1], P. Zemankova[1], M. Borecka[1], L. Stolarova[1], M. Vocka[2], P. Kleiblova[1,3], F. Lhota[4], M. Koudova[4], S. Tavandzis[5], L. Foretova[6], E. Machackova[6], Z. Kleibl[1].* 1) Institute of Biochemistry and Experimental Oncology, Charles University, First Faculty of Medicine, U Nemocnice 5, 12000 Prague 2, Czech Republic; 2) Department of Oncology, General Faculty Hospital and First Faculty of Medicine, U Nemocnice 2, 12000 Prague 2, Czech Republic; 3) Institute of Biology and Medical Genetics, General Faculty Hospital and First Medical Faculty, Charles University, Albertov 4, 12800 Prague 2, Czech Republic; 4) GENNET laboratories Prague, Kostelní 9 170 00 Prague 7, Czech Republic; 5) AGEL laboratories Novy Jicin, Purkyňova 2138/16, 741 01 Novy Jicin, Czech Republic; 6) Department of Epidemiology and Tumor Genetics, Masaryk Memorial Cancer Institute, Zluty kopec 7, 656 53 Brno, Czech Republic.

Hereditary mutations in cancer-susceptibility genes are responsible for tumor development in limited but clinically important subgroup of breast cancer (BC) patients. The carriers of hereditary mutations face a high life-time risk of cancer, which often develops at an early age. Tailored care improving life expectancy in these high-risk individuals requires an unequivocal identification of causative mutations in hundreds of known cancer-susceptibility genes. We have designed the multigene targeted next-generation sequencing panel CZECANCA (CZEch CAncer paNel for Clinical Application) enabling simultaneous analysis of 219 cancer susceptibility or candidate genes that we used for the analysis of 744 high-risk BC patients. The primary analysis revealed 238 truncating (indels, nonsense, splicing, and large genomic rearrangements) germline variants in 78 genes that affected 210 (28.2%) patients. Over half of truncating variants was identified in only seven genes, that included mutations in *BRCA1* (in 6.6% of analyzed patients), followed by mutations in *BRCA2* (3.6%), *PALB2* (2.0%), *CHEK2* (1.3%), *PMS1* (0.9%), *RECQL* (0.8%), and *ATM* (0.7%). Seven patients carried a combination of two truncating variants in these genes. Only a single truncating germline variant was detected in 48 genes that in majority coded for DNA repair proteins. The mean age of all truncating variant carriers, carriers of mutations in seven the most frequently altered genes and carriers of mutations in other 71 genes were nearly identical. However, the mean ages at BC diagnosis (ranging between 39.8 – 60.0 years) differed among carriers of mutations in the most frequently mutated genes. The analysis show that mutations in the most frequently altered genes affected 113/744 (15.2%) patients, but other 13% patients carried mutations in other BC susceptibility or candidate genes. We are aware that the further studies will need to be performed in order to characterize biological importance and clinical significance of variants in non-conventional BC susceptibility genes. However, the early age of BC diagnoses in carriers of variants in these genes warrants such effort. The data obtained from analyzed patients will be used for the development of national-wide database of genotypes and associated phenotypic characteristics for diagnostic purposes. Acknowledgement: This work was supported by grants AZV MZCR 15-28830A and Charles University Projects PROGRES Q28/LF1 and SVV 260367.

668T

Results of *BRCA1/2* testing in 1339 high risk Israeli patients with breast and/or ovarian cancer who were tested negative for the founder mutations. *I.M. Kedar[1], S. Hartmayer[1], N. Ginzach[1], L. Bazak[1], M. Lifshitc-Kalis[1], N. Magal[1], L. Basel-Salmon[1], E. Reinstein[2,4], T. Shapira[2], R. Bernstein Molho[3], Y. Goldberg[1,2].* 1) Recanti Genetics, Rabin Medical Center, Petah Tikva, Israel; 2) Macabbi Health Services; 3) Department of genetics and Department of Oncology, Tel Aviv Medical Center, Ichilov Hospital, Tel Aviv , Israel; 4) Department of Genetics, Meir Medical Center, Kfar Saba, Israel.

   Introduction: We report results of *BRCA1/2* testing in 1339 Israeli breast or ovarian cancer patients who were referred for comprehensive gene testing. **Methods:** During the years 2014-2017 patients who had at least 10% chance of being a *BRCA1/2* carrier by various probability models (BRCAPRO, PENNII or Boadicea) and were tested negative for founder mutations, were offered *BRCA1/2* full sequencing. Non-Ashkenazi patients also had deletion/duplication testing. Patients from ethnic groups with unknown founder mutations were referred directly to sequencing. **Results:** Fifty-three patients (4%) carried a pathogenic or likely pathogenic variant. Sixty-six (5%) had variants of unknown significance. Four of the carriers (8%) had various gene rearrangements. Overall, we identified 46 different pathogenic variants in 53 patients. One likely pathogenic mutation was detected in 4 unrelated Arab patients from the same village. Four variants occurred twice in unrelated patients. 3% of full Ashkenazi Jews, 9% of half Ashkenazi Jews, and 2% of the Non-Ashkenazi Jews had a non-founder mutation. Most of the carriers (92%) had family history of cancer. **Conclusions:** Detection rate of a *BRCA1/2* mutation was lower than expected by probability models in the various Jewish ethnic subgroups. There was a relatively high rate of 16% *BRCA1/2* carriers among Arab patients. This is the first comprehensive report of the yield of *BRCA1/2* sequencing in the Israeli population. Mapping the *BRCA1/2* variant spectrum in the local population should be used to further facilitate and guide *BRCA1/2* testing and interpretation. These findings also suggest that different probability models should be used for patients tested negative for founder mutations.

669F

Whole exome sequencing reveals genes with elevated germline rare variants burden in myeloid malignancy patients. *S. Li, S. Grandhi, C. Bosworth, J. Wang, Z. Li, T. LaFramboise.* Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH.

   Myeloid malignancy is a clonal disorder which is found in hematopoietic stem or progenitor cells. Because genetic changes can contribute to the disease by disturbing key aspects in cells such as self-renewal, proliferation and differentiation, it is important to know which genes are associated with myeloid malignancy in order to increase our understanding, assess disease risk, and develop therapeutic strategies to treat the disease. However, there has been no significant genetic association reported through genome wide association studies (GWAS) for myeloid malignancy. As GWAS has typically focused on common variants, the association of rare variants and susceptibility to myeloid malignancy remains unclear. We have analyzed germline whole-exome sequencing data from a total of 850 myeloid malignancy patients. We first explored the truncating and disruptive rare variants in genes that are found to have crucial roles in cancer development or known to be frequently mutated in myeloid malignancy patients. We found that there are a number of cancer causing genes which are having rare but disruptive variants. We then conducted rare variant association tests and observed that some genes have elevated rare variant burden in myeloid malignancy patients. We also saw a number of genes showing higher-than-expected numbers of patients harboring both germline and somatic disruptions in the same gene. Lastly, we looked at the read depth ratio of reference and alternate alleles in tumor cells for regions that have elevated rare variant burden in the matched normal cell. We observed that regions which bearing rare variants tend to selectively remain in tumors. This may suggest that retaining those regions can have an advantageous effect on tumor progression by unmasking the effect of germline rare variants. Our results show that inherited rare variants can be crucial for susceptibility to myeloid malignancy.

670W

Large rearrangement analysis in *GREM1* and the identification of novel deletions and duplications. *D. Mancini-DiNardo, E. Mundt, T. Judkins, J. Holladay, K.R. Bowles, B.B Roa. Myriad Genetic Laboratories, Inc., Salt Lake City, UT.*

   Background: Duplication of the *GREM1* regulatory region leading to changes in expression has been reported in patients with Hereditary Mixed Polyposis Syndrome (HMPS), a rare condition that has been identified in a small number of families to date. A 40kb duplication in the *GREM1* upstream regulatory region was reported as an Ashkenazi Jewish founder mutation in HMPS families. Individuals with HMPS develop multiple colorectal polyps of varied types at young ages and have a significantly increased risk for colorectal cancer. Here we report on our laboratory's experience with genetic testing for *GREM1* large rearrangements (LRs) as part of a hereditary cancer panel.   **Methods:** *GREM1* was incorporated into multi-gene pan cancer panel testing in July 2016. Next generation sequencing (NGS) dosage analysis was designed to detect LRs in *GREM1* and the upstream region overlapping the adjacent gene *SCG5*. All LRs were confirmed using MLPA. The 40kb duplication was further confirmed by breakpoint-specific PCR and sequencing analysis. LRs in *GREM1* detected between July 2016 and May 2017 were assessed here.   **Results:** Seven unique *GREM1* LRs (4 deletions, 3 duplications) were detected in 26 individuals. The only previously reported LR detected here is the 40kb duplication, which spans a region upstream of the *GREM1* 5'UTR and includes exon 2 of the adjacent *SCG5* gene. Two other duplications have also been detected, which extend beyond the endpoints of the founder duplication. The 40kb duplication is classified as pathogenic because of the collective supporting data; however, the pathogenicity of the other two duplications and four deletions has not been established. Therefore, those LRs are currently classified as variants of unknown significance.   **Conclusions:** LR analysis of *GREM1* in this population has demonstrated more variations than previously reported. As such, care must be taken in the analysis and interpretation of LRs involving *GREM1*. Assessment of the regions upstream and downstream of *SCG5* exon 2 is critical to distinguishing between similar duplications, which may differ in clinical significance. Proper identification and interpretation of LRs detected in *GREM1* is essential to appropriate clinical management.

671T

Identification of genetic variants associated with lung cancer risk among European and African Americans with COPD. *V.L. Martucci[1,2], Y. Bradford[3], W.J. Blot[4,5], P.P. Massion[6,7], M.C. Aldrich[1,2,4]. 1) Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 3) Biomedical and Translational Informatics Program, Geisinger Health System, Danville, PA; 4) Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN; 5) International Epidemiology Institute, Rockville, MD; 6) Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN; 7) Veterans Affairs Medical Center, Nashville, TN.*

   Chronic obstructive pulmonary disease (COPD) is an established risk factor for lung cancer (LC). Candidate genes contributing to risk of each disease have been identified in independent investigations of COPD and LC. Yet genetic variants associated with LC among individuals with concomitant COPD and among African Americans (AA) remain understudied. The goal of this study was to identify genetic variants contributing to LC development among European American (EA) and AA individuals with COPD. We conducted a nested case-control study of COPD and LC cases (N=257) and COPD controls (N=342) among ever smokers from the Southern Community Cohort Study, a large prospective cohort of EAs and AAs in the Southeastern US. Samples were genotyped using the Affymetrix UK Biobank Axiom Array. After quality control assessment, a total of 363,487 SNPs and 125 cases/151 controls remained for AAs and 363,512 SNPs and 125 cases/151 controls remained for EAs. Candidate genes were selected based on prior associations with COPD, LC, and/or tobacco use. Biofilter software was used to identify all SNPs within the candidate genes (+/-10 kb) (N=4,127 SNPs). In AAs, the three most significant results were rs76641146 (OR=2.06, 95% CI: 1.33-3.19) and rs7276417 (OR=1.86, 95% CI: 1.24-2.80) in the *DSCAM* region on chromosome 21 and rs11080470 (OR=0.54, 95% CI: 0.37-0.79) in the *FAM38B* region on chromosome 18. In EAs, the most significant results were rs7237913 (OR=2.03, 95% CI: 1.37-2.99) and rs600695 (OR=0.54, 95% CI: 0.37-0.78) in the *LAMA1* region on chromosome 18 and rs12495100 (OR=0.48, 95% CI: 0.31-0.74) in the *RBMS3* region on chromosome 3. All four gene regions have been previously associated with LC, primarily in European populations, and all except *FAM38B* have been associated with smoking. However, none of these regions passed multiple testing correction using a false-discovery rate (FDR) of 15%. Focused analyses of candidate SNPs from prior GWAS of LC, COPD, or smoking identified a statistically significant association with LC among EA (after applying a FDR correction) with a prior smoking-associated SNP rs9365334 (OR=2.47, 95% CI: 1.54-3.97) in *PARK2* on chromosome 6q26. *PARK2* has been implicated as a tumor suppressor gene in prior studies of LC. While genetic variants may contribute to LC risk among individuals with comorbid COPD, larger sample sizes in diverse populations are needed to confirm or refute potential associations.

672F

Genome-wide homozygosity and risk of four non-Hodgkin lymphoma subtypes. *A. Moore[1], J. Vijai[2], A.R. Brooks-Wilson[3,4], N. Camp[5], K.E. Smedby[6,7], C.F. Skibola[8], G. Salles[9,10,11], D. Albanes[1], P. Cocco[12], G.G. Giles[13,14], E. Kane[15], B.K. Link[16], L.R. Teras[17], C.M. Vajdic[18], Y. Zhang[19], Q. Lan[1], S. Chanock[1], J.R. Cerhan[20], N. Rothman[1], S.I. Berndt[1], InterLymph NHL GWAS Initiative.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA; 2) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA; 3) Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada; 4) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada; 5) Division of Hematology and Hematologic Malignancies, Department of Internal Medicine, Huntsman Cancer Institute and University of Utah School of Medicine, Salt Lake City, Utah, USA; 6) Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden; 7) Hematology Center, Karolinska University Hospital, Stockholm, Sweden; 8) Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, Georgia, USA; 9) Department of Hematology, Hospices Civils de Lyon, Pierre benite Cedex, France; 10) Department of Hematology, Université Lyon-1, Pierre benite Cedex, France; 11) Laboratoire de Biologie Moléculaire de la Cellule UMR 5239, Centre National de la Recherche Scientifique, Pierre benite Cedex, France; 12) Department of Public Health and Clinical Molecular Medicine, Università di Cagliari, Cagliari, Sardinia, Italy; 13) Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia; 14) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia; 15) Department of Health Sciences, University of York, York, United Kingdom; 16) Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, Iowa, USA; 17) Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA; 18) Centre for Big Data Research in Health, University of New South Wales, Sydney, New South Wales, Australia; 19) Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, USA; 20) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA.

Genome-wide association studies (GWAS) have successfully uncovered multiple germline variants contributing to the risk of specific subtypes of non-Hodgkin lymphoma (NHL); however, much of the heritability remains unknown, even after considering all common SNPs simultaneously as predictors and estimating the variance explained. Thus far, little consideration has been given to the potential role of recessive genetic effects on disease risk. Interest in examining runs of homozygosity (ROHs) across the genome, as one aspect of recessive genetic inheritance, has grown in recent years. Identification of ROHs in an outbred population may provide insight into the contribution of recessive effects on disease risk. We evaluated the association between genome-wide homozygosity and NHL risk by estimating the fraction of ROHs (FROH) across the autosomes in eight previously conducted GWAS of NHL of European ancestry, including 2,761 chronic lymphocytic leukemia (CLL) cases, 3,715 diffuse large B-cell lymphoma (DLBCL) cases, 2,563 follicular lymphoma (FL) cases, and 720 marginal zone lymphoma (MZL) cases and up to 7,287 controls. Identified ROHs were a minimum of 1.5 megabases in length. We performed logistic regression to estimate odds ratios and 95% confidence intervals for the association of FROH with risk of NHL subtype, controlling for age, sex, and top ten principal components of ancestry. Regression results for each NHL subtype by study were combined using fixed-effect meta-analysis. Preliminary results indicate a positive association between FROH and risk of CLL (β = 23.8, 95% CI = 13.1-34.4, p = 1.19×10^{-6}), corresponding to an OR of 1.27 for every 1% increase in FROH. We also observed a positive association between FROH and risk of FL (β = 12.9, 95% CI = 2.31-23.5, p = 0.017), corresponding to an OR of 1.14 for every 1% increase in FROH. We did not find evidence to support an association between FROH and risk of DLBCL (p = 0.15) or MZL (p = 0.13). Our findings indicate a role for recessive genetic effects in the etiologies of CLL and FL. These results may help explain some of the missing heritability of these complex diseases. More in-depth analyses of specific regions of homozygosity may lead to additional insight into genetic susceptibility to both CLL and FL.

673W

Whole-genome sequencing analysis of HPV31 and HPV35 reveals variability in cervical cancer risk. *M. Pinheiro[1], Y. Xiao[1], M. Yeager[1,2], M. Cullen[1,2], J.F. Boland[1,2], R. Burk[4,5], Z. Chen[6], K. Yu[1], N. Wentzensen[1], M. Steinberg[1,2], S. Bass[1,2], L. Burdett[1,2], T. Raine-Bennett[3], T. Lorey[3], P.E. Castle[4], M. Schiffman[1], L. Mirabello[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick, MD; 3) Regional Laboratory and Women's Health Research Institute, Division of Research, Kaiser Permanente Northern California, Oakland CA; 4) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 5) Departments of Pediatrics, Microbiology and Immunology, and Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY; 6) Department of Microbiology, The Chinese University of Hong Kong, Hong Kong.

Background: High-risk human papillomavirus (HPV) infection causes virtually all cervical cancers. HPV31 and HPV35 are attributed to ~4% and 2% of cervical cancers worldwide, respectively. They are the closest genetically related to HPV16, the most carcinogenic type causing ~60% of cancers. Within each HPV type there are Evolutionary-derived variant lineages, which are linked to strong differences in carcinogenicity for HPV16, but this is not known for HPV31 and HPV35. We do not know why HPV types and variant lineages differ profoundly in risk of cancer, but viral genetic variation is likely part of the answer. We conducted the largest case-control study of HPV31 and HPV35 genetic associations with risk of cervical precancer/cancer using HPV whole-genome sequencing data. **Method**: 2,073 HPV31 and 608 HPV35 cervical cell specimens from the NCI-KPNC PaP study were whole-genome sequenced using custom Ion AmpliSeq panels. The cohort included CIN2 (HPV31: 438; HPV35: 179) and CIN3 precancer (HPV31: 342; HPV35: 94) and cancer (HPV31: 7; HPV35: 1) cases, compared with controls with benign infections ($\leq$CIN1; HPV31: 1,286; HPV35: 334). Logistic regression models were used to estimate CIN3/cancer (CIN3+) risk among HPV variant lineages, each individual SNP, and to determine if race/ethnicity modified risk. Burden tests assessed the combined effects of rare variants (MAF<0.03). **Results**: HPV31 A and B lineages had higher risks of CIN3+ compared to the C lineage (OR=1.7, 95%CI=1.2-2.3; OR=1.6, 95%CI=1.1-2.4, respectively); the A2 sublineage had the highest risk (OR=2.1, 95%CI=1.2-3.6). Risks were not modified by a woman's race/ethnicity (P>0.1). Eighty-nine individual SNPs were significantly associated with CIN3+ (P<0.05), 27 of which mapped to the viral regulatory region (URR) which regulates both viral replication and transcription (strongest SNP, OR=5.3, 95%CI=1.7-16.7). Rare non-silent SNPs in the L1 gene were significantly more frequent in the controls (Burden P=0.03). For HPV35, there was a statistically significant difference among lineages by a woman's race/ethnicity (P=2.8×10^{-8}). African-American women had a higher risk of CIN3+ compared with European women if infected with A2 (OR=11.7, 95%CI=1.2-114). For A1, European women had a higher risk (OR=3.4, 95%CI=1.1-10.2). Individual SNP analyses revealed 8 SNPs associated with CIN3+ (P<0.05). **Conclusion:** HPV31 and HPV35 genetic variation influences risk of cervical cancer, and HPV35 risk is modified by a woman's race.

674T

Height and body mass index as modifiers of breast cancer risk among 22,588 carriers of *BRCA1* or *BRCA2* mutations: A Mendelian randomization study. *F. Qian[1], J. Mitchell[2], G. Chenevix-Trench[3], M.A. Rookus[4], O.I. Olopade[1], T. Rebbeck[5], D. Huo[6] on behalf of Consortium of Investigators for the Modifiers of BRCA1/2 (CIMBA).* 1) Department of Medicine, University of Chicago, Chicago, IL; 2) Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 3) Department of Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 4) Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands; 5) Department of Epidemiology, Dana Farber Cancer Institute, Boston, MA, USA; 6) Department of Public Health Sciences, University of Chicago, Chicago, IL, USA.

   Background Inheritance of *BRCA1/2* mutations confers a high life-time risk of breast cancer, and there are potential modifiers of this risk. Prospective cohort studies and Mendelian randomization analysis have shown height and body mass index (BMI) to affect breast cancer risk in the general population, but it remains unclear whether these associations are causal in *BRCA1/2* mutation carriers and their respective strengths. **Methods** We used the Mendelian randomization method to evaluate the association between height/BMI and breast cancer in the Consortium of Investigators for the Modifiers of *BRCA1/2* (CIMBA), consisting of 14,676 *BRCA1* and 7,912 *BRCA2* mutation carriers, with 11,451 cases of breast cancer. We created a height genetic score (height-GS) using 586 height-associated genetic variants and a BMI-GS using 93 BMI-associated genetic variants, weighted by their respective effect estimates in previous genome-wide association studies. We modeled the association between each GS as an instrumental variable and breast cancer risk using Cox proportional hazards models. **Results** We found body height was positively associated with breast cancer risk (hazard ratio [HR]: 1.07 per 10 cm increase in height, 95% confidence interval [CI]: 1.01–1.12; P=0.01) and BMI was inversely associated with breast cancer risk (HR: 0.95 per 5 kg/m² increase in BMI, 95% CI: 0.92–0.98; P=0.001). The height-GS explained 13.4% of the variation in height and the BMI-GS explained 2.6% of variation in BMI. The height-GS was positively associated with breast cancer risk, though this was not statistically significant (HR: 1.06 per 10 cm increase in genetically predicted height, 95% CI: 0.98–1.15; P=0.13). The BMI-GS was found to be inversely associated with breast cancer risk (HR: 0.91 per 5 kg/m² increase in genetically predicted BMI, 95% CI: 0.83–0.99; P=0.04). We found similar results when evaluating the pooled HR of individual height or BMI-associated genetic variants using an inverse-variance fixed-effects meta-analysis. The associations were similar for *BRCA1* and *BRCA2* carriers. **Conclusion** The observed associations between height/BMI and breast cancer risk in *BRCA1/2* carriers corroborate findings from prior studies in the general population. Under Mendelian randomization assumptions, these associations could be interpreted as causal. Further studies are warranted to elucidate the precise biological pathways that mediate the relationship between height/BMI, *BRCA1/2*, and breast cancer.

675F

A *cis*-eQTL genetic variant of the cancer-testis gene *CCDC116* is associated with risk of multiple cancers. *N. Qin[1,2,3], C. Wang[1,2,3], Q. Lu[1,2,3], T. Huang[1,2,3], J. Dai[1,2,3], H. Ma[1,2,3], G. Jin[1,2,3], C. Wu[4], D. Lin[4], H. Shen[1,2,3], Z. Hu[1,2,3].* 1) Department of Epidemiology and Biostatistics, Nanjing Medical University, Nanjing, Jiangsu, China; 2) State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing, Jiangsu, China; 3) Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, Jiangsu, China; 4) State Key Laboratory of Molecular Oncology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

   Recent studies have found that cancer-testis (CT) genes, which are expressed predominantly in germ and cancer cells, may be candidate cancer drivers. Because of their crucial roles, genetic variants of these genes may contribute to the development of cancer. Here, we systematically investigated the association of common variants in CT genes and their promoters for the risk of lung cancer in our initial GWAS (2,331 cases and 3,077 controls), followed by *in silico* replication using additional 10,512 lung cancer cases and 9,562 controls. We found a significant association between rs3747093 located in the *CCDC116* promoter and lung cancer risk (OR=0.91, P_{meta}=7.81×10⁻⁶). Although *CCDC116* was expressed at lower levels in somatic tissues compared to the testis, the protective allele A of rs3747093 was associated with decreased *CCDC116* expression in many normal tissues, including the lung (P=8.1×10⁻¹³). We subsequently genotyped this variant in another four commonly diagnosed cancers (gastric, esophageal, colorectal and breast cancers), as we found expression quantitative trait locus (eQTL) signals for rs3747093 and *CCDC116* in their corresponding normal tissues. Interestingly, we observed consistent associations between rs3747093 and multiple cancers (gastric cancer: OR=0.85, P=2.21×10⁻⁴; esophageal cancer: OR=0.91, P=2.57×10⁻²; colorectal cancer: OR=0.80, P=1.85×10⁻⁶; breast cancer: OR=0.87, P=1.55×10⁻³). Taken together, the A allele of rs3747093 showed significant protective effects on cancer risk (OR=0.88, P_{pool}=6.52×10⁻¹³) in an Asian population. Moreover, our findings suggest that low abundance expression of CT genes in normal tissues may also contribute to tumorigenesis, providing a new mechanism of CT genes in the development of cancer.

676W

Multiple-gene sequencing revealed novel mutation characteristics beyond BRCA1/2 in Chinese women with familial breast cancer. *Y. Shi[1], W. Li[1], Y. Deng[1], L. Yang[1], T. Zhou[1], X. Li[1], Y. Wu[1], Q. Zhi[1], Y. Yu[1], H. Weng[1], Z. Lan[1], D. Zhao[1], H. Qin[1], G. Sun[1], M. Swift[2], B. Xu[3].* 1) TopGene Tech Co., Ltd, Guangzhou, Guangdong , China, China; 2) Top Medical (Europe) Co., Ltd. Green Pastures, Stockport, SK4 3RA, United Kingdom; 3) Department of Medical Oncology at the Cancer Hospital and Institute, Chinese Academy of Medical Sciences(CAMS), and Peking Union Medical College(PUMC) in Beijing, China.

   Background: Germline mutations on BRCA1/2 are predisposing risk for a woman to develop breast or/and ovarian cancer. However, mutations on BRCA1/2 genes could not account for all the familial breast cancers. As a result, mutations on other genes which is associated with familial breast or/and ovarian cancer should be considered. **Methods:** In this study, a total of 412 Chinese women patients with familial primary breast cancer were enrolled from 56 Chinese hospitals. A pre-designed multiple-gene panel (22 genes including BRCA1/2) was used to enrich target sequences. PE150 sequencing reads were produced by NextSeq CN500 platform. Germline mutations were identified by strict bioinformatics method and were further validated by Sanger sequencing. For each validated variant, an in-house developed method which is superior to previous reported methods was used to evaluate its clinical effect. **Results:** 71 out of 412 patients (17.23%) were identified carrying at least one deleterious mutation (classified as likely pathogenic or/and pathogenic). Among these 71 patients, 49(69.01%) carried BRCA1/2 mutations, 14(19.72%) carried MMR mutations, 3(4.23%) carried PALB2 mutations and the remaining 5 patients each carried mutation from RAD50 (1.41%), ATM (1.41%), CHEK2 (1.41%), PTEN (1.41%), TP53 (1.41%) . Of all the deleterious mutations, one likely pathogenic mutation identified as BRCA1: c.5470_5477delATTGGGCA(NM_007294) causing frameshift at codon 1824 in one patient (proband(III.2)) was also found in other 4 family members from the same pedigree. For the remaining 70 patients with deleterious mutation(s), 69 (98.59%) out of it, same mutation(s) in proband could be found in at least one other family member in the same pedigree. In addition to the above, our results also showed that about 3% patients were concluded as carrying mutation(s) with "Uncertain Significance (VUS)" on BRCA1/2 genes, while 10% patients were identified with VUS mutations on <u>non</u>-BRCA1/2 genes. **Conclusions:** Multiple-gene sequencing technology has been confirmed as an efficient solution for identifying novel mutation characteristics beyond BRCA1/2 gene in Chinese women with familial breast cancer. In our study population, about 40% susceptible mutations occurred in non- BRCA1/2 gene which broaden the mutation spectrum for familial breast cancer in Chinese women. In the future, independent research in large sample size of patients is worth conducting to investigate non-BRCA1/2 genes.

677T

Genetic variants in the 8q24 region are associated with prostate cancer risk in Mexican men. *B. Silva[1], R.A. Aragon[2], G. Castillo[2], O.S. Frausto[3], D.I. Ibarra[2], M.B. Calao[2], E. Torres[2], E. Macias[1], K. Penuelas[4], M.A. Bermudez[4].* 1) Immunogenetics, CIBIN-IMSS, Monterrey, NUEVO LEÓN, México; 2) UMAE 25, IMSS, Monterrey, NUEVO LEÓN, México; 3) HG Z33, IMSS, Monterrey, NUEVO LEÓN, México; 4) Molecular Biology, CIBIN-IMSS, Monterrey, NUE-VO LEÓN, México.

   Prostate cancer (PCa) is the second most common cancer diagnosed in males in the developed world. Several independent genetic association studies identify single nucleotide polymorphisms (SNPs) within the 8q24 region as a risk factor for PCa. The case control study was performed in a Mexican male population from the Mexican Social Security Institute (IMSS) consisting of 185 patients with biopsy confirmed adenocarcinoma of the prostate and abnormal serum prostate-specific antigen (PSA) and 384 male volunteer controls older than 45 years who had no history of PCa. We selected 4 SNPs at 8q24, those most strongly associated with PCa risk in several studies (rs 16901979C/A, rs6983267G/T, rs1447295 C/A, rs7837328 G/A), and they were genotyped by 5' exonuclease TaqMan assays. Statistical analysis was performed using: Epi Info V-7, HWE calculator and SNP Stats software. We found significant association between PCa and three SNPs in 8q24, the rs16901978 AC genotype (OR=2.21; 95% CI=1.3-3.7, Pc=0.01), rs1447295 AC genotype (OR=2.07; 95% CI=1.37-3.13, Pc=0.002), and rs7837328 AG genotype (OR=1.64; 95% CI=1.14-2.34, Pc=0.03). Pairwise linkage disequilibrium (LD) analysis was conducted. We found evidence of strong LD (D' greater than 0.5) within the 8q24 SNPs. Haplotype analysis revealed haplotypes containing the risk alleles, confer high risk to PCa. In conclusion, our study established the association between SNPs in 8q24 with increased susceptibility to PCa in Mexican male population. However, more validation studies and functional assays should be performed to prove true effect in the susceptibility to disease.

678F

Genetic variation and gastric cancer susceptibility in Western Honduras. *G. Tavera[1], R.L. Dominguez[3], J. Bartlett[1], B.G. Schneider[2], C.B. Cherry[2], D.A. Israel[2], J. Romero-Gallo[2], M.B. Piazuelo[2], R.M. Peek[2], K.T. Wilson[2], S.M. Williams[1], D.R. Morgan[2].* 1) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Division of Gastroenterology, Hepatology, and Nutrition, Vanderbilt University, Nashville, TN; 3) Hospital de Occidente, Santa Rosa de Copán, Copán, Honduras.

Gastric cancer is the third leading cause of cancer mortality, and the leading infection-associated cancer. Central America has one of the highest global incidences and mortality rates of gastric cancer in the world (mortality-to-incidence ratio of 0.82). We analyzed samples from gastric biopsies taken from a population-based, case-control study in rural Western Honduras (n = 672 cases and 722 controls), to evaluate 94 single-nucleotide polymorphisms (SNPs) in distinct inflammatory and nutritional genes that may be implicated in the development of gastric cancer. Quality control excluded 11 SNPs with a minor allele frequency of less than 0.05 and one SNP was removed, which was out of Hardy–Weinberg equilibrium ($p < 1.0 \times 10^{-16}$). Initial association tests identified rs1927914 on chromosome 9, in the toll-like receptor-4 (TLR4) gene and the C allele with gastric cancer risk (p = 0.000674; OR = 1.31). This SNP has been associated with risk of prostate cancer in multiple studies. The SNP rs530537 on chromosome 11, in the caspase-1 gene (CASP1), with the A allele, was associated with mild protection from gastric cancer risk (p = 0.000925; OR = 0.78). Caspase-1 is involved in protein cleavage of inflammatory cytokines. Further analysis will be performed exploring the existence of potential gene networks. Identification of SNPs related to inflammatory and nutritional processes may point to mechanisms of pathogenesis and targets for therapy and preventative screening for gastric cancer.

679W

A model averaging approach for improved *in silico* variant prediction. *Y. Tian, S. Li, M.H. Black, A. Elliott, H.M. Lu, D. Qian.* Bioinformatics and Computational Biology, Ambry Genetics, Aliso Viejo, CA.

A variety of *in silico* scores are currently utilized in missense variant assessment as evidence of pathogenicity. A Bayesian Integrated Evaluation method was developed to leverage the information from these scores, in which *in silico* predictors are initially selected manually or by stepwise regression. However, the models produced often contain uninformative predictors and do not quantify their gene-specific importance. A robust and stable gene-specific model is needed to calibrate information from *in silico* tools. Using data from 129,691 patients who underwent genetic testing at a single diagnostic laboratory in 2012-2016, we assembled a collection of 1,292 missense variants in 8 genes (*BRCA1*, *BRCA2*, *CDH1*, *TP53*, *MLH1*, *MSH2*, *MSH6* and *PMS2*), each classified according to a five-tier classification system per ACMG guidelines. Nine *in silico* scores (Polyphen2, SIFT, Grantham, CADD, AGVGD, MutationAssessor, phastCons, phyloP and REVEL) were collected for all variants. Gene-specific Model Averaging Logistic Regression (GMALR) models were trained on known classified variants using the nine *in silico* scores as predictors. We assessed each predictor using relative variable importance (RVI) defined as the sum of Akaike weights over all candidate models (range 0 to 1). The predicted rate of variants of unknown significance (VUS), positive predictive value (PPV), negative predictive value (NPV), and coefficient mean squared error (MSE) with 10-fold cross validation, were compared for GMALR vs. Gene-specific Stepwise Logistic Regression (GSLR) models. GMALR results indicated that the relative importance of each *in silico* predictor varied by gene. For example, AGVGD, CADD, phyloP and REVEL had the highest RVI in *BRCA1*, while MutationAssessor, phastCons and phyloP had the highest RVI in *BRCA2*. Predicted VUS rates were 39.5% and 45.2% for GMALR and GSLR models respectively; PPV and NPV were ~99% for both models. Thus, GMALR reduced the predicted VUS rate and had similarly low misclassification rates compared to GSLR. GMALR models also improved model robustness by a meaningful margin compared to GSLR models; mean ± SD of MSE across all genes was 1.7 ± 1.2 and 5.0 ± 4.9 for GMALR and GSLR models, respectively. Our results underscore the improved prediction performance and model robustness of the GMALR approach when evaluated in the context of *in silico* variant prediction and as compared to stepwise regression methods such as GSLR.

680T

SNP-SNP interactions associated with prostate cancer aggressiveness in African Americans. *H. Tung, H. Lin.* Biostatistics, LSU Health New Orleans, New Orleans, LA.

African American (AA) men tend to have a higher incidence rate of prostate cancer (PCa) and are also more likely to have aggressive PCa. Although single nucleotide polymorphisms (SNPs) associated with PCa aggressiveness in AAs have been reported, impact of SNP-SNP interactions are under studied. The study objective is to identify SNP-SNP interactions and build a prediction model to predict aggressive PCa in AAs. A total of 1,933 PCa patients from the Multiethnic Cohort (MEC) Study were included. We evaluated 5,064 SNPs in the four pathways (angiogenesis, mitochondria, miRNA, and androgen metabolism). PCa aggressiveness is defined as prostate cancer with a Gleason score >= 8/undifferentiated or regional/distant SEER staging. All models were adjusted for study site and the first four principal components for population stratification. SNP main effects associated with PCa aggressiveness were evaluated using logistic regressions. For SNP-SNP interactions, the SNP Interaction Pattern Identified (SIPI), a novel statistical method, was used. SIPI take advantages by considering the inheritance modes (additive, dominant, recessive), code direction and the non-hierarchical models structure. None of these SNP findings reached the stringent the Bonferroni significant level. The promising SNP main effects (p-value<1×10^{-3}) and SNP-SNP interaction pairs (p-value<1×10^{-6}) were candidates for building the prediction model. The model with the identified SNP-SNP interactions has a good discrimination ability for PCa aggressiveness than the one without these interactions. The values of area under the Receiver Operating Characteristic curve (AUROC) are >0.8 and 0.65, respectively. Among 12 SNPs and 56 SNP pairs in the final model, the most important SNP is rs3789889 in *SYK*, which was shown as the main effect and was involved in several leading interaction terms. The top SNP interaction pair is rs3789889 in *SYK* and rs7358800 in *COL4A2* (p=9.3×10^{-9}). *SYK* has been reported to be a candidate kinase target for the treatment of advanced prostate cancer previously. This study demonstrated that SNP-SNP interactions may have an important impact on predicting PCa aggressiveness in AAs.

681F

DNA methylation reveals distinct signatures in profiling tumor-infiltrating immune cells. *X. Wang[1], C. Liang[2], L. Wang[3].* 1) Department of Biostatistics and Bioinformatics, Moffitt Cancer Center and Research Institute, Tampa, FL; 2) Yale University; 3) University of Texas MD Anderson Cancer Center.

Introduction: Immunotherapy has produced promising results in treating cancers. Further understanding of the crosstalk between tumor and its immune microenvironment is likely to have an immediate impact on therapeutic intervention as well as the development of markers of drug resistance. One main difficulty in quantifying the components of tumor microenvironment is that currently there are considerable technological and analytical barriers to assessing and measuring tumor immunity in situ. **Methods**: The TCR-regions were sliced from RNA-seq bam files from all TCGA major cancer types. The CDR3 sequences were identified and annotated, and the diversity of TCR repertoire was then calculated and normalized. We use Tumor Immune Estimation Resource to estimate relative abundance of major tumor-infiltrating immune cell types (i.e., CD8 T cell, CD4 T cells, B cells, neutrophils, macrophages, and dendritic cells) from RNA-seq data. The penalized regression together with stability selection procedure (1000 runs) was employed to detect top epigenetic markers that are associated with immune phenotypes. Because it is hard to apply elastic net method directly to the all the features in the methylation arrays, we developed a novel methylation feature screening procedure based on the distance correlation (DC-SIS) to get a smaller set of informative features. The advantage of DC-SIS is that it does not require model specification for responses and it can be used directly to screen grouped predictor variables. **Results:** We applied our integrated deconvolution method and observed that methylation panel provides a strikingly accurate prediction of immune cell abundance, especially in estimating T cells and dendritic cells (with r>0.9), as well as the diversity of TCR repertoire. The results were benchmarked with existing NanoString immune panel and flow cytometry data in a local skin cancer data set. We also demonstrated the potential of using methylation panel to explore T cell exhaustion profiles. Results from this project suggest that the DNA methylome adds a new dimension to the pan-omic snapshot of cancer immunity and tumor microenvironment, and the identification of driver epigenetic events would potentially promulgate a novel paradigm for immunotherapy of cancer.

682W

Risk factors associated with primary cutaneous melanoma of the scalp and neck. *S.V. Ward[1,4], R.P. Wood[1], J.S. Heyworth[2], M.J. Millward[3], I. Orlow[4], P.A. Kanetsky[5], M. Berwick[6], N.E. Thomas[7], C.B. Begg[4], E.K. Moses[1].* 1) Centre for Genetic Origins of Health & Disease, University of Western Australia, Perth, Western Australia, Australia; 2) School of Population and Global Health, University of Western Australia, Perth, Western Australia, Australia; 3) School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia; 4) Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York; 5) Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida; 6) Department of Internal Medicine, University of New Mexico Cancer Center, University of New Mexico; 7) Department of Dermatology, University of North Carolina, Chapel Hill, North Carolina.

   Background: Scalp and neck melanoma (SCNM) has a lower survival rate and worse prognosis than melanoma of other anatomic sites. Various prognostic and demographic risk factors for SCNM have been identified but little is known about other factors associated with melanoma at this site. This study aimed to investigate the phenotypic, environmental and genetic factors associated with SCNM, compared to all other anatomic sites of melanoma (OSM).   **Methods**: Two large, population based collections of melanoma cases formed the sample for this study; the Western Australian Melanoma Health Study (WAMHS, n=1200) and the international Genes, Environment and Melanoma study (GEM, n=2943)**.** Pooled analyses in all cases used logistic regression models, adjusted for study centre, to investigate the non-genetic factors associated with SCNM *vs.* OSM. Genetic association analyses between anatomic site and 34 single nucleotide polymorphisms (SNPs) from known melanoma susceptibility loci were also performed. Participants from the WAMHS formed the discovery sample, with significant associations validated in the GEM study. An exploratory genome wide association study (GWAS) was also conducted in the WAMHS, to identify potential novel variants associated with SCNM.   **Results**: Females had a significantly decreased odds of SCNM (OR= 0.52, 95% CI: 0.39−0.68), as did those who had used a sunbed more than once (OR=0.63, 95% CI: 0.42−0.92). A borderline association for an increased odds of SCNM was observed in those older than 70 years (OR= 1.02, 95% CI: 1.00−1.05) and with a higher degree of weekend sun exposure in childhood (OR= 1.07, 95% CI: 1.002−1.15) and adolescence (OR= 1.06, 95% CI: 1.00−1.13). An initial associations was observed between SNP rs45430 (*MX2* gene) and a decreased odds of SCNM but after adjustment for multiple testing, no SNPs remained significantly associated in the candidate gene analyses. Suggestive results for the association of SCNM with novel SNPs were observed in the GWAS, including a peak on chromosome 12 in the *UBC* (rs7309653: p**=**5.44 x 10[-7], rs7135151: p=1.14 x 10[-7]) and *RNU6-927P* (rs7304293: p**=**3.87 x 10[-6]) genes.   **Conclusions**: Novel associations were identified between environmental risk factors and SCNM. Known melanoma risk SNPs may not play a role in the site-specific development of melanoma but other, potentially novel genomic loci were identified. These results now require validation in an independent population.

683T

Tumor profiling of separated carcinomatous and sarcomatous components from uterine carcinosarcoma biopsies provides insights into their development. *Z. Weber[1,2], Y. Liu[3,4], F.A. San Lucas[4], A. Deshpande[5], R. Sulaiman[6], M. Fagerness[6], N. Flier[6], J. Sulaiman[6], C. Davis[1], J. Fowler[4], G. Davies[1], D. Starks[6], L. Rojas-Espaillat[6], P. Scheet[4], E. Ehli[1].* 1) Avera Institute for Human Genetics, Sioux Falls, SD; 2) University of South Dakota, Sanford School of Medicine, Sioux Falls, SD; 3) Graduate School of Biomedical Sciences, University of Texas at Houston Health Science Center, Houston, TX; 4) University of Texas MD Anderson Cancer Center, Houston, TX; 5) Weill Cornell Graduate School of Medical Sciences, New York, NY; 6) Avera Cancer Institute, Sioux Falls, SD.

   Motivation: Uterine carcinosarcoma (UCS) is a rare and aggressive form of uterine cancer. It is bi-phasic, exhibiting histological features of both malignant epithelial (carcinomatous) and mesenchymal (sarcomatous) elements. Studies have indicated that UCS arises from sarcomatous differentiation of high-grade carcinoma while others have suggested a bi-clonal nature. Given these differences, we sought to separate the carcinoma and sarcoma elements of UCS to try to understand their molecular differences and gain further insights into how these tumors develop.   **Method:** We macrodissected carcinomatous, sarcomatous, and normal cells from formalin fixed paraffin embedded (FFPE) uterine samples of 10 UCS patients. SNP microarrays and deep sequencing (~6000x) of 26 cancer genes were performed. RNA-sequencing was also performed to quantify gene expression. Allelic imbalance (AI), mutation and gene expression profiles were compared between carcinomatous and sarcomatous components.   **Results:** AI profiling suggests that sarcomatous samples have higher genomic aberration loads compared to their carcinomatous counterparts, while mutation profiling indicates that carcinomatous and sarcomatous components share similar mutated gene patterns. Two of 10 patient tumors exhibited strong evidence of intra-tumor heterogeneity, showing mutation and AI profile differences between their paired-tumor compartments. These potentially more aggressive tumors were classified to be mesenchymal-like through gene expression profiling.   **Conclusion:** For the first time, we investigate differences between the carcinomatous and sarcomatous components of UCS from multiple molecular perspectives including AI, mutation and gene expression profiling. From these perspectives, both carcinomatous and sarcomatous components share similar profiles suggesting that they originated from a single clone. Where differences can be identified, carcinomatous aberrations appear to be a subset of those found in sarcomatous tumor profiles, with sarcomatous samples having elevated aberration loads. This suggests that the sarcomatous cells originated and evolved from the carcinomatous cells.

684F

Rare germline variants in known cancer predisposition genes in sporadic chordoma. *Y. Xiao[1], B. Zhu[1], K. Jones[1,2], A. Vogt[1,2], L. Burdette[1,2], W. Luo[1,2], B. Zhu[1,2], B. Hicks[1,2], N. Freedman[1], S. Chanock[1], D. Parry[1], A. Goldstein[1], X. Yang[1].* 1) Human Genetics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA.

Chordoma is a rare bone cancer that occurs in the skull base and spine with an incidence rate less than 0.1 per 100,000 in the United States. We previously identified germline *T* gene duplications as a major susceptibility mechanism in familial chordoma. Common and rare germline variants in the *T* gene were also found to be associated with both familial and sporadic chordoma. However, underlying genetic susceptibilities for the majority of chordoma cases remain unknown. To systematically characterize rare germline variants in established cancer predisposition genes in chordoma, we extracted whole exome sequencing (WES) data from 133 unrelated sporadic chordoma cases recruited from the United States and Canada for 114 known cancer predisposition genes. Rare variants were defined as <0.1% in the 1000 Genomes Project, Exome Sequencing Project, and ≤ 2 families from our in-house database of 1,000 cancer-prone control families. Among the 114 known cancer predisposition genes, we identified 172 rare nonsynonymous variants of which 13 are loss-of-function (LOF) variants. To assess the overall genetic burden of rare germline variants, we performed a rare variant burden test (SKAT statistics) comparing the chordoma cases to 598 unrelated population controls from two cohort studies (Cancer Prevention Study [CPS]-II and Prostate, Lung, Colorectal and Ovarian Screening Trial [PLCO]) that were sequenced using the same platform as the chordoma cases. Cases had increased frequencies of rare germline variants for *MET*, *UROD*, *ERCC4*, *TRIM37*, and *BRCA2* compared to controls (p-value < 0.01). After Bonferroni correction, the *MET* gene remained significant (p-value < 4x10⁻⁴). Of particular interest, a missense variant in *MET* (chr7:116397716) was shared by 4 chordoma cases but none of the controls. This variant, which was not reported in any public or in-house control datasets, is predicted to be deleterious by all 13 *in silico* programs we evaluated. Our results suggest that rare germline variants in *MET* may be associated with susceptibility to sporadic chordoma.

685W

BRIP1, RAD51C, and RAD51D ovarian and breast cancer associations observed in a multi-gene panel testing cohort. *L. Yackowski, L.R. Susswein, K.J. Arvai, R.I. Torene, M.L. Marshall, S.R. Solomon, P.D. Murphy, R.T. Klein, K.S. Hruska.* GeneDx, Gaithersburg, MD.

Statement of Purpose: This analysis examines the association of ovarian (OC) and breast (BC) cancer with pathogenic or likely pathogenic variants (collectively, PV) in *BRIP1*, *RAD51C*, and *RAD51D* in a cohort of patients undergoing multi-gene hereditary cancer panel testing. The association of these genes with OC has been well-studied and results are consistent with an increased risk. The BC association with these genes, however, is tenuous and needs further study. **Methods**: We performed a retrospective case-control study to assess whether cancer risk was associated with each gene. Cases were women with BC or OC referred for panel testing including up to 62 genes. Controls were reportedly healthy mothers of individuals referred for whole exome sequencing. We reviewed genetic testing results and personal cancer history for both groups. Women with more than one PV in these or another gene were excluded from the analysis. Adjusted odds ratios (ORs) were generated by multivariate logistic regression analysis controlling for age. **Results**: Ovarian cancer: *BRIP1*, *RAD51C*, and *RAD51D* are associated with a statistically significant increased risk of ovarian cancer (Table 1).

Table 1

Gene	Ovarian Cancer Cases	Controls	OR (95% CI)	p-value
BRIP1	0.7% (32/4296)	0.1% (16/12887)	3.14 (1.46-6.95)	0.004
RAD51C	0.7% (31/4285)	0.05% (6/12887)	13.10 (5.09-38.81)	<0.0001
RAD51D	0.4% (17/4282)	0.02% (3/12887)	11.16 (2.61-63.35)	0.003

Breast cancer: *BRIP1, RAD51C*, and *RAD51D* are associated with a statistically significant increased risk of breast cancer (Table 2).

Table 2

Gene	Breast Cancer Cases	Controls	OR (95% CI)	p-value
BRIP1	0.3% (89/30047)	0.1% (16/12887)	1.93 (1.10-3.59)	0.029
RAD51C	0.1% (45/30003)	0.05% (6/12887)	2.72 (1.18-7.43)	0.030
RAD51D	0.1% (31/29993)	0.02% (3/12887)	4.06 (1.29-18.13)	0.032

Conclusions: Pathogenic variants in *BRIP1*, *RAD51C*, and *RAD51D* are well-established with respect to OC risk and are further confirmed in this cohort. Although recent *BRIP1* case-control data show no evidence of an increased risk for BC, earlier reports suggest an increase in BC risk. This study supports a modestly increased BC risk with *BRIP1* PV in a large, independent cohort. For *RAD51C* and *RAD51D*, the consensus in the literature is that PV seem to be rare in BC-only families. In this case-control study, we observed a statistically significant association with BC risk and *RAD51C* and *RAD51D* PV. These results add to the rapidly evolving knowledge of cancer risks associated with PV in *BRIP1*, *RAD51C*, and *RAD51D*.

686T

Rare germline variants in the promoter region of *CDKN2B* may cause melanoma predisposition. *R. Yang[1], P. Hyland[1], J. Choi[1], A. Vu[1], Y. Xiao[1], K. Jones[1,2], A. Vogt[1,2], L. Burdette[1,2], B. Zhu[1,2], M. Yeager[1,2], B. Hicks[1,2], N. Freedman[1], L. Beane-Freeman[1], S. Chanock[1], T. Andresson[2], M. Tucker[1], K. Brown[1], A. Goldstein[1].* 1) div Cancer Epidemiology & Genetics, NCI/NIH, Bethesda, MD; 2) Frederick National Laboratory for Cancer Research, Frederick, MD.

Approximately 10% of cutaneous malignant melanoma (CMM) cases occur in a familial setting and known high-risk CMM genes (*CDKN2A* being the most frequent one) account for melanoma risk in less than 40% of melanoma-prone families. To identify additional high-risk susceptibility genes for familial CMM, we performed whole exome sequencing in germline DNA from 144 CMM patients/obligate gene carriers from 76 American melanoma-prone families. We identified two rare missense variants in *CDKN2B* that showed cosegregation with CMM in two families. The first variant (g.22008763 C>T), with a frequency of 3.6×10^{-5} in the Exome Aggregation Consortium (ExAC) among non-Finnish Europeans, was seen in 6 of 7 sequenced CMM patients in a family with a total of 13 CMM patients. The second variant (g.22008767 C>G), which was not reported in any public database or in-house controls, was seen in all 3 patients (including a pair of identical twins) sequenced in a smaller family. We subsequently sequenced the entire coding region of *CDKN2B* in 795 CMM cases and 807 matched controls from two cohort studies (Prostate, Lung, Colorectal and Ovarian Screening Trial and Agriculture Health Study) and found a third *CDKN2B* variant that was present in a single CMM patient. This variant (g.22008772 G>C) was also not reported in any database or any controls we sequenced. All three variants are located only a few bases away from each other and within the intronic region of the major *CDKN2B* transcript expressed in melanocytes. Based on experimental data from the NIH Epigenomics Roadmap project and ENCODE, these variants map to a CpG island and an active promoter in normal primary melanocytes and skin fibroblasts as well as open chromatin in different cell lines. Most interestingly, these variants locate to a canonical motif and DNA footprint for CCCTC-binding factor (CTCF) in melanoma cell lines. Moreover, preliminary results from a semi-quantitative CTCF binding assay using variant or wild-type *CDKN2B* constructs and melanoma cell extracts showed that both variants found in families had reduced CTCF binding compared to the wild-type. CTCF has an important role in regulating chromatin architecture suggesting that the germline *CDKN2B* variants may alter CTCF regulatory function and gene expression important for melanoma predisposition in both familial and sporadic CMM. However, future functional studies are needed to establish the link between these variants and melanoma susceptibility.

687F

Combining weak signals by rank truncated products of p-values with application to detection of simultaneous closely-spaced mutations in cancer genomes. *D. Zaykin[1], F. Hu[1], L.J. Klimczak[1], D. Gordenin[1], O. Vsevolozhskaya[2].* 1) NIEHS, Durham, NC., USA; 2) U of Kentucky.

Methods for combining top ranking statistics or P-values are useful for analysis of high dimensional genomics data, because they aggregate top ranking signals into a single score and at the same time adjust for multiple comparisons. The Rank Truncated Product method (RTP) by Dudbridge and Koleman (2003) gained popularity in many extensions, including its adaptive versions, where the number of top-ranking P-values is varied to maximize power. However, all previously known theoretical forms of the RTP distribution are unwieldy, involve complicated multiple integration, and result in page long equations. We developed several new ways of evaluating the RTP distribution that greatly simplify the analytic form of the combined P-value, down to a single line of R code. Our approach leads to new powerful methods for combining top-ranked, possibly correlated statistics. We applied our methods to analysis of spatial distribution of mutations caused by single-strand (ss) DNA specific APOBEC cytidine deaminases in cancer genomes. These mutations often occur simultaneously and in close proximity to each other, i.e., in clusters. Clusters formed by APOBEC mutagenesis in ss DNA would have only C or only G mutated in a given strand (C- or G-coordinated). Simultaneous mutations in transient long stretches of ss DNA in a cluster are unlikely to be broken by somatic recombination or genome rearrangements, therefore allele frequencies of mutations within a cluster are expected to be the same. Analysis using our methods confirms that allele frequencies within C- or G-coordinated clusters are closer to each other compared to frequencies within other cluster types.

688W

Significance of secondary genetic findings in a large prospective population sample. *K. Kristiansson[1,2], P. Salo[1,2], V. Salomaa[1], P. Jousilahti[1], H. Kääriäinen[1], M. Perola[1,2,3].* 1) National Institute for Health and Welfare, Helsinki, Finland; 2) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 3) University of Tartu, Estonian Genome Centre, Tartu, Estonia.

   Introduction: Genetic studies are encouraged to return secondary findings from clinically important genes to the study participants. The rare but highly penetrant mutations in these genes typically lead to serious consequences, such as cancer or serious heart dysrhythmia, but there are known medical interventions with potential life-saving value. Although classified as pathogenic in web-based databases, the impact of most risk variants at a population level remains unclear. **Material and Methods:** We studied variation in 59 genes of the American College of Medical Genetics and Genomics (ACMG) secondary findings recommendation. Impact of the risk variants on clinical outcomes was evaluated in 25-74 year old FINRISK study participants (N~23 000) using up to 23 years of follow-up. The variants, genotyped with genome-wide microarrays and imputed using population specific reference panel, were analyzed in time-to-event and logistic regression models. **Results:** The expected number of risk allele carriers for 75 ACMG risk variants varied between 4 and 300 based on publicly known allele frequencies and similar numbers were seen in FINRISK. Although some variants associated with increased risk of an event during follow up (e.g. carriers of a risk variant in a cardiomyopathy associated gene had a 3.7 times higher risk of heart failure compared to non-carriers in Cox survival analysis (p=0.0014)), for a large majority we observed no clinical significance. **Conclusions:** Before secondary findings from clinically important genes are returned to study participants, the significance of the variants needs to be carefully evaluated at a population level.

689T

A esophageal adenocarcinoma susceptibility locus at 9q22 also confers risk to esophageal squamous cell carcinoma by regulating the function of BARX1. *C.W. Yan[1,2], Y. Ji[3], T.T. Huang[1], Y. Gao[4], Y.Y. Gu[1], Q. Qi[1], H.X. Ma[1], G.F. Jin[1,2].* 1) Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China; 2) Jiangsu Key Lab of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Medicine, Nanjing Medical University, Nanjing, China; 3) Department of Cardiothoracic Surgery, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China; 4) Department of Medical Oncology, The Affiliated Huaian First People's Hospital of Nanjing Medical University, Huai'an, China.

   A genome wide association study (GWAS) identified genetic variants at 9q22 significantly associated with the risk of esophageal adenocarcinoma (EAC)/ Barrett's esophagus (BE), which were different from those loci discovered by previous GWAS on esophageal squamous cell carcinoma (ESCC). It is important to evaluate whether these susceptibility loci for EAC/BE are also implicated in ESCC development. In the current study, we analyzed genetic variants at 3p13, 9q22, 16q24 and 19p13 in a case-control study including 2139 ESCC patients and 2463 cancer-free controls in a Chinese population, and further characterized the function relevance of genetic variants by functional assays. We found that the G allele of rs11789015 at 9q22 as compared with A allele, was significantly associated with a decreased risk of ESCC with a per-allele odds ratio of 0.77 (95%CI, 0.65-0.90; *P*=1.38×10[-3]), whereas the other three loci were not associated with ESCC risk. We further found that rs11789015-G allele correlated with decreased mRNA levels of *BARX1*. Dual-luciferase reporter gene assay revealed that the A > G change at rs11789015 significantly decreased the promoter activity of *BARX1*. Both the mRNA and protein levels of *BARX1* were significantly higher in ESCC tumor tissues compared with the corresponding normal tissues. Moreover, the deletion of *BARX1* substantially reduced ESCC cells growth, migration and invasion. In conclusion, these results suggest that genetic variants at 9q22 are associated with the risk of both EAC/BE and ESCC, possibly by regulating the function of *BARX1*.

690F

Detection of epigenetic field defects using weighted epigenetic distance-based method. *Y. Wang[1], A. Teschendorff[2], M. Widschwendter[2], S. Wang[1]. 1) Department of Biostatistics, Columbia University, New York, NY; 2) Department of Women's Cancer, University College London, London, UK.*

Detecting molecular alterations early in carcinogenesis is important for cancer early detection. One common approach in detecting early molecular alterations, known as field defects, is to compare normal tissues from healthy individuals to the normal tissues adjacent to tumors. Due to its important role in many biological processes and cancer progression, studies have identified epigenetic field defects, notably early DNA methylation alterations using current EWAS (epigenome-wide association studies). Here we propose a complementary approach based on characterization of similarity/dissimilarity of DNA methylation of a set of individuals at multiple CpG sites on a gene or a genetic region. This similarity-based approach has been proven to be powerful in genome-wide association studies (GWAS) and gene expression studies, but hasn't been investigated in epigenetics. We propose a weighted epigenetic distance-based regression on gene or genetic region with several site-level weights considered: (1) p-value testing for mean differences between normal tissues from healthy individuals to the normal tissues adjacent to tumors; (2) site-level correlation capturing average correlation to the rest of the sites on this gene. We demonstrated the superior performance of the proposed weighted distance-based method over non-weighted version and standard site-level EWAS method. Application to DNA methylation data of breast cancer from Gene Expression Omnibus (GEO) comparing normal tissues adjacent to tumors and normal tissues from independent age-matched cancer-free women shows that many novel epigenetic field defects in breast cancer across the genome that were not previously reported and were missed by standard EWAS method in this analysis are cancer-related.

691W

Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population. *Y. Li[1], X. Xiao[1], Y. Han[1], O. Gorlova[1], D. Christiani[2], M. Johansson[3], J. McKay[3], P. Brennan[3], R. Hung[4], C. Amos[1] OncoArray Consortium. 1) Biomedical Data Science Department, Dartmouth College, United States; 2) School of Public Health, Harvard University, United States of America; 3) Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, France; 4) The Lunenfeld-Tanenbaum Research Institute, Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Canada.*

Non-small cell lung cancer (NSCLC) is the most common histological type of lung cancer. Both environmental and genetic risk factors contribute to lung cancer carcinogenesis. Here, we conducted a genome-wide interaction analysis between SNPs and smoking status (never vs ever smokers) in a European-descent population. We adopted a two-step analysis strategy in the discovery stage: we first conducted a case-only interaction analysis to assess the relationship between SNPs and smoking behavior using 13,336 NSCLC cases. Candidate SNPs with p-value less than 0.001 were further analyzed using a standard case-control interaction analysis including 13970 controls. The significant SNPs with p-value less than 3.5×10^{-5} (correcting for multiple tests) from the case-control analysis in the discovery stage were further validated using an independent replication dataset comprising 5377 controls and 3054 NSCLC cases. We further stratified the analysis by histological subtypes. Two novel SNPs, rs6441286 at gene *IL12A-AS* (chr13q25.33) and rs17723637 at gene *ZNF462* (chr9q31.2), were identified. The interaction odds ratio and meta-analysis p-value for these two SNPs were 1.24 with 6.96×10^{-7} and 1.37 with 3.49×10^{-7}, respectively. Additionally, interaction of smoking with rs4751674 at gene *AFAP1L2* (chr10q25.3) was identified in squamous cell lung carcinoma with an odds ratio of 0.58 and p-value of 8.12×10^{-7}. This study is by far the largest genome-wide SNP-smoking interaction analysis reported for lung cancer. The results from our study reinforce that gene-smoking interactions play important roles in the etiology of lung cancer and account for part of the missing heritability of this disease.

692T

Type 2 diabetes susceptibility variants contribute to breast cancer risk. *M.H. Black, S. Li, B.A.J. Sarver, A. Elliott, H.M. Lu.* Bioinformatics and Computational Biology, Ambry Genetics, Aliso Viejo, CA.

Type 2 diabetes mellitus (T2DM) is a known risk factor for breast cancer (BC), but whether the conditions share a common genetic etiology remains to be elucidated. Analyses to date have focused mainly on T2DM GWAs-identified intronic or intergenic variants. Recently, exome analysis in ~100,000 individuals detected 21 coding variants predisposing to T2DM. We performed whole exome sequencing in 9,639 BC cases and 3,988 controls, among patients referred for genetic testing at a single laboratory in 2012-2015, and leveraged this vast collection of genotype and phenotype data to investigate the contribution of these coding variants to risk for and clinical features of BC. We also used 1000 Genomes (G1K) as a reference population to confirm our findings. All models were conducted within genetically inferred G1K-defined populations and adjusted for sex and 3 ancestry PCs. Among Europeans (n=7,235 BC, 2,666 controls), *PPARG* P12A Pro (RAF=94.2%) was associated with BC at genome-wide significance (OR_{add}=1.49, p=3.9x10^{-10}). BC was also associated with *PPIP5K2* S1207G (RAF=2.5%, p=2.4x10^{-6}), *PAM* D563G (RAF=3.0%, p=9.5x10^{-5}) and *ASCC2* V123I (RAF=93.8%, p=8.0x10^{-4}). P12A was significantly associated with HR$^+$, HR$^-$, and HER2$^-$ BC. D563G and S1207G were significantly associated with HER2$^-$ and V123I with HER2$^+$ BC only. P12A results comparing cases to G1K (n=501) were consistent (OR_{add}=1.78, p=1.5x10^{-6}). Risk estimates from G1K comparisons for the other 3 variants were nearly identical to those from internal controls, but not significant due to sample size. Among Mexicans, Central and South Americans (n=809 BC, 768 controls, 334 G1K), *PPARG* Pro (RAF=94.3%) was associated with BC compared to controls (OR_{add}=1.98, p=1.9x10^{-6}) and G1K (OR_{add}=1.90, p=0.001). Among South Asians (n=105 BC, 99 controls, 479 G1K), *PPARG* Pro (RAF=95.9%) was linked to BC in comparisons with G1K (OR_{add}=3.38, p=0.003); internal control sample size was too small for risk estimation. No associations were detected among Africans (n=968 BC, 293 controls, 639 G1K) or East Asians (n=522 BC, 162 controls, 502 G1K). Results were similar when excluding males and/or those carrying cancer gene mutations. Genetic risk scores assessing cumulative effects across 21 variants were not associated with BC in any racial/ethnic group. These data suggest that *PPARG* P12A and other T2DM coding variants confer risk for BC and its clinical subtypes, with potentially important implications for clinical screening and management.

693F

Novel approach to construction of causal gene regulatory network. *R. Jiao[1], Z. Hu[2], M. Xiong[1].* 1) Biostatistics, University of Texas Health Science Center, Houston, TX; 2) School of Life Science, Fudan University, Shanghai, China.

The popular methods for construction of gene regulatory networks are correlation analysis, independent tests and Bayesian networks. Correlation analysis is unable to discover regulatory direction. The independent tests and Bayesian networks, these traditional causal inference only can identify up to Markov equivalence class and cannot find the unique causal solutions or unique gene regulatory network. To uniquely discover the causal relationships and causal gene regulatory networks, novel functional additive noise models (ANMs) for causal inference where smoothing spline regression was used to fit the functional model have been developed and applied to construction of causal gene regulatory network. The proposed ANMs were applied to Wnt signaling pathway with RNA-Seq of 145 genes measured in 447 tissue samples. Fifty most significant causal relations were selected to test detection power that was defined as the proportion of correctly inferred paths, given that the paths in the KEGG pathway database were taken as the true paths. For comparison, the score-based linear structural equation models (SEMs) with integer programming (IP) and Glasso were also included in the analysis. Given directed paths in KEGG, results showed that 38% of detected paths by the ANM agreed with the paths in the KEGG, while SEMs only reached 22% accuracy. Given both directed and undirected paths, the results showed that the ANM could reach 46% accuracy in path detection while SEMs and Glasso could only reach 29% and 24% respectively. The ANM outperformed the score-based linear SEMs and the Glasso algorithms.

694W

Rate of reclassification of cancer genetic variants differs by race/ethnicity, depending on gene. *L.R. Van Tongeren[1], T.P. Slavin[1], C.E. Behrendt[2], I. Solomon[1], C. Rybak[1], B. Nehoray[1], L. Kuzmich[1], M. Niell-Swiller[1], K. Blazer[1], S. Tao[3], K. Yang[1], J. Culver[4], S. Sand[1], D. Castillo[1], J. Herzog[1], S.W. Gray[1], J.N. Weitzel[1].* 1) City of Hope, Division of Clinical Cancer Genomics, Duarte, CA; 2) City of Hope, Department of Information Sciences, Duarte, CA; 3) City of Hope, Integrative Genomics Core, Duarte, CA; 4) University of Southern California Norris Comprehensive Cancer Center, Cancer Genetics Program.

BACKGROUND: Efforts are ongoing to improve interpretation of cancer genetic tests, especially for racial/ethnic minority populations, whose genetic tests disproportionately indicate variants of uncertain significance. We hypothesized that rate of variant reclassification differs for individual racial/ethnic minorities relative to non-Hispanic Whites (NHW). Because *BRCA* genes have been studied most extensively, we considered whether findings differed by gene. METHODS: We studied participants enrolled with informed consent from two sites in the Clinical Cancer Genomics Community Research Network from 9/96-12/16. Participants self-identified as primarily Hispanic White, NHW, Asian, African or Native American. Non-Benign variants in actionable genes were followed for reclassifications through 2/17 or until reclassification to Benign. Duplicate within-family reports (same laboratory, gene and variant) were excluded. A log-linked generalized linear model of reclassifications per years of follow-up tested 4 categories of minority race/ethnicity, taking into account interaction by gene (*BRCA* vs non-*BRCA*) and correlation among tests at the same laboratory. The model adjusted for covariates (laboratory, year, and test result as a time-dependent variable) and multiple hypothesis testing. RESULTS: Of N=1,817 tests studied, 315 (17%) were reclassified at least once (maximum 3 times), for a total of 359 reclassifications in 10,508 years. Median time to first reclassification was 3.58 years. Rate of reclassification differed significantly from NHW for 3 of 4 racial/ethnic minorities, but only for non-*BRCA* tests. (Table)

Table. Covariate-Adjusted* Ratio (95% CI) of Reclassification Rates.			
Race/Ethnicity	Non-*BRCA* Variants	*BRCA1* or *BRCA2*	Holm p
African American	2.45 (1.61-3.74)	1.45 (0.96-2.19)	0.006
Asian American	0.76 (0.63-0.91)	1.12 (0.97-1.30)	0.034
Hispanic White	1.73 (1.59-1.89)	1.15 (1.02-1.30)	<0.001
Native American	1.91 (0.42-8.64)	0.80 (0.17-3.76)	0.903
Non-Hispanic White	1	1.21 (1.09-1.34)	

*Rate ratios shown are adjusted for laboratory, result and year of original test. CONCLUSIONS: For non-*BRCA* variants only, reclassification rates were higher for Hispanic and African American, and lower for Asian American, than NHW tests. Higher rates of reclassification may stem from efforts to improve representation of racial and ethnic minorities in reference genomic databases. However, that improvement does not appear to extend to variants found among Asian Americans.

695T

Genes associated with pancreatic cancer predict survival and prognosis. *A. Gonzalez-Reymundez[1], S.Y. Lunt[2], A.I Vazquez[1].* 1) Department Epidemiology & Biostatistics, Michigan State University, East Lansing, MI; 2) Department of Biochemistry and Molecular Biology, Michigan State University, MI.

Pancreatic cancer, which causes >40,000 deaths per year in the U.S., has a dismal 5-year survival rate of ~7%. Despite some improvements in patient care, the survival rate for most pancreatic cancer patients at the time of diagnosis is less than one year, and our ability to predict which patients will benefit from surgery and treatment remains inadequate. One of the main limitations for predicting patient response to surgery and treatment is the poorly understood molecular basis of pancreatic cancer. Here, we address the challenge of predicting pancreatic cancer survival and prognosis by identifying key genes that differentiate tumor grade and patient survival. To achieve this goal, we used a homogenous sample of pancreatic cancer patient data from TCGA, consisting of primary ductal pancreatic ductal carcinomas from white Caucasian individuals, all having received a Whipple procedure and without history of prior malignancies or metastasis (n=103: 44 and 59 females and males, respectively; 60 and 43 alive and deceased patients at the moment of the last follow up, respectively). We first determined the association between gene expression with tumor grade and survival at six months via multiple logistic regression models, using the patient's age, gender, tumor stage, and log transformed RNA reads by 13511 genes (Affymetrix HT Human Genome U133A Array), as predictors. We further sub-selected de-regulated genes in terms of their predictive power by applying a forward variable selection coupled with a linear discriminant analysis. We found several genes that were previously confirmed to be involved in pancreatic cancer: *PVRL1* (involved on cell-cell contacts; its downregulation is associated with poor prognosis and tumor proliferation), *ESAM* (involved on cell aggregation; frequently up-regulated in intestinal type gastric cancer), and *RNASEN* (maintenance at adult pancreatic acinar cells; involved on *KRAS*-driven pancreatic neoplasia; regulates differentiation and viability during mouse pancreatic cancer initiation). We also found several novel genes with functions that could potentially be linked with pancreatic cancer: *GABRA1* (downregulation associated with insulin secretion), *METAP1* (required for normal progression through cell cycle), and *CAMTA2* (potential tumor repressor). Our results suggest these genes are good candidates for further studies on target therapy.

696F

Efficient gene-by-treatment interactions test to develop predictive bio-markers on genome-wide studies using multidimensional hierarchical mixture models. *T. Otani[1,2], H. Noma[2,3].* 1) Risk Analysis Research Center, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan; 2) CREST, Japan Science and Technology Agency, Tachikawa, Tokyo, Japan; 3) Department of Data Science, The Institute of Statistical Mathematics, Tachikawa, Tokyo, Japan.

In the development of molecular diagnostics for optimizing individualized treatment, efficient exploration of predictive biomarkers is of particular importance to predict responses to therapeutics effectively. In this assessment, an efficient test of gene-by-treatment interactions is crucial, because conventional statistical analyses for detecting interaction effects based on standard regression models lack sufficient statistical power especially on large-scale genome-wide studies. In this study, we illustrate an efficient multi-subgroup gene screening method using a multidimensional semi-parametric hierarchical mixture model and an optimal discovery procedure (ODP) developed to overcome this issue, with application to breast cancer and stroke randomized clinical trials with genomic data. This method reveals the underling effect size distribution of disease-related genes across treatment and control groups via empirical Bayes estimation under the hierarchical mixture model. Then, based on the optimal gene ranking and posterior probabilities of association for each gene derived from the fitted model, disease-related genes are detected by the ODP with control of false discovery rate levels. We applied this method to randomized clinical trials with molecular data to detect interactions between single nucleotide polymorphisms (SNPs) and treatments, and detected new 6 SNPs being associated with progression-free survival in breast cancer patients and new 3 SNPs being associated with homocysteine levels in blood that are strongly associated with the risk of stroke. We present estimated effect size distributions which could provide a clue for understanding biological mechanisms of diseases, detected disease-related SNPs with their characteristics, and a result of genomic annotation of those using publicly available tools and databases for biological investigation and for validation of the method. Also, we provide a comparison of efficiency with the conventional statistical analyses for detecting gene-by-treatment interactions based on the standard regression models.

697W

Two susceptibility variants associated with osteosarcoma in the HLA class II region. *C. Zhang[1,2], K. Walsh[2,1].* 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA.

Osteosarcoma is the most commonly diagnosed primary malignant bone tumor in adolescents and young adults. A genome-wide association study previously identified a risk variant in *GRM4* (6p21.3), approximately 1 Mb from the human leukocyte antigen (HLA) class II region. Although they detected additional signals in the HLA class II region, the risk variant in *GRM4* was on a separate haplotype block. Therefore, further analyses were necessary to determine whether there is an independent risk locus in the HLA class II region. We performed genetic analyses of HLA class II genes on chromosome 6 in 207 cases and 696 controls using dried newborn bloodspots stored by the California Department of Public Health. Genotype data were generated with the Affymetrix Axiom Array supplemented with ~18,000 additional probes across the MHC region. We imputed HLA alleles using a reference panel of 5,225 individuals in the Type 1 Diabetes Genetics Consortium who underwent high-resolution HLA typing via next-generation sequencing. All subjects were of European-ancestry, and case-control comparisons were adjusted for population stratification using ancestry-informative principal components. Using four-digit classical HLA alleles, which distinguish HLA gene variants at the amino acid level, 27 commonly observed HLA class II alleles (frequency>0.05) were analyzed, and three highly correlated variants were associated with osteosarcoma risk. The most significantly associated variant in the discovery dataset, *HLA-DQB1*02:01* (OR=0.51, P=0.003), was replicated in an independent dataset of non-overlapping 657 osteosarcoma cases and 1183 controls from dbGaP (OR=0.70, P=0.018) (OR$_{meta}$=0.62, P_{meta}=1.5E-4). The meta-analysis also revealed a second independent signal at *HLA-DQA1*01:01* (OR$_{meta}$=1.33, P_{meta}=0.001). Both HLA class II association signals are independent of the *GRM4* variant identified in the prior genome-wide association analysis, and showed no linkage disequilibrium (r^2<0.01). Our findings suggest a role for HLA class II variants in contributing to the risk of osteosarcoma. Additional work is needed to clarify potential causal mechanisms underlying these associations.

698T

HPV16 genomic insights into cervical carcinogenesis: Viral oncogene E7 conservation is critical to carcinogenesis, and relation to human APO-BEC3 activity. *L. Mirabello[1], M. Yeager[1], K. Yu[1], G. Clifford[2], B. Zhu[1], Y. Xiao[1], M. Cullen[1], J.F. Boland[1], N. Wentzensen[1], T. Raine-Bennett[3], Z. Chen[4], S. Bass[1], L. Song[1], M. Steinberg[1], L. Burdett[1], M. Dean[1], D. Roberson[1], T. Lorey[5], S. Franceschi[2], P.E. Castle[6], J. Walker[7], R. Zuna[7], R.D. Burk[6,8], M. Schiffman[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD; 2) Infections and Cancer Epidemiology Group, International Agency for Research on Cancer (IARC), Lyon, France; 3) Women's Health Research Institute, Division of Research, Kaiser Permanente Northern California, Oakland, CA; 4) Department of Microbiology, The Chinese University of Hong Kong, Hong Kong; 5) Regional Laboratory, Kaiser Permanente Northern California, Oakland, CA; 6) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 7) University of Oklahoma Health Sciences Center, Oklahoma; 8) Departments of Pediatrics, Microbiology and Immunology, and Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY.

   Objective: To determine if HPV16 genetic variation influences risk of cervical precancer and cancer, and if there is a relation to the antiviral activity of human APOBEC3 cytidine deaminases, we conducted a large case-control study of HPV16 genomes from a total of 5,720 women. **Methods:** HPV16 genomes were sequenced from 3,215 cervical samples from NCI-KPNC PaP Cohort women (1,107 controls [≤CIN1], 906 CIN2 precancer, 1,093 CIN3 precancer, 109 cancer), and each HPV16 SNP and the combined effects of rare nonsynonymous variants in each viral gene region were evaluated for associations with risk of precancer/cancer (CIN3+). We replicated these findings by HPV16 genome sequencing 996 samples from a cross-sectional U.S. SUCCEED population and 1,509 invasive cancers collected internationally by IARC. HPV16 mutation signatures were evaluated to determine if observed variants matched an APOBEC3-induced signature. **Results:** Forty-nine SNPs were associated with CIN3+, two SNPs located in the regulatory region (URR) were strongly associated with a reduced risk and remained significant after adjustment for multiple tests (OR 0.04, P=1.6x10^{-9}; OR 0.06, P=7.2x10^{-5}). Unexpectedly, controls had significantly more HPV16 nonsynonymous/nonsense variants throughout the genome (Beta of -0.13/mutation, P=6.5x10^{-07}) compared to cases. Strikingly, the oncogene, E7, was devoid of non-silent variants in precancers/cancers compared to higher levels in the controls (P=1.0x10^{-07}); we confirmed this in cancers from around the world and E7 was significantly less variable than all other gene ORFs in cancers, even compared to the other oncogene E6 (P=6.1x10^{-5}). In a mutation signature analysis, 24% of variants were potentially induced by APOBEC3 in all samples, and these variants were not evenly distributed across the HPV16 genome. Controls had significantly higher numbers of APOBEC3-induced variants compared to precancer cases (P=5.6x10^{-5}), particularly in E7 (P=0.009), consistent with an antiviral mechanism; whereas, cancer cases had a uniquely high level of APOBEC3-induced variants (P=0.004), which may be related to their longer infection time and activation of APOBEC in cancer. **Conclusions:** There is evidence of APOBEC3 editing throughout the HPV16 genome related to infection outcome. Our data indicate that strict conservation of E7, which disrupts Rb function, is critical for HPV16 carcinogenesis, presenting a highly specific target for etiologic and preventive research. .

699F

Proteomics and genomics integration to predict ovarian cancer survival. *U. Ozbek[1], C. Conley[2], J. Peng[2], P. Wang[1].* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) University of California at Davis, Davis, CA.

   Ovarian cancer is the fifth leading cause of cancer death. It's estimated that there will be 22,440 new cases and 14,080 deaths of ovarian cancer in the US in 2017. To improve the treatment of ovarian cancer patients, it is of great importance to accurately identify the patients with relatively low survival. In this study, our goal is to integrate multiple omics data sets, including DNA copy number alterations (CNA), mRNA expressions, and protein expression profiles, using network analysis tools to derive gene expression signatures useful to predict survival outcomes of ovarian cancer patients. We hypothesize that genes (1) being influenced by CNA events, (2) serving as hub nodes in gene regulatory networks, and (3) impacting activities of many proteins, shall play important roles in shaping the cell phenotypes and are particularly useful for predicting patients' survival. To test this hypothesis, we first construct a regulatory network of CNAs and mRNA expressions based on genomic profiles of 293 ovarian cancer samples from The Cancer Genome Atlas (TCGA) using a newly developed statistical tool, spaceMap, a conditional graphical model, which learns the conditional dependency relationships between two types of nodes through a penalized multivariate regression framework. Based on the inferred regulatory network, we identify a collection of hub genes that are influenced directly by CNA and are interacting with a large number of other genes. We then select a subset of these genes whose expressions also appeared to have a big impact on global protein activities based on both mRNA and proteomics profiles of 100 ovarian tumors from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) study. In the end, we demonstrate that these selected genes are in general more informative to predict patient survival outcome. We expect such a strategy could lead to better prediction models for disease outcome in clinical practice. The candidate genes identified from the analysis could also lead to novel drug targets to improve the patient disease outcome.

700W

Survey of microRNA SNPs identifies novel breast cancer susceptibility loci in a case-control population-based study of African American women. *J. Bensen[1], M. Graff[1], K. Young[1], P. Sethupathy[2], J. Parker[1], C. Pecot[1], K. Currin[1], S. Haddad[3], E. Ruiz-Narvaez[3], C. Haiman[4], C-C. Hong[5], L. Sucheston-Campbell[5], Q. Zhu[5], S. Liu[5], S. Yao[5], E. Bandera[6], L. Rosenberg[3], K. Lunetta[3], C. Ambrosone[5], J. Palmer[3], M. Troester[1], A. Olshan[1].* 1) University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Cornell University, Ithaca, NY; 3) Boston University, Boston, MA; 4) University of Southern California, Los Angeles, CA; 5) Roswell Park Cancer Institute, Buffalo, NY; 6) Rutgers University, New Brunswick, NJ.

   BACKGROUND: MicroRNAs (miRNAs) regulate gene expression and influence cancer. Primary transcripts of miRNAs (pri-miRNAs) are poorly annotated and little is known about the role of germline variation in miRNA genes and breast cancer (BC). We sought to identify germline miRNA variants associated with breast cancer risk and tumor subtype among African American (AA) women. **METHODS:** Genotyping and imputed data from four studies of breast cancer in AA women were combined into a final data set containing 224,188 miRNA gene SNPs for 8,350 women; 3,663 cases and 4,687 controls. Primary miRNA sequence was identified for 566 miRNA genes expressed in ENCODE Tier 1 cell types and human pancreatic islets. Association analysis was conducted using logistic regression for BC status overall and by tumor subtype. **RESULTS**: A novel breast cancer signal was localized to an 8.6kb region of 17q25.3 by 4 SNPs (rs9913477, rs1428882938, rs28585511, and rs7502931) and remained statistically significant after multiple test correction (OR (95%CI) = 1.44 (1.26,1.65); p-value=3.15×10^{-7}; FDR=0.03). Primarily ER positive and PR positive cases drive this signal. These SNPs reside in a genomic location that includes both the predicted primary transcript of non-coding miRNA gene *MIR3065,* and first intron of the gene coding for brain-specific angiogenesis inhibitor 1-associated protein 2 (*BAIAP2*). Additionally, miRNA associated SNPs on chromosomes 1p32.3, 5q32, and 3p25.1 were the strongest signals for hormone receptor, luminal vs. basal and her-2 enrichment status, respectively. A second phase of genotyping (1,397 BC cases, 2,418 controls) that included rs9913477 was used for replication and meta-analysis. Replication of this 17q25.3 SNP confirmed its positive association with BC (OR (95%CI) = 1.16 (1.12, 1.76); p-value=3.56×10^{-7}). Furthermore, when meta-analyzed with the original data set, the association of rs9913477 with BC (OR (95%CI) = 1.39 (1.31, 1.57); p-value=4.28×10^{-7}) remained significant. **CONCLUSION:** Germline genetic variation in this region indicates that *MIR3065* may play an important role in BC development and heterogeneity among AA women. Further investigation to determine the potential functional effects of these SNPs is warranted. This study contributes to our understanding of BC risk in AA women and highlights the complexity in evaluating variation in gene-dense regions of the human genome.

701T

Mosaic truncating *PPM1D* mutations are age-related but lack a strong association with breast cancer risk. *T.A. Myers[1], M.J. Machiela[1], C. Hautman[2], R. Koster[1], W. Figg[1], L. Jessop[1], L.M. Colli[1], M. García-Closas[1], S.J. Chanock[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD.

   Mosaic protein truncating variants (PTVs) in exon 6 of the protein phosphatase, Mg2+/Mn2+ dependent 1D (*PPM1D*) gene in blood-derived DNA have been associated with breast and ovarian cancer risk. Recent studies in patients with ovarian cancer have suggested mosaic *PPM1D* PTVs are selected for by chemotherapy and may not confer cancer risk. Our aim was to perform ultra-high depth (>2,500x target coverage) sequencing of the *PPM1D* mutation cluster region to evaluate whether *PPM1D* PTVs are associated with breast cancer risk or if these mutations were selected for by cancer treatment. We investigated blood-derived DNA from 1,734 breast cancer cases and 1,129 controls from the Polish Breast Cancer Study (PBCS), a breast cancer case-control study from Poland, and 2,158 breast cancer cases and 1,932 controls from the Prostate Lung Colorectal and Ovarian Cancer Screening Trial (PLCO), a prospective trial in the United States. We performed multiplex amplicon PCR using two sets of target-specific primers designed to amplify the mutation cluster region (MCR) in exon 6 of *PPM1D* and incorporate Illumina adapter overhang sequences. Low-cycle PCR was performed to add Illumina sequence adapters and dual indices to the amplicons allowing for pooling of 384 samples per Illumina MiSeq run. Data analysis was performed by the MiSeq Reporter 2.6.2.3 PCR Amplicon Workflow and variants were called using the Illumina Somatic Variant Caller, designed to detect low-frequency mutations below 5%. All samples were required to pass a minimum QC filter of 250x coverage across 80% of the MCR for subsequent analysis. Detected *PPM1D* PTVs were independently validated using a Nextera XT library preparation protocol. In total, we identified 50 of 6,953 (0.9%) women with a mosaic *PPM1D* PTV variant frequency between 1.0% and 30.6%. A higher frequency of mosaic *PPM1D* PTVs was observed with increasing age (P_{trend}=0.005); however, there was no significant difference in variant frequency between breast cancer cases and controls (OR=1.44, 95% CI=0.81-2.59, P-value=0.22). Furthermore, we found no significant effects of treatment prior to blood collection (e.g., chemotherapy, radiation or hormone exposure) or smoking on risk of developing *PPM1D* PTVs. Our study indicates *PPM1D* PTVs are present in women from two studies of breast cancer and suggests frequencies increase with age, but finds limited evidence supporting a strong association between *PPM1D* PTVs and breast cancer risk or related treatments.

702F

Gender differences in germline mutations in bladder cancer participants from the DiscovEHR study. *L. Bang[1], M. Shivakumar[1], D. Lavage[2], T. Garg[3,4], D. Kim[1,5] on behalf of the DiscovEHR collaboration.* 1) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA; 2) Phenomics Core, Geisinger Health System, Danville, PA; 3) Department of Urology, Geisinger Health System, Danville, PA; 4) Department of Epidemiology and Health Services Research, Center for Health Research, Geisinger Health System, Danville, PA; 5) The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA.

Sex-related differences in bladder cancer are clinically well-known and merit further genetic characterization. Bladder cancer affects men more frequently than women (3:1); however, women are diagnosed at higher stages and have worse overall survival. Of urologic cancers in The Cancer Genome Atlas (TCGA), bladder cancer had the most differences by sex and extensive sex-biased molecular signatures. The DiscovEHR study, a cohort with electronic health records (EHR) linked to the MyCode biobank of Geisinger Health System, provided whole exome sequencing (WES) and SNP array data for consented patients. We examined a cohort diagnosed with urothelial carcinoma (N=355) as well as an age, sex, and smoking status matched non-cancer control cohort (N=710). The influence of rare germline variants (MAF<0.05) from WES data on differential presentation of bladder cancer based on sex was characterized by rare-variant binning and applying a dispersion-based test using Bin-KAT, adjusting for age and smoking status. Rare germline variants in *ARSD* and *CCNI2* were significantly associated with sex-related discrepancies in bladder cancer occurrence. Females with bladder cancer were more likely to have variants in *CCNI2* than males (Bonferroni-corrected P =0.045); males with bladder cancer were more likely to harbor mutations in *ARSD*, a pseudoautosomal X-chromosome gene (Bonferroni-corrected P = 0.0032). This association with *ARSD* was separately confirmed in SNP array data for the same cohort of patients, with common missense variant *rs211653* within *ARSD* being suggestively significant (P = 2.0x10-6) in males with bladder cancer as compared to females. *ARSD* is a gene in the sulfatase family involved in X-linked skeletal dysplasia and warfarin metabolism; significant differential activity of Arylsulfatase A in patients with genitourinary tract disorders such as bladder cancer has been found previously. Our results demonstrate that germline mutation analysis in an unselected hospital-based cohort improve our understanding of the etiology of sex disparities in bladder cancer. A larger DiscovEHR cohort in the future will support the replication of these results.

703W

Integrative Bayesian group bridge regularization analysis in multiple heterogeneous high-dimensional survival data. *Y. Li.* School of Statistics, Shanxi University of Finance & Economics, Taiyuan, China.

In the analysis of survival data in high-throughput genomic studies, results generated in the analysis of a single dataset may suffer a lack of reproducibility because of small sample sizes. Integrative analysis provides an effective way of pooling information across multiple independent studies with comparable designs, and can be more effective than single-dataset analysis and meta-analysis. In this study, we conduct integrative bayesian group bridge regularization analysis for the sparsity structures of multiple heterogeneous high-dimensional survival data. Our study also advances from the existing studies by conducting systematic simulations and direct comparisons of multiple methods. The proposed approach can effectively accommodate heterogeneity among multiple studies and identify markers with consistent effects across studies. You may contact the author (during and after the meeting) at liyi@sxufe.edu.cn.

704T

Inflated genome-wide *de novo* mutation rate in carriers of *TP53* germline mutations. *X. Pan[1,2], J. Bojadzieva[1], E. Dodd[1], B. Li[3], D. Wheeler[4], V. Johnson[2], L. Strong[1], W. Wang[1].* 1) MD Anderson Cancer Center, Houston, TX; 2) Texas A&M University, College Station, TX; 3) Vanderbilt University, Nashville, TN; 4) Baylor College of Medicine, Houston, TX.

Germline mutations are the driving force behind genome evolution and etiology of diseases. It has been well established that the number of *de novo* germline mutations is positively correlated with the age of the father. However, it remains largely unknown what factors other than perhaps paternal age would affect the initiation and accumulation of germline mutations. This question is particularly important for Li-Fraumeni Syndrome (LFS), a cancer predisposition syndrome mostly caused by germline *TP53* mutations, in which there is an unusually high prevalence of *de novo* mutations (7-20%). To investigate this issue, we generated whole-genome sequencing data (at a 30x read depth) from five quartets of LFS probands, healthy siblings and their parents, and tried to compare the genome-wide *de novo* mutation number between the probands and their healthy siblings. We selected all available *de novo* probands who had unaffected siblings within close age range (1-4 years) in order to reduce the age difference between the paired proband and sibling as much as possible. All the *de novo* probands had at least two types of cancer, the most common being osteosarcoma and breast cancer. None of the siblings or parents of probands had any record of cancer. We applied the *de novo* mutation calling software Polymutt and additional stringent filters to identify genome-wide *de novo* mutations in the 5 probands with *TP53* mutations and 5 healthy siblings. As a result, we observed that in three of the five families, the number of *de novo* SNVs differed widely between the probands and their siblings (53 vs. 36, 48 vs. 19 and 54 vs. 34), while in the other two there was no difference (44 vs. 45 and 36 vs. 33). Overall, the probands had significantly higher *de novo* mutation number than their normal siblings (p = 0.035, paired t-test). This suggests the possibility that factors other than the age of the father may have a strong impact on the *de novo* mutation number within families, and that these factors may put individuals at increased risk for deleterious mutations.

705F

Detection of signal regions in whole genome genotyping and sequencing association studies using scan statistics. *Z. Li[1], X. Lin[1,2].* 1) Biostatistics Dept, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Statistics Dept, Harvard University, Cambridge, MA.

Large scale Whole-genome sequencing (WGS) studies, such as the TOPMed and Genome Sequencing Programs, are currently being conducted to identify susceptible rare variants associated with human diseases or traits. The classical single-marker association analysis for rare variants has limited power and SNP-set based analysis are commonly used in analyzing rare variants. A major challenge in analyzing WGS data is how to handle non-coding variants. These existing SNP-set based tests for rare variant analysis, such as burden and SKAT, require to pre-specify gene regions and hence are not directly to WGS data, as analysis units and window sizes (if a moving window method is used), are not well defined and unknown in WGS analysis, because of a large number of non-coding variant regions. We propose a scan statistic based method to optimal detect association signal segment sizes and locations by scanning the genome continuously. The proposed method accounts for the LD among genetic variants, and allows for signal regions to have both causal and neutral variants, and causal variants whose effects can be in different directions. The proposed procedure properly controls for family-wise error rates. Through analysis of simulated data, we showed that the proposed scan procedure can effectively detect association signal regions. We applied the proposed method to analyze WGS datasets to identify the genetic regions that are associated with heart- and blood-related traits. .

706W

A novel Bayesian multiple testing approach for region-based analysis of next generation sequencing (NGS) data. *J. Xu[1,2], W. Xu[1,3], R. Hung[1,2], G. Liu[1,3], L. Briollais[1,2].* 1) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 2) Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, Canada; 3) University of Health Network, Princess Margaret Cancer Centre, Toronto, ON, Canada.

The discovery of rare genetic variants through Next Generation Sequencing (NGS) is becoming a very challenging issue in the human genetic field. We have recently proposed a novel region-based statistical test based on a Bayes Factor (BF) approach to assess evidence of association between a set of rare variants located on this region and a disease outcome. Gene-based inference can be performed using permutation testing. For genome-wide inference (i.e. multi-regions test), we introduce a Bayesian False Discovery Rate (BFDR) control procedure inspired by the work of Efron (2005). In this setting, it is critical to estimate the null distribution of the BF for the non-associated regions. We used an empirical estimate of the null distribution and studied its properties by simulation and analytically. In particular, we showed that the choice of priors can impact seriously the BFDR procedure. Our BF approach has been applied to a study of lung cancer from Toronto including 262 cases and 261 controls with whole exome sequencing data. We applied the BFDR control procedure to >13.7K genes and estimated the FDR for the top 5, 10 and 20 genes to be respectively, 2.9%, 7.5% and 12.8%. Our top genes included for example TERT and TLR6. In conclusion, the use of empirical Bayes priors along with a Bayesian control of FDR offer a comprehensive framework to make genome-wide statistical inference about the important chromosomal regions associated with the disease of interest in the context of NGS data.

707T

Population-based breast cancer risk estimates associated with mutations in cancer predisposition genes from the CARRIERS study. *F.J. Couch[1], C. Hu[1], S.N. Hart[2], J. Lilyquist[2], K.Y. Lee[1], C. Gao[3], E.C. Polley[2], H. Anton-Culver[4], P. Auer[5], L. Bernstein[6], C. Haiman[7], E.M. John[8], L. Kolonel[9], J. Lacey[6], J.E. Olson[2], J.R. Palmer[10], K.J. Ruddy[11], D.P. Sandler[12], M.E. Sherman[13], J. Taylor[12], A. Trentham-Dietz[14], C.M. Vachon[2], C. Weinberg[12], S. Yao[15], M.M. Gaudet[16], S.M. Domchek[17], J.N. Weitzel[6], D.E. Goldgar[18], P. Kraft[3], K.L. Nathanson[17].* 1) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Harvard School of Public Health, Boston, MA; 4) University of California Irvine, Irvine, CA; 5) University of Wisconsin-Milwaukee, Milwaukee, WI; 6) City of Hope Cancer Institute, Duarte, CA; 7) University of Southern California, Los Angeles, CA; 8) Cancer Prevention Institute of California, Freemont, CA; 9) University of Hawaii, Manoa, HI; 10) Boston University, Boston, MA; 11) Oncology, Mayo Clinic, Rochester, MN; 12) NIEHS, Research Triangle Park, NC; 13) Laboratory Medicine and Pathology, Mayo Clinic, Jacksonville, FL; 14) University of Wisconsin, Madison, WI; 15) Roswell Park Cancer Institute, Buffalo, NY; 16) American Cancer Society, Atlanta, GA; 17) University of Pennsylvania, Philadelphia, PA; 18) University of Utah, Salt Lake City, UT.

Clinical germline genetic testing of cancer predisposition gene panels is now widely used to identify women at increased risk for breast cancer. The identification of pathogenic mutations in established predisposition genes, such as *BRCA1* or *BRCA2*, may result in improved risk management of breast cancer for tested patients and their family members through tailored screening, prophylactic surgeries, or chemoprevention. The efficacy of these clinical interventions shows the promise of identifying pathogenic mutations in predisposition genes. However, the utility of results from hereditary cancer panel testing of genes other than BRCA1 and BRCA2 has been limited because the levels of risk for breast cancer associated with mutations in several of the panel genes have not been clear. Recent studies of women eligible for clinical hereditary panel testing have likely overestimated the risks associated with mutations in each gene because of the presence of genetic modifiers associated with a family history of breast and/or ovarian cancer. Here we report interim results from the "CAnceR RIsk Estimates Related to Susceptibility" (CARRIERS) study, which aims to estimate breast cancer risks associated with mutations in hereditary cancer panel genes in the general population. CARRIERS involves sequencing all coding sequence and splice consensus sites from 37 cancer predisposition genes along with 126 common genetic risk factors for breast cancer in 40,000 breast cancer cases and 35,000 matched controls from cohort-based nested breast cancer studies and population-based case-control studies. We will report associations with risk of breast cancer for individual pathogenic mutations and combined mutations in each gene, as well as interactions with strength of family history. Population-based risk estimates for breast cancer for each of the hereditary cancer panel genes will be useful for risk management of patients and for development of breast cancer risk prediction models.

708F

Clinicopathological and prognostic significances of EGFR, KRAS, BRAF and PI3KCA mutations in biliary tract cancer. *H. Lee[1], S. Kang[1], T. Park[3], J. Kim[2].* 1) hematology-oncology, Ajou university school of medicine, Suwon, Gyeonggido, South Korea; 2) pathology, Ajou university school of medicine, Suwon, Gyeonggido, South Korea; 3) biochemistry, Ajou university school of medicine, Suwon, Gyeonggido, South Korea.

Biliary tract carcinomas (BTCs) are more common in Korea in western countries. It is more difficult to treat. Epidermal growth factor receptor (EGFR) is a therapeutic target for the cancer. Mutations of the EGFR gene and the activation of its downstream pathways, including KRAS and BRAF, is important for targeted therapy. This study aims to analyze the EGFR, KRAS, BRAF and PI3KCA mutations in BTCs and their association with clinical outcomes. Paraffin-embedded specimens containing 113 resected BTCs at the Ajou University Hospital between 2009 and 2013 were analyzed. The exons 18-21 of EGFR gene, the codon 12, 13 and 61 of KRAS gene, BRAF V600E, and PI3KCA mutation were analyzed by cobas 4800. We tested the correlation between these mutations and the overall survival, tumor location, stage, and lymphocyte numbers at diagnosis in BTCs. There is no mutation of EGFR and BRAF mutation in BTC patients. 23(20.3%) BTC patients had the codon 12, 13 of KRAS mutation, 12(10.6%) patients had the codon 61 of KRAS mutations while 7 (6.1%) patients had PI3KCA mutations. Factors influencing survival on univariate analysis were tumor positive margin, lymphocyte, perineural invasion and bilirubin level. On multivariate analysis, tumor positive margin and perineural invasion were independent prognostic factors. A correlation between KRAS mutations and survival tend to be observed(p=0.051). PI3KCA mutation was not correlated with survival(p=0.956). EGFR and BRAF mutations are not shown in BTCs. KRAS mutation showed 34 patients (31%) and 7 (6.1%) patients had PI3KCA mutations. Tumor positive margin and perineural invasion were independent prognostic markers in BTCs.

709W

Evolutionary genomic analysis of a longitudinal series of prostate cancer bone metastases and xenografts from one patient revealed selection of progressively therapy resistant metastatic clone. *T. Gaasterland[1,3,4], M. Muldong[2], A. Gallegos[3,4], C. Wu[2,5], T. Mendoza[2,5], W. Zhu[2,5], L. Edsall[1,3], M. Liss[6], A. Kulidjian[2,7], C. Jamieson[2,5].* 1) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 2) Moores Cancer Center, University of California San Diego, La Jolla, CA; 3) Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA; 4) Bioinformatics and Systems Biology Program, University of California San Diego, La Jolla, CA; 5) Urology, Department of Surgery, University of California San Diego, La Jolla, CA; 6) Department of Urology, University of Texas Health Sciences Center, San Antonio, TX; 7) Department of Orthopedic Surgery, University of California San Diego, La Jolla, CA.

Background: Prostate cancer (PrCa) metastasizes to bone in 50-90% of patients with advanced disease, yet little is known about genome-wide alterations in these metastases. To understand genome changes and progressive therapy resistance, we examined genome variation and gene expression in a 3-year longitudinal series of surgical PrCa bone metastasis (mets) samples and xenografts derived from the same patient. Methods: Surgical specimens collected during orthopaedic repair surgery resulted in a longitudinal sample series including blood, mets 1 (post-ADT, radiation), mets 2 (+10.5 mon. post-ADT, radiation, docetaxel), mets 3 (++12 mon, post-abiraterone, radiation, cabazitaxel). Intra-femoral xenograft mouse model PCSD1 was established from mets 1. We used copy number variation (CNV) from SNP arrays (Affymetrix Oncoscan), whole exome sequencing (WES), and gene expression from all samples to examine longitudinal changes, impact of treatment, and xenograft genome integrity. Results: CNV profiles of mets 1 and a low passage PCSD1 xenograft revealed extensive alterations highly conserved in both, with copy number gains across the majority of the genome and scattered regions of loss. In contrast, comparison of CNV profiles for mets 1 and 2 showed significantly different patterns with some regions of conservation. Unexpectedly, CNV profiles were highly conserved between mets 2 and high passage PCSD1 xenografts derived from mets 1. Whole exome analysis identified shared and unique point mutations across all samples. Phylogenetic dendrogram analysis of these mutations together with CNV profiles indicated reduced heterogeneity over time, and selection of one sub-population both in patient samples and xenografts. WES also identified germline alterations in DNA repair genes (BRCA2, ATM and CHEK2) and tumor suppressor genes (TP53, PTEN). Treatment of the xenograft model with anti-androgen bicalutamide induced expression of a neuronal gene signature. Conclusions: CNV and WES revealed that selection of a similar therapy-resistant sub-population occurred in both patient metastases and xenografts derived from the same patient, providing the first direct evidence that a therapy-resistant sub-clone was already present in early metastasis. Analysis of surgical metastases at different stages of treatment and progression in this patient provides a foundation to profile genomic diversity in recurrent bone metastatic PrCa.

710T

The early evolutionary signatures of clonal hematopoiesis leading to blood based cancers and cardiometabolic conditions. *K. Skead[1], S. Abelson[2], J. Wang[2], M. Minden[2,3], L. Shlush[2,4], J. Dick[2,5], P. Awadalla[1,5].* 1) Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Princess Margaret Cancer Centre, Toronto, Ontario, Canada; 3) Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada; 4) Weizmann Institute of Science, Rehovot, Israel; 5) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

Population aging has emerged as a major factor influencing patterns of health and shifting disease profiles. Recent studies have shown that hematopoietic stem cells (HSCs) accumulate somatic mutations over time as a function of regular aging. Certain private mutations confer a slight fitness advantage causing particular HSCs to undergo clonal expansions and contribute disproportionately more to the mature blood cell pool in addition to impacting disease susceptibility in a process referred to as clonal hematopoiesis. Most notably, these clonal expansions have been found to associate with blood-based cancers and various cardiometabolic conditions. However, clonal hematopoiesis has been observed in individuals both in the absence of a disease diagnosis and as a precursor state to disease development; a phenomenon termed clonal hematopoiesis of indeterminate potential (CHIP). As such, there is a clear need to capture and characterize the evolutionary trajectory from normal blood cell composition, through to CHIP and, more significantly, disease state clonal hematopoiesis. For the purposes of this study, we have leveraged two large population cohorts, the Canadian Partnership for Tomorrow Project (CPTP) and the European Prospective Investigation into Cancer and Nutrition (EPIC) study, to identify incidental cases of cardiometabolic conditions, hematological cancers, "healthy" CHIP and corresponding biological samples collected prior to diagnosis. Following targeted deep sequencing and the genomic characterization of the pre-hematological cancer, pre-cardiometabolic and "healthy" CHIP samples, we are using a Bayesian approximation approach to construct quantitative evolutionary framework with which to assess the forces shaping the accumulation of somatic mutations in blood, and interpret site frequency spectrums with respect to clonal expansions and clinical outcomes. Effectively, we will discern early stage patterns of clonal hematopoiesis indicative of 1) hematological cancers, 2) cardiometabolic conditions, from 3) "healthy" CHIP, that is, benign clonal patterns not found to be a precursor state to disease. Ultimately, this study provides a platform through which we are able to explore the predictive utility of clonal hematopoiesis in order to differentiate individuals more likely to be susceptible to age associated diseases in addition to functioning as an earlier diagnostic tool.

711F

Similar frequency of ACMG-59 gene secondary findings in a large whole exome sequenced cancer cohort and ethnicity-matched controls. *J. Kim[1], M. Wang[2,3], W. Luo[2,3], MM. Gaudet[4], SM. Gapstur[4], B. Zhu[2,3], AM. Goldstein[1], DR. Stewart[1], NCI DCEG Cancer Genomics Research Laboratory, NCI DCEG Cancer Sequencing Working Group.* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 3) Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; 4) Epidemiology Research Program, American Cancer Society, Atlanta, GA.

Intro. The frequency of pathogenic (P) or likely pathogenic (LP) secondary findings (SF) in ACMG-59 genes is 1-3%, largely from clinical laboratories and analysis of publicly available databases. We investigated the frequency of ACMG-59 SF variants in a large, heterogeneous, family-based cancer-research cohort and in cancer-free ethnicity-matched controls. *Methods.* Whole exome sequencing was conducted on 1320 cases and 994 cancer-free controls from a variety of cancer-based studies. Examination of ethnicity-informative variants ensured comparable case/control population structure. We matched capture region and read depth for case-control comparisons. Pathogenicity of ACMG-59 variants (MAF < 1%) was determined using InterVar (a semi-automated application of ACMG/AMP Guidelines) and a hotspot-, ClinVar-, HGMD-DM- and LoF-based tiering scheme (tier 1: most pathogenic). *Results.* Using InterVar, in the 24 ACMG-59 cancer genes, we observed 6 P/LP unique variants (22 total; 2.2%) in controls and 5 P/LP unique variants (27 total; 2.0%) in cases (P=0.77). In the 35 ACMG-59 non-cancer genes, we found 9 P/LP unique variants (9 total; 0.9%) in controls and 12 P/LP unique variants (18 total; 1.0%) in cases (P=0.34). There were 696 unique VUS in cases and 590 unique VUS in controls. Alternatively, using the tiering scheme, in the 24 ACMG-59 cancer genes, we found 76 unique tier 1 variants (146 total, 15%) in controls and 78 tier 1 variants (229 total, 17%) in cases (P=0.09). However, 52% of tier 1 variants had conflicting P/LP status between ClinVar and HGMD-DM. No significant differences were observed in the ACMG-59 non-cancer genes. Similar results were observed when restricting the analysis to 1 cancer patient/pedigree. *Discussion.* Using InterVar, a conservative method, overall, we observed ~3% P/LP variants and no significant differences in P/LP variant frequencies between cancer cases and controls for both ACMG-59 cancer and non-cancer genes. Similarly, there were no significant differences in tier 1 variant frequencies in cases and controls. We acknowledge limited power inherent in studying rare variants, especially for single genes. For both methods used, the large number of VUS require detailed database and literature review, a time- and labor-intensive process hampered by the difficulty in interpreting conflicting P/LP determinations. Our data suggest that resolving numerous VUS will likely impose a significant burden on the return of ACMG-59 SF.

712W

Genetic diversity and rare mutations in the Puerto Rican population. *E.J. Torres Gonzalez[1], W. Wolfsberger[2], T.K. Oleksyk[2], M. Dean[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD; 2) Caribbean Genome Center, University of Puerto Rico at Mayaguez, PR.

Modern Puerto Ricans are an admixture of African, European and Native American ancestral populations. Previous studies such as 1000 Genomes have reported on common mutations and the distribution of admixture in the island. To identify rare (<1 %) and potentially disease-related mutations we carried out exome sequencing of a diverse Puerto Rican sample set. From a collection of 673 individuals from 66 municipalities we selected 120 individuals from 10 municipalities, based on either high expected European, African or Native American genetic ancestry. A preliminary analysis yielded some known disease mutations in the population, such as Phenylketonuria (PKU), G6PD deficiency, Stargardt macular dystrophy (STGD), Familial hypercholesterolemia (FH), HFE-associated hereditary hemochromatosis, Familial Mediterranean fever (FMF), Sickle-cell disease (HbS), and the Duffy and Diego blood groups. Variant calling was carried out with GATK and LoFreq. These samples will be joined to 107 individuals from the 1000Genomes (PUR population) and 2 targeted sequencing panels.   .

713T

Functional role of intragenic methylation in alternative splicing in cancer. *Y. Lee, D. Kim, S. Han.* Biomedical Informatics, University of Utah, Salt Lake City, UT.

Background: DNA methylation in promoter regions leads to transcriptional silencing, such as of tumor suppressor genes in cancer. However, the functions of DNA methylation in the gene body (intragenic, i.e. exons and introns) have not yet been completely elucidated. Recently, hypermethylation has been implicated in enhancing exon recognition by recruiting methyl—CpG—binding protein (MECP2) to hypermethylated sites. In this study, we examined intragenic methylation in splicing regulatory elements (SREs: exonic splicing enhancer (ESE), exon splicing silencer (ESS), intronic splicing enhancer (ISE), intronic splicing silencer (ISS)) as a mechanism to understand how epigenetic factors contribute to alternative splicing events in cancer. **Hypothesis**: Methylation patterns of SREs are involved in directing alternative splicing events in cancer. **Methods**: We developed a splicing decision model to identify actionable methylation loci potentially affecting splicing events (i.e. exon skipping) among CpG sites in putative intragenic SRE regions. We used DNA methylation and RNA—seq data for breast and lung cancer samples and normal cases from The Cancer Genome Atlas (TCGA). We additionally used hyper- and hypomethlyation data published in Selamat et. al (2013) as an independent lung cancer dataset. To identify regions hyper— and hypomethylated in cancer, we performed an unpaired t-test for differential methylation status (beta value) between cancer and normal cases. Multiple testing was accounted for by controlling the false discovery rate ($p < 0.05$ with FDR< 0.1). **Results**: We investigated whether specific patterns of methylation were enriched in SRE regions affecting exon skipping. We found that for breast cancer samples, differential methylation status was enriched in SRE regions but not in non—SRE regions. Particularly, hypomethylation in cancer showed a greater enrichment in ESE regions with putative exon skipping events. Similar results were observed in the both the TCGA lung cancer cases and an independent lung cancer dataset. These data suggest that hypomethylation in SRE regions may cause failure to recruit MECP2, therefore leading to exon skipping. **Conclusion**: Our study suggests that intragenic methylation status is important for splicing regulation and may improve our understanding of how intragenic epigenetic markers play a role in gene regulation by affecting alternative splicing.

714F

Systematic evaluation of copy number variations: Towards rational personalized cancer therapy. *S. Appenzeller[1], A. Gesierich[2], A. Thiem[1,2], A. Hufnagel[3], H. Kneitz[2], A. Rosenwald[4], R. Bargou[1], M. Goebeler[2], S. Meierjohann[3].* 1) Comprehensive Cancer Center Mainfranken, University Hospital Würzburg, Würzburg, Germany; 2) Department of Dermatology, Venereology and Allergology and Skin Cancer Center, University Hospital Würzburg, Würzburg, Germany; 3) Department of Physiological Chemistry, University of Würzburg, Würzburg, Germany; 4) Institute of Pathology, University of Würzburg, Würzburg, Germany.

   With the identification and functional characterization of an increasing number of tumor oncogenes, the inhibitor pools of pharmaceutical companies have significantly increased in the last decades. Nowadays, numerous inhibitors e.g. targeting a large number of kinases are either approved by FDA and EMA or are currently evaluated in clinical trials. However, usage of these inhibitors is largely restricted to tumors, which carry a defined single activating somatic mutation, e.g. BRAF (p.V600E/K) in melanoma or EGFR (p.L858R) in non-small cell lung carcinoma. Copy number variations, such as deletions or amplifications of tumor-relevant genes, are rarely investigated, although they may be of relevance and potentially targetable by pharmacological intervention. We used targeted deep panel sequencing to identify genetic alterations in melanoma. We used DNA isolated from paired tumor and fresh-frozen blood (as non-pathological control) samples in order to be able to detect somatic single nucleotide variants (SNVs) as well as somatic copy number variations (CNVs) in a panel of more than 50 melanoma relevant genes. While the detection of somatic mutations and small indels using targeted panel sequencing was feasible, CNV detection turned out to be challenging with the restricted genomic information of the gene panel. This is mainly due to the fact that thresholds for calling a CNV are largely undefined and existing bioinformatic tools give unsatisfactory and conflicting results. To address this issue, we developed a bioinformatic pipeline, which is able to reliably identify CNVs even from small sequencing panels. The results were evaluated and confirmed using an array-based assay, the current gold standard for CNV analysis. This analysis will be a valuable diagnostic and prognostic device, which can be used in the clinical setting for personalized therapy decisions at comparatively low cost.

715W

Single-molecule resolution of haplotype-specific, megabase-scale and complex oncogenic rearrangements in metastatic cancers. *S. Greer[1], B. Lau[2], J. Chen[1], C. Xia[1], H. Ji[1,2].* 1) Oncology, Stanford University, Stanford, CA; 2) Stanford Genome Technology Center, Stanford University, Stanford, CA.

   Resolution of complex structural variants (SVs) remains one of the last major challenges of NGS data analysis, despite the ubiquity and significance of SVs in both normal and disease genomes. SVs are difficult to resolve due to the inability of short sequencing reads to span across complex rearrangements, leading to resolution only by indirect evidence. To address this challenge, we have developed a method that leverages barcoded sequence reads derived from single molecules to resolve complex genomic structures without the need for assembly. With this methodology, we have resolved complex rearrangements involving multiple classes of SVs over a range of many megabases which are not evident even with long read sequencing technologies.   We have successfully resolved complex rearrangements at megabase scale across many types of malignancies, including diffuse gastric cancer, colorectal cancer, and cholangiocarcinoma. All of this was accomplished at single molecule accuracy and was possible even in samples with tumor fractions of less than 30%. Our initial analysis involving metastatic samples from a patient with primary diffuse gastric cancer has been published and highlights our ability to identify unique rearrangement motifs that have undergone segmental duplication. Further, clustered somatic rearrangements could be assigned to specific parental haplotypes. This was possible without the need for long read sequencing.   Further, we used haplotype-informed local assembly to validate haplotype-specific rearrangement motifs. We boosted the performance of assembling tumor-specific SVs by the bioinformatic extraction of somatic regions at single molecule resolution. Sequence assembly thus yielded approximately 100 kilobase contigs that confirmed the results of our SV resolution method. Overall, we demonstrate the utility of SV resolution and assembly to improve our understanding of genomic instability in the context of cancer genome analysis.

716T

NGSEA: Network-augmented Gene Set Enrichment Analysis. *H. Han, S. Lee, I. Lee.* Biotechnology, Yonsei University, Seoul, South Korea.

Gene set enrichment analysis (GSEA) is a widely used computational method to test a statistical association between a gene set and rank-ordered expression signatures. Despite the popularity, there are shortcomings in this method. For examples, perturbation of disease genes in the patient would give expression signature of the responsive genes rather than the causal genes, resulting in limited associations between the patient expression signature and the relevant disease. Furthermore, the mixed signature of up- and down-regulations for the same disease gene set could diminish the statistical significance of the association between the expression signature and the disease. To overcome such limitations, we present a modified GSEA, network-augmented GSEA (NGSEA), which uses re-ranked signature genes based on differential expressions of neighbors in a functional gene network. In this method, genes with differentially expressed neighbor genes are more weighted. We found NGSEA performed better than GSEA in identifying associations between KEGG disease pathways and expression signatures derived from the patient with a relevant disease. NGSEA also shows higher performance than GSEA in identifying drugs for patients with relevant diseases, suggesting that NGSEA can be applied to drug repositioning. With these results, we demonstrated that incorporating information of gene functional networks can improve the GSEA. NGSEA should be a useful resource, as network-augmented scoring allows us to identify disease-related gene sets which were undetectable with expression signatures only.

717F

Methodologies for FMS-like tyrosine kinase (FLT3) internal tandem duplication (ITD) detection from two types of next generation sequencing data such as whole exome sequencing and amplicon sequencing. *D. Kim[1], Y. Hong[2], Y. Koh[3], K. Ahn[4], J. Shin[3], H. Yun[2], S. Yoon[2,1].* 1) Seoul National UnversiCancer Research Institute, Seoul National University College of Medicine, 101 Daehagro, Jongno-gu, Seoul 110-799, Koreaty, Seoul, South Korea; 2) Development Group, Healthcare Analytics Team, Samsung SDS, SDS Tower West Campus 11F, Olympic-ro 35-gil, Songpa-gu, Seoul 138-240, Korea; 3) Department of Internal Medicine, Seoul National University Hospital, 101 Daehagro, Jongno-gu, Seoul 110-744, Korea; 4) Functional Genome Institute, PDXen Biosystem Inc. Hwarang Ro, KIST venture town, Seoul, Republic of Korea.

We developed new ITD detection algorithm (ITDetect) and applied two type of next generation sequencing data based on capture library construction method (WES) and PCR amplification (amplicon sequencing). ITDetect is based on BWA and is specified for ITD detection including FLT3-ITD. We validated and compared result of ITDetect with other ITD detecting algorithms using nested polymerase chain reaction (PCR) method. In 81 acute myeloid leukemia patients with WES data, FLT3-ITD was positive in 11 patients (13%) when called with ITDetect, all of whom were validated with nested PCR. Meanwhile FLT3-ITD was positive only in 7/81 patients by conventional PCR. The concordance rate of ITDetect and nested PCR was 95% (77/81). ITDetect showed better ITD detection performance when compared with previously reported ITD callers. In large AML cohort (n=213), patients who were positive for FLT3-ITD with nested-PCR but not with conventional PCR had shorter survival outcomes than patients who were negative for FLT3-ITD with nested PCR, suggesting clinical significance of sensitive FLT3-ITD detection. We also performed amplicon based sequencing in separated AML cohort (n=41) and applied ITDetect and Pindel. In this cohort, Pindel was shown better performance than ITDetect. Among 5 FLT3-ITD positive patients in 41 AML cohort, 2 of them were FLT3-ITD positive with ITDetect but, 4 of them were FLT3-ITD positive with Pindel. In conclusion, It is very careful when ITDetect is applied for amplicon based sequencing data. Amplicon based sequencing data was shown irregular soft clipping reads, because amplicon based sequencing is logically likely to include PCR error. However, we developed more sensitive detection methods for FLT3-ITD based on WES data that is clinically meaningful and low performance will be overcome after optimizing for amplicon sequencing data.

718W

Genomic analysis of PDX sequencing data. *J. Kim, S. Kim, A. Choi, S. Lee.* Ewha Womans University, Seoul, South Korea.

   PDX mouse model is an emerging platform to test treatment responses for cancer patients in preclinical setup, thus providing ample opportunities to realize the personalized and precision medicine. Detailed histological and molecular characterization is essential to confirm the compatibility of PDX mice with the patient tumor. We have established 220 PDX models representing 180 lung cancer patients and generated whole exome and transcriptome sequencing data. Ideally, trio samples of patient normal, patient tumor, and mouse tumor tissues are required to identify somatic mutations in PDX mouse that are concordant with patient tumor. Here we describe computational pipelines for analyzing deep sequencing data for PDX mice. First, we present a procedure for determining the compatibility between PDX mice and the patient tumors based on intricate mapping and accurate detection of somatic variants. Resulting somatic variants and copy number variations (CNVs) were compared in 70 cases with trio samples. We classified the PDX mice as good models when >10% of somatic variants in patient tumor sample were replicated in the PDX mice. We obtained 40 good models out of 70 cases (success rate of 57%). The CNV profiles of patient and PDX tumors were similar in successful cases. Next, we developed a computational pipeline to identify somatic mutations for exome sequencing data of PDX mouse when patient normal or tumor data were missing. Extensive filtering processes were used to remove mouse-originated mutations and germline mutations. We tested our pipeline for 70 trio cases of lung cancer, and demonstrated that we could retrieve most of genuine somatic mutations without too much of false positives without patient reference tissues. Our results highlight the power, cautions, and limits of PDX sequencing data.

719T

A tailored topic model integrates both nucleotide context and genomic location heterogeneity in mutational process profiling. *S. Li[1], M. Gerstein[1,2,3].* 1) Computational Biology and Bioinformatics, Yale University, New Haven, CT; 2) Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 3) Department of Computer Science, Yale University, New Haven, CT.

   Multiple endogenous and exogenous mutational processes modify cancer genome and leave distinct fingerprints over time, including somatic alterations both before the cancer initiation and during cancer development. Noticeably, they exhibit characteristic genomic location preferences and mutational nucleotide context biases. In a generative model, multiple latent processes generate mutations drawing from their corresponding nucleotide context distributions ("mutation signature"). By sequencing cancer sample, we observe mutations accumulated from mixed mutation processes that are unevenly operative in different regions of the genome. Several mutation processes have been recognized by both computational and experimental studies and associated with known etiologies. Mutation is the cornerstone of tumorigenesis and understanding the fundamental underlying processes helps gain insights of cancer initiation and development.   One key issue is to disentangle the mixed mutational processes in cancers mutation profiles. Previous work assumed these processes are evenly distributed when learning them, although post-analysis found they show genomic location preference. Here we propose a topic model that organically encodes mutation heterogeneity of both genomic location and mutational nucleotide context. Our model allows mutational processes smoothly evolving as dictated by the genomic location and features. Each mutational process has a characteristic nucleotide context distribution to reflect the underlying chemical and biological biases. Additionally, to avoid the nucleotide context complexity blowing up exponentially, our model parsimoniously builds the context tree by pruning less informative subtrees. By simultaneously learning mutational processes from both genomic features and nucleotide context, our model leads to robust and biologically interpretable mutational processes inference. Moreover, consolidating these two dimensions of heterogeneity gives us a chance to better portray mutational processes and investigate the vibrant interplay between them. We demonstrated the performance of our model on both simulated data and real world cancer data.

720F

Copy number segmentation with left-to-right hierarchical Dirichlet process hidden Markov model and segment clustering. *K. Liao[1], W. Hsieh[2].* 1) Department of Computer Science, National Tsing-Hua University, Hsinchu City, Taiwan; 2) Institute of Statistics, National Tsing-Hua University, Hsinchu City, Taiwan.

Hidden Markov model (HMM) is a powerful statistical tool and is often used to analyze copy number data. A limitation of the standard HMM is that the number of states is fixed. Although the number of states can be selected using model selection methods such as Bayesian information criterion, they are mostly computationally expensive and there is no consensus on the optimum method (Harati Nejad Torbati et al., 2016). Hierarchical Dirichlet process hidden Markov model (HDPHMM) is a variant of hidden Markov model which allows unbounded number of states. The number of states can be automatically decided with Gibbs sampling methods. However, standard HDPHMM only allows ergodic structure with time-homogeneous state transition probability and this leads to relatively uniform lengths of segments with the same state. We developed a variant of HDPHMM called left-to-right HDPHMM with segment clustering (LRC-HDPHMM), which allows the same state to change state transition probability over time. Our method is based on a variant of HDPHMM called LR-HDPHMM which implements non-ergodic structure by disabling state transition from right to left. Although LR-HDPHMM is free from relatively uniform segment length, the same state cannot be revisited again. We solved this problem by clustering segments using another Dirichlet process which generates the state space of interest. Simulation study showed that LRC-HD-PHMM is better than the standard HDPHMM in that it allows relatively variable lengths of segments with the same state. We applied our method to the analysis of subclonal copy number data. Compared to other HMM-based methods, our model has more flexible state transition probability and allows unbounded number of subclones and copy number. Furthermore, our method can be easily extended to jointly analyze multiple samples from the same tumor.

721W

Heterogeneous deconvolution of mixed tumor expression - DeMix-Py. *R. Liu[1,2], W. Wang[3], C. Huang[4], H. Zhu[1].* 1) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Statistics, Rice University, Houston, TX; 3) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 4) Department of Biostatistics, UNC at Chapel Hill, Chapel Hill, NC.

Introduction: Tumor samples normally contain non-unique tissues such as immune or healthy tissues, which leads to estimation error of gene expression signatures associated with cancer diagnosis and main therapy. The existence of mixed tumor samples consequentially affects gene expression profiling commonly performed under RNAseq and based transcriptomics. The physical purification is generally infeasible, therefore, in silico dissection of mixed tumor samples is an expecting tool that needs to be precedent to analyze gene expressions. **Methods and Results:** We present a frequentist model (DeMix-Py) applicable for RNAseq data follow a negative-binomial distribution. In DeMix-Py, the gene expressions are assumed to be observable from both unmatched pure normal tissue samples and mixed tumor tissue samples (mixtures of tumor tissue and health tissue), where the normal tissue and tumor tissue are assumed to follow the negative-binomial (NB) distribution. In order to estimate the mixture proportion and the parameters in NB distribution, our DeMix-Py proposes a three-step strategy: (i) estimating parameters in NB distribution for normal tissues; (ii) estimating the mixture proportion and parameters in NB distribution for tumor tissues; and (iii) calculating the normal tissue expression and tumor tissue expression. Simulation studies show that DeMix-Py performs well in both deriving the mixture proportion of tumor tissue in various gene expression datasets and calculating the tumor tissue expression for each gene samples with extremely heterogeneity. Moreover, DeMix-Py outperforms the existing Bayesian method in term of computational complexity for large-scale data analysis. **Conclusion:** The proposed frequentist DeMix-Py method provides an effective deconvolution way of mixed tumor expression. It is a natural route that offers uncertainty measures for proportions through posterior inference.

722T

Identification of germline copy number variations (CNVs) using targeted sequencing data on 6q in hereditary lung cancer families. *D. Mandal[1], D. Tran[1], K. Wood[1], A. Musolf[2], M. Badoo[3], M. de Andrade[4], C. Gaba[5], P. Yang[4], M. You[6], M. Anderson[7], A. Schwartz[8], S. Pinney[7], C. Amos[9], J. Bailey-Wilson[2].* 1) Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 3) Tulane Cancer Center, New Orleans, LA; 4) Mayo Clinic, Rochester, MN; 5) University of Toledo Dana Cancer Center, Toledo, OH; 6) Medical College of Wisconsin, Milwaukee, WI; 7) University of Cincinnati College of Medicine, Cincinnati, OH; 8) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 9) Geisel School of Medicine, Dartmouth College, Lebanon, NH.

Family history is a significant risk factor for lung cancer (LC). About 10% of lung cancer cases (22,000 cases per year in the U.S.) have at least one first-degree relative affected with lung cancer, and 25% of cases have at least one first- or second-degree affected relative. Previous multipoint linkage analyses of hereditary lung cancer families (≥3 LC cases/family) collected by the Genetic Epidemiology of Lung Cancer Consortium detected significant linkage to the 6q25 region. To identify germline genetic variants, targeted sequencing of 6q23-6q27 was obtained. Using these sequencing data we have performed analyses to identify germline copy number variations (CNVs). CNVs have been reported on lung cancer cell lines and somatic samples; however, no germline specific CNVs have been reported in high-risk LC families. The objective of the present study is to detect germline CNVs from the targeted sequencing data in hereditary lung cancer (HLC) families that might be responsible for increased lung cancer susceptibility in these families. In addition to identifying CNVs, transmission of germline CNVs in these families will be detected in multiple generations using targeted sequencing data on affected and informative unaffected family members in future. The nine most informative high-risk HLC families (≥4 LC cases/family) on whom we have targeted sequencing data included 25 affected and 55 informative unaffected family members. Phenotype, genotype, and smoking data are available on all family members. A CNV-specific algorithm incorporated in CANOES (http://www.columbia.edu/~ys2411/canoes/) was used in order to successfully call and identify germline CNVs. To confirm, CNVs were then visualized independently using Integrative Genomics Viewer (IGV). We have identified germline CNVs on five genes, including *EYA4*, *PARK2*, *RGS17*, *ARID1B*, and *NOX3*, in four LC families. These genes were identified earlier as candidates using either linkage analysis of the HLC families or by targeted sequencing analysis of 6q23-25. CNVs on these genes are present in more than one HLC family. Our findings suggest that the identified genes warrant further investigation to unveil functional mechanisms in the HLC families. In the long run, identification and hereditary transmission of these germline CNVs might be useful in conjunction with the identification of other genomic variants in the detection of high-risk individuals with multiple cases of lung cancer in the families.

723F

ViFi: Virus integration and fusion identification in tumor samples. *N. Nguyen, V. Deshpande, J. Luebeck, V. Bafna.* Computer Science and Engineering, University of California, San Diego, La Jolla, CA.

Viral infection followed by integration into the human genome has been shown as a driver of tumorigenesis, particularly in cervical cancer and hepatocellular carcinoma (HCC). Detecting viral integrations from NGS technology, however, is a challenging task. One problem hindering existing tools is separating out true integrations from false positives integrations. For example, integrations from the Hu et al., Nat. Genet. 2015 study have been challenged as false positives integrations as a result of sequencing artifacts (Dyer et al., Nat. Genet. 2015). We developed Virus Integration and Fusion Identification (ViFi), a tool for detecting viral integrations from WGS data and human-virus fusion mRNA from RNASeq data. ViFi uses a combination of reference-based mapping and phylogenetic modeling to identify viral reads, and incorporates the mappability scores to reduce false positive detections. The end result is a tool that can accurately detect viral integrations, even when the viruses are highly mutated or are not found in the reference genomes. We present a comparison of ViFi to a competing tool, VERSE, on both simulated and biological datasets with experimentally verified integrations. In a simulation study, both ViFi and VERSE have high precision and recall on the easy model conditions. However, on the hardest model condition where the integrated virus is only 65% similar to the closest related reference virus, the use of phylogenetic modeling allows ViFi to accurately detect viral integrations with a precision and recall of 88%, while VERSE fails to detect any integration. On the biological datasets, ViFi detects an additional integration compared to VERSE. We applied ViFi to 68 cervical cancer samples with matched WGS and RNA-Seq data. We found integrations in 51 samples, of which 49 contained fusion mRNA sequences. Importantly, genomic segments with integrations had increased expression of nearby human mRNA (2.5 fold increase), with significantly higher expression when human-viral fusion mRNA are present (5.5 fold increase). Surprisingly, we found that integration also results in the expression of proximal LTR and LINE elements (3.5 to 5.3 fold increased expression when fusion mRNA is present). Guided by previous results that implicate fusion HBV-LINE1 with increased tumor susceptibility in HCC (Lau et al. Cancer Cell 2014), our findings highlight the need to explore the roles of LINE and LTR elements in HPV pathogenesis in cervical cancer.

724W

Expanding GEMINI to annotate and prioritize subclonal mutations in heterogeneous tumors. *T. Nicholas[1,2], B. Pedersen[1,2], X. Huang[1,2], Y. Qiao[1,2], G. Marth[1,2], A. Quinlan[1,2].* 1) Human Genetics Department, University of Utah, Salt Lake City, UT; 2) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT.

DNA sequencing has unveiled extensive tumor heterogeneity in several different cancer types, with many exhibiting substantial clonal substructures. Considerable genetic variation exists within individual tumors and between metastatic sites from shared phylogenies, highlighting ongoing tumor growth and evolution. These differences can be further magnified over time by selective pressures introduced by different treatments. Properly identifying and tracing this variation throughout the expansion and progression of a tumor represents a significant challenge. Furthermore, being able to prioritize and focus on mutations most likely to contribute to tumor evolution or that could serve as potential therapeutic targets represents an ongoing problem, but is crucial towards determining possible clinical care. GEMINI is an effective tool for exploring genetic variation by enabling the querying of multiple genome annotations into a single database. Here we describe our progress towards an adaptation to the GEMINI framework that is more ideally suited for the complex patterns of genetic variation inherent to the study of tumor heterogeneity across multiple tumor biopsies. This is accomplished by enhancing GEMINI variant annotation to include tumor clonal specifications. Additionally, by incorporating existing tools and databases that facilitate the interpretation of cancer mutations (e.g., CIViC, DGIdb, VEP, CRAVAT) into the GEMINI framework, the identification of mutations that may be driving clonal evolution is simplified. The advances to the GEMINI framework that we are developing will better enable cancer scientists and research oncologists to discern genetic details, at a clonal level, of how a tumor evolves over time, and better interpret these mutations in the context of potential patient care.

725T

Identification and characterization of novel oncogene candidates in invasive breast carcinoma. *D.G. Piqué[1,2], J.M. Greally[2], J.C. Mar[1,3].* 1) Dept. of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY; 2) Dept. of Genetics, Albert Einstein College of Medicine, Bronx, NY; 3) Dept. of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY.

Invasive breast carcinoma (IBC) is the most common cancer in women worldwide. The overexpression of certain oncogenic mRNA isoforms, such as an *ERBB2* isoform that skips exon 16, drives resistance to the anti-cancer therapeutic trastuzumab and promotes the development of metastatic breast lesions. One characteristic of an oncogenic mRNA isoform is that it is highly expressed in a subgroup of tumors relative to normal tissue. However, both the prevalence and regulation of oncogenic mRNA transcript isoform expression in RNA-sequencing (RNA-seq) data from breast tumors remain largely unknown. Here, a novel statistical approach was developed for identifying oncogene candidate mRNA isoforms (OCs) based on the observation that oncogenes are overexpressed in a subset of tumors relative to normal tissue. From applying our approach to RNA-seq data from 113 paired tumor-normal breast tissue samples from TCGA, we identified five OCs, two of which have known roles in oncogenesis. In addition, several mRNA isoforms of the well-known breast cancer oncogene *ERBB2* share statistical properties with a theoretical OC, which shows that our statistical approach reflects patterns of known oncogene expression. To understand the factors that drive the overexpression of OCs in certain breast tumors, multivariate logistic regression that incorporated both clinical and molecular covariates such as DNA methylation and copy number variation (CNV) was conducted. Our results suggest that DNA methylation, particularly within introns, is a major factor associated with the overexpression of the full-length isoform of *CBX2*, an oncogene candidate that was identified using our method and that promotes metastasis in prostate cancer. *CBX2* overexpression in primary breast tumors was associated with the upregulation of mRNA isoforms involved in cell cycle-related processes, consistent with its potential role as an oncogene. Due to the low levels of *CBX2* mRNA expression in normal adult female tissue and availability of CBX2 chromodomain inhibitors, pharmacologic inhibition of CBX2 activity may serve as a therapeutic target in aggressive subtypes of breast cancer. The identification of oncogene candidates based on mRNA expression profiles between tumor and normal samples may yield fundamental insights into the molecular regulation of transcriptional programs in cancer and uncover new therapeutic targets to treat aggressive subtypes of cancer.

726F

Pathogenicity of Mutation Analyzer (PathoMAN): A fast, automation of germline genomic variant curation in clinical sequencing. *V. Ravichandran, K. Offit, J. Vijai.* Clinical Genetics Service, Memorial Sloan Kettering Cancer Center, New York, NY.

   Multiplexed gene sequencing for clinical testing, generates large number of genetic variants. The guidelines for variant curation provided by the ACMG requires the aggregation of multiple lines of genomic data evidences from diverse sources and familiarity with the scientific literature and publicly available databases. We know of no computational tools that automates this, which would provide uniformity to the process and significantly speed up the interpretive process. We present *Pathogenicity of Mutation Analyzer* **(PathoMAN)**, a tool that automates germline genomic variant curation from clinical sequencing. PathoMAN aggregates multiple tracks of genomic, protein and disease specific information from public sources such as ClinVar, ExAC, UniProt, 1000 genomes, dbNSFP and locus specific databases. Variant and gene specific annotations are used to classify per ACMG rubric. We analyzed 2500 manually curated and classified, high quality variants in 180 genes from three large, recently published studies to quantify the performance of PathoMAN. The dataset consisted of 242 pathogenic/likely pathogenic (P/LP), 1272 benign/likely benign (B/LB) and 1261 variants of uncertain significance (VUS). We report the summary of PathoMAN classifications in 4 categories: concordance, discordance, discovery and loss of resolution when contrasted against the manual curation. As new data, we present the automated curation of 76 cancer predisposition genes using PathoMAN for the gNOMAD dataset, useful for gene burden tests. For the 3 published datasets on cancer, PathoMAN achieves an average of 75% and 60% concordance and 1.5% and 0.1% discordance for P/LP and B/LB respectively. PathoMAN resolves 12% of reported VUS as P/LP or B/LB. Although the average false positive rate was <3%, the lack of resolution of 30% of the variants, reflected the absence of evidence available in the public databases to classify genomic variants as pathogenic or benign based on the ACMG guidelines. We describe the bottlenecks to automation of variant curation and propose new standards. We propose, a new code scheme for 5 ACMG classes that were not utilizable in the present scenario, which we hope will be used by the community to enhance the utility of the ACMG guidelines. In addition, the algorithm provides clinical and research labs to perform rapid identification of pathogenicity calculation of variants discovered and could be a valuable tool for quick turnaround of clinical and research results.

727W

Weighted similarity network fusion through integrating genomic functional annotation. *P. Ruan[1], S. Wang[2].* 1) Department of Statistics, The George Washington University, Washington, DC; 2) Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY.

   Recent technology developments have made it much easier to collect various kinds of genome-wide data, which provides scientists a great chance to utilize the information from different data sources to solve complex problems such as classification of diseases subtypes or identification of disease related genes. Many data-integration methods have been developed to meet such needs. Among these methods, similarity network fusion method, which effectively fuses the similarity networks of samples for each data sources into one fused network, has shown great potential for capturing similarities of samples and further identifying disease subtypes. Here, we propose a weighted similarity network fusion method, which adds weights to similarity networks by integrating genomic functional annotations, thereby improving the method's performance of disease subtypes classification. With information combined from gene expression and DNA methylation when both weighted by genomic functional annotations, we classified cancer subtypes more accurately across multiple TCGA (the Cancer Genome Atlas) cancer datasets than existing data-integration approaches. Further investigation on survival data confirmed the improved cancer subtype classifications.

728T

Telomere length dynamics from whole genome sequencing using Telomeasure in progressing and non-progressing Barrett's esophagus.

J.M. Shelton[1], T.G. Paulson[2], M. Shah[1], K. Arora[1], R. Sanghvi[1], X. Li[2], C.A. Sanchez[2], K. Oman[2], B. Hubert[1], B. Riley-Gillis[1], E. Venturini[1], A. Corvelo[1], N. Robine[1], D.M. Oschwald[1], P. Galipeau[2], B.J. Reid[2]. 1) New York Genome Center, 101 Avenue of the Americas, New York, NY 10013, USA; 2) Fred Hutchinson Cancer Research Center PO Box 19024, Mailbox C1-157 Seattle WA 98109.

Barrett's esophagus (BE) is a generally benign metaplasia that develops as a protective adaptation to chronic acid reflux, but 5-10% of BE patients progress to esophageal adenocarcinoma (EAC) [1]. Because EAC is resistant to multiple treatment options [2] and currently has a poor patient survivorship [3], understanding the transition from BE to EAC would allow for better surveillance and treatment strategies. It has been hypothesized that telomere shortening may be involved in the development of chromosomal instability and the transition to EAC [4], but telomere dynamics have not been well characterized prior to a diagnosis of EAC. To identify differences in telomere length between EAC progressors and non-progressors, we performed whole-genome sequencing of esophageal epithelium from two groups of BE patients; forty who progressed to EAC and 40 with stable benign BE, matched by follow-up time. Purified BE epithelium from two locations in the esophagus and from a minimum of two times points, along with constitutive blood samples were collected and sequenced. To determine average telomere length in each sample, we developed a novel method – Telomeasure – that uses chromosome arm-level coverage to infer the number of telomeres and alignment to mock-human telomeres to estimate telomeric DNA content. Despite showing a strong correlation with estimates from a previously published tool (Telseq [4]), Telomeasure is less affected by sequencing errors, variation in read length, and higher degrees of genomic instability, a frequent occurrence in BE samples. Consistent with previous studies, BE samples have in general shorter telomeres compared to their matching normal controls, which likely reflect higher cell turnover in the reflux-exposed esophagus. EAC progressors show significantly higher variability in telomere length and in telomere lengthening/shortening rate between T1 and T2 timepoints compared to non-progressors. Additionally, several telomere-associated genes were predominantly mutated in progressor samples. Copy number variants overlapping TERT, BCOR, RB1, CCNE1, TP53 and PDGFRA and mis-sense mutations in ASF1B, RTEL and TP53 were almost exclusively found in progressor samples. These results demonstrate the utility of Telomeasure for characterizing telomere dynamics from whole genome sequence data. [1] doi:10.1038/nrc2773 [2] doi: 10.1101/gr.214296.116. [3] doi: 10.1097/JTO.0b013e3182397751 [4] doi: 10.1038/ncomms6224. [5] doi: 10.1093/nar/gku181.

729F

Practical noninvasive biomarkers: Identification of blood gene signatures for diagnosis of lung cancer.

B. Song[1,2], L. Shi[1,2], Z. Yi[1], W. Zhang[1]. 1) Department of Medicine, Icahn School of Medicine at Mt. Sinai Hospital, New York City, NY; 2) Horace Mann High School, Bronx, NY.

Lung cancer is a leading cause of cancer related death with almost 160,000 deaths in the United States per year. One factor that significantly decreases a patient's chance of survival is late diagnosis, as there are minimal therapeutic treatments that are effective in late stage lung cancer. Even early in cancer development, however, lung cancer manifests itself with several distinctive genetic and genomic signatures. One of these is abnormal expression of genes which can affect cell growth and proliferation. Current studies of these focus on cancer tissues that require invasive biopsies from tumors. The approach presented in this study looked for mRNA biomarkers for lung tumor tissue that could be found in the blood, allowing for diagnosis from a simple blood test even at an early stage of the cancer development. Robust meta-analysis was performed to analyze mRNA expression profiles in blood from lung cancer patients. Using a training set of 4 mRNA expression lung cancer datasets with a total of 646 patients, 152 meta-genes were significantly over or under expressed in cancer patients compared to heathy individuals , many of these having functions in intracellular signaling cascades and glycerolipid metabolic processes. When these were compared to 2308 metagenes identified from lung tumor tissue expression profiles in 7 datasets with a total of 1636 patients, they yielded 33 overlapping metagenes. These 33 metagenes could effectively diagnose lung cancer status with an average AUC of 0.878 in the training set. They were validated in an independent lung cancer cohort with an average AUC of 0.939. The metagenes were also then applied in a cox regression model with patient survival data, and were shown to produce significant association with prognosis in 4 datasets ($p=0.00406-2.68 \times 10^{-8}$). The blood biomarkers that were identified in this study could be a breakthrough that allows for development of an effective and quick assay for early diagnosis of lung cancer in a non-invasive and cost-effective manner.

730W

Identification of somatic tumor-only variants on on 1120 solid tumor cases through intelligent variant filtration. *S. Van Vooren[1], T. Stockley, PhD, FCCMG, FACMG[2], M. Sukhai, PhD[2], S. Kamel-Reid, PhD, FACMG[2].* 1) Agilent Technologies, Santa Clara, CA; 2) Genome Diagnostics, University Health Network, Toronto, Canada.

Tumor molecular profiling by next-generation sequencing (NGS) aims to use information on somatic tumor-specific variants to direct cancer treatment. A significant challenge in tumor molecular profiling is identifying tumor-specific variants in the presence of germline variants which are also identified during tumor testing. The majority of germline variants are benign changes not contributing to the cancer occurrence, although rare germline cancer predisposition variants may also be identified during the testing of tumor tissue.   In order to identify true somatic tumor-specific variants, a common approach is the sequencing of both tumor and normal tissue (often peripheral blood), with variants from each tissue source used to classify variants as either somatic or germline. However paired tumor-normal testing leads to increasing costs due to the need to test both tissue sources. In addition paired tumor-normal testing may identify potentially unwanted incidental germline cancer predisposition variants.   In this study, we present a solution that can improve the identification of somatic tumor-specific variants obtained from NGS testing of tumor tissue only, without the need to sequence normal tissue for comparison to germline variants. We evaluate the use of the Agilent Alissa platform (formerly Cartagenia Bench Lab NGS) to implement the lab's filtration strategy and assess performance of the approach on 1120 solid tumor cases for which paired tumor-normal NGS variant results were available at the Genome Diagnostics Lab at the University Health Network.

731T

Integrated somatic mutation detection from tumor-normal sequencing data using multiple calling methods. *Y. Wang, J. Li.* Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Multiple computational methods have appeared for calling somatic mutations based on next-generation sequencing data of matched tumor-normal pairs. The idea underlying most of these methods is to identify new alleles in the tumor sample or altered allele frequencies in the tumor that are statistically robust given the sequencing depth. However, several commonly adopted methods, such as *Mutect*, *Varscan* and *Strelka*, often detect very different sets of somatic variants. We sought to understand the source of such differences and found that some of these methods failed to consider the case of loss-of-heterozygosity (LOH), thus could not report Ref/Ref (or Alt/Alt) genotype in the tumor when the normal is Ref/Alt. Others relied on the Ref/Ref as the baseline genotype and failed to call the Ref/Alt in the tumor when the normal is Alt/Alt. Further, many of these methods assumed no tumor-normal mixing and would not leverage read count data to detect quantitative differences between the normal and tumor allele frequencies. We used an in-house exome sequencing data for a T-N pair at an average depth of 300X to compare Mutect, Varscan, and Strelka, and characterized their striking differences in the category of variants detected and their different cutoff values. We have since established a unified approach that combines the respective strengths of these methods. Our approach is to create a union list of called variants from these methods, combined with a parallel series of calls based on T-N swapping, and re-test the T-N allele frequencies using the Fisher's Exact test. The results thus obtained covered a much more complete catalog of somatic variants in all possible T-N allelic combinations and detected changes that occurred only in a fraction of the cells. Our approach is superior to those that simply combines multiple variants lists by majority vote, as it is built on the actual underlying allele count differences. Its sensitivity is directly linked to allele frequency alterations and the sequencing depth and averted the problem that different methods have distinct sensitivities and sometimes incomplete consideration of possible mutation types.

732F

Access and discover pathways from Pathway Commons. *J.V. Wong[1], I. Rodchenkov[1], O. Babur[2], A. Luna[3,4], E. Demir[2], C. Sander[3,4], G. Bader[1].* 1) Department of Computational Biology, University of Toronto, Toronto, Ontario, Canada; 2) Department of Computational Biology, Oregon Health & Science University, Portland, OR; 3) Dana-Farber Cancer Institute, Boston, MA; 4) Department of Cell Biology, Harvard Medical School, Boston, MA.

Pathway Commons (www.pathwaycommons.org/) serves researchers by integrating data from public pathway and interaction databases and disseminating it in a uniform fashion. The knowledge base is comprised of metabolic pathways, genetic interactions, gene regulatory networks and physical interactions involving proteins, nucleic acids, small molecules and drugs. Alongside attempts to increase the scope and types of data, a major focus has been the creation of user-focused tools and resources that facilitate access, discovery and application of existing pathway information to facilitate day-to-day activities of biological researchers. For those wishing to browse and discover pathways within the collection, we offer a web-based 'Search' application that enables users to query by keyword and visualize ranked search results. 'PCViz' is a web tool that accepts gene names and returns a customizable interaction network visualization based upon pathway data resources. These complement existing desktop software add-ons linking Pathway Commons to the Cytoscape (CyPath2) network analysis tool and the R (paxtoolsr) programming language. To facilitate analysis and interpretation of experimental data - for instance, enrichment studies that distill pathway alterations from underlying gene expression changes - pathway data file downloads can be directly used in software tools such as Gene Set Enrichment Analysis. For those wishing to learn more about pathway resources and analysis, an online 'Guide' includes case studies and guided workflows. Ongoing development of web apps will enhance the accessibility to pathways and integrate support for visualization and interpretation of experimental data.

733W

Identification of germline copy number variations (CNVs) using whole-exome sequencing data in Caucasian and African American men with hereditary prostate cancer. *K. Wood[1], D. Tran[1], M. Baddoo[2], M. DeRycke[3], S. Riska[3], S. McDonnell[3], D. Schaid[3], S. Thibodeau[3], D. Mandal[1].* 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) Tulane Cancer Center, New Orleans, LA; 3) Mayo Clinic, Rochester, MN.

Prostate cancer (PCa) is the 2nd leading cause of cancer-specific deaths among American men, and there is a significant racial disparity between African American (AA) and Caucasian (CAU) affected men, with AA men having a 60% higher incidence rate than their CAU counterparts. Due to limited genomic data on AA men with PCa, no ancestry specific genetic variants have been detected to date. So far, somatic copy number variations (CNVs) have been documented in PCa tumors, but changes in CNV have not been studied in germline DNAs from hereditary PCa (HPC) cases, which accounts for approximately 15% of PCas. The goal of the present study is to identify germline CNVs in whole exome sequencing (WES) data on cases from HPC families (≥3 PCa affected/family) and population controls. We hypothesize that rare germline CNVs in actionable genes contribute to higher incidence of PCa in AA HPC men. We have WES data on 13 affected cases from 13 independent HPC families and 10 population controls. Of the 13 cases, 11 are CAU and 2 are AA, with the age at diagnosis ranging from 51 to 75 years. A CNV-specific calling algorithm, CANOES (http://www.columbia.edu/~ys2411/canoes/), was used to successfully detect germline CNVs by modeling read counts using a precise distribution. To confirm, CNVs were then visualized independently using Integrative Genomics Viewer (IGV). CANOES called 466 germline CNVs on CAU, 303 of which were confirmed using IGV; 18 of them are located within cancer-related genes. Seven of those cancer-related genes harboring germline CNVs were also previously shown to be associated with PCa tumors, including *GSTM1*, *DMBT1*, *GSTT1*, *POTEM, UGT2B17, MTUS1* and *PTGER*. Four of the 7 germline CNVS in cancer-related genes were found only among the CAU HPC cases and not in the controls. CANOES called 132 germline CNVs from the AA WES datasets, 30 of which were confirmed using IGV, and 3 of those were within cancer-related genes. Two of these 3 germline CNVs were within genes that have previously been associated with PCa tumors in CAU, including *SPAG11A* and *CKMT1B*. The preliminary analysis in our dataset shows no overlap between germline CNVs identified in AA and CAU, further suggesting that ancestry plays a significant role in PCa health disparity. Documentation of these CNVs in PCa tumors emphasizes the importance of studying these genes further. In general, the identification of germline CNVs within HPC families will facilitate PCa screening in the future.

734T

Optimal design of single cell studies for detecting and quantifying clonal subpopulations. *J. Yu[1], C. Li[1], O. De La Cruz C.[2].* 1) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Department of Mathematical Sciences, Kent State University, Kent, OH.

   Motivation. Cancer cell populations often evolve, generating distinct clonal subpopulations. Characterizing the clonal structure of a tumor is valuable for understanding the biology of the disease. Recent advances in single-cell DNA sequencing technologies enable the characterization of clonal architecture in heterogeneous cell populations. In studies of clonal detection with single-cell sequencing, there is usually a trade-off between the number of cells assayed and sequencing depth per cell. This study aimed to provide a guideline for determining the optimal number of cells and sequencing depth needed for clonal detection. **Methods**. We mimicked experimental designs with various numbers of cells and overall sequencing depths by subsampling cells and paired reads from a tumor single-cell sequencing dataset for 6 patients with childhood acute lymphoblastic leukemia (Gawad et.al, 2014). We repeated the subsampling 5 times and results were averaged over the 5 replicates. **Results**. **1)** Our results suggest that the number of cells has a higher impact on clustering results than sequencing depth, and that under a fixed sequencing budget, more cells with lower depth tend to yield the correct number of clusters with higher probability, as well as lower misclassification rate. **2)** To evaluate how low the depth can be without losing information on clone detection, we identified the depth at which ≥80% of the mutants present in a cell can be detected in ≥90% of the cells. We used an exponential model to fit data for each cell. Our results show that the depth needed for mutant detection varies from patient to patient, ranging from 30X to 88X. **3)** We also evaluated a few potential factors affecting the robustness of clustering results. We found that detection of a small clone tends to require a high number of cells and relatively high depth, that large clones tend to be split into two clusters at a low depth, and that results tend to be more sensitive to reduction in number of cells and depth when the clustering is driven by relatively few somatic mutations.

735F

Evaluating relationships between pseudogenes and genes: From pseudogene evolution to their functional potentials. *Y. Zhang, T. Johnson, S. Yu, K. Huang.* Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH.

   Pseudogenes are fossil relatives of genes. They have long been thought of as "junk DNAs", since they do not code proteins in normal tissues. Although most of the human pseudogenes do not have noticeable functions, ~20% of them exhibit transcriptional activity. There has been evidence showing that some pseudogenes adopted functions as lncRNAs and work as regulators of gene expression. Furthermore, pseudogenes can even be "reactivated" and translated into proteins in some conditions, such as cancer initiation. All the above have shown that pseudogenes could have functional potentials in the genome. In this study, we developed a novel approach integrating graph analysis, sequence alignment and functional analysis to evaluate pseudogene-gene relationships in human genome: (a) Find candidate regulatory target genes. Two known transcriptional level regulatory mechanisms of pseudogenes are RNA interference, and transcription products serving as competitive miRNAs binding targets. Both of these two mechanisms rely on sequence similarity between the pseudogenes and regulatory target genes. It is intuitive to find the candidate target genes by sequence similarity. Thus, we built a database of pseudogene-gene (PGG) families. Each family contains one or multiple pseudogenes and homologous genes (potential target genes) as members. We have developed the approach for constructing PGG families using sequence alignment and network analysis, and we are optimizing the pipeline for efficient compute performance using GPUs. (b) Detect key mutation signatures within each family. This database will also provide a resource for identifying mutation signatures that switch on/off the function of pseudogenes. For example, loss-of-function mutations that silenced pseudogenes, or frameshift mutations that coded abnormal peptides. (c) Infer functional annotations of pseudogenes. Functional annotations (e.g. GO terms) will be overlaid on the PGG networks, and functions of pseudogenes will be inferred from known annotations of its neighbors (i.e. other nodes in the same network). In summary, evaluating the relationships between pseudogenes and their gene counterparts could help us reveal the evolutionary path of pseudogenes and associate pseudogenes with functional potentials. It also provides an insight into the regulatory networks involving pseudogenes with transcriptional and even translational activities.

736W

Prediction of genome-wide DNA methylation in locus-specific repetitive elements. *Y. Zheng[1,2], B. Joyce[1], L. Liu[1], Z. Zhang[1], W. Kibbe[3], W. Zhang[1], L. Hou[1].* 1) Center for Population Epigenetics, Northwestern University, Chicago, IL; 2) Division of Health and Biomedical Informatics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 3) Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, MD.

DNA methylation in repetitive elements (RE) suppresses their mobility and maintains genomic stability, and decreases in it are frequently observed in tumor and/or surrogate tissues. Averaging methylation across RE in the genome is widely used to quantify global methylation. However, methylation may vary in specific RE and play diverse roles in disease development, thus averaging methylation across RE may lose significant biological information. The ambiguous mapping of short reads by and high cost of current bisulfite sequencing platforms make them impractical for quantifying locus-specific RE methylation. Although microarray-based approaches (particularly Illumina's Infinium methylation arrays) provide cost-effective and robust genome-wide methylation quantification, the number of interrogated CpGs in RE remains limited. We report a random forest-based algorithm (and corresponding R package, *REMP,* available in Bioconductor) that can accurately predict genome-wide locus-specific RE methylation based on Infinium array profiling data. We validated its prediction performance using alternative sequencing and microarray data. Testing its utility with The Cancer Genome Atlas data demonstrated that our algorithm offers more comprehensively extended locus-specific RE methylation information that can be readily applied to large human studies in a cost-effective manner. We found that hypomethylated RE loci in cancer may play a role in transcription activation, potentially impair chromatin stability, and regulate key cancer-related pathways. Specifically, we observed a significant positive correlation between intronic locus-specific RE methylation and the host gene expression, which supported the hypothesis that DNA methylation in intronic regions may potentially silence RE to maintain a gene's efficient transcription. Locus-specific RE methylation also achieved significantly higher power than global methylation in discriminating tumor from normal tissues (area under the ROC curve = 98.3% vs 74.1%, respectively, p<0.001), which underscored the valuable information added by our algorithm. Our work has the potential to improve our understanding of the role of global methylation in human diseases by leveraging methylomic data from large-scale human studies and identify novel RE loci that may exert important biological and pathological effects for cancer early detection and diagnosis.

737T

Incorporating multiple NGS read features enables detection of transposon insertions across the genome. *A. Zimmer, A. Robertson, Z. Mahkovec, G. Mishne, J. van den Akker.* Color Genomics, Burlingame, CA.

Large insertions can disrupt coding sequences and gene function. Genomic insertions can arise from multiple sources, including errors in homologous recombination and the transposition of mobile elements (transposons). Transposons, genetic elements that can transpose from one location to another, make up a large fraction of the human genome. While the majority of transposons have lost the ability to be transposed, some are still active and several such variants with clinical relevance have been published, including a Portuguese founder mutation in *BRCA2* (c.156_157insAlu). Detecting insertions, and transposon insertions in particular, poses several technical challenges in capture-based NGS panels: (i) large insertions are a major disruption to the genomic sequence and can affect the hybridization between targeting probes and the sample DNA; (ii) a transposon inserted sequence is present many times in the genome, causing problems with alignment (iii) the inserted sequence is generally longer than NGS reads, so no reads span the inserted sequence; (iv) transposon insertions often include A-rich linker sequences which cause sequencing artifacts. Here, we describe the development of a novel method to identify insertions in NGS data. Our algorithm to detect structural variants is based on reads where one end aligns to the region of interest, but the other does not, i.e. soft-clipped reads. For positions where there are a significant number of soft-clipped reads at the same location, we compute several features of the reads and the locus which can be used to distinguish true insertions from NGS artifacts. The most important features were: number of soft-clipped reads, fraction of soft-clipped reads, sequence agreement of the soft-clipped reads, GC content of the locus, sequence uniformity of the locus, number of reads supporting a nearby large indel, and number of inverted reads. A second step uses local assembly to estimate the inserted sequence; for insertions at the size of an Alu element (~300 bp), this is typically sufficient to uniquely identify the exact element in public databases. We validated this method by sequencing samples identified from the 1000 Genomes data set that carry insertions in the genes of the Color Test. In conclusion, this caller is able to identify large insertions based on a single breakpoint and independent of the inserted sequence. It can therefore be used to call insertions across the genome.

738F

A novel algorithm to identify somatic copy-number alterations which delivers high accuracy in targeted resequencing of cancer genes from tumor specimens. *F.M. De La Vega[1,2], S.A. Irvine[3], K. Gaastra[3], Y. Pouliot[2], R.T. Koehler[2], D. Mendoza[2], P. Cavales[2], A. Vilborg[2], L. Trigg[3].* 1) Stanford University School of Medicine, Stanford, CA, USA; 2) TOMA Biosciences, Foster City, CA, USA; 3) Real Time Genomics, Ltd., Hamilton, New Zealand.

   Next-generation, deep sequencing of gene panels is being adopted as a diagnostic test to identify actionable mutations in samples from cancer patients. Targeted sequencing involves positive selection of gene exons typically involving PCR at some stage in the protocol. PCR creates GC-content and length biases that skew the representation of DNA fragments within the sequenced library. Combined with the small counts of reads in some of the targeted exons, this results in reduced sensitivity to the identification of copy number alterations (CNA). Here we present a novel CNA calling algorithm optimized for targeted resequencing data delivering high sensitivity and specificity by comparing test data to a negative control or to a model constructed from normal samples. Coverage is corrected for G+C content, a median normalization is applied, and the Mumford-Shah energy function is used for segmentation. We benchmarked and validated this algorithm with both simulated data and with real sequencing data obtained with an oligonucleotide-selective sequencing (OS-Seq) enrichment assay for a 130-cancer gene panel with minimal PCR bias. We sequenced to a >1,000X on-target median depth a set of cancer cell lines with orthogonally verified CNAs and a new reference material comprised of synthetic CNA sequences spiked on a well-known Genome-in-a-Bottle cell line background (SeraCare, Inc). A distinctive feature of our method is its ability to filter raw calls to a specified sensitivity and specificity trade-off by selecting empirical score thresholds based on the analysis of the reference datasets. We show that we can detect all true hemizygous deletions and amplifications from about 0.6 to more than 150 excess copies over the diploid background, with a FDR <5% across the 130 genes in the panel. We find an excellent correlation (r^2=0.97) between the copy number estimated by our method and the results from digital droplet-PCR assays. We compared our calls with those from the CNVkit and VEGAWES programs using ROC curve analysis. Our results show that our CNA caller outperforms these other programs in terms of ROC area under the curve and overall accuracy. To our knowledge this stringent evaluation with an absolute ground truth from defined reference materials has never been performed to evaluate NGS CNA algorithms and assays. We suggest that our benchmarking protocol could be used to validate clinical sequencing analysis pipelines that provide CNA calls.

739W

AmpliconArchitect: On the fine structure of focal amplifications in cancer. *V.B. Deshpande[1], J. Luebeck[2], M. Bakhtiari[1], N. Nguyen[1], K.M. Turner[3], P.S. Mischel[3,4,5], V. Bafna[1].* 1) Department of Computer Science and Engineering, University of California at San Diego, La Jolla, California 92093, USA; 2) Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, California 92093, USA; 3) Ludwig Institute for Cancer Research, University of California at San Diego, La Jolla, California 92093, USA; 4) Moores Cancer Center, University of California at San Diego, La Jolla, California 92093, USA; 5) Department of Pathology, University of California at San Diego, La Jolla, California 92093, USA.

   Focal amplifications or amplicons are copy number(CN) gains of small portions of the genome, as opposed to large-scale aneuploidies. We recently showed that extrachromosomal DNA (ecDNA) drives oncogene amplification and tumor growth in as many as 40% of all human cancer (Turner et al, Nature 2017). ecDNA breaks the eukaryotic norm of tumor evolution and is interconvertible with intrachromosomal amplification as homogeneously staining regions (HSRs). Amplicons often undergo multiple rearrangements and understanding their structure sheds light on the nature of ecDNA. However, traditional structural variant callers are incapable of analyzing and representing the complex structures. We present AmpliconArchitect (AA), a tool which predicts the structure of complex focal amplifications from whole genome sequencing of cancer samples. AA performs CN-aware detection of amplified regions and rearrangements to construct a breakpoint graph and decomposes the graph into cycles. Finally, AA provides an interactive visualization of the cycles in the genomic context to explore all candidate amplicon structures. AA accurately predicted complex structures in 960 diverse simulations with 85% accuracy even at low sequencing coverage (≥1×). We applied AA to 117 samples from 13 cancer subtypes and found 157 amplicons in 55 samples. We compared their distribution with the CN gains from 3919 SNV array samples from TCGA. Amplicon size (μ≈1.8Mbp) and CN (μ≈3.6), each, showed exponential distributions. Amplicons containing multiple genomic regions were larger than localized amplicons (p<5e-4). FISH analysis of 37 amplicons showed that while ecDNAs tend to be smaller (<5Mbp) with high CN (up to 200) and HSRs can be larger (up to 20Mbp), there was no clear separation in their distributions as expected from their interconvertible nature. Amplicon structure was preserved across biological replicates and cancer subtypes were enriched for specific oncogenes. In a second study, analyzing 69 cervical cancer samples infected with human papillomavirus (HPV), we found HPV integration into human genome in 56 samples and 61 chimeric human-viral amplifications in 42 samples. While there is oncogene enrichment, these cases form a small portion of the amplicons. A majority of the amplicons have "bifocal" circular integration as opposed to the intuitive notion of viral integration at a single cut site, suggesting human-viral chimeric ecDNAs containing segments from multiple chromosomes.

740T

Tumor mutation burden (TMB) as a marker for DDR and IO combination.
Z. Lai[1], R. McEwen[2], D. Hodgson[2], B. Dougherty[1], J.C. Barrett[1], J. Dry[1]. 1) AstraZeneca, Waltham, Ma; 2) AstraZeneca, Cambridge, UK.

   Both immune checkpoint inhibitors (IO therapies) and PARP inhibitors have been shown to greatly benefit patients in certain cancers, though many patients eventually relapsed to the therapies. There are growing evidences that TMB might be a tumor agnostic marker for predicting IO therapy responses. Deficiencies in DDR (DNA Damage Response) can lead to damaged DNA un-repaired or mis-repaired, which would contribute to increasing tumor mutation burden. Tumors with certain DDR deficiencies, such as BRCA1 and BRCA2 mutations, have been shown to be sensitive to PARP inhibitors, both pre-clinically and clinically. We hypothesize that patients with DDR deficiencies, defined as those harboring germline or somatic mutations in HRR (Homologous recombination repair) genes, will have higher TMB and more immune active, and can potentially receive greater benefit from the combination of DDR and IO therapies. We investigated the relationship between DDR deficiencies and TMB across TCGA (The Cancer Genome Atlas) disease cohorts, as well as some internal datasets from DDR clinical studies. In addition, we also examined the relationship between TMB and immunoscores, which quantify the in situ immune infiltrate and have been demonstrated to be a prognostic factor. We found across many cancer types, patients with DDR deficiencies have significantly higher TMB and higher immunoscores. The result supports that TMB is a potential marker for selecting patients for DDR and IO combination therapy.

741F

Genomic instability phenotypes in multidimensional genomic cancer studies. *B.N. Lasseigne, S.J. Cooper, G.M. Cooper, R.M. Myers. HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, AL.*

   The ability to infer molecular signatures from various genomic data sets can aid in prioritizing future experiments and provide biological insight into disease-associated pathways and mechanisms. In particular, chromosomal instability (CIN, altered chromosome number and structure) and the CpG island methylator phenotype (CIMP, widespread altered promoter methylation) are molecular phenotypes representing distinct cancer etiologies and chemotherapy responses. These genomic instabilities also demonstrate overlapping information content across data types because gross alterations in one feature set results in consistent changes in others (e.g. CIN is linked to widespread DNA hypomethylation and characteristic gene expression changes). In this proof of concept study in kidney cancer (n=291), we compared several definitions each of CIN (total number of breakpoints, number of genes with significant CNV, percent of bases with CNV, or total functional aneuploidy) and DNA methylation instability (the raw number of CpGs methylated, percent of CpGs methylated in CpG islands, shores or non- islands, and the density of methylated CpGs to non-methylated CpGs), characterizing their relationship to each other and clinical phenotypes like patient survival. Our results suggest these metrics are capturing distinct biological properties to a considerable extent. For example, we show that CIN better predicts long-term surviving patients (p=0.0559) while DNAm instability better predicts non-surviving patients (p=0.00017). Additionally, using machine learning methods we predict both CIN (LASSO model ROC AUC=0.866) and DNAm instability (LASSO model ROC AUC=0.8817) status from gene expression data alone. Our findings will facilitate prioritization of experiments in future studies, improve interpretation of these instability signatures for both basic biology and clinical use, and allow their inference from each of the major genomic data types.

742W

Predictive, discriminative versus associated or prognostic biomarker? Comparisons of discriminant, predictive and association and network analysis methods for mass spectrometry data from ovarian cancer. *Y. Liang[1], A.G. Kelemen[2], H. Zhang[3].* 1) 1Department of Family and Community Health, University of Maryland, Baltimore, Baltimore, MD; 2) 2Department of Organizational Systems and Adult Health, University of Maryland, Baltimore; 3) Department of Pathology, Johns Hopkins Medical Institutions, Baltimore.

Ovarian cancer is the deadliest gynecologic malignancy with majority patients diagnosed in later stage. Antineoplastic therapeutics is vital to treating serious ovarian patients but have heterogeneous responses. Mass spectrometry (MS) based proteomic technologies have enabled global expression profiling at the protein level to examine the linkages between glycoproteins and ovarian cancer as a network and to determine which glycoproteins linked to cancer subtypes. However, there are various key challenges analyzing MS glycoproteome data such as small/weak effects, heterogeneity, high percentage missing, skewness, which make most algorithms and modeling based approaches not applicable. In this study, we conduct multi-stage and multilayer systematic analysis of MS glycoproteins from ovarian tumor sample obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and The Cancer Genome Atlas (TCGA). One hundred twenty two samples from 9606 proteins are analyzed, and ovarian status is defined based on the putative homologous recombination deficiency (HRD) (67 HRD positive, 55 non-HRD/ negative). Multilayer systematic analysis include 1) statistical process control for correcting technical variations and examine heterogeneity; 2) missing data imputations with SVD; 3) two-way regression based sliding window analysis for identifying the significant and predictable biomarkers; 4) Discriminative analysis and neural network approach for discriminative power and predictive accuracy; 5) Ovarian subtype identification using clustering analysis (mixture model, HC, PCA); 6) Correlation network comparisons and visualization (Cytoscape); and pathway and function analysis (Ingenuity pathway analysis) for validating the findings and understanding key mechanisms. Results reveal a marked difference when select the glycoproteins based on predictive, discriminative, associative or prognostic biomarkers, indicating the subtypes of proteomic profiles with functional differences and distinct subpopulations within ovarian patients. The proposed sequential multilayer approaches provide different perspective of the important biomarkers linked to the important biological processes influencing the ovarian progression, heterogeneity, and efficacy of platinum therapeutics and drug resistance, which are validated with associated pathway and function analysis.

743T

Tissue-specific feature of whole genome sequencing aids tissue-mapping in plasma. *H. Liang[1,2], X. Zhou[1,2], S. Qiao[1,2], G. Xie[1,2], X. Shi[1,2], X. Zhao[1,2], K. Wu[1,2], F. Li[1,2], H. Liang, F. Li,.* 1) BGI-Shenzhen, Shenzhen 518083, China; 2) China National Genebank-Shenzhen, BGI-Shenzhen, Shenzhen, 518083, China.

Tissue-specific genome polymorphism generated during histodifferntiation marks tissues with different patterns which requires mass whole genome sequencing data to mine. With plenty of samples of multiple kinds of tissues provided by public genomic databases, The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) represented, we use statistical methods to mine those potential tissue-specific patterns, basing on which we structure a model to track samples' origin with their sequencing data, and then try to apply this model on cfDNA to evaluate the risk of suffering tumor under the conception that as the tumor's growth will accelerate cell apoptosis, the extra DNA fragments released into plasma during this period would take tissue-specific signals, through detecting those patterns we can find the subclinical damage of the relevant tissue. We test multiple kinds of tissues, the result shows great accuracy in both tests of tissue and plasma. Our model detects the niduses of most patients with their plasmas correctly, besides with the low false positive of misjudging health controls, suggests that our model has great potentiality on early cancer screening.

744F

Discovering resistance mechanisms of mutant *NRAS* melanoma towards MEK inhibitor treatment from patient derived tumors. *G. Moriceau, W. Hugo, RS. Lo.* Division of Dermatology, UCLA.

Despite the recent improvements of targeted therapies and immunotherapies in metastatic human melanoma, primary and acquired resistance arises and limits patient survival benefit. Development of combinations based on these foundational therapies should yield further survival benefits in the near future. MEK inhibitor monotherapy has limited clinical activity against [MUT]*NRAS* melanoma but can serve as a foundation for rational combinations. In this context, we investigate the genetic/non-genetic and immune mechanisms of acquired MEKi resistance in [MUT]*NRAS* melanoma through analyzing pre- and post-tumors (with a total of 15 tumors) from seven [MUT]*NRAS* melanoma patients. To enable mechanistic study, we also generated >20 patient derived xenograft (PDX) MEKi-resistant tumors from two different [MUT]*NRAS* melanoma patients. These PDX tumors were treated with MEKi at three different dosages to study the correlates of different resistance mechanisms with multiple MEKi dosing. We utilized the nCounter® Vantage 3D™ Assays to analyze two slides from each patient-derived and PDX tumor: one for immune RNA profiling and another for Protein Solid Tumor Assay for FFPE. We combine the data from this analysis with whole exome data (with patient-matched normals) from the same set of tumor tissues. A cross-samples comparison using the nCounter Vantage 3D Assays provides systematic assessment of genetic, immune subset and protein changes associated with MEKi resistance within actual patient tumors. These observations will be validated both *in vitro* (using isogenic pairs of [MUT]*NRAS* cell lines pre-/post-MEKi resistance) and *in vivo* (using the PDX and PDX derived cell lines).   .

745W

Access, visualize and analyze pediatric genomic data on St Jude Cloud. *S. Newman[1,6], X. Zhou[1,6], C. McLeod[1,6], G. Wu[1], E. Sioson[1], S. Wang[1], J.R. Michael[1], A. Patel[1], A. Frantz[1], T. Chang[1], R.I. Davidson[3], S. Ma[4], I. McGuire[2], N. Robison[2], E. Suh[2], L. Tanner[2], J. McMurry[2], K. Perry[2], M. Rusch[1], J.R. Downing[1,5], J. Zhang[1].* 1) Department of Computational Biology, St Jude Children's Research Hospital, Memphis, TN; 2) Department of Information Services, St. Jude Children's Research Hospital, Memphis, TN; 3) Microsoft Research, Redmond, WA; 4) DNAnexus, Mountain View, CA; 5) Department of Pathology, St Jude Children's Research Hospital, Memphis, TN; 6) Contributed Equally.

Cloud computing is poised to revolutionize genomics as it facilitates secure sharing and collaborative analysis of huge datasets. Many pediatric diseases are rare and, as data and expertise are widely dispersed, a central focus for sharing and collaboration will be vital for pediatric oncology research moving forward. Our new cloud platform, built in collaboration with DNAnexus, centers around three core concepts: Data, Tools and Visualization.   First, we make all data from the Pediatric Cancer Genome Project (PCGP) securely available having remapped it to the *GRCh38* reference genome using a highly optimized BWA/GATK pipeline powered by Microsoft Azure. The initial data offering is comprised of approximately 700 tumor/germline whole genome pairs from common and rare pediatric cancers. Non-identifiable data can be accessed immediately whereas raw data can be securely obtained through a one-time streamlined access request procedure.   We next offer highly optimized cloud-based implementations of our extensively validated analysis pipelines. This allows users to quickly and privately run complex workflows on their own data regardless of computational expertise. Our initial workflows include differential gene expression, structural variant prediction, neoepitope prediction, variant annotation and pathogenicity classification. Users may upload their own data and build a "virtual" cohort integrated with PCGP data or reanalyze PCGP data using their own tools and workflows. Data and results can be securely shared inside of the platform.   We provide innovative cohort-level visualizations for data mining including protected access browsing of germline variants, and open access interactive browsing of somatic variants, copy number abnormalities, and structural alterations. These interactive visualizations allow investigators to drill down on interesting loci to identify new gene variants potentially linked to a disease. Of particular interest are the thousands of non-coding somatic and germline variants detected in PCGP samples. We overlay ATAC-seq data to help users assess the chromatin state of potential regulatory non-coding variants.   By bringing analytic tools to the world's largest set of pediatric genomics data, St Jude Cloud will enable innovative genomic analysis, data sharing and development of new methods that will further our knowledge of pediatric cancer and support precision medicine initiatives across the globe.

746T

Expression variability is associated with breast tumour subtype. *J.F. Pearson[1], G.A.R. Wiggins[2], M. Black[3], A. Dunbier[3], L.C. Walker[2].* 1) Biostatistics and Computational Biology Unit, University of Otago Chrsitchurch, Christchurch, New Zealand; 2) Department of Pathology, University of Otago Christchurch, Christchurch, New Zealand; 3) Department of Biochemistry, University of Otago Dunedin, New Zealand.

Inherited mutations in *BRCA1* are associated with greater risk of developing breast and ovarian cancers. BRCA1 is important for several functions in cells throughout the body, however mutations in *BRCA1* are strongly associated with breast and ovarian cancer with only minor effects in other tissues. Approximately, 70% of BRCA1-related breast tumours display a basal-like phenotype (eg. high grade and oestrogen receptor negative), although the gene networks that are critical for the mechanism remain unclear. Gene expression studies have extensively interrogated breast cancers, including tumours from *BRCA1* mutation carriers, however results from these studies lack consensus, with inconsistency in candidate genes and pathways. Importantly, these studies have focused on differentially expressed genes by comparing the mean of each group and determining statistical significance. We demonstrate the application of an alternative approach, which is to assess expression data using differential variability analysis, thereby discovering genes with a significant difference in variance between two sample groups. Expression data from 74 familial breast cancers and a 2116 sporadic breast cancer meta-cohort were interrogated for gene expression variability. We found that basal tumours and BRCA1 related tumours showed 39.9% (95% CI 39.4-40.3%) and 13.2% (95% CI 12.6-13.7%) increased global variability, respectively. Differential variability analysis identified gene sets that were enriched in pathways associated with epithelial cell function and development. Furthermore, we identified two genes, *PAX6* and *FOXA1*, involved in these pathways with significantly greater expression variability in basal or BRCA1 related tumours, in both study cohorts. These findings are novel and suggest epithelial cell development pathways may be more tightly regulated in non-basal and non-*BRCA1* related tumours.

747F

Transcriptome-based classification of primary melanoma identifies tumor subtypes that predict outcome in stage I. *R. Thakur[1], J. Nsengimana[1], J. Laye[1], M. Lauss[2], A. Filia[1,3], S. O'Shea[1], M. Chan[1], G. Jönsson[2], D.T. Bishop[1], J. Newton-Bishop[1], J.H. Barrett[1].* 1) Section of Epidemiology and Biostatistics, University of Leeds, Leeds, United Kingdom; 2) Department of Oncology and Pathology, University of Lund, Lund, Sweden; 3) National Heart and Lung Institute, Imperial College London, London, United Kingdom.

The 5-year disease-specific survival rate for AJCC (American Joint Committee on Cancer) stage I melanoma patients is around 95%, while for stage IV it is only ~15%. Even though most stage I patients survive their disease, some do not, and the desire is therefore to identify prognostic biomarkers in addition to clinical predictors within stage I melanoma. Previous transcriptome analyses from metastatic melanoma (mainly stage IV) have generated molecular classification schemes for identifying tumor classes that predict outcome. However, when these schemes were applied to primary melanomas from the Leeds cohort (n=702), the molecular classes showed different predictive properties across tumor stages (interaction P=0.03), failing to offer prognostic value in stage I tumors. In this study, a robust consensus-based PAM (Partitioning Around Medoids) method was used to cluster whole genome transcriptomes from 702 primary melanomas from the Leeds melanoma cohort (34% stage I, 51% stage II, 15% stage III). Seven distinct tumor classes were identified based on cluster separation and a number of other criteria. The 7 classes showed a strong association with outcome after adjusting for AJCC stage. When stratified by stage, in stage I tumors the 7 classes still showed borderline evidence of association with survival (P=0.06), and one particular class was strongly predictive of poor outcome (HR= 6.7, 95% CI: 1.4-31.7, P=0.016). Differentially expressed genes across the 7 classes were identified, and biological pathway enrichment was explored using ReactomeFIViz. The class with the worst prognosis was associated with upregulation of pathways related to cell cycle and downregulation of immune response pathways, whilst the inverse of this was observed in the best survival group. In keeping with previously reported studies, tumors from the worst survival group were histologically thicker, more frequently ulcerated and had more mitoses. The clinical relevance of the tumor classes was validated using an independent cohort of 223 primary melanomas. In conclusion, tumor classes derived from gene expression have been identified of demonstrated prognostic value, and importantly these are pertinent in stage I melanoma. These classes could serve to improve clinical decision-making and allow patient treatment to be tailored according to prognosis.

748W

A novel framework for tumor classification which uses sufficient dimension reduction for feature selection and Bayesian networks for integrating CT image and epigenomic. *Y. Wang, M. Xiong, N. Lin.* University of Texas Health Science Center at Houston, Houston, TX.

Next generation of genomic, sensing and image technologies will produce deeper and deeper genomic, epigenomic, imaging and phenotypic data with millions of features. They are increasingly used for disease prediction and prevention, especially for cancer. Discovering biomarkers from miRNA-seq and methylation-seq profiles is a high-dimensional, small sample size problem in machine learning. A fundamental question is how to efficiently extract epigenomic variants of clinical utility and to develop novel unified approach to classification analysis of epigenomic data. The developed sufficient dimension reduction, multitask feature selection algorithms for large-scale machine learning will offer a powerful tool for developing a general framework for efficiently and jointly extracting genomic and epigenomic variants of clinical utility and shift paradigm for cancer marker selection from independent, rank-based selection to powerful systematically and jointly genome-wide searching selection. Using Bayesian network to develop novel and highly discriminating algorithms for combining miRNA, methylation and imaging classifiers will substantially improve our ability to accurately detect cancer. We applied the proposed methods to TCGA ovarian cancer data set with 112 samples (80 tumor tissue samples and 32 tumor free samples) for ovarian cancer prediction . The results showed that the average prediction accuracy for each individual dataset is 74% for the CT image dataset, 69% for the mi-RNA dataset, 72% for the DNA Methylation dataset. After using Bayesian Network to combine different types of the datasets, the average prediction accuracy increased to 79%.

749T

Multiregion high-depth whole exome sequencing of matched primary and metastatic tumors revealed inter- and intra-individual genomic heterogeneity and polyclonal seeding in colorectal cancer metastasis. *Q. Wei[1], Z. Ye[2], X. Zhong[3], L. Li[2], C. Wang[2,6], R. Myers[2], J. Plazzo[5], A. Yan[2], S. Waldman[4], X. Chen[2], B. Jiang[5], B. Li[1], H. Yang[2].* 1) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN 37235, USA; 2) Department of Medical Oncology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA 19107, USA; 3) Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37235, USA; 4) Department of Pharmacology and Experimental Therapeutics , Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA 19107, USA; 5) Department of Pathology, Thomas Jefferson University, Philadelphia, PA 19107, USA; 6) Department of Environmental Health, School of Public Health, Nantong University, Nantong, Jiangsu 226000, China.

Purpose: Colorectal cancer (CRC) has high prevalence and mortality in many developed countries. Distant metastasis accounts for about 90% of CRC deaths. Intratumor heterogeneity (ITH) has been extensively identified in many solid malignancies; however, ITH in primary and metastatic CRC tumors, as well as the primary to metastatic tumor evolution are largely underexplored. **Patients and Methods:** We conducted high-depth whole exome sequencing in multiple regions of matched primary and metastatic tumors from four CRC patients. Using a total of 28 tumor, normal, and lymph node samples, we analyzed inter- and intra-individual heterogeneity, inferred the subclonal architecture of each tumor, and depicted the subclonal evolutional route from primary to metastatic tumors for each patient. **Results:** CRC has significant inter-individual heterogeneity but relatively limited intra-individual heterogeneity. Genomic landscapes were more similar within primary, metastatic or lymph node tumors than across these types. Remarkably, all metastatic tumors inherited multiple genetically distinct subclones from primary tumors, supporting a polyclonal seeding mechanism for metastasis. Nonetheless, metastatic tumors exhibited less ITH than primary tumors, indicating that single-region sequencing may be adequate to identify important metastasis mutations to guide treatment. Analysis of one patient with the trio of primary, metastatic, and lymph node tumors supported a mechanism of synchronous parallel dissemination from the primary to metastatic tumors that were not mediated by lymph nodes. **Conclusion:** In CRC, metastatic tumors have different but less heterogeneous genomic landscapes than primary tumors. CRC has a polyclonal seeding mechanism which suggests that different subclones migrate to distant organs and promote metastasis. These findings demonstrated the rationale and feasibility for identifying and targeting primary tumor-derived metastasis-potent subclones for the prediction, prevention, and treatment of CRC metastasis. .

750F

The Seven Bridges Cancer Genomics Cloud: Enabling discovery from petabyte-scale human genomic data resources. *E.H. Williams, G. Kaushik, E. Lehnert, R. Malhotra, A. Sethi, Z. Onder, B. Davis-Dusenbery. Seven Bridges, Cambridge, MA.*

Next-generation sequencing has catalyzed the generation of petabytes of genomic data from thousands of patients, creating immense research opportunities but also immense challenges related to organizing, sharing, and analyzing large and often heterogenous data volumes. The Cancer Genome Atlas (TCGA) network alone has produced more than 2.5 petabytes of data. Downloading the complete TCGA dataset to a local data store can take several weeks or more, and integrated analysis requires resources available only to a limited number of researchers with access to large institutional compute clusters. With the aim of democratizing access to TCGA data, the National Cancer Institute in 2015 launched the Cancer Genomics Cloud Pilots program to colocalize data and the computational resources necessary for its analysis in the cloud. Funded by this program, the Seven Bridges Cancer Genomics Cloud (CGC; cancergenomicscloud.org) is a customizable and scalable data access and analysis platform that connects users via the web to extensive public datasets, including TCGA, the Simons Genome Diversity Project, and the Cancer Cell Line Encyclopedia, and enables collaborative, reproducible analysis across both public and private cohorts. The CGC also gives users access to a public toolkit containing nearly 250 common analytical tools and workflows, and additional resources including the open-source CGC Software Development Kit and Data Cruncher facilitate custom tool prototyping and deployment on the platform. Since its launch in early 2016, more than 1700 researchers from more than 150 institutions in 30 countries have used the CGC to deploy more than 5,000 applications and perform more than 80,000 analyses, representing almost 100 years of computation time. To demonstrate how the CGC can help researchers unlock insights into the relationship between genomics and human health, we will present data analysis case studies - beginning with raw sequencing reads and ending with population-scale genomic and clinical data that can be fed into statistical models for biomarker discovery - conducted through this secure, compliant cloud environment. For example, we will show how to generate biomarkers for TCGA cancer types using machine learning in Python and R to classify cancer subtypes from gene expression data that has been reprocessed using gold-standard RNA sequencing workflows described in the Common Workflow Language.

751W

CliP: Fast subclonal architecture reconstruction from whole-genome sequencing data. *K. Yu[1], S.J. Shin[2], H. Zhu[1], W. Wang[2], the Evolution and Heterogeneity Working Group of the Pan-Cancer Analysis of Whole Genomes Initiative. 1) Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX; 2) 1.Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston,TX.*

Introduction: Tumors as well as various human tissues usually consist of different subpopulations (subclones) that are characterized by somatic mutations. The composition of such subpopulations may affect disease prognosis and treatment efficacy. And understanding the subclonal structure helps infer the evolutionary history of cells which can further guide the discovery of driver mutations such as those in cancer studies. Currently most subclonal reconstruction methods are Dirichlet Process (DP) based, requiring expensive computing since MCMC algorithm is commonly adopted to solve the problem, and careful post-processing due to the fact that the number of clusters scales with number of single nucleotide variations (SNVs) in the DP setting. Methods and Results: We present CliP (Clonal structure identification through penalizing pairwise difference), a fast and minimum post-processing algorithm for calling subclonal structures using whole-genome sequencing data. CliP deterministically solves a penalized likelihood problem, making execution much faster. In a simulated sample with 15,000 SNVs, it takes CliP less than 1 hour to finish the subclonal reconstruction, which is about 100 times faster than other methods. CliP is also capable of handling sequencing data at low coverages (30X-60X). Based on 125 simulated samples, CliP showed comparable prediction performance with DP-based methods. In the application to ICGC whole-genome sequencing data from 2,703 tumor samples from 39 cancer types, CliP provided results within 4 days and was able to analyze the hypermutator (>100,000 SNVs) samples correctly. Conclusion: We have proposed CliP, a fast and accurate subclonal reconstruction method that requires minimum post-processing. It is especially suitable for quickly processing large cohort containing thousands of whole-genome sequencing data to analyze cancer heterogeneity and their evolutionary histories.

752T

Dissecting tumor-immune system interaction in non-small cell lung cancer using TCGA data. *X. Yu, X. Wang.* Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL.

Immunotherapies have recently revolutionized the treatment of diverse types of cancer, including non-small cell lung cancer (NSCLC). While treatment with PD-1 blockers induce durable response in approximately 20% of patients with advanced NSCLC, majority of patients still fail to response or do not derive substantial clinical benefits. Efforts have been made to study the tumor-immune interaction and to develop predictive genomic markers for immune checkpoint blockage in varied cancer types. Resent studies have linked the higher response rate to elevated mutational load or predicted class-1 neo-antigens, increased CD8+ tumor infiltrating lymphocytes (TILs) and clonally expanded T-cell populations. However, much less is known about how the interaction between tumor and immune system impact NSCLC patient outcome. In this work, we comprehensively studied the association between tumor genomic landscape, immunogenicity, and composition of the tumor immune microenvironment in TCGA NSCLC patients utilizing a wide range of computational genomic tools. We found that the mutational load and neoantigen burden were not significantly associated with immunogenicity or TIL pattern. Using the gene expression measured from RNAseq, were able to identified makers only associated with high mutation low cytolytic activity. In addition, strong correlations were observed between tumor cytolytic activity, T cell receptor (TCR) diversity and TIL compositions, however a group of patients who have high T-cell infiltrating and low immunogenicity showed the worst survival, regardless of their mutational burden, indicating a potential target group for PD-1 axis blocker treatment. .

753F

RADAR: A *RNA* bin*D*ing protein regulatory network resource for c*A*ncer Research. *J. Zhang[1,2], J. Liu[1,2], D. Lee[2], L. Lochovsky[1,2], M. Gerstein[1,2,3].* 1) Molecular Biophysics & Biochemistry, Yale University, New Haven, CT; 2) Program in Computational Biology & Bioinformatics, Yale University, New Haven, CT; 3) Department of Computer Science, Yale University, New Haven, CT.

In the cell, numerous RNA binding proteins (RBP) interact with single or double stranded RNAs to play a series of critical post-transcriptional regulatory roles. Malicious regulatory activities of these RBPs, either from abnormal expression of RBPs or disruption of binding sites by mutations, may affect various types of RNA processing and result in aberrant expressions of target genes. Hence, it is important to investigate how regulatory changes of RBPs are associated with phenotype alterations in various diseases, such as cancer. However, currently, there is no systematic assessment of RBP regulatory networks in various cell types. Hence, we collected 318 RBP binding profiles from public data and endeavored to make it a suitable resource for cancer research. Specifically, we first investigated the functional importance of the RBP binding sites by studying their selection constraints from single-nucleotide polymorphisms (SNPs). We identified categories of RBP binding sites to be significantly enriched with rare SNPs, even to a degree similar to the coding regions, strongly indicating that they are playing important roles in gene regulation. We then set up a high-confidence RBP regulatory network by linking RBP binding profiles from multiple experiments with gene annotations. To pinpoint key RBPs that shape tumor-specific expression patterns, we performed differential expression analysis in 24 cancer types using data from TCGA data and evaluated the regulation power of each RBP by investigating their target gene expression changes. Specifically, we identified several RBPs, such as DDX55 and SUB1, to significantly drive tumor-to-normal differential expression. In addition, we performed a de novo motif scan on each RBP binding profile and compared it with RBPs reported from previous investigations using SELEX experiments. The majority of our discovered RBP motifs are in concordance with previous results, indicating the high quality of the binding profiles we collected. Based on these motifs, we further examined the RBP motif gain and loss events in multiple cancers and assigned their significance by comparing them against robustly simulated somatic variants As a culmination of our efforts, we release our tool, RADAR, for prioritizing and inferring RBP's regulatory potential for cancer research. It is aimed to interpret patient gene expression profiles using our identified RBP network and evaluate the RBP regulatory changes introduced by somatic mutations.

754W

Domain retention in transcription factor fusion genes and its biological and clinical implications: A pan-cancer study. *Z. Zhao[1,2], P. Kim[1].* 1) School of Biomedical Informatics, University of Texas HSC Houston, Houston, TX; 2) Human Genetics Center, School of Public Health, University of Texas HSC Houston, Houston, TX.

Genomic rearrangements involving transcription factors (TFs) can form fusion proteins resulting in either enhanced, weakened, or even loss of TF activity. Functional domain (FD) retention is a critical factor in the activity of transcription factor fusion genes (TFFGs). A systematic investigation of FD retention in TFFGs and their outcome (e.g. expression changes) in pan-cancer has not been done yet. Here, we examined the FD retention status in 386 TFFGs across 13 major cancer types and identified 151 TFFGs involving 95 TFs that retained functional domains. To measure the potential biological relevance of TFs in TFFGs, we introduced a major active isofusion index (MAII) and built a prioritized TFFG network using both MAII scores and the observed frequency. Interestingly, the four TFFGs (*PML-RARA*, *RUNX1-RUNX1T1*, *TMPRSS2-ERG*, and *SFPQ-TFE3*) having the highest MAII scores showed 50 differentially expressed target genes (DETGs) in fusion-positive versus fusion-negative samples. Gene set enrichment tests of these DETGs revealed genes involved in tumorigenesis-related processes in each cancer type. *PLAU*, which encodes plasminogen activator urokinase and serves as a biomarker for tumor invasion, was found to be consistently activated in the samples with highest MAII scores. Among the 50 DETGs, 21 are drug targetable and 14 are expressed in acute myeloid leukemia (AML) samples. Accordingly, we constructed an AML-specific TFFG network, which included 38 DETGs in *RUNX1-RUNX1T1* or *PML-RARA* positive samples. In summary, this pan-cancer study revealed several critical TFFGs and their potential target genes, and provided insights into the clinical implications of TFFGs.

755T

Integrative approach to cancer driver gene discovery from somatic mutations. *S. Zhao, X. He, M. Stephens.* Human Genetics, University of Chicago, Chicago, IL.

Knowledge of driver genes whose mutations lead to tumor-genesis is important for un- derstanding the mechanisms of cancer and for identifying promising drug targets. Despite intensive sequencing and computational efforts, our knowledge of driver genes in most tumor types remain far from complete. Computational methods are essential to leverage cancer sequencing data to identify driver genes. Existing methods typically rely on some pattern of somatic mutations in the genes: driver genes would have more somatic mutations than expected, or the mutations in driver genes are more likely to be deleterious and spatially clustered. Nevertheless, most existing methods would utilize one feature a time. In this work, we develop a novel integrative approach for driver gene discovery that incorporates multiple features of driver genes in a single framework. As part of our approach, we come up with a mixed effect model to capture the baseline mutation rates, that incorporates both global features predictive of mutation rates and local variations of mutation rates. Our model also allows us to put more emphasis on more deleterious mutations (e.g. nonsense mutations) when evaluating the mutational burden of a gene. Finally, our model is designed to reward genes with mutations that are closely clustered. We evaluate our approach using TCGA data, and we found that in almost all tumor types, it outperforms MutSigCV, the current leading method for driver gene discovery.

756F

Identification of tissue-specific regulatory networks. *M. Bilow, E. Eskin.* Computer Science, UCLA, Los Angeles, CA.

   A great number of gene-regulatory networks and pathway analyses have been performed, beginning with the advent of microarray technology 20 years ago, and these number and complexity of these networks continues to grow. Many of these analyses have been performed in a single tissue only, for example blood or brain, and this has resulted in the publication of gene interaction networks that are insufficiently general. The Gene-Tissue Expression Project has changed this by sampling from many tissues belonging to the same individual. This advance makes it possible to identify tissue-specific regulatory networks and fit a Bayesian network to this data. The Bayesian network has three primary components: tissue-transcript effects: the influence by each tissue on the quantity of each transcript; variant-specific effects: relationships between each genetic variant and each transcript; and transcript-transcript interactions: networked causal relationships between transcripts. Our algorithm follows an expectation-maximization paradigm and is able to provide confirmation and evidence for multi-tissue gene regulatory networks, many of which extend previously-known gene networks.

757W

A comprehensive characterization of tumor profiles using custom SureSelect targeted panels. *A. Ashutosh, A. Vadapalli, L. Forsmark, H. Tao, C. Lecocq, L. Niu, A. Khare, B. Arezi, G. Amparo, M. Ramaiah, M. Ruvolo, C. Pabon, J. Ghosh, J. Jin, D. Roberts.* Diagnositics and Genomics (DGG), Agilent Technologies, Santa Clara, CA.

   Somatic copy number variations (CNV) play a critical role in oncogenesis and in tumor progression. However, due to tumor heterogeneity and plurality of clones in tumors, the problem becomes complex. Additionally, aberration size, copy number and prevalence vary widely within and across tumor types. Further, formalin fixed paraffin embedded (FFPE) tissues present additional experimental challenges such as poor sample quality, crosslinking, and deamination. Traditional molecular profiling of cancer samples (e.g. FISH and SNP arrays) is gradually being replaced by next-generation sequencing (NGS) since in addition to detecting CNVs, it provides a way for rapid characterization of single nucleotide variations, short insertions and deletions, and translocations in a single assay. Until now CNV detection has been with either whole exome or large gene NGS panels containing hundreds of genes. Several current computational approaches for estimating both copy number and cancer clone fraction work better on large panels, exomes or WGS than they do on small panels.In this abstract, we address the challenge of detecting copy number variations at both gene and chromosome scales with NGS using small gene panels. The success of the analysis method is tied to the strategic selection of Agilent SureSelect probes that cover not only the coding regions of the genes of interest but also span the surrounding regions to provide additional allelic copy number information. This analysis utilizes either a matched normal or standard control DNA to compute read depth ratios. The aberrant regions are identified by jointly segmenting the read depth ratios and SNP allele frequencies. The clonal structure is discerned by evaluating two Bayesian cancer models in a hierarchical manner. The contribution of the major aberrant clone is separated out by the first model and a more detailed second model provides a refined clonal structure if present. We validate the results by an orthogonal technique-(FISH), samples with known aberrations and dilution series experiments for sensitivity. As an example, we show detection of 3 copies of MET at 20% tumor fraction and 16 copies of ERBB2at 2%.

758T

GATK4 adds germline and somatic *copy number variant* plus somatic *SNV and indel* calling. *S.H. Lee, Data Sciences and Data Engineering (DSDE) Team.* The Broad Institute of MIT and Harvard, Cambridge, MA.

The Genome Analysis Toolkit software is well established for germline SNP and indel discovery from whole genome (WGS) and exome short-read sequencing data. The command-line-interface program has 46K users worldwide[1] and offers comprehensive support through extensive documentation[2], an active online forum[3], and interactive workshops[4]. Major version 4 (GATK4) is re-engineered to compute and scale more efficiently. It is completely open source[5] and has new and improved collections of computational tools towards (i) somatic *copy number variant* discovery, (ii) somatic *SNV and indel* calling with an enhanced Mutect2, (iii) germline *copy number variant* discovery (experimental workflow), and (iv) germline *SNP and indel* calling with Haplo-typeCaller. We will focus on the newer workflows (i–iii). Additionally, we will touch briefly upon functionality and resources that enable researchers to study large genomic data. For example, GATK4 supports use of the GenomicsDB datastore[6] towards scalable analyses. This functionality, contributed by collaborators at Intel Health and Life Sciences, enables calling on large datasets exemplified by the joint variant calling of 15K WGS samples that resulted in the gnomAD resource[7]. In addition, the GATK4 engine supports Apache Spark parallelization of compute for faster results, and can directly read from and write to Google Cloud Storage. Finally, resources include the Cromwell and WDL pipelining solution that enables reproducibility of analyses as well as easy parallelization of compute, e.g. scattering across samples and over genomic intervals[8].

_____ [1] GATK user metrics *https://software.broadinstitute.org/gatk/blog?id=7736* [2] GATK documentation *https://software.broadinstitute.org/gatk/documentation* [3] Recent forum discussions *http://gatkforums.broadinstitute.org/gatk/discussions* [4] Information on workshops *https://software.broadinstitute.org/gatk/blog?id=8622* [5] Open source announcement *https://software.broadinstitute.org/gatk/blog?id=9645* [6] GenomicsDB *https://github.com/Intel-HLS/GenomicsDB/wiki* [7] GnomAD website *http://gnomad.broadinstitute.org* [8] WDL is the workflow definition language that the Cromwell engine reads. Cloud Cromwell-WDL user story *https://software.broadinstitute.org/gatk/blog?id=8646* and documentation *http://gatkforums.broadinstitute.org/wdl*.

759F

Detecting pathogenic structural variants with long-read PacBio SMRT Sequencing. *A.M. Wenger, L. Hickey, Y. Li, B. Lerch, P. Peluso, J. Korlach.* Pacific Biosciences, Menlo Park, CA.

Most of the basepairs that differ between two human genomes are in intermediate-sized structural variants (50 bp to 5 kb), which are too small to detect with array comparative genomic hybridization or optical mapping but too large to reliably discover with short-read DNA sequencing. Long-read sequencing with PacBio Single Molecule, Real-Time (SMRT) Sequencing platforms fills this technology gap. PacBio SMRT Sequencing detects tens of thousands of structural variants in a human genome with approximately five times the sensitivity of short-read DNA sequencing. Effective application of PacBio SMRT Sequencing to detect structural variants requires quality bioinformatics tools that account for the characteristics of PacBio reads. To provide such a solution, we developed pbsv, a structural variant caller for PacBio reads that works as a chain of simple stages: 1) map reads to the reference genome, 2) identify reads with signatures of structural variation, 3) cluster nearby reads with similar signatures, 4) summarize each cluster into a consensus variant, and 5) filter for variants with sufficient read support. To evaluate the baseline performance of pbsv, we generated high coverage of a diploid human genome on the PacBio Sequel System, established a target set of structural variants, and then titrated to lower coverage levels. The false discovery rate for pbsv is low at all coverage levels. Sensitivity is high even at modest coverage: above 85% at 10-fold coverage and above 95% at 20-fold coverage. To assess the potential for PacBio SMRT Sequencing to identify pathogenic variants, we evaluated an individual with clinical symptoms suggestive of Carney complex for whom short-read whole genome sequencing was uninformative. The individual was sequenced to 9-fold coverage on the PacBio Sequel System, and structural variants were called with pbsv. Filtering for rare, genic structural variants left six candidates, including a heterozygous 2,184 bp deletion that removes the first coding exon of *PRKAR1A*. Null mutations in *PRKAR1A* cause autosomal dominant Carney complex, type 1. The variant was determined to be *de novo*, and it was classified as likely pathogenic based on ACMG standards and guidelines for variant interpretation. These case studies demonstrate the ability of pbsv to detect structural variants in low-coverage PacBio SMRT Sequencing and suggest the importance of considering structural variants in any study of human genetic variation.

760W

Integrated search for multi-omics data using extended GA4GH Genomics API. *S. Kawano[1], Y. Suzuki[2], S. Sugano[3].* 1) Database Center for Life Science, Joint Support Center for Data Science Research, Research Organization of Information and Systems, Kashiwa, Chiba, Japan; 2) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan; 3) Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan.

In recent years, diagnosis using genome sequence information has been a major method for rare diseases and cancers. It is important that these information are efficiently shared between hospitals in order to find similar patients and help medical treatment. Global Alliance for Genomics and Health (GA4GH), which is an international collaboration accelerating health-data sharing, provides several tools such as Genomics API, Beacon, and Matchmaker Exchange in order to share the genomic and phenotypic information. The GA4GH tools mainly focuses on single nucleotide variants (SNVs) in genotypic information. However, other factors such as epigenome (DNA methylations and histone modifications measured by bisulfite sequencing (BS-seq) and chromatin immunoprecipitation sequencing (ChIP-seq), respectively), transcription start sites (TSS) identified by TSS-seq, and amounts of gene expression determined by RNA-seq (hereinafter collectively called multi-omics) also affect phenotypes. It is convenient to be able to search not only SNVs but also multi-omics data at the same time in order to obtain more detailed information about relationships between genomic information and phenotypes. Here, we chose Genomics API to integrate multi-omics data. Because the Genomics API does not support to multi-omics data, we extended it's data model: we defined extra classes such as BS-seq, ChIP-seq, TSS-seq, and RNA-seq in addition to the variant class. As sample datasets, we used multi-omics data including SNVs data measured from 26 lung cancer cell lines, and forcibly converted them into Variant Call Format (VCF) files. The VCF files were loaded into a Genomics API repository for searching. As a result, multi-omics data was searchable in an integrative way. For example, our API returned a list of SNVs, DNA methylation sites, and histone binding sites etc. from a specific region (e.g. 10000 - 15000) on a specific chromosome (e.g. chr.1). Application of this API enables correlation analysis as well as similarity search between samples (patients) using multi-omics data. In conclusion, we extended the standard technology provided by GA4GH while ensuring compatibility, and enabled integrated search of multi-omics data.

761T

MicroRNA eQTL analysis in pancreatic cancer with efforts towards functional validation. *A. Jermusyk[1], L. Rost[1], M. Zhang[1], R. Atrah[1], J.W. Hoskins[1], B. Zhu[2], W. Xiao[3], I. Collins[1], N. Chatterjee[3,4], J. Shi[2], L. Amundadottir[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA; 2) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA; 3) Division of Bioinformatics and Biostatistics, National Center for Toxicology Research, FDA, Jefferson, Arkansas 72079, USA; 4) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, USA.

MicroRNAs (miRNA) are small, non-coding RNAs of about 21 nucleotides in length that bind the 3' UTR region of messenger RNAs (mRNAs) preventing translation and causing destruction of the mRNA. The roles of individual miRNAs in cancer remains to be fully uncovered. Our aim in the current study is to identify miRNAs whose expression is genetically regulated in pancreatic cancer. We performed a miRNA eQTL analysis using data from TCGA pancreatic cancer tumor samples (n=115) and employing the linear regression program, Matrix eQTL. After determining significant SNP-gene pairs through running 100 permutations of our analysis and establishing a FDR of 10%. At a *P*-value threshold of $P < 8.5e-7$ (FDR<0.1), two SNP-gene pairs were significant; miR-577 and rs35182026, and miR-3130-1 and rs10664575. We then predicted the mRNAs they potentially regulate using predictive binding databases. We investigated mRNAs that were consistently reported in multiple prediction packages and possibly involved in cancer. Based on these analyses: *GOLFPH3*, *TFAP2A*, and *HOXA1* are probable targets of mir-577 and *FAS*, *SIM1*, and *RAB14* are probable targets of mir-3130-1. We then assayed the expression of two of the most significant miRNAs with eQTLs (miR-3130-1 and mIR-577) in 30 pancreatic cancer cell lines in order to select cell lines to overexpress and downregulate these miRNAs. We identified IMIMPC-2, MiaPaCa-2, and AsPC-1 as ideal cell lines to overexpress and/or downregulate miR-577 and miR-3130-1. We will up- and downregulate these two miRNAs in these cell lines and then assess the expression of the predicted target mRNAs. Ultimately, we are working towards determining whether we can verify the miRNA-mRNA binding predictions, and determine what are the implications of these binding events for cancer. We have identified miRNA eQTLs in pancreatic cancer tissues and are working towards identifying their target mRNAs. We hope to uncover the relevance of these miRNA in pancreatic cancer biology through overexpression and downregulation, and examine the resulting phenotypic effects.

762F

Designing for success: The right CRISPR design strategies for the right experiment. *L. Brody, N. Humphryes-Kirilov, V. Dillard, R. Doyle.* Desktop Genetics, 28 Hanbury Street, London, E1 6QR, United Kingdom.

CRISPR design is a complex multivariate problem. The weight of each variable differs depending on the experiment under consideration. Current methods of designing CRISPR experiments are based on a single-variable, resulting in poor experimental outcomes. Our current scoring methods were tested and validated using an sgRNA library of over 3,500 sgRNAs in essential genes with a variety of predicted sgRNA activities and specificities. The predicted 'good' sgRNAs achieved significantly higher read depletion for essential genes, with fewer failing sgRNAs, compared to sgRNAs that didn't meet the scoring criteria. Our scoring algorithm incorporates multiple published observations for choosing effective sgRNAs, which differs from other CRISPR tools that tend to focus on one factor such as sgRNA sequence or specificity. In addition, we have incorporated variant databases which provide a basis for the the investigation of the molecular and cellular activity of genetic variation, which is increasingly important for understanding the mechanistic function of the human genome. At Desktop Genetics, we have developed a CRISPR design platform accounting for genomic variants including single nucleotide polymorphisms, small insertions/deletions, translocations and copy number variants. Using NGS data and our optimized CRISPR design rules, we demonstrate a bespoke methodology for *in vitro* functional assays in cultured human cells of a defined genotype. We show how model-specific genomes in conjunction with customizable guide RNAs can be used to precisely target genomic sites of interest to induce loss-of-function mutations and validate predicted variant function. This may provide a unique approach for downstream variant interpretation and interrogation of their role in the pathogenesis of rare and common genetic diseases.

763W

Whole genome sequencing signatures for early detection of cancer via liquid biopsy. *B.G. Kermani.* Crystal Genetics, Inc., Los Altos, CA.

The enthusiasm around early detection of cancer using next-generation sequencing (NGS) has placed this goal in the spotlight, particularly in the recent years. Liquid biopsy is often defined as the modus operandi for early detection, as taking biopsies from the actual organs is not practical for a wide-spread screening test. Liquid biopsy comprises cell-free tumor DNA (ctDNA) and circulating tumor cell (CTC) approaches. While ctDNA is more popular, mostly due to the ease of operation, it suffers from low signal-to-noise ratio (SNR). CTC, on the other hand, provides the ability to interrogate single cells with high SNR. However, finding such cells, especially at the earlier stages of cancer, has been challenging. In addition to the targeted gene panels, whole exome sequencing (WES) and whole genome sequencing (WGS) have been considered in the past, for CTC applications, albeit primarily on prognosis (and not diagnosis). The common ideas hinge upon correlating the count of the CTCs or the discovered copy number variations (CNVs) with the state of the disease or lack thereof. In this work, our approach has been focused on using WGS for cancer diagnosis, although other NGS modalities may also be considered. We have identified proprietary signatures that have shown promise in identifying cancer versus normal tissues, in specific cancer types such as breast cancer. Some of these signatures have certain properties that would make them portable to the CTC domain. Since most of publicly available data on CTC work has been on metastatic cancers, we have shown that some signatures hold for such data. Moreover, considering the error modes of CTCs, e.g., allele dropout (ADO), there appears to be a path to maintain the integrity of some of these signatures, although less efficiently, in CTCs from the earlier stages of cancer. Currently, based on limited data, our approach has shown promise at the WGS tissue level, with a detection rate of ~90% for Stage I and Stage II of breast cancer. In order to calculate the upper-bound on the sensitivity of this method using liquid biopsy, the tissue-derived number would have to be multiplied by the detection rate of the CTCs, which is currently low to medium, depending on the technology and the cancer type. However, as the CTC detection rate improves, given the R&D efforts in this area, we anticipate that this method would gain more significance in the early detection of cancer.

764T

Developing validated phenotypic cancer cohorts for molecular stratification and susceptibility assessment, a use case: Patients diagnosed with early versus late stage non-small cell lung cancer. *B.R. Johnson[1], N. Fillmore[1], A. Zimolzak[1], Y.L. Ho[1], D. Elbers[1], B. Katcher[1], D. Gagnon[1,2], F. Meng[1,3], M. Brophy[1,4], L. Fiore[1,4], A. Lesse[5,6], J. Concato[7,8], J.M. Gaziano[1,9], N. Do[1,10], P.L. Elkin[5], K. Cho[1,9], The Million Veteran Program, Department of Veterans Affairs Office of Research & Development.* 1) Massachusetts Veterans Epidemiology Research and I, VA Boston Healthcare System, Boston, MA., Select a Country; 2) partment of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; 3) ical Imaging and Informatics Group, Department of Radiological Sciences, David Geffen School of Medicine, UCLA, LA, CA, 90024 USA; 4) School of Medicine, Boston University, Boston, MA 02118, USA; 5) Department of Medicine, VA Western New York Healthcare System, Buffalo, NY 14203; 6) Department of Internal Medicine, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY 14203, USA; 7) CERC, VA Healthcare System, West Haven, CT; 8) School of Medicine, Yale University, New Haven, CT; 9) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 10) Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY 14203, USA.

Background: We identified a large, curated source population of U.S. veterans with non-small cell lung cancer (NSCLC). This enables us to a) train text-based machine learning case-finding algorithms, and b) derive correlations between genetic variants, comorbidities, exposures, and cancer presentation. Identification of this cohort will support accurate, timely screening patients and building prediction algorithms from structured and unstructured data, to drive better clinical care. We evaluated structured medical codes (ICD codes) for phenotypes to compare to curated cases in the cancer registry (CR), which includes histology data. We also identified the subset of patients in the Million Veteran Program (MVP). The cohort is categorized by tumor stage at diagnosis, allowing us to test initial workflow and algorithms, as well as validate genetic tools relative to known susceptibility loci (eg, *ALK*, *EGFR*, *KRAS*, Chr. 15,6,5), comorbidities (eg, COPD), and exposures (eg, smoking). **Methods:** Patients with early (stage I and II) and late (III and IV) stage onset of cancer at diagnosis were identified in the national corporate data warehouse (CDW) and CR. We used the CMS chronic disease data warehouse definitions of >20 ICD codes, with >1 encounter observed in an electronic health record, to identify subjects in both CDW and MVP. Data were integrated with date of diagnosis, date of first treatment, histology, stage, and primary tumor type from the CR. We included only patients with adenocarcinoma, squamous cell carcinoma, or large cell carcinoma specified. We excluded patients who had conflicts in stage or histology, and we only included patients once if described in the CR multiple times. **Results:** Using the CMS definition, we identified 399,504 distinct patients within the CDW, and 14,435 in MVP. Whereas, queries of the manually curated CR identified 211,899 patients, of whom 6,206 are in MVP. We found a significant difference (p<0.001) in mean time to treat NSCLC patients diagnosed at early (52.8 days) versus late (41.8 days) stage. **Conclusion:** By integrating clinical data with a manually curated CR, we have identified a large cohort of NSCLC subjects, including thousands with germline variants and survey data collected. We also demonstrate a difference by stage in time to the first course of treatment. This vetted cohort allows the correlation of susceptibility genes, comorbidities, and exposures to clinical presentation (stage and histology).

765F

Somatic mutation identification through haplotype discrepancy in tumor-only sequencing without matched normal DNA. *W. Chen, G. Wu, S. Wang, J. Wen.* St. Jude Children's Research Hospital, Memphis, TN.

The major interests in cancer research is to identify mutations that are of clinical importance, especially tumor-specific (somatic) alterations that help disease prognosis, diagnosis and treatment. To achieve high specificity, matched normal/germline samples are always required to distinguish somatic mutations from human germline variants. However, very often the matched normal specimens are not available, limiting our ability to take advantage of larger cohort to detect recurrent somatic mutations. The challenge becomes more significant for rare tumors that might harbor novel driver mutations. To analyze the tumor-only samples, currently there are two approaches. The first approach relies on large human polymorphism databases to eliminate germline variants and cancer mutation databases to recognize somatic mutations. The second approach utilizes the fact that most tumor specimens contain normal cell contamination that cannot be removed in surgery, and compute the likelihood of somatic status by assessing the variant allele frequency deviation. Both approaches, however, use indirect evidences from either unmatched samples or statistical inferences, which cannot validate the true state of somatic mutations. To capture direct evidences of somatic mutations, we described a novel approach called SMITH (somatic mutation identification through haplotype) that utilizes the nature of normal contamination in tumor samples and the barcode feature in 10x Genomics Chromium data to call haplotypes in tumor-only sequencing. We hypothesized that variants with >2 reliable haplotypes are from somatic origins, otherwise from germline origins. We evaluated our approach using one low grade pediatric diffuse astrocytoma. First, the paired tumor and normal samples were sequenced using Illumina WGS. Using paired analytical approach, we identified 58 somatic mutations to serve as positive control and 1949 rare germline variants as negative control. We then sequenced the tumor sample using the Chromium platform, and applied SMITH. Out of the 58 somatic mutations, 51 were successfully classified as somatic, while only 2 out of the 1949 germline variant were mis-classified as somatic by SMITH (precision=96.2%, recall=87.9%). Overall these results suggest that SMITH is an effective approach capturing direct evidence for somatic mutations, and highlighting its potential application including, but not limited to, tumor-only somatic mutation or sub clonal somatic mutation discovery.

766W

Leveraging protein coding gene expression profiles to accurately impute lncRNA transcriptome of uncharacterized samples. *A. Nath, P. Geeleher, R.S. Huang.* Medicine, University of Chicago, Chicago, IL.

Long non-coding RNA's (lncRNAs) play an important role in gene regulation and are increasingly being recognized as crucial mediators of disease pathogenesis. Accurate profiling of the lncRNA transcriptome however remains a challenge owing to significantly lower expression compared to protein coding genes (PCGs). Thus, unlike PCGs, high-quality lncRNA transcriptome data remains sparsely available for a vast array of phenotypes. Here we propose a lncRNA expression imputation (LEXI) framework to reconstruct the lncRNA transcriptome of uncharacterized samples using their PCG profiles. LEXI is based on learning patterns of PCGs associated with each lncRNA across a diverse cohort of samples, and then imputing lncRNA transcriptome profiles in the uncharacterized sample. Specifically, we developed LEXI by evaluating the performance of various machine-learning algorithms benchmarked in a cross-validation study across a cohort of cancer cell lines. Furthermore, we demonstrate that LEXI accurately imputes lncRNA profiles not only in cell lines, but in cancer and normal tissues as well with models generated using independent RNAseq and microarray datasets. The goal of LEXI is to harnesses the enormous wealth of publicly available PCG expression data and enable discovery of novel lncRNA associations in various phenotypes, including disease susceptibility, prognosis and drug sensitivity. As a proof of concept, we corroborated differential expression patterns of oncogenic lncRNAs *MALAT1* and *HOTAIR* in human cancers. These results suggest that indeed LEXI can enable profiling of lncRNA expression in various uncharacterized samples and therefore allow follow-up exploration of lncRNA function.

767T

Subtype-specific expression of long noncoding RNAs in b-cell acute lymphoblastic leukemia. *C. Nodzak[1], J. Wen[1], G. Centoducatte[2], J.A. Yunes[2], X. Shi[1].* 1) Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC; 2) Laboratorio de Biologia Molecular, Centro Infantil Boldrini, Campinas, Sao Paulo, Brazil.

Precursor B-cell acute lymphoblastic leukemia (ALL) is a hematological cancer that commonly afflicts children and adolescents. One of the first subtypes of ALL described was the Philadelphia chromosome positive (Ph+) ALL, characterized by the reciprocal translocation between the long arms of chromosomes 9 and 22 with an expressed fusion product of the BCR and ABL1 kinases. Many other subtypes of ALL have been defined by various other gene fusion products (ETV6/RUNX1, TCF3/PBX1, AF4/MLL, among others) and abnormal ploidy levels. Recently, a poor prognosis type of ALL, known as Philadelphia-like (Ph-like), was defined by clustering with the (Ph+) ALL samples in global gene expression analysis. Long noncoding RNAs (lncRNAs) are known to play a role in the regulation of genes involved in multiple physiological processes and recent findings have continued to provide evidence that aberrant expression of these molecules can affect genes implicated in aspects of cancer hallmarks. Here, using data from the NCI-TARGET initiative we investigated subtype-specific co-expression networks between lncRNAs and mRNA across ten ALL subtypes followed by enrichment analysis to identify pathologically associated pathways. From the same source, a subset of Ph-like samples with matched RNA-seq and SNP 6.0 arrays were then tested for allele specific effects at heterozygous sites within these networks using an FDR corrected binomial test on the allelic ratios of uniquely mapped reads. The findings from our analysis show subtype-specific lncRNA expression in ALL, as well as an allelic effect on genes that belong to co-expression networks with lncRNAs in the Ph-like subtype.

768F

The NantOmics Pharmacogenomics Test: An integrative panomic approach to pharmacogenomics screening. *C. Schwartz[1], J. Little[1], C. Vaske[1], S. Benz[1], P. Soon-Shiong[3], S. Rabizadeh[2], J.Z. Sanborn[1].* 1) NantOmics, Santa Cruz, CA; 2) NantOmics, Culver City, CA; 3) NantWorks, Culver City, CA.

Numerous cancer therapies contain pharmacogenomics warnings on their FDA labels, yet pharmacogenomics screening is not yet used in routine clinical practice. Pharmacogenomics testing can be used to tailor therapies to the patient's genotype to reduce the chance of drug-induced toxicities, improve patient outcomes and reduce treatment costs. Here we present the NantOmics pharmacogenomics test, which uses whole genome (DNA) and whole exome sequencing data from FFPE tumors and matched normal samples to predict and advise physicians of possible toxicities. Using FDA labels and guidelines from the Clinical Pharmacogenomics Implementation Consortium (CPIC), we developed a clinical pharmacogenomics panel including 46 markers in 11 genes linked to toxicities from 14 cancer therapies, in addition to a research panel of 13 gene-drug pairs reported in primary literature. The test screens for both germline and somatic variants to determine how an oncology patient will respond to potential therapies in the clinical setting. The test was run on over 890 patient samples, 96.5% of which contained a variant in at least one gene from our panel. Furthermore, 8% of patients had genomic variants that could have resulted in severe or life threating drug toxicities. For all alleles in our clinical panel, we observed similar alleles frequencies to those reported in the ExAC database. Our test was validated on a cohort of patients previously genotyped by an independent CLIA-validated PCR-based panel, as well as on a set of synthetic data. Our validation studies demonstrated that the test is able to detect each variant in our panel, and is able to correctly determine patient genotype in all studied cases. Given the high number of patients with potential treatment-altering genomic variants, these results underscore the need for pharmacogenomics screening for oncology patients in a clinical setting.

769W

Cepip: Context-dependent epigenomic weighting for prioritization of regulatory variants and disease-associated genes. *J. Wang[1], M.J. Li[2,3,4], Z. Liu[3,5], B. Yan[3], Z. Pan[3], D. Huang[4], Q. Liang[4], D. Ying[3], F. Xu[3], H. Yao[3], P. Wang[1], J.P.A. Kocher[1], Z. Xia[3,5], P.C. Sham[3], J.S. Liu[2].* 1) Department of Health Sciences Research and Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ 85259, USA; 2) Department of Statistics, Harvard University, Cambridge, Boston, MA 02138-2901, USA; 3) Center for Genomic Sciences, The University of Hong Kong, Hong Kong SAR, China; 4) Department of Pharmacology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China; 5) Department of Anaesthesiology, The University of Hong Kong, Hong Kong SAR, China.

It remains challenging to predict regulatory variants in particular tissues or cell types due to highly context-specific gene regulation. By connecting large-scale epigenomic profiles to expression quantitative trait loci (eQTLs) in a wide range of human tissues/cell types, we identify critical chromatin features that predict variant regulatory potential. We present cepip, a joint likelihood framework, for estimating a variant's regulatory probability in a context-dependent manner. Our method exhibits significant GWAS signal enrichment and is superior to existing cell type-specific methods. Furthermore, using phenotypically relevant epigenomes to weight the GWAS singlenucleotide polymorphisms, we improve the statistical power of the gene-based association test.

770T

Negative binomial model-based clustering: Discover novel molecular subtypes of ovarian cancer. *Q. Li[1], J.R. Noel-MacDonnell[2], D. Koestler[3], E.L. Goode[4], B.L. Fridley[1].* 1) Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL; 2) Department of Health Services and Outcomes Research, Children's Mercy Hospital, Kansas City, MO; 3) Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Unsupervised clustering refers to the grouping of subjects and/or genomic features based on their similarity and represents one of the widely applied methods in analysis of high-throughput 'omic data. A variety of unsupervised clustering methods have been proposed for RNA-Seq count data, however most of the methods currently applied do not capture the over-dispersed nature of this data. We propose a Negative Binomial Model-Based (NBMB) method to cluster samples based on RNA-Seq data using a modified Expectation-Maximization algorithm, and is available in an R package named *NB.MClust*. Results from an extensive simulation study found that NBMB without transform outperforms Gaussian Mixture Model (GMM) with different transformations, based on Adjusted Rand Index (ARI).When sample size is N=150 or larger, ARI of NBMB is better than or equal to that of GMM in over 95% of simulated datasets, half of which show strictly better ARI in NBMB. We further assessed the ability of transcriptome wide NBMB clustering to determine molecular subtypes using two tumor RNA-Seq studies of epithelial ovarian cancer (EOC), the first from Mayo Clinic and the second from TCGA. We compared the cluster assignments from NBMB to the existing "CLOVAR" classifier. In both datasets, NBMB discovered similar but slightly distinct clusters or subtypes based on the CLOVAR classifier. Three of NBMB subtypes in TCGA data were associated with progression-free survival. In order to clarify prognostic relevance, future work would apply NBMB clustering on more EOC tumors and integrating NBMB-based subtypes with known prognostic factors.

771F

Identification of potential LREA regions in prostate cancer cell lines using multi-omic analysis features of Strand NGS software. *P. Karuna, S.K. Sivamani, S. Ramachandrula, S.C. Cherukuri.* Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India.

Epigenetic gene deregulation in cancer commonly occurs through chromatin repression and promoter hypermethylation of tumor-associated genes. However, this mechanism of the epigenetic-based gene activation in carcinogenesis is still poorly understood.   Here, using Strand NGS bioinformatics software, we provide an illustrative example by integrating gene expression, chromatin and DNA methylation genome-wide profiles in prostate cancer and were able to identify 35 LREA (long-range epigenetic activation) domains that harbor 251 genes, including multiple gene families, and tumor-related genes. As a part of this study, we re-analyzed the publicly available datasets GSE38685. This study highlights the identification of LREA regions by Expression profiling, Methylation profiling, Genome variation profiling, and Genome binding by high throughput sequencing in an efficient and powerful way using Strand NGS. The clustering analysis of these activated domains showed significant enrichment in several gene families, including the *MAGE*, *UGT2*, and *KLK* families and the differential expression of the gene clusters were identified. The Genome binding analysis and the methylation analysis were performed to assess whether these activated gene clusters also exhibit epigenome changes. In prostate cancer cells, the entire *KLK* family shows a remarkable feature of switching between the gain of the repressive histone modifications H3K9me2 and H3K27me3 along with loss of active Histone modifications H3K9ac and H3K4me3. We also found examples of CpG island methylation specifically at the *KLK* locus that resulted in transcriptional activation. An overlay of the Expression, Methylation, and the ChIP-Seq data using WikiPathways displayed that *KLK* genes play a vital role in the Prostate cancer signaling pathway. Our results clearly demonstrate that cancer-specific DNA methylation of CpG islands contribute to deregulation of promoter usage, and can result not only in gene repression but also in cancer associated gene activation.

772W

Matched tumor/germline samples aid in detecting genomic instability in multiple myeloma using linked-read whole genome sequencing without the need for high molecular weight DNA. *C. Ashby, M.A. Bauer, A. Mikulasova, O.W. Stephens, R.G. Tytarenko, F.E. Davies, G.J. Morgan, B.A. Walker.* Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA.

   Introduction: Multiple Myeloma is a hematological malignancy of plasma cells in the bone marrow characterized by translocations, copy number aberrations and mutations. However, the interaction of these events has not been fully explored due to technological inadequacies. We sought to gain new understanding by using 10X Genomics linked-read whole genome sequencing (WGS). However, archived myeloma DNA samples do not meet the high molecular weight (HMW) recommendation by 10X Genomics. **Materials and Methods**: 32 pairs of tumor and germline samples from myeloma patients were analyzed using 10X Genomics WGS, of which 27 had known translocations involving the Ig loci. Sequences were aligned to hg38 using Longranger (v2.1.0). A novel algorithm was developed to detect somatic translocation events by statistically comparing shared barcodes across identified breakpoints between tumor and germline. Similarly, normalized depth and barcode ratio, haplotype and B allele frequency comparison was used for copy number variant detection. **Results**: Samples had a median molecule length (ML) of 25 kb, with only 7 of the 64 (11%) having the 10X Genomics recommended ML. Samples were sequenced to a median depth of 43.7x. Longranger's structural variant (SV) caller identified a median of 105 SV calls per sample but, when compared with germline, only a median of 22 were somatic events (range 3-145). Of these somatic events, a median of 6 (27%) had less than 2 supporting split and/or paired reads (as reported by Longranger), suggesting that they would not have been detected using standard WGS. There was a median of 5 (range 1-20) inter- and 14 (range 1-142) intra-chromosomal SVs. Of the 27 samples with expected Ig rearrangements, copy number breakpoints or SVs were detected in 24 (89%). Nine had an event that did not pass filter, but were verified by lack of overlapping barcodes in the germline sample. Of the discrepant samples, one had lower ML and mean depth (7 kb;32x) and one had lower mean depth (34x). Interestingly, two samples with lower ML had copy number events at the expected translocation breakpoint. **Conclusions**: Assisted by matched control sample, even low MW samples have sufficient barcodes to find somatic SV events revealing that the genome of myeloma is heterogeneous with each sample showing a wide range of SVs. It is clear that the linked read approach will help characterize the wide variance of genome instability in myeloma, and this in turn may impact outcome.

773T

Expression-based Variant Impact Phenotyping (eVIP) for determination of somatic mutation function in cancer. *A. Berger[1], L. Jerby[2], A. Regev[2], J. Boehm[2], A. Brooks[3].* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Broad Institute of MIT & Harvard, Cambridge, MA; 3) UC Santa Cruz, Santa Cruz, CA.

   Variants of unknown significance (VUS) pose a major challenge to the implementation of precision medicine. A bottleneck in functional characterization of genetic variants is the development of optimized and highly tailored assays for each gene of interest. We hypothesized that functionally impactful genetic changes will directly or indirectly affect the cellular transcriptional program, and gene expression profiling might therefore represent a general and scalable assay for characterization of mutation function.   To test this hypothesis, we generated an open reading frame (ORF) library of 194 missense or in-frame indel variants in 53 different genes. The mutations selected were somatic variants observed in primary lung adenocarcinomas. The library was introduced into human lung cell lines via lentiviral transduction. 96 hours after transduction, transcriptional signatures were generated by L1000 profiling.   We built a publicly-available computational pipeline, eVIP, that compares wild-type and mutant transcriptional signatures to classify mutations as impactful (gain-of-function, change-of-function, loss-of-function) or likely neutral. eVIP identified 69% of the mutations as impactful, versus 31% as likely neutral. eVIP correctly classified 21 known oncogene mutations as impactful and identified dozens of never-characterized gain- and loss-of-function mutations. We performed two complementary screens to determine the effect of the variants on drug resistance and xenograft tumor formation phenotypes. We validated rare, oncogenic variants in *ARAF*, *BRAF*, *EGFR*, *ERBB2*, *KRAS*, and *RIT1,* and identified small molecule MEK inhibitors as able to reverse the phenotypes induced by these rare variants.   To adapt eVIP to a pooled screen format, we sought to apply eVIP to single-cell RNA sequencing data. Ten wild-type or mutant ORFs were used to generate a pooled population of A549 cells. Single-cell RNA-seq transcriptional profiles were highly correlated to the bulk signatures generated by L1000, demonstrating the promise of single-cell RNA-seq to significantly expand eVIP throughput by enabling pooled screens.   Here we present eVIP as a generalizable method for characterization of genetic variants that requires no prior knowledge of gene function. Continued high-throughput functional characterization of mutations should illuminate the role of VUS in cancer and other diseases.

774F

Fix-C: A novel experimental and computational method for structural variation detection and *in silico* long range phasing from FFPE tumor tissue. *H.A. Costa[1], M. Blanchette[2], C.D. Bustamante[1,3], R.E. Green[4], P.D. Hartley[2], C. Kunder[5], N. Putnam[2], B. Rice[2], C. Troll[2], J.L. Zehnder[5].* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA, 94305 USA; 2) Dovetail Genomics, LLC, Santa Cruz, CA, 95060 USA; 3) Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, 94305 USA; 4) Department of Biomolecular Engineering, UC Santa Cruz, Santa Cruz, CA, 95064 USA; 5) Department of Pathology, Stanford University School of Medicine, Stanford, CA, 94305 USA.

Introduction: The clinical management and therapy of many solid tumor malignancies is dependent on the utilization of genomic tools for comprehensive detection of medically actionable or diagnostically relevant genetic variation. However, a principal challenge in tumor sequencing is the use of archival formalin-fixed paraffin-embedded (FFPE) tissue which produces highly fragmented low-molecular weight nucleic acid. This sub-optimal input specimen, coupled with the short read nature of DNA and RNA next-generation sequencing protocols, proves challenging to generate long-range (Mbp+) genomic sequence data that are able to accurately and robustly provide information on underlying large-scale structural variation and phasing from these specimens. **Methods**: As a proof-of-principle we have developed a Hi-C chromosome conformation capture methodology for FFPE tissue called "Fix-C" which yields phased read pairs spanning distances up to full chromosomes and enables unambiguous structural variation detection and variant phasing in archival specimens. We have applied our method to 15 clinical lung adenocarcinoma and sarcoma specimens spanning a broad range of tumor purity for structural variation fusion detection and somatic variant phasing. **Results**: Our method achieves a 100% concordance rate with concurrent FISH testing, which is currently used as the clinical gold standard. Additionally, this global structural detection approach enables unbiased whole-genome perspective for structural variation detection, and identifies novel structural events in our cohort missed by FISH and traditional RNA and DNA sequencing on these specimens. Lastly, our novel approach allows us to produce long phased sequences from FFPE tissue with an N50 of 132 megabases (full chromosome scale) in these specimens—enabling accurate and haplotype-aware somatic mutation phasing. **Conclusions**: Here we present a novel sequencing approach that leverages the genome scale structural resolution of Hi-C with the throughput of traditional Illumina sequencing methods. By *in silico* analysis we are able to reconstruct long allele haplotypes from fragmented FFPE specimens, and demonstrate the ability to identify known and novel structural events in tumor tissue. This advance is a paradigm shift in FFPE nucleic acid analysis and will enable high-detailed resolution of global structural variation underling solid tumor malignancies while simultaneously allowing unambiguous somatic variant phasing.

775W

Using liquid biopsies for low frequency variant detection and tissue-of-origin exploration. *K. Cunningham, J. Lenhart, A. Wood, S. Sandhu, C. Schumacher, L. Kurihara, V. Makarov, T. Harkins.* Swift Biosciences, Ann Arbor, MI.

Liquid biopsy assays enable non-invasive profiling of circulating, cell-free DNA (cfDNA) and circulating tumor cell DNA to assist in early-stage diagnosis of disease and monitoring treatment response. Since high sequencing depth is required to profile cfDNA variants, most liquid biopsy assays use targeting to cost-effectively achieve deep coverage of target loci for pathogenic variant detection as low as 1% allelic fraction. An assay that produces uniform, comprehensive coverage from low DNA quantities is critical for obtaining the necessary sensitivity. We developed a liquid biopsy workflow for low frequency variant detection from 10 mL of blood combined with Accel-NGS® library preparation. In addition, we explored whether methylation patterns of cfDNA possess information about the tissue-of-origin. Whole blood samples were collected in Streck Cell-Free DNA BCT® vials from patients with late stage cancer and cfDNA was extracted with the Promega Maxwell RSC. A total of 10 ng cfDNA was used to make an Accel-NGS 2S Hyb library followed by hybridization capture using IDT xGen® Pan Cancer probes. Molecular barcodes were incorporated to label each library molecule uniquely prior to amplification. Sequencing was performed to a minimum of 8000X mean bait coverage. In parallel, 10 ng cfDNA samples were used as input to the Accel-NGS Methyl-Seq Kit to study methylation patterns. Extraction yielded 8-32 ng cfDNA with a size peak of 170 bp and a mean qPCR integrity score of 0.22, characteristic of high quality cfDNA lacking cellular DNA. The Accel-NGS 2S Hyb Kit exhibited 90% conversion with cfDNA, providing highly complex libraries with uniform target coverage (>99% of bases covered >100X). Molecular barcodes enabled removal of PCR duplicates while preserving fragmentation and strand duplicates to maximize coverage. Sensitive and precise detection of variants was achieved down to 0.5% allele frequency. In addition, using whole genome bisulfite sequencing (WGBS) and *a priori* knowledge of differentially methylated regions characteristic of different human tissues, we identified predominant tissue sources of cfDNA in blood. We developed a liquid biopsy workflow for low frequency variant detection from clinically relevant quantities of cfDNA and a WGBS workflow showing the epigenetic status of cfDNA can be used to identify tissue-of-origin. Together, these approaches provide powerful methods for detecting, identifying, and monitoring disease.

776T

Obtaining high quality transcriptome data from formalin-fixed, paraffin-embedded diagnostic prostate tumour specimens. *L.M. FitzGerald[1,2], C.H Jung[3], E.M Wong[4], J. Joo[4], J.A. Gould[5], V. Vasic[5], J.K. Bassett[2], N. O'Callaghan[4], T. Nottle[6], J. Pedersen[6], G.G. Giles[2,7], M.C. Southey[2,4].* 1) Cancer, Genetics and Immunology, Menzies Institute for Medical Research, Hobart, Tasmania, Australia; 2) Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, VIC, 3004; 3) Melbourne Bioinformatics, University of Melbourne, Parkville, VIC, 3010; 4) Department of Pathology, The University of Melbourne, Parkville, VIC, 3010; 5) Monash Health Translation Precinct, Medical Genomics Facility, Hudson Institute of Medical Research, Clayton, VIC, 3168; 6) TissuPath Specialist Pathology, Mount Waverley, VIC, 3149; 7) Centre for Epidemiology and Biostatistics, School of Global and Population Health, The University of Melbourne, Melbourne, VIC, 3010.

Prognostic genomic biomarkers that can be measured at diagnosis to aid choice of treatment options are unavailable for most common cancers. This has been due in part to the poor quality and quantity of available diagnostic specimens and to limitations in genomic technologies. Recent technical advances now provide the ability to perform high density molecular genetic analyses using suboptimal biological specimens. Here we describe the optimisation of a gene expression assay for use with formalin-fixed, paraffin-embedded (FFPE) diagnostic prostate cancer (PrCa) specimens. We applied the Ion AmpliSeq Transcriptome Human Gene Expression Kit (AmpliSeq Kit) to RNA samples extracted from 36 tumour-enriched and 16 adjacent normal tissues (ADJ$_{NT}$) from 37 FFPE PrCa specimens over a series of eight pilot studies, incorporating protocol modifications in studies 2 to 5. Data quality was measured by 1) the total number of mapped reads; 2) the percentage of reads that mapped to AmpliSeq target regions (OnTarget%); 3) the percentage of genes on the AmpliSeq panel that had a read-count ≥10 (TargetsDetected%); and 4) a comparison of the gene read-count distribution of the prostate tissue samples with the median distribution of control cell line-derived RNA samples. Modifications incorporated into pilot study 5 provided high quality gene expression data equivalent to that from control cell line-derived RNA samples. These modifications included the use of freshly cut slides for macrodissection; increased tissue section thickness (8μm); RNA extraction using the RecoverAll Total Nucleic Acid Isolation Kit for FFPE (ThermoFisher); 18 target amplification cycles; and processing six samples per chip. Preliminary Differential Gene Expression (DGE) analysis on TMM normalised data from 15 ADJ$_{NT}$ and 26 prostate tumour samples identified 2,630 differentially expressed genes (FDR <0.01), of which 1,275 were up-regulated and 1,355 were down-regulated. A modified protocol for the AmpliSeq Transcriptome Human Gene Expression Kit can generate high quality, transcriptome data from suboptimal FFPE-derived RNA. This protocol will facilitate the discovery of prognostic biomarkers for cancer by allowing researchers to exploit previously under-utilised diagnostic FFPE specimens.

777F

Digital gene expression from low sample input: Highly multiplexed and robust profiling of formalin-fixed paraffin-embedded (FFPE) and fresh frozen samples from as little as 1 ng of RNA using the nCounter® Platform. *D. Hanson, M. Bailey, L. Dennis, M. Eagan, N. Elliot, G. Meredith, S. Warren, A. White.* NanoString Technologies, Seattle, WA.

Gene expression analysis from tissue biopsies and formalin-fixed paraffin-embedded (FFPE) samples are an industry standard and invaluable to scientists who seek to answer biological questions in the fields of translational and biomarker research. However, one of the major challenges faced by scientists and clinicians is one of sample conservation. Researchers working with samples of low RNA content, biopsies with limited tissue amounts or rare cell types are often faced with the difficult choice of how to run multiple analytical tests and generate as much biologically relevant data as possible from limited available sample material. Furthermore, samples such as FFPE are often degraded or crosslinked due to fixatives used in the tissue archival process. For this reason, it is important for molecular technologies to not only be able to generate quality data but to do so while consuming the least amount of sample possible. NanoString's nCounter System is an automated platform that, by using a digital molecular barcoding technology, enables counting of unique transcripts within a sample. The standard nCounter protocol recommends RNA inputs of 100 ng for successful expression profiling. Here we demonstrate a method using a pre-amplification approach upfront of the nCounter workflow that significantly reduces RNA input requirements 10 to 100-fold depending on sample type. RNA extraction was performed on a sample set containing both F/F (fresh/frozen) and FFPE samples from spleen, brain, lymph node and tumor tissues. Samples were profiled on the nCounter across NanoString gene expression panels using both 100 ng of un-amplified input RNA with the standard workflow, and either 1 ng (F/F) or 10 ng (FFPE) RNA with the Low RNA Input workflow. Profiles were generated from all samples and the overall number of genes detected showed high concordance (r>0.85) between samples prepared using the standard (un-amplified) and Low RNA Input (amplified) methods. The data presented demonstrates successful validation of a Low RNA Input nCounter workflow enabling a 10x-100x reduction in sample input while maintaining robust expression profiling with an increased sensitivity of detection of low expressing genes.

778W

Multiplexed molecular characterization of non-small cell lung cancer PDX models with NanoString's nCounter® Vantage 3D™ DNA:RNA:Protein Solid Tumor assay. *D. Hinerfeld[1], S. Airhart[2], A. White[1], T. Hood[1], C. Bult[2].* 1) Nanostring Technologies, Seattle, WA; 2) The Jackson Laboratory, Bar Harbor, ME.

   Background: While significant progress has been made in the development of targeted and immune-based therapies, non-small cell lung cancer (NSCLC) frequently acquires resistance resulting in disease progression and ultimately death. Decreasing this high mortality rate requires the development of new therapeutic strategies and rational combinations of new and existing drugs. Part and parcel of these efforts is development of biomarkers that can predict both efficacy and potential side effects of these novel NSCLC therapies. Patient Derived Xenografts (PDX) are powerful models for pre-clinical research into therapeutic efficacy and mechanism of action. However, drug studies in PDX models are time consuming and resource intensive. Leveraging these models to gain a deeper understanding of the heterogeneity of the PDX tumors as well as the molecular determinates of therapeutic response or resistance adds significant value beyond determining the change to tumor volume alone. **Methods:** Somatic variants in 25 different clinically actionable genes as well as the expression of 770 cancer-relevant RNAs plus 26 proteins and phosphoproteins were simultaneously analyzed in twelve NSCLC models using the NanoString Vantage 3D research assay. This new multi-analyte technology for mutational status and gene expression was validated against previous DNA and RNAseq analysis of all tumors. **Results:** NanoString SNV and indel calls revealed consistent results when comparing to the DNAseq data for the primary patient tumor and the PDX model. For all twelve tumors, the average correlation between the RNAseq (frozen) and NanoString (FFPE) analysis was R^2 of 0.62 ± 0.05, which is similar to what has been previously reported, particularly considering the difference in sample type. The expression profile of 26 proteins and phospho-protein targets reveals heterogeneity in pathway activation across the different models. **Conclusion:** Even amongst samples with the same driver mutations, there is significant heterogeneity in both RNA and protein expression. This heterogeneity is particularly interesting with respect to two PDX lineages of a single patient tumor where the simultaneous analysis of RNA and protein reveals clues into their observed differences in sensitivity to treatment with erlotinib.

779T

Development of a lysate-based multiplex proteomics platform using nCounter. *J. Lee[1], G. Demirkan[2], G.K. Geiss[2], C.P. Vellano[1], B. Filanoski[2], Y. Lu[1], Z. Ju[3], S. Yu[1], H. Guo[1], L. Bogatzki[2], W. Carter[2], R. Meredith[2], S. Warren[2], Z. Ding[1], A.A. Sahin[4], J.M. Beechem[2], G.B. Mills[1].* 1) Systems biology, UT MD Anderson Cancer Center, Houston, TX; 2) NanoString Technologies,Inc., Seattle, WA; 3) Bioinformatics and computational biology, UT MD Anderson Cancer Center, Houston, TX; 4) Pathology, UT MD Anderson Cancer Center, Houston, TX.

   Comprehensive understanding of the functional proteome has helped characterize many cancers and identified treatment regimens that translate to improved outcomes. Integration of genomics and transcriptomic with functional proteomics will provide the most information for understanding cancer biology and improving patient outcomes. Recent protein measurement technologies have expanded in plexity; however, accurate protein measurement using antibody-based multiplex methods is still challenging due to signal amplification steps, which decreases robustness and accuracy. In this study, we developed multiplex DNA-oligo-labeled antibody panels and associated nitrocellulose or ELISA plate-based protocols for lysate samples based on the digital nCounter technology that assesses single molecules without amplification. Targets encompassed major cancer signaling proteins and phospho-proteins, with an emphasis on breast cancer markers. We implemented a rigorous pipeline to validate barcoded antibodies using Western blotting, ELISA and cross-platform comparison with Reverse Phase Protein Array (RPPA), an orthologous high-throughput proteomics technology. Results from the antibody panel were highly reproducible both in technical and biological replicates ($r = 0.99$ and 0.92 respectively) using lysate samples from multiple cancer cell lines with different perturbations. Importantly there was no evidence for "batch effects" across samples sets caused by non-biological factors ($r = 0.99$). To assess potential clinical utility, we processed lysates from frozen breast cancer patient tumors and demonstrated accurate breast cancer subtyping. Taken together, we show highly quantitative utility for protein measurement in relevant conditions. Importantly, the multiplex proteomics platform is highly suited to the clinical setting due to its robustness and reproducibility. .

780F

Functional validation of pleiotropic suspectibility loci for breast and ovarian cancer using chromosome conformation capture technology.
J. Plummer[1], S. Kar[2], Q. Li[3], M. Freedman[3], P. Pharoah[2], S. Gayther[1]. 1) The Center for Bioinformatics & Functional Genomics, Department of Biomedical Science, Cedars Sinai Medical Center, CA; 2) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 3) Department of Medical Oncology, Dana-Farber Cancer Institute.The Eli and Edythe L. Broad Institute, Cambridge, MA.

Breast and ovarian cancers share common etiologies and similar genetics, such as mutations in *BRCA1/2* are responsible for most multi-case breast/ovarian cancer families. This shared genetic liability is also manifested in terms of common genetic susceptibility variants that suggests widespread pleiotropic effects on both cancer types. Previous GWA and cis-eQTL analyses identified *RCCD1* as a likely candidate gene for both breast and ovarian cancer risk (Cancer Discovery 2016). Chromosome conformation capture (3C) methods are powerful tools to identify physical interactions between target genes and risk-associated SNPs. We have used circularized chromosome conformation capture (4C) assays in both breast and ovarian normal (MCF10A and FT246) and cancer cell lines (BT549, MCF7 and UWB1.289, KURA-MOCHI) to identify interactions between the *RCCD1* promoter and possible candidate causal variants at the 15q26 pleiotropic risk locus. Of interest, rs763280402 and rs774617992 showed 4C interactions with the *RCCD1* promoter in cancer cells. Given regulatory elements are the likely functional targets of most risk-associated alleles, we used regulatory feature annotation (Statehub) to show both variants lie in predicted enhancers of normal breast cells (HMEC), a likely functional target. In fact, ovarian cancer patients who have high RCCD1 expression (P<0.004) show prolonged survival curves (KM Plotter). These findings demonstrate that using an integrated functional genomics to comprehensively evaluate interactions of a given breast and ovarian cancer risk locus can identify candidate susceptibility genes for both cancer types.

781W

Hereditary predisposition to asynchronous bilateral breast cancer: Going beyond BRCA1, BRCA2 and PALB2. *M. Tischkowitz[1], E. Fewings[1], J. Redman[1], M. Goldgraben[1], J. Hadfield[2], P. Concannon[3], J. Bernstein[4], D. Conti[5], A. Larionov[1], The WECARE Consortium.* 1) University of Cambridge, Cambridge, United Kingdom; 2) Genomics Core, Cancer Research UK Cambridge Institute, Cambridge, United Kingdom; 3) Genetics Institute and Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, United States; 4) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, United States; 5) Division of Biostatistics, Department of Preventive Medicine, University of Southern California, Los Angeles, CA, United States.

Background: BRCA1/2-negative asynchronous contralateral breast cancer (CBC) is associated with familial history, suggesting a heritable component in CBC beyond the BRCA1/2 mutations. However, no datasets have yet been generated to study germline variants associated with BRCA1/2-negative CBC. **Methods:** DNA extracted from blood was obtained from 256 CBC and 256 matched unilateral breast cancer (UBC) patients of European ancestry enrolled in the Women's Environmental Cancer and Radiation Epidemiology (WECARE) study. Women with *BRCA1*, *BRCA2* or *PALB2* mutations were excluded. Whole exome sequencing (WES) was performed using Illumina library preparation and sequencing technologies. VCF files were generated following GATK Best Practices workflow. Rare variants were aggregated per gene and statistical associations were assessed using SKAT R package. Trends in crude allelic frequencies (AFs) were evaluated by comparing AFs in CBC, UBC and non-Finish European cases from 1-thousand genomes project. **Results:** 235 CBC and 245 UBC were successfully sequenced to ~60x average depth on targets. Focused analysis in established breast cancer genes suggested that *ATM* and *CHEK2* showed crude AFs trends consistent with CBC association. Small sample and effect sizes precluded exome-wide detection of candidate genes that would pass multiple testing-correction in SKAT analysis. Combining results from the SKAT analysis, the crude AFs assessments and manual prioritising by biological functions allowed for the selection of candidate genes, which could be used for further genotyping in a larger cohort or for experimental validation. **Conclusion:** The study reports a new WES dataset for the identification of candidate genes potentially associated with asynchronous contralateral breast cancer.

782T

Genomic features of gastric cancer patient-derived xenograft (PDX) models. *C. Zhang[1], S-Y. Cho[2], D. Na[2], C. Sung[3], J. Chae[4], W. Kang[1], E. Cerveira[1], M. Ryan[1], J. Lee[5], J. Kang[5], S. Min[5], A. Lee[5], E. Kwak[5], J. Kim[5], B. Choi[5], H. Kim[1], J. Chuang[1], S. Park[6], J. Roh[3], M-S. Cho[6], C-S. Park[3], S. Park[7], Y. Ju[7], Y-S. Suh[8], S-H. Kong[8], H-J. Lee[8,9], H. Park[10], J-I. Kim[4,9,11], H-K. Yang[8,9], C. Lee[1,5].* 1) Jackson Laboratory for Genomic Medicine, Farmington, CT; 2) Ewha Institute of Convergence Medicine, Ewha Womans University Mokdong Hospital, Seoul, Korea; 3) Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 4) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; 5) Department of Life Science, Ewha Womans University, Seoul, Korea; 6) Department of Pathology, Ewha Womans University College of Medicine, Seoul, Korea; 7) Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea; 8) Department of Surgery, Seoul National University College of Medicine, Seoul, Korea; 9) Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea; 10) Department of Biomedical Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, Korea; 11) Medical Research Center, Genomic Medicine Institute (GMI), Seoul National University, Seoul, Korea.

Gastric cancer is the second leading cause of cancer-related deaths worldwide. Approximately 1,000,000 individuals are diagnosed with gastric cancer annually worldwide, resulting in 800,000 deaths. Gastric cancer is a highly heterogeneous disease and novel therapeutic regimens based on the genomic data from gastric cancer are still limited. We have successfully generated 69 patient-derived xenograft (PDX) mouse models from Korean gastric cancer patients using immunodeficient NSG mice. These PDX models cover different types of gastric cancers, including the diffuse, intestinal, and mixed types based on the Lauren classifications. We performed whole-exome sequencing and RNA-seq on 47 primary gastric cancer patient samples and their corresponding PDX tumor samples. Systemic analysis and comparison of the mutational profiles in the primary and PDX samples indicated that the mutational process is preserved throughout tumor growth in the mice, and in most tumors, this is due to the CpG→TpG mutation mechanism. The gene expression patterns were also highly conserved between the primary and corresponding PDX tumors, with expression-based TCGA subtype classifications being maintained. Consistent with the loss of immune-related functions and human stroma in the PDX establishment, expression signatures associated with immune-related and stromal cell function were decreased in the PDX tumors compared with the matched primary tumors. In addition, we observed the development of 27 human EBV⁺-Diffuse large B-cell lymphoma (DLBL) during the engraftment of gastric cancer into the immunodeficient mice and analysis of these tumors are ongoing. The PDX models established in this study will become important resources for studies of tumor heterogeneity, clonal evolution, drug sensitivity and resistance, molecular pathogenesis of EBV-induced DLBL lymphoma as well as the development of novel target therapy for gastric cancer precision medicine.

783F

Comparative analysis between gene expression profile and genomic profile in adrenocortical carcinoma samples. *F.P. Fortes[1,2], L.M. Canto[2], H. Kuasne[2], F.A. Marchi[2], K.C. de Andrade[1,2], S.R. Rogatto[3], M.I. Achatz[1].* 1) Division of Cancer Epidemiology and Genetics- DCEG, National Cancer Institute/ National Institutes of Health (NCI/NIH), Bethesda, MD; 2) International Research Center (CIPE)- AC Camargo Cancer Center, São Paulo, São Paulo, Brazil; 3) Department of Clinical Genetics, Vejle Hospital, SLB, University of Southern Denmark, Vejle - DK.

Li-Fraumeni syndrome (LFS) is a cancer predisposition disorder that increases the risk of developing tumors at an early age as well as multiple primary tumors. Pathogenic germline *TP53* mutations are the underlying cause of LFS. In Southern and Southeastern Brazilian population the syndrome is present in 0.3% due to the presence of a founder mutation, p.R337H *TP53*. This mutation was initially described as tumor-specific, predisposing individuals to adrenocortical carcinoma (ADR). ADR is a rare neoplasm worldwide, but pediatric ADR is 10-15 times more incident in p.R337H mutation carriers. To date, the factors that lead to the development of ADR in some carriers of *TP53* germline mutations while others carriers remain asymptomatic remain unclear. Our hypothesis is that the presence of the p.R337H mutation may not be sufficient for ADR development. The aim of this study was to evaluate the copy number alterations (CNA) of ADR patients who were either positive (N=3) or negative (N=4) for p.R337H mutation, and to compare the results with global gene expression analysis data from our group. Analysis of gains and losses was evaluated by CytoScan HD Array (Affymetrix, Santa Clara, CA, USA). Analysis of global gene expression was evaluated by microarray (4x44K, Agilent Technologies). Genomic alterations that distinguished tumors that were positive for the p.R337H mutation from those that were negative for it were identified using Fisher's exact test, where regions with p <0.05 were regarded as statistically significant. For the gene expression analysis, Student's t-test was used to evaluate differences between p.R337H-positive and p.R337H-negative groups. There were 25 alterations involving chromosomes 1, 10, 11, 12 and 17. Most alterations were located on chromosomes 12 (11/25) and 17 (9/25). A total of 25 genes were identified as having altered transcript expression. Of these, 17 presented gain and increased expression when comparing positive vs. negative cases, and eight presented loss and decreased expression for the same analysis. Despite the small sample size, ADR patients who were positive for the p.R337H mutation presented a lower level of genomic instability than p.R337H-negative. This data suggests that patients with the p.R337H mutation may present a carcinogenic process that is distinct from that of patients without mutations in the *TP53* gene.

784W

Using NGS to detect mutations below 1% allele frequency in circulating cell free DNA and associated tumors. *A. Wood, S. Sandhu, V. Kelchner, J. RoseFigura, J. Lenhart, L. Kurihara, V. Makarov, T. Harkins. Swift Biosciences, Ann Arbor, MI.*

The use of circulating, cell-free DNA (cfDNA) for early detection and monitoring of disease is rapidly growing. This necessitates accurate variant detection at and below 1% allele frequencies due to a low population of tumor DNA within cfDNA samples. Reliable, low-frequency variant detection by next-generation sequencing (NGS) is challenging due to non-specific background noise from PCR and sequencing errors. We have employed molecular identifiers (MIDs) to uniquely label individual DNA molecules prior to amplification, facilitating the distinction of true variants from PCR and sequencing errors. MID incorporation also results in increased data retention by removing PCR duplicates while preserving fragmentation and sister strand duplicates. Here, we have applied MIDs to both our Accel-NGS® 2S whole genome library prep followed by hybridization capture with various cancer panels, and to our Accel-Amplicon™ library prep that uses multiplex PCR ideal for small panels that are amenable to deep sequencing. To validate this technology, we performed low frequency spike-in experiments at 1% and 0.5% with genomic DNA as well as various cfDNA samples. 2S libraries were prepared and hybridization capture was performed with an 800kb pan cancer panel. Libraries were sequenced to greater than 8000x coverage and a consensus sequence was generated with BMFtools. All known variants present at 1% and 0.5% allele frequencies were maintained in the resulting data. Further, true variants were preserved while PCR and sequencing errors were removed, demonstrating improved specificity using MIDs. We also validated variant calling below a 1% allele frequency using the Accel-Amplicon *EGFR* Pathway Panel with MIDs. After validation of our MID technology, libraries were prepared with cfDNA and tumor samples from individuals with ovarian, liver, stomach, and colon cancers. Libraries were sequenced to a minimum of 13,000x coverage, and we determined data retention after de-duplication with and without the use of MIDs. We observed a significant increase in data retention that led to a 2 to 5-fold increase in coverage using MIDs. Variant calling identified pathogenic mutations in all cfDNA samples, including those present in a corresponding tumor sample when available. This study highlights the ability of MID technology to enable variant detection at and below 1% allele frequencies, critical to track known variants and identify novel pathogenic mutations in cfDNA samples.

785T

Functional integration of genomic and transcriptomic data using Strand NGS explains drug resistance in basal cell carcinoma. *S. Kapoor, S. Ramachandrula, S.C. Cherukuri. Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India.*

Understanding a phenomenon as complex as cancer needs a comprehensive understanding and interpretation of the integrative events at genomic, transcriptomic, and proteomic levels. Currently, NGS has brought a major change in the way multifaceted diseases like cancer are perceived and treated. Just understanding and interpreting the data for base level changes or expression level changes is not enough to decode a disease like cancer. Therefore, simultaneous and integrated comprehension of the events is the need of the hour as it can provide an insight into the biological processes operational in complicated diseases, and this would be instrumental in planning and providing personalized cure to the affected individuals. Strand NGS is an integrated NGS data analysis tool that has strong analysis workflows and allows one to perform integrated analysis of various omics data. This tool has analysis and interpretation capability called MOA (Multi-omic analysis) which allows one to integrate results across different experiment types as well as multiple organisms. The visualization of experimental data overlaid on curated pathways or literature derived networks allows one to get an exhaustive picture of the biologically significant happenings. Using publicly available genome and transcriptome NGS data set GSE58377, we demonstrate integrated analysis in Strand NGS in an attempt to explain the Drug resistance in basal cell carcinoma and explore the plausible approaches for treating basal cell carcinomas (BCCs). Using Strand NGS we could find various SNPs and CNVs in genes *TP53*, *PTCH1*, and *SMO* in resistant BCCs. This case study emphasizes on the need of integration of sequence variation data with expression data to understand and interrogate biological problems like cancer that could be helpful in diagnostics, precision medicine, and might lead to discovery of cancer biomarkers. This study highlights the competence of Strand NGS in carrying out such analyses in a single tool with ease and providing clinicians and researchers with fast and accurate results, excellent visualizations, and allowing them to comprehend their data further by carrying out downstream analysis of their NGS data.

786F

DNA repair improves sequencing accuracy in FFPE DNA samples. *F. Stewart, C. Song, P. Liu, L. Chen, L. Ettwiller, E. Dimalanta, T. Davis, T. Evans.* New England Biolabs, Inc., Ipswich, MA.

Targeted cancer therapy based on genomic alterations can be remarkably effective. Currently, cancer genome profiling using next generation sequencing (NGS) is routinely applied in cancer therapies to guide personalized treatment. The accuracy of this profiling directly impacts therapeutic choices and the outcomes of patient care. Although samples of blood and other bodily fluids are being actively explored for early disease diagnosis and treatment monitoring, DNA isolated from FFPE samples is currently the main source for NGS-based genetic profiling. However, sequencing DNA from FFPE samples is challenging due to limited quantities and poor quality, a result of DNA damage incurred during fixation and storage of the tissue sample. We previously showed that false positive variants are abundant and can account for a major fraction of identified somatic variations in publicly available datasets (Chen, L., et al., Science 2017). These false positive variants show signs of mutagenic DNA damage. We further demonstrated that enzymatic DNA repair increases sequencing quality by lowering damage-induced background noise. Therefore, enzymatic DNA repair has the potential to improve sequencing accuracy, avoiding incorrect somatic variant calls and consequently reducing incorrect diagnostic conclusions, especially from damaged FFPE DNA samples. In this study, we have optimized a second generation of FFPE DNA repair enzyme mix (v2) that is specifically designed to work on the most abundant types of DNA damage in FFPE samples. FFPE DNA isolated from tumor samples were treated with or without the DNA repair enzyme mix v2 prior to library preparation. Target enrichment of a panel of 151 cancer genes was performed, and deep sequencing data was obtained for variant analysis. A subset of candidate variants identified by sequencing was further validated using droplet digital PCR (ddPCR) assay. The original FFPE DNA input was also analyzed using ddPCR to confirm the variant calls detected via deep sequencing. Both deep sequencing and ddPCR showed no difference in high frequency variant abundance between the control and repair groups. However, the number of false positive variant calls was reduced in the repair group. ddPCR results from the original FFPE DNA samples confirmed the real positive variant calls in the repair group. Our data demonstrates that DNA repair significantly increases sequencing accuracy without altering the frequency of actual mutations in tumor samples.

787W

Longitudinal integrative omics of rituximab treatment on primary B cells. *L.R.K. Brooks, G.I. Mias.* Michigan State University, East Lansing, MI.

We will present results from our investigation of longitudinal profiling in primary B-cells following their perturbation through treatment with Rituximab(RTX). RTX is a monoclonal antibody anticancer drug, targeting the B cell marker CD20, used to treat different types of blood cancers. This project aims to further investigate RTX's pathway of action, the possible gene interactions precipitated by RTX treatment within B cells, and changes in cell signaling mediated by exosomes following multiple time points after RTX treatment. Our approach employs the integration of transcriptomic and proteomic molecular components derived from treated and untreated cells, and uses the dynamics to classify temporal responses and deduce inter- and intra- cellular interactions over time. Primary B cells cultured in serum free media were treated and followed over time. Following drug treatment, cells were assayed over multiple time points (6 and 24) in the span of 24 hours including the RTX treatment. At each time point, RNA and protein were extracted and prepared for RNA-Sequencing and mass spectrometry proteomics respectively. Additionally, for 6 time points exosomes were extracted from the media (secreted from the cells) and exosomal small RNA was prepared for sequencing. The data were analyzed using our MathIOmica framework to identify temporal trends and statistically significant overrepresentation of KEGG pathways and gene ontology (GO) terms. Our findings indicate activation of genes, showing specific patterns of activation, and corresponding over-representation (p<0.01) in the Nuclear Factor Kappa B (NFκB) signaling pathways, Fc gamma R-mediated phagocytosis and PI3K-AKT pathway, including many genes involved in apoptotic processes, negative regulation of cell proliferation and extracellular exosomes. Overall our time series profiling approach identified known and novel gene and protein expression responses of B cell populations to RTX, and allowed us to construct interaction networks of the different molecular components. The focus on NFκB signaling pathway shows clear response following drug treatment, and offers an example of the method's utility. Finally, our approach is generalizable to any treatment and dynamic monitoring of cells. LRKB is supported by a Michigan State University Mavis Richardson Travel Grant. GIM and research in this investigation were supported by the National Human Genome Research Institute (NHGRI) under grant number R00 HG007065.

788T

Use of Bionano Optical Maps to identify medically-relevant genomic variation. *A.W.C. Pang, J. Lee, S. Chan, W. Andrews, T. Anantharaman, A. Hastie, H. Cao.* Research and Development, BioNano Genomics, San Diego, CA.

Structural variation (SV) detection is fundamental to understand genetic diseases. While karyotyping and conventional molecular detection approaches are robust, they can be manually intensive, biased towards targeted loci, and cannot determine the copy number of long repeats. Bionano Genomics® with its Saphyr™ System offers an unbiased and sensitive method to detect large variants. Bionano's next-generation mapping (NGM) technique, relies on NanoChannel arrays to linearize DNA, image DNA with high resolution and haplotype-aware *de novo* assemble megabase long optical maps. This method can capture over 4,000 insertions and deletions (>1kbp), translocations and inversions in a human sample. We introduce here a software pipeline that analyses Bionano SV calls to identify candidate disease-associated SVs. To do so, the software searches for high-confidence calls, and evaluates quality scores of the assembly, alignments, and SVs detected. To further ascertain relevant candidates, it filters the remaining calls with over 443,000 SVs collected from 144 individuals with no observable disease phenotypes, and compares them with calls from disease-specific databases. Using this pipeline, whose total runtime takes only a few hours, we can efficiently select the few dozens of SVs for further analysis. We ran specimen from individuals with hematological malignancies, applied the candidate-finding pipeline, and identified known rearrangements such as the t(9;11)(q21.3;q23.3) in acute myeloid leukemia AML and t(9;22)(q34.12;q11.23) in chronic myeloid leukemia CML. Moreover, we uncovered novel mutations ranging in size from a small 4.2kbp insertion disrupting an ALL gene *CNOT3* to a large 9.6Mbp deletion at 17q25. Also, we applied NGM on rare and currently undiagnosed diseases. Using a family-based study design, we looked for rare or *de novo* SVs in affected individuals. In one quartet family, we identified a 174kbp heterozygous deletion shared by the two affected children and the carrier father. This deletion overlaps *ATCD3*, a member of the ATP binding cassette gene family. Also, in one affected sample, we detected a rare 364 kbp heterozygous inversion, which was also inherited from the father. Further experimental validation of these candidate SVs and comparison with the phenotypes are needed to establish disease-association. In conclusion, Bionano maps can significantly aid the discovery of functional SVs, and improve our understanding of the mechanisms of diseases.

789F

Familially-inherited fusion genes as a new-class of cancer predisposition genes. *D. Zhuo.* SplicingCodes.com, Biotailor Inc, Palmetto Bay, FL.

Genetic predisposition to cancer has been well documented for several centuries and has initially observed via unusual cancer occurrences of familial clusterings. Studying cancer-prone families has lead to identify over 100 Mendelian cancer predisposition genes. However, All these CPGs are derived from known genes carrying point mutations and none of them are derived from gene fusion. Recently, by systematic analysis of RNA-seq data of glioblastoma, prostate cancer, lung cancer, breast cancer, multiple meyloma, and lymphoma from different regions of the World, we have found that *KANSARL* (*KANSL1-ARL17A*) fusion transcripts are detected in are rarely detected in the tumors of individuals from Asia or Africa. In contrast, they exist in 30 - 52% of the tumors from North Americans cancer patients. Further analyses have revealed that *KANSARL* is a familially-inherited fusion gene is specific to 28.9% of the population of European ancestry origin. Therefore, we have confirmed that *KANSARL* is the first cancer predisposition fusion gene associated with genetic backgrounds of European ancestry origin. Our"Big Data" approaches, we have identified hundreds of thousands fusion transcripts generated by copy number variations (CNVs), inversions, duplications, deletions and read-through. Since many of these genetic alternations are familially-inherited, the familially-inherited fusion genes may be much more common and widespread. Familially-inherited fusion genes contribute genetic predisposition of cancer and other human complex diseases. Large-scale and systematic validations of these potential familially-inherited fusion transcripts will be discussed.

790W

The identification of biomarkers for EGFR-TKI-induced interstitial lung disease through whole genome sequencing analysis. *H. Zembutsu[1,2], C. Udagawa[1,2], H. Horinouchi[3], Y. Ohe[3].* 1) Cancer Institute, Tokyo, TOKYO, Japan; 2) National Cancer Center Research Institute; 3) National Cancer Center Hospital.

The drug-induced interstitial lung disease (ILD) are one of the critical issues of chemotherapy for patients with cancer because the mortality of this adverse event is extremely high. The incidence of EGFR-TKI-induced lung injury is more common in Japan (3.98%) than in many other countries (0.3% in USA), and the EGFR-TKI-associated ILD mortality rate is reported to be about 33%-50%. Hence, the predictive marker for EGFR-TKI-induced ILD is expected to be developed to prevent the tragedy for the patients with cancer, however, it is not yet available. Until today, no association study has identified genetic marker for EGFR-TKI-induced ILD. In this study, we carried out whole-genome sequencing (WGS) to identify not only common-variants but also rare-variants related to EGFR-TKI-induced ILD using 27 cases who had been diagnosed as having EGFR-TKI-induced ILD from clinical information and CT image by radiologist or respiratory oncologist independently. More than four million single nucleotide variants (SNVs) and >1 million insertion-deletion (INDEL) were called in each sample after quality check. We perform association study between these genetic variants in 27 EGFR-TKI-induced ILD cases and those in general population (control; 1000 Genomes Project data). In this association study, > 9 million variants called by the WGS were tested by Fisher's exact test for the three genetic models (allelic, dominant and recessive mode). More than 29,000 variants showed minimum adjusted $P < 0.1$ in the three modes. Four INDELs (2 insertions and 2 deletions) showed strong association with EGFR-TKI-induced ILD ($P < 1.92 \times 10^{-50}$). These findings suggest that these genetic variants may contribute to variation in development of EGFR-TKI-induced ILD. Validation study of these findings may allow for the identification of patients at increased risk of EGFR-TKI-induced ILD and inform the use of an alternative to EGFR-TKI and/or the clinical management of this toxicity.

791T

Novel sequencing adapters resolve index-hopping with unique, dual-matched barcoding and enable low frequency mutation detection with consensus analysis. *M. Light[1], M. McNeill[2], K. Lai[1], K. Giorda[1], K. Shah[1], S. Groenewold[2], M. Jarosz[1].* 1) Research and Development, Integrated DNA Technologies, Redwood City, CA; 2) Research and Development, Integrated DNA Technologies, Coralville, IA.

Rapid growth in next generation sequencing throughput is driving demand for higher sample multiplexing and enabling applications that require deeper sequencing. Historically, combinatorial indexing has been used to increase multiplexing capacity during hybridization capture and sequencing using a minimal number of sample barcodes (8-i5 indices x 12-i7 indices for 96 samples). This combinatorial approach is vulnerable to sample cross-talk, where a single index misassignment causes a read from one sample to be assigned to another sample. There are many sources of sample cross-talk, such as adapter cross-contamination, sequencing or PCR errors, and index hopping during multiplexed capture or cluster amplification. A set of adapters with unique, dual-matched sample barcodes have been developed to filter out mis-assigned barcodes. Further, molecular tags (UMIs) have been incorporated for *in silico* error correction. This combination of dual-matched indices and UMIs enables highly accurate low-frequency variant detection. Libraries were made using custom adapters (described above) and commercially available library construction kits, and enriched using the IDT xGen® AML Panel. To highlight the utility of dual-matched indices, index hopping was measured for capture pools of 1, 4, 8, and 16 libraries using the frequency of mismatched index 1 and index 2 sequences. The baseline rate of barcode mismatches was low for single-plex captures (0.09%) but increased with higher multiplexing (up to 0.39% for 16 samples). Next, consensus analysis error correction was demonstrated using a 1% mixture of Coriell DNA samples NA12878/NA24385 (25 ng input, 0.5% minimum allele frequency) to model low-frequency mutations. A 75-kb custom xGen Lockdown® Panel was used for target enrichment, and libraries were sequenced on the Illumina MiSeq® sequencer. Sensitivity and positive predictive value (PPV) were assessed across the Genome in a Bottle high-confidence region (35 kb) with a variant calling threshold of 0.2%. UMI consensus analysis combined with mutation-specific thresholds for oxidative damage improved the PPV for low-frequency (<1%) variants from 28% to 96% while retaining a sensitivity of 87% for true positives. This consensus approach was then used demonstrate error correction in formalin-fixed, paraffin-embedded (FFPE) and cell-free derived tumor DNA.

792F

Highly efficient double-stranded molecular tagging empowers improved accuracy of ultra-low frequency mutation detection. *J. Wang, K. Lai, K. Giorda, M. Jarosz.* Integrated DNA Technologies, Redwood City, CA.

Molecular diagnostics for precision medicine are becoming a reality due to rapid advances in next generation sequencing. Yet the development of non-invasive methods to detect and monitor tumors during treatment using liquid biopsies spotlights the challenges remaining for ultra low-frequency (<17%) mutation detection of circulating tumor DNA. These subclonal mutations are difficult to distinguish from low-level errors introduced during sample preparation, target enrichment, and sequencing. Molecular barcoding strategies, which tag both strands on input DNA molecules, allow for exquisite error correction, but these adapters are challenging to manufacture and their low library conversion efficiency have limited their utility for clinical research. IDT has developed unique duplex adapters that are compatible with standard library preparation methods but offer significantly higher library conversion for low input samples. Mixtures of Coriell DNA samples NA12878 and NA24385 were used to model low-frequency mutations (1.0–0.1% minimum alternative allele frequencies). Novel complementary unique molecular index (UMI) adapters were used with standard library conversion kits to generate libraries. Target enrichment libraries were prepared with a custom 75 kb IDT xGen® Lockdown® panel and sequenced on MiSeq®, NextSeq® and/or HiSeq® sequencing platforms (Illumina). Using 25 ng input, library conversion efficiency was comparable for standard and UMI adapters, with ~1,400X and ~1,500X de-duplicated coverage, respectively. Our analysis pipeline generated consensus reads based on duplicate read base agreement, sequencer base quality, and top and bottom strand observations. Consensus reads were used for variant calling with a 0% threshold with VarDict. Sensitivity (true positive rate) and positive predictive value (PPV) were measured across the genome in a bottle high confidence region (35 kb portion of custom panel). Standard adapters provided 90% sensitivity for variants >0.2% but called >5,000 false positive variants for a PPV of just 5%. In contrast, duplex UMI adapters and consensus analysis had excellent sensitivity of >90% but contained no false positives for 100% PPV. Duplex UMI adapters virtually eliminate false positive errors. Unparalleled accuracy for ultra low-frequency variant calls at 0.1% allele frequency combined with highly efficient library conversion will provide robust tools for more sensitive diagnostics and improved patient care.

793W

New methods for high-throughput nucleic sequencing and diagnostics using a thermostable group II intron reverse transcriptase (TGIRT). *C.D. Wu, Y. Qin, J. Yao, R.M. Nottingham, A.M. Lambowitz.* Institute for Molecular and Cellular Biology, Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX.

We have developed new methods for high-throughput RNA and DNA sequencing based on the use of thermostable group II intron-encoded reverse transcriptases (TGIRTs). TGIRT enzymes have higher fidelity and processivity than commonly used retroviral reverse transcriptases, along with a novel template-switching activity that enables attachment of sequencing adapters to nucleic acid template sequences without tailing or RNA ligation. The latter activity enables RNA-seq library construction from small amounts of degraded RNA samples, such as FFPE tumor slices or cell-free (cf) RNAs in human plasma, in <2 h. Validation of the method using fragmented human reference RNAs with ERCC spike-ins demonstrated advantages compared to the widely used TruSeq v3 method, including higher strand-specificity, more uniform 5'- to 3'-gene coverage, and detection of more splice junctions, particularly near the 5' ends of genes, even from fragmented RNAs. Importantly, TGIRT-seq enables the quantitative profiling of small non-coding (nc) RNAs in the same RNA-seq as protein-coding and long ncRNAs and gives full-length, end-to-end reads of tRNAs and other structured small ncRNAs, neither of which is possible with other RNA-seq methods. TGIRT-seq of cfRNAs in human plasma revealed RNA fragments derived from protein-coding genes and lincRNAs, as well as tRNAs, Y RNAs and most other classes of structured small ncRNAs, with many tRNAs being full-length transcripts rather than fragments as reported previously. In a separate study, TGIRT-seq of RNAs present in highly purified HEK-239T cell exosomes in collaboration with the Schekman lab (UC Berkeley) showed that the predominant membrane-encapsulated RNA cargos are full-length tRNAs and other small ncRNAs, along with smaller amounts of spliced mRNAs, which can vary with cell type. Finally, TGIRT single-stranded (ss) DNA-seq of cell-free DNA in human plasma enabled analysis of (i) nucleosome positioning, (ii) transcription factor occupancy, (iii) DNA methylation sites, and (iv) tissues-of-origins comparably to conventional ssDNA-seq methods, but with a more streamlined workflow that enables precise mapping of DNA ends. Current efforts focus on the use of TGIRT-seq for analysis of patient primary tumor tissues, PBMCs, and plasma samples for diagnostic applications.

794T

Cell cycle specific copy number profiling from parallel single cell genomics and transcriptomics. *R. Rahbari[1], M. Teng[1], I. Macaulay[2], P. Kumar[3], T. Voet[1,3].* 1) WellcomeWellcome Trust Sanger Institue, Hinxton, CB10 1SA, UK Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) 2Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK; 3) Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, Belgium.

Current large-scale single-cell genome sequencing projects have delivered novel insight in intra-tumour subclonal architectures, acquired somatic mutations in normal tissue as well as in diseases such as neurological disorders. However, in these studies, the genetic composition of the cells is often studied without considering confounding biological factors imposed by the cell cycle status. In this study, we have utilised the powerful parallel Genomic and transcriptomic single cell sequencing method (G&T-seq1) to build a novel approach for cell-cycle specific copy number profiling. We demonstrated that DNA-replication perturbs conventional single-cell DNA copy number analysis, leading to the detection of focal read depth differences across the genome that may be interpreted as false structural DNA imbalances, but in fact oscillate according to the DNA replication program. Following single-cell G&T-seq, we identified the cell state of each single cell based on their transcriptomic profile. This cell cycle state classification then informed our single-cell DNA copy number analysis pipeline in order to normalise DNA read depth signals of S-phase cells against those from G1-phase cells, exposing the early and late DNA replication domains of the S-phase cell. In addition, comparing the DNA copy number profiles of the cells in "early" S-phase to those in "late" S-phase, revealed variation in replication timing between cells of the same type, which may be due to stochastic origin firing in individual cells. Further, this method can help in better understanding of links between DNA replication timing and other cellular processes, including transcription, DNA-mutational burden, genome instability and the aetiology of genetic disease. Key words: Single cell, cell type, replication timing, CNV profiling Reference: 1. Macaulay, I.C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods* **12**, 519-22 (2015).

795F

A novel NGS target enrichment technology: Improved speed, selectivity, and uniformity. *J. Pel.* Boreal Genomics, Vancouver, BC, Canada.

Targeted Next Generation Sequencing (NGS) has become common practice in many research and clinical applications, as a faster, cheaper, and often simpler approach to equivalent depth whole-genome or whole-exome sequencing. Target capture methods are especially useful when covering large regions beyond the capacity of multiplexed PCR, such as multi-gene liquid biopsy panels. Current target capture methods, however, typically have long, multi-day workflows that limit their utility in a clinical setting. In addition, small panels (below ~100kb), typically have high off-target capture rates, resulting in an increased cost of sequencing to reach a desired depth of coverage. We have developed a novel library construction method that replaces typical multi-day 'PCR-capture-PCR' workflows with a single 'PCR/capture' step. This approach, which uses physically linked probes and primers, reduces the time and complexity of the workflow, eliminating long capture and clean-up steps and enabling the capture to be completed in less than a day with minimal hands on time. Additionally, PCR and capture are combined in such a way that off-target DNA is reduced due to multiple capture stages, reducing the cost of sequencing. Performance of multi-gene cancer panels using cfDNA is presented, along with variant detection and estimates of cost reduction.

796W

Cryptic forms of mutant splicing detected by cBROCA. *S. Casadei[1], S. Gulsuner[1], M.K. Lee[1], J.B. Mandell[1], B.H. Shirts[3], C.C. Pritchard[3], M.-C. King[1,2], T. Walsh[1].* 1) Medical Genetics Dept, University of Washington, Seattle, WA; 2) Genome Sciences Dept, University of Washington, Seattle, WA; 3) Laboratory Medicine Dept, University of Washington, Seattle, WA.

Mutations that alter splicing most frequently lead to exon skipping or introduction of new splice sites, but can also result in intron retention, activation of pseudoexons, or degeneration of exonic sequence. *In silico* tools flag variants that may alter splicing, but are less able to predict sequences of mutant transcripts or their relative abundance. To detect and characterize unconventional mutant transcripts, we used cBROCA, our targeted RNA sequencing approach, to sequence cDNA from patient LCLs for 41 cancer genes. Because cBROCA targets the entire locus of each gene, all transcripts are detected and their relative abundance can be determined. In families severely affected with breast or ovarian or colon cancer, we quite frequently encounter examples of unusual consequences of splice mutations. For example, *BRCA1* c.4185(+1) insAdel21 was predicted to disrupt a canonical splice site and to lead to skipping of exon 12, which is small, inframe, and expendable. However, analysis by cBROCA indicates that most transcripts from this mutant allele retain the 402bp intron 11, which includes multiple truncation codons. A second example is *ATM* c.5674(+1)G>T, which leads to both skipping of a small inframe exon and retention of 1645bp intron 36 with multiple truncation codons. A very different example is *APC* c.532(-1000)delGT, which was revealed by cBROCA analysis to disrupt an intronic splicing silencer, leading to exonification of 165bp of previously intronic DNA and a premature stop. In our experience, RNA sequencing using cBROCA enables discovery and quantitative characterization of splice mutations of all classes, including relative abundance of multiple different transcripts from the same mutant allele. Supported by the Komen Foundation, by NIH R35CA197458, and by NIH R01CA175716.

797T

Exome sequencing reveals a novel germline gain-of-function *EGFR* mutation in a young adult with bilateral adrenocortical carcinoma. *S. Akhavanfard[1,2], L. Yehia[1,3], I. Smith[1], T. Romigh[1], C. Eng[1,2,4,5,6].* 1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH; 2) Department of Molecular Medicine, Lerner College of Medicine, Case Western Reserve University School of Medicine, Cleveland, OH; 3) Department of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH; 4) Taussig Cancer Institute, Cleveland Clinic Foundation, Cleveland, OH; 5) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 6) CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

Adrenocortical carcinoma (ACC) is a rare endocrine tumor with poor overall prognosis and slight predilection for females. Five-year survival is approximately 38% if the tumor is confined to the adrenal gland and 7% in the presence of distant metastasis. In children, ACC is commonly associated with inherited cancer syndromes, predominantly Li-Fraumeni syndrome, with 50-80% of childhood ACC harboring *TP53* germline mutations. ACC in adolescents and young adults are rarely due to germline *TP53* mutations. Other less common germline mutated genes in ACC include *IGF2*, *PRKAR1A* and *MEN1*. The majority of sporadic ACC tumors occur in adults and their somatic mutations occur mainly in the same genes as germline mutations. We are interested in the subset of patients with clinical features of inherited ACC but who are wild-type for all known germline mutated genes. As part of a larger prospective case-series of the whole-exome sequencing of fifty children, adolescents and young adults with rare tumors, we found a novel missense heterozygous variant in the gene encoding C-terminal tail of epidermal growth factor receptor (*EGFR* [OMIM 131550]; p.Asp1080Asn) in a 21-year-old female with bilateral ACC. This patient was wild-type for known ACC germline mutations including in *TP53*. We functionally confirmed that *EGFR* p.Asp1080Asn variant is an activating gain-of-function mutation which causes increased phosphorylation of EGFR at Tyr1045 position. Increased phosphorylation of ERK1/2 was also observed, confirming the activation of downstream signaling pathways by this mutation. While our observation should be validated through a larger series, it suggests that *EGFR* can be an underlying germline predisposition factor for ACC, especially in young adults. Importantly, identifying a targetable gene for ACC has the potential to improve molecular diagnosis and treatment of this disease, which is known to have limited therapeutic options.

798F

Trans-eQTLs in prostate cancer risk. *M. Bicak, M. Middha, R.J. Klein.* Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai Hospital, New York, NY.

   GWAS successfully identified numerous prostate cancer risk SNPs to date. However we still lack understanding of how such SNPs function to alter an individual's risk of prostate cancer, which is further hindered by insufficient sample sizes. We conducted eQTL association analysis on tumor and normal datasets from multiple cohorts to test if prostate cancer risk SNPs are correlated with gene expression changes in nearby and remote genes by using linear regression. Meta-analysis was applied to resulting cis and trans-eQTLs to increase the power to detect risk loci that are shared among different cohorts. As a result, previously unreported cis-eQTLs were found, such as NOTCH4 and rs30967026. Furthermore, 66 trans-eQTLs were found at an FDR of 5%, for which some SNPs seem to be trans-eQTLs for many genes, suggesting to be prostate cancer associated trans-eQTL hotspots. These include: (i) SNPs that are at loci for prostate-secreted proteins, such as rs10993994 (MSMB) and rs17632542 (KLK3/PSA), as well as (ii) SNPs that are at loci with transcription factors, such as rs12653946 which is confirmed as cis-eQTL with IRX4; rs1512268 which is near NKX3-1, a transcription factor known to play a role in prostate cancer; and rs339331 which has been shown to associate with RFX6 activity. Ongoing work is performing mediation testing and causal inference analysis to predict if the expression changes of the trans-eQTLs are mediated through cis-eQTLs.

799W

Interrogating key RECQL4 related genomic and epigenomic alterations in osteosarcoma. *H. Horn, P. Meltzer.* Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD.

   Osteosarcoma is the most common type of bone cancer (or malignancy), affecting children and adolescents. One of the key hallmarks of OS is the high rate of structural and copy number variation. The DNA helicase RECQL4 is required for normal osteoblast formation and mutations in RECQL4 have been associated with osteosarcoma pathogenesis. In this study, we investigate the underlying mechanisms of RECQL4 deficiency by detecting genomic and epigenomic alterations in two OS cell lines; one in-house RECQL4 deficient and one highly rearranged osteosarcoma cell line. Both cell lines were sequenced using 10x linked-read, whole–genome and exome sequencing technologies. The combination of multiple sequencing platforms allowed us to accurately discover and validate a comprehensive catalogue of RECQL4 related rearrangements. One step further, genomic information was integrated with methylation ChIP-seq data towards detection of enhancer associated epigenomic rearrangements. Our holistic approach of combining multiplatform genome sequencing together with epigenomic information has provided us a comprehensive characterization of RECQL4 related rearrangements with possible clinical impact in osteosarcoma.

800T

Ion AmpliSeq™ TERT promoter sequencing. *J.M. Kilzer, S. Wang, M. Manivannan, A. Broomer, A. Kraltcheva, D. Mazur, C. Van Loy, M. Andersen.* Thermo Fisher Scientific, Carlsbad, CA.

TERT promoter mutations are highly recurrent in over 50 cancer types. Two independent hotspot mutations, TERT C228T and C250T, in this region have been identified as the most common point mutations in several tumor types. Studies suggest that identification of TERT promoter mutations might be used as potential clinical biomarkers for predicting survival rates for some cancers. However, this region has been shown to be challenging to sequence for both forward and reverse strands. Herein, we describe the design optimization of a TERT promoter assay for use in highly multiplexed target amplification reactions to generate libraries for analysis with the Ion Torrent™ NGS platform. In this study, 16 amplicon primer designs of sizes ranging from 125-200bp were generated to target the TERT promoter region containing two known hotspot mutation sites. These designs were combined with larger Ion AmpliSeq™ DNA research panels to generate highly multiplexed libraries. Libraries were used for template preparation on the Ion Chef™ Instrument and subsequently sequenced using the Ion S5™ System. Multiple options were tested to determine optimal conditions for TERT assay performance, including changes to templating reagents, multiple chip types, and changes in nucleotide flow order during sequencing. The resulting sequencing data was used to evaluate coverage of known common point mutations by each design for each condition tested. We generated several primer designs targeting common point mutations in the TERT promoter region that resulted in improved amplicon coverage using the Ion S5™ System. Sequenced strand bias ranged from 99% forward reads for poorest performers to as low as 57% for best performers. End-to-end read through of reverse reads was significantly improved from as low as 0% up to 53% in two of the 16 amplicons tested. Comparison of available Ion system options demonstrated a large range in performance, though trends were similar between amplicons across all platforms. Optimal results were generated using Ion 530 chips in combination with modifications to the Ion Chef™ template preparation chemistry and an alternative nucleotide flow order for all amplicons. In conclusion, we were able to generate and identify new amplicon designs that demonstrated a significant improvement over prior results. In combination with optimized Ion system conditions, these amplicons resulted in 62% and 60% forward reads and 55% and 51% reverse reads reading end-to-end.

801F

The role of antioxidants in the context of carcinogen induced chromosome aberrations. *Y.C. Li, Z.C Hung.* Department of Biomedical Sciences, Chung Shan Medical University, Taichung, Taiwan.

As we know generally, the antioxidant is considered as an agent for cancer prevention because it would scavenge the ROS (reactive oxygen species) which cause DNA damage and turn to initiate the development of cancer. Several studies showed that chewing betal nut were associated with the high risk of oral cancers. Arecoline is a major alkaloid in betal nut and suggested as a probably carcinogen because it had been evidenced to cause DNA damage, cell death, and ROS. However, it is unclear whether scavenging ROS by antioxidants could prevent carcinogen itself inducing DNA damage. In this study, we would establish the kinetic relation of arecoline-induced toxicities including DNA damage, cell death, and ROS and to further understand the role of antioxidants in the context of arecoline induced DNA damage. We analyzed the chromosome aberrations (CAs), ROS and cell death under CHO-K1 cells treated with arecoline alone or combination with antioxidants in different concentration and duration. We observed arecoline cause CAs in four stages as dose- and time-dependent manner. At the beginning, arecoline would directly cause primary CAs. Secondary CAs occurred following time and can be inhibited by GSH. Once CAs over a threshold, CAs would induce ROS and then ROS would cause the tertiary CAs and cell cycle arrest. After CAs and ROS accumulated a certain amount, the cell death initiated and then caused the fourth stage of CAs. In our study, we also found that antioxidants, like catalase, inhibited the ROS induced tertiary CAs via scavenging ROS but neither primary nor secondary CAs. Taken together, it suggested that ROS was induced under CAs over a threshold and then cause cell death to prevent abnormal cells with CAs alive. Moreover, the antioxidant can only prevent ROS induced CAs, but not arecoline (a carcinogen) itself induced CAs. In turn, the cells with CAs are survival at antioxidants treatment. The survival aberrant cells would be a high risk factor of carcinogenesis. Therefore, it should be more carefully evaluated for the use of antioxidant in cancer prevention.

802W

HPV16 integrated genomic and molecular characterization of cervical cancer in Guatemala. *H. Lou[1], J. Boland[3], W. Zhou[3], J. Mitchell[3], D. Roberson[3], L. Garland[1], M. Herrera[4], D. Wells[1], J. Troyer[1], F. Pinto[2], S. Bass[3], X. Zhang[3], M. Yeager[3], E. Gharzouzi[2], M. Dean[4].* 1) LTG-DCEG, NCI-Leidos Biomedical Research, Inc., Gaithersburg, MD. USA; 2) Instituto de Cancerologia, Guatemala City, Guatemala; 3) CGRL-DCEG, NCI-Leidos Biomedical Research, Inc., Gaithersburg, MD. USA; 4) LTG-DCEG, NCI, Gaithersburg, MD USA.

Viral integration into the host genome is a characteristic genetic event in cervical carcinogenesis. HPV integration occurs into different genomic locations within the vicinity of important cancer related genes, in some tumors. Here, we compared HPV16 integration frequencies in 96 Guatemalan cervical cancer patients by capture of HPV DNA on Nimblegen array and DNA sequencing. A total of 65 % (62/96) of the cervical cancer had detectable HPV16 integration. HPV16 can be divided into four main subtypes including A, B, C and D, differing in the high conserved L1 sequence by less than 10%. HPV16 subtype A showed a higher rate of integration than D2 and D3. HPV16 and its subtype A subjects have a significantly younger age in integrated cervical cancer patients. Patients with HPV16 subtype D2 are younger than subtype A. The analysis of the HPV integrated cervical cancer patients and *PIK3CA* mutations demonstrate that there is a significantly older age in HPV16 integrated patients with *PIK3CA* mutations as compared to without *PIK3CA* mutations. HPV integration sites were identified in 62 patients. At least 13 genes including *GAPDH*, *LRP1B*, *MACROD2*, *NAALADL2*, *GLI2*, *WWOX*, *BCL11B*, *INPP4B*, *CSMD1*, *PGAP3-ERBB*, *PTPN13* and *FHIT* were reported in previous studies. In additional, we found HPV16 integrated into the promoter of host *IRF4* (interferon regulatory factor4) that plays an important role in the regulation of the interferon response to virus infection and is as an oncogenic biomarker for hematological malignancies. Our data indicated that the age, specific HPV16 subtype and somatic *PIK3CA* mutations are associated with viral integration in HPV16 positive cervical cancers and the genes at the specific sits of viral integration may play important roles in cervical cancer progression and metastasis.

803T

Effects of 744ins20 - ter240 *BRCA1* mutation on breast/ovarian carcinogenesis and role of curcumin in telomerase inhibition. *M. Pongsavee.* Department of Medical Technology, Faculty of Allied Health Sciences, Thammasat University, Patumthani, Thailand.

Breast cancer is correlated with high estrogen level. *BRCA1* gene mutation increases risk of hereditary breast/ovarian carcinogenesis. Telomerase can protect telomere shortening. Curcumin is a yellow component in Curcuma longa Linn and has antioxidant properties. Study the effect of 744ins20 - ter240 *BRCA1* mutation at exon 10 about DNA repair function by MTT dye reduction and study about Telomerase inhibition caused by curcumin with Telomerase activity assay and Trypan blue exclusion assay were done. The results showed that this mutation caused DNA repair defect ($P<0.05$) and curcumin could inhibit telomerase function and affected cancer cell progression. 744ins20 - ter240 *BRCA1* mutation involved in DNA repair defect and curcumin could inhibit telomerase function. .

804F

Rhesus macaques with mutations in MLH1 and MSH6 develop Lynch syndrome colorectal cancers. *M. Raveendran[1], R. Alan Harris[1], B. Dray[2], P. Gillespie[2,3], D. Brammer[2,4], S. Gray[2], L. Williams[2], D. Muzny[1], R. Gibbs[1], C. Abee[2], J. Rogers[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) MD Anderson Cancer Center, ME Keeling Center for Comparative Medicine, Bastrop, TX; 3) Eli Lilly & Co, Indianapolis, IN; 4) University of Houston, Houston, TX.

Colorectal cancer (CRC) is among the most common forms of human cancer. Lynch syndrome is a hereditary form of CRC caused by mutations in any of four genes involved in DNA mismatch repair (*MLH1*, *MSH2*, *MSH6* and *PMS2*). In humans, lifetime risk of both colorectal and endometrial cancer is dramatically increased in heterozygotes with these mutations. We identified a series of genealogically related rhesus macaques (*Macaca mulatta*) exhibiting CRC near the ileocecal junction or in the ascending colon, proximal to the hepatic flexure, as in human Lynch syndrome. Histopathological exams of resected tumors found invasive colonic adenocarcinomas with desmoplastic features. We performed a series of genetic analyses to test the hypothesis that this is a model of Lynch syndrome. Sanger sequencing of *MLH1* in affected animals found several carrying a 2-base pair deletion in the *MLH1* promoter. In vitro analyses show this promoter mutation significantly reduces *MLH1* expression in human SW480 cell lines. Whole genome sequencing of 16 affected macaques identified a premature stop codon in *MLH1* in 3 individuals carrying the 2 bp deletion. Other affected animals not carrying *MLH1* mutations are heterozygous for a missense mutation in *MSH6* that is scored by CADD as probably damaging. Ten of the 16 CRC affected animals we sequenced carry identifiable damaging mutations in *MLH1*, *MSH6* or both, establishing this as a genetic model of Lynch syndrome, with equivalent cancers attributable to spontaneous mutations in genes that cause Lynch syndrome in humans. These studies are part of a larger program designed to discover new disease models through whole genome sequencing in rhesus macaques. Five hundred thirty macaques have been sequenced using Illumina. Genotype calling employs GATK with annotation through Ensembl VEP. We have identified 72,746,387 unique SNVs. The average SNVs per macaque individual is 9,476,124, substantially more than in humans. Predicted variants include 8,556 stop codons gained, 1,346 stops lost, 340,104 missense and 65,707 splicing variants, as well as 17,286,051 small indels (1 to 60bp). Rhesus macaques provide many important animal models of human disease. The documentation of Lynch syndrome in this species underscores the value of naturally occurring functional genetic variation as a resource for development of genetically equivalent translational models.

805W

Detection of viral sequences and integration sites in HPV-positive (HPV+) recurrent/metastatic head and neck cancer (RMHNC) patients. *D. Thach, G. Sulkowski, M. Ahmad, R. Iyer, J. Deeken.* Inova Translational Medicine Institute, Inova, Fairfax, VA.

Background: The cause for the rising incidence of HPV+ HNC remains unexplained, and while patients can be cured by combined modality treatment, more than 1 in 4 patients will suffer recurrence. Understanding the basis for recurrence risk with respect to virus sublineages and their interactions with the host genome could be critical to personalizing initial treatment. Previously, we have genotyped virus sublineages via SNP markers and also found an unexpected high rate of *TP53* mutations in our cohort. Here, we detect viral lineages via whole genome sequencing, and searched for their sites of integration into the host genome. **Methods:** FFPE DNA from resected recurrent tumors of HPV+ RMHNC patients were analyzed for viral sequences and integration into the host genome. The Accel-NGS 2S kit was used to prepare whole tumor libraries that were sequenced on the Illumina Hiseq 2500. VirusFinder2.0 was used to detect viral sequences by searching non-human reads against a database of known virus genomes. Detected viruses were further analyzed for chimeric reads supporting their sites of integration into the host genome. **Results:** Recurrent tumors from 3 patients were analyzed. High quality HPV16 contigs of sizes 7614, 8445, and 1269 were detected in all 3 samples, and the contigs were more than 99% identical to known HPV isolates. The spiked-in positive control, phiX DNA, was detected as well. No HPV contigs were detected in a negative control tumor; instead a 1399 contig with 99.9% identity to HHV-4 was detected. In all samples and controls, low quality contigs of unexpected viruses were repeatedly detected, suggesting background sequence homologies or reagent contamination. Novel HPV16 integration sites was found at *3q26.33*, disrupting *LINC01206*, a long intergenic non-protein coding RNA of unknown function in the host. The breakpoints in the HPV16 genome was in the E1, L1, and upstream of E6 genes. **Conclusions:** We have detected and sequenced HPV genomes in recurrent tumors, and have found novel integration sites disrupting a host lincRNA, which are unlikely to be inhibited by new-targeted therapies. More research is needed to identify viral-tumor integration and mutational profiles based on viral integration to optimize precision medicine strategies.

806T

An integrative detection and analysis of structural variation in cancer genomes. *J. Xu[1], J. Dixon[2], V. Dileep[3], Y. Zhan[4], F. Song[1], V. Le[2], G. Yardim-ci[5], A. Charkraborty[6], D. Bann[1], Y. Wang[1], C. Ernst[9], S. Hadjur[8], D. Odom[9], J. Stamatoyannopoulos[7], J. Broach[1], R. Hardison[1], F. Ay[6], W. Noble[5], J. Dekker[10], D. Gilbert[3], F. Yue[1].* 1) Penn State University, Hershey, PA; 2) Salk Institute for Biological Studies; 3) Florida State University; 4) University of Massachusetts; 5) University of Washington; 6) La Jolla Institute for Allergy and Immunology; 7) Altius institute for Biomedical Sciences; 8) University College London; 9) University of Cambridge; 10) University of Massachusetts Medical School.

Structural variants can contribute to oncogenesis through a variety of mechanisms, yet, despite their importance, the identification of structural variants in cancer genomes remains challenging. Here, we present an integrative framework for comprehensively identifying structural variation in cancer genomes. For the first time, we apply next-generation optical mapping, high-throughput chromosome conformation capture (Hi-C), and whole genome sequencing to systematically detect SVs in a variety of cancer cells. Using this approach, we identify and characterize structural variants in up to 29 commonly used normal and cancer cell lines. We find that each method has unique strengths in identifying different classes of structural variants and at different scales, suggesting that integrative approaches are likely the only way to comprehensively identify structural variants in the genome. Studying the impact of the structural variants in cancer cell lines, we identify widespread structural variation events affecting the functions of non-coding sequences in the genome, including the deletion of distal regulatory sequences, alteration of DNA replication timing, and the creation of novel 3D chromatin structural domains. These results underscore the importance of comprehensive structural variant identification and indicate that non-coding structural variation may be an underappreciated mutational process in cancer genomes. .

807F

Functional characterization of a novel prostate cancer candidate gene at 2q37 in normal human prostate cell line. *C. Cieza-Borrella[1], S. Wakerell[1], I. Whitmore[1], Q. Karlsson[1], E. Saunders[1], R. Eeles[1,2], Z. Kote-Jarai[1].* 1) The Institute of Cancer Research, London, UK; 2) The Royal Marsden Hospital NHS Trust, London, UK.

Prostate Cancer (PrCa) is the most frequently diagnosed cancer among men in developed countries. The effectiveness of PSA screening remains controversial and new molecular biomarkers are needed to improve screening and clinical management of disease. Genome Wide Association studies and consequent fine-mapping analyses, identified four independent association signals at 2q37 represented by intronic and coding variants in the *NGEP/ANO7* gene. This gene encodes a cell membrane protein involved in cell-cell signalling and only expressed in prostate tissue. To investigate its role in PrCa we knocked out *NGEP/ANO7* in the diploid human epithelial prostate cell line RWPE-1 using CRISPR/Cas9 technique by transfecting cells with Cas9 protein and three different sgRNAs using Lipofectamine RNAiMAX. We also overexpressed this gene using a pCMV6-AC-GFP TrueORF vector. RNA and protein expression of the clones were checked by rt-qPCR and western blot. *NGEP* knockout, overexpressed and non-transfected clones were seeded in 3D cultures and their response to various stimuli was measured. Our results show that only those clones with both alleles knocked out resulted in complete loss of *NGEP* mRNA expression. A homozygous in-frame deletion of 24bp found in exon 1, within the N-terminal cytoplasmic tail, resulted in loss of *NGEP* mRNA expression compared to the 'no gRNA/Cas9 only' and 'non-transfected' control clones. This CRISPR'd clone forms spheroids which are smoother and more compact in appearance. It possesses a significantly reduced rate of spheroid proliferation and initial results suggest that it may also be more resistant to staurosporine induced apoptosis than the non-CRISPR'd cells. To assess the differential expression of other genes when repressing or overexpressing *NGEP* we will show results of RNAseq from positive clones. Our observations so far indicate that *NGEP/ANO7* plays a key role in prostate cells' physiology and analyses are ongoing to elucidate the biological mechanisms behind its function.

808W

Optical mapping reveals a higher level of chained fusion events in human cancer. *V.M. Hayes[1,2], A.T. Papenfuss[3], D.C. Petersen[1], D.M. Thomas[4], E.K.F. Chan[1].* 1) Human Comparative and Prostate Cancer Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia; 2) Central Clinical School, University of Sydney, Camperdown, NSW, Australia; 3) Bioinformatics Division The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria, Australia; 4) Cancer Division, Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW, Australia.

Chromosomal rearrangements are common in cancer, with potential to drive cancer progression and treatment response. These rearrangements can be complex, resulting in fusions of multiple chromosomal fragments as well as derivation of supernumerary chromosomes, such as neochromosomes. These fusion events may be driven by single or combination of processes, leading to the observed molecular and clinical complexity. Liposarcomas for example, are characterized by neochromosome formation, mostly likely a result of chromothripsis and hundreds of cycles of breakage-fusion-bridge amplifications. Short read sequencing is largely unable to assay all regions of the genome and generally unable to resolve relationships between distant rearrangements (e.g. chained rearrangements). In this study we show the utility of optical mapping using the Bionano Genomics next generation mapping technology to directly observe complex chained fusion events in a well-characterized liposarcoma cell line 778. From a total of 2,486 structural variations over 1Kb in length, we describe 778 as significantly translocated, observing a total of 151 chromosomal translocations. Importantly, we identified 101 fusion-specific genome maps harboring 282 large complex rearrangements, including 117 inter-chromosomal translocations and capturing much of the previously described neochromosomes. Using these large intact genome maps (range 189 Kb to 6.2 Mb) we were able to reconstructed optical maps of chained fusions, including the order, orientation, content and size of rearranged fragments. We show that optical mapping resolves fusion events marked by large intervals, revealing a higher level of chained rearrangements than detectable by sequencing alone. Furthermore, our data supports evidence for a single catastrophic genomic event early during the evolution of 778, followed by additional random, plausibly stepwise, rearrangement processes. We conclude that optical mapping is an important complement to deep sequencing for detecting and reconstructing complex genomic rearrangements.

809T

Comprehensive whole-genome analysis of the primary ENCODE cell line K562. *B. Zhou[1], S. Byeon[2], J. Arthur[3], X. Zhang[1], X. Zhu[1], R. Pattni[1], S. Ho[1], N. Ben-Efraim[1], C. Purmann[1], M.S. Haney[1], W.H. Wong[3], G. Song[2], A. Abyzov[4], A.E Urban[1].* 1) Department of Psychiatry and Behavioral Sciences and Department of Genetics, Stanford University School of Medicine, 3165 Porter Drive, Palo Alto, CA, 94304, USA; 2) School of Computer Science and Engineering, Pusan National University, Building 313 Room 308, 2 Busandaehak-ro 63 beon-gil, Geumjeong-gu, Busan 609-735, South Korea; 3) Department of Statistics, Department of Biomedical Data Science, Bio-X Program, Sequoia Hall, 390 Serra Mall, Stanford University, Stanford, CA 94035, USA; 4) Department of Health Sciences Research, Center for Individualized Medicine, Mayo Clinic, 200 1st Street S.W., Rochester, MN 55905, USA.

K562 is an immortalized myelogenous leukemia cell line derived from a 53-year-old female, one of the three tier-one cell lines of ENCODE, and one the premier cell lines used for CRISPR/Cas9 large scale gene editing screens. As such it is a major 'work horse' cell line that is frequently used in many laboratories around the world. However, although transcriptomic and epigenomic characteristics of K562 are extensively studied and documented, the genome sequence has never been determined and higher order structural features were only cursorily known. We performed deep short-insert whole genome sequencing, long-insert mate-pair sequencing, 10X Genomics linked-read sequencing, and used a combination of novel and established computational tools to call and catalog a wide spectrum of genomic variants: single nucleotide variants (SNVs, copy number corrected), small deletions and insertions (indels), phased SNP haplotype blocks, large copy number variants (CNVs) by chromosomal region, retrotransposon insertions, and structural variants (SVs), many of which are phased and a significant fraction of which are the result of complex genomic rearrangements. A subset of the SVs were validated using a combination of PCR and Sanger sequencing. In addition, we obtained karyotyping and performed genomewide microarray analysis using highdensity oligonucleotide arrays to validate the large CNVs resolved by our deep whole-genome sequencing based approach. The sequencing data and variant calls will be released through the ENCODE Portal. We imagine that they may be useful for a wide variety of purposes, e.g. for projects that use this cell line for CRISPR screens, where knowledge of the SNVs can extend the number of editing targets while knowledge of aberrant copy numbers will allow for the avoidance of non-diploid regions; or for studies that carry out in-depth analyses of the vast amount of ENCODE epigenomics data available for the K562 cell line, e.g. to determine whether a potential or known regulatory sequence element has been altered by SNVs, indels, retroinsertions or a gain or loss of copies of that given element and whether this can be correlated with signal changes on the epigenomic or transcriptome level. Furthermore, this study may serve as a technical example for advanced, integrated whole genome sequence and structure analysis beyond standard short read/short-insert whole-genome sequencing.

810F

De novo inference of enhancer-gene networks in diverse cellular contexts reveals the long-range regulatory impact of disease-associated variants. *J. Wang[1], P. Greenside[2], A. Kundaje[2], M. Kellis[3].* 1) Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI; 2) School of Medicine, Stanford University, Stanford, CA; 3) Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA.

Non-coding variants implicated in genome-wide association studies (GWAS) are enriched in enhancer elements active in disease-relevant cellular contexts. Identifying context-specific target genes and downstream pathways affected by enhancers harboring regulatory variants remains a challenge. We developed a novel mixed-membership probabilistic model that leverages the modular dynamics of gene expression and enhancer chromatin state across 56 diverse human cell types and tissues from the Roadmap Epigenomics Project to infer highly-connected, tissue-specific enhancer-gene networks. Chromatin conformation maps and expression QTLs validate the superior accuracy and tissue-specificity of our predicted networks compared to existing approaches. While half of the predicted links involve elements less than 50 kilobases apart, only a third associate enhancers with their nearest genes. Linked enhancers significantly improve tissue-specific regression models of gene expression. Distal co-association of regulatory sequence motifs suggests synergistic regulation of genes by multiple enhancers with a key role for protein-protein interactions between lineage-specific transcription factors in mediating enhancer-promoter interactions. Networks of cooperating enhancers with shared motif composition and target genes are depleted of disease-associated variants, suggesting regulatory buffering mechanisms. We demonstrate the utility of our context-specific enhancer-gene links to predict putative target genes, biological processes and pathways of non-coding variants associated with diverse traits and diseases, including colorectal cancer. Subsets of our predictions also demonstrate promising functional relevance in mouse breast cancer experiments.

811W

Exosomes in cancer: Small vesicular transporters for cancer development and metastasis, biomarkers in cancer therapeutics. *A. Abak[1], E. Sakhinia[2], A. Abhari[3].* 1) Tabriz University of Medical Sciences and Department of Biochemistry and Clinical Laboratory, Division of Medical Genetics, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran, Email: Atefeh. abak@gmail.com; 2) Tuberculosis and Lung Disease Research Center and Department of Bioechemistry and Clinical Laboratory, Division of Medical Genetics, Faculty of Medicine, Tabriz University of Medical Sciences, Golgasht Street, Tabriz, Iran; 3) Department of Bioechemistry and Clinical Laboratory, Division of Clinical Biochemistry, Faculty of Medicine, Tabriz University of Medical Sciences, Golgasht Street, Tabriz, Iran.

Cancer progression is a multistep procedure in which exosomes play substantial roles. Exosomes are small, lipid bilayer membrane vesicles of endocytic origin with a diameter of 40-100 nm. They can transport bioactive molecules such as mRNAs, proteins, DNA fragments, and non-coding RNAs from a donor cell to recipient cells, causing of the change of genetic information and reprogramming of the recipient cells. Many diverse cell types such as mesenchymal cells, immune cells, and cancer cells release exosomes. Increasing evidence illustrates that cancer derived exosomes which may contribute to the tumor initiation, tumor progression, metastasis, and drug resistance. Exosomes are secreted nanovesicles that play significance roles in cell-cell communication through shuttling proteins and nucleic acids to recipient cells and tissues. In this review, we discussed multiple mechanisms related to biogenesis, load, and shuttle of exosomes. Also we illustrated the multifaceted roles of exosomes in cancer development, tumor immunology, angiogenesis, and metastasis. Exosomes may act as the promising biomarkers for the prognosis of various types of cancers which suggested a new pathway for anti-tumor therapy of exosomes and progress exosome-based cancer diagnostic and therapeutic programs. Our objective was to meta-analysis systematically Review the literature from 2005 to 2017 with the terms "exosome" and "cancer", which included breast cancer, pancreatic cancer, gastric cancer, colorectal cancer, hepatocellular carcinoma , prostate cancer, bladder cancer, papillary thyroid carcinoma, glioblastoma, melanoma and leukemia. Inclusion Criteria is all the study designs relating to topic of interest, containing different cell lines, tissue specimens of several cancers, plasma samples of different cancer patients, using different methods (qRT-PCR and Western blotting). Exclusion criteria are any articles that not related to role of exosomes in cancer.

Exosomal Cargos	Cancer cell Types	Methods	Gene Expression Levels in Tumor Samples by Real Time-PCR	References
MiR-96-5p, MiR-182-5p, MiR-149	Human Tumor Colon Tissue Samples, blood specimens, and Colon Carcinoma Cell Lines HT-29, and HCT-116	qRT-PCR Analysis, and Western blotting	1)MiR-96-5p Expression in Normal Colon Tissues compared with CRC Tumor Tissues: 0.25% Decrease, p<0.001 2)MiR-149 Expression in Normal Colon Tissues compared with CRC Tumor Tissues: 0.31% Decrease, p<0.001 3)MiR-182-5p Expression in Normal Colon Tissues compared with CRC Tumor Tissues: 1.48% Increase, p<0.01	Jian Li et al.,2017
lncARSR (lncRNA Activated in RCC with Sunitinib Resistance)	Human RCC Cell Lines 786-O, 7su3rd, ACHN, and ACSu3rd	qRT-PCR Analysis, and Western blotting	1)Expression of lncARSR in RCC Tissues from Patients who Relapsed during Sunitinib Therapy compared with Patients who Benefited from Neoadjuvant Sunitinib Therapy: 2% Increase, p<0.01 2) Expression of lncARSR in the Plasma of RCC Patients After Tumor Resection compared with Before Tumor Resection: 0.62% Decrease, p<0.01	Li Qu et al.,2016
Curcumin (anti-cancer drug for lung cancer remedy)	Human Lung Cancer Cell lines BEAS-2B, A549, PC9, and H1299	qRT-PCR Analysis, and Western blotting	Relative Exosomal TCF21 Expression Levels of Normal lung Cell Correlated with the Aggressiveness of Cancer Cells: 0.625% Decrease, p<0.01	Hao Wu et al.,2016
let-7a, MiR-23b, MiR-27a/b, MiR-21, let-7, and MiR-320b	Human Breast Cancer Cell Lines MCF7, MDA-MB-231, BT20 and ZR-75-1	qRT-PCR Analysis, and Western blotting	Expression of let-7a, MiR-23b, MiR-27a/b, MiR-21, let-7, and MiR-320b in the DHA-Treated MCF7 Exosomes was measured in the Exosomes from the other Breast Cell Lines: 4.17% Increase, p<0.05	Bethany N et al.,2015
CD9, CD63, Fibronectin 1 (FN1),), and Laminin Gamma 1 (LAMC1)	Human Gastric Cancer Cell Lines Kato III (RCB2088), MKN45 (RCB1001), and MKN74 (RCB1002)	qRT-PCR Analysis, and Western blotting	Expression Levels of FN1 and LAMC1 in TEX-internalized Met-5A Cells: 34.75% Increase, p<0.05	Tomohiro Arita et al.,2016
GAIP interacting protein C terminus (GIPC)	Pancreatic cancer cell lines AsPC-1 and PANC-1	qRT-PCR Analysis, and Western blotting	Glut1 Expression in Level upon GIPC knockdown in AsPC-1 and PANC-1 Cells: 0.68% Decrease, p<0.01	Santanu Bhattacharya et al.,2014
MiR-9, CD44, CD45, CD105	Glioblastoma Cell Lines U87, T98G BT145 and BT164 Cells and GBM tissue specimens	qRT-PCR Analysis, and Western blotting	1)Expression of MiR-9 in U87 and T98G, Untreated and Treated with Temozolomide (TMZ): 2.6% Increase, p<0.05 2)Exosomal MiR-9 from GBM Cells was Analyzed for MiR-9 Levels using qRT-PCR Analysis: 0.96% Increase, p<0.05	Jessian L Munoz et al.,2013

812T

Epigenetic regulation of *POLG1* in breast cancer. *P. Bajpai[1], B. Singh[1], K.K. Singh[1,2,3].* 1) Department of Genetics,University of Alabama at Birmingham, Birmingham,Alabama, 35294, USA; 2) Department of Genetics, Departments of Pathology2, Environmental Health2, Center for Free Radical Biology2, Center for Aging2 and UAB Comprehensive Cancer Center2, University of Alabama at Birmingham, Birmingham,Alabama, 35294, USA; 3) Birmingham Veterans Affairs Medical Center3, Birmingham, Alabama, 35294, USA.

Nuclear-encoded polymerase POLG1 performs critical function of mtDNA replication and repair in mitochondria. We have previously reported that reduction of mtDNA content induces cancer progression. Methylation-dependent alterations in genes expression are associated with cancer progression. Since alteration in *POLG1* expression impacts mitochondrial DNA (mtDNA), which in turn leads to decreased mitochondrial OXPHOS, one of the most common phenotypes of cancer cells, we wanted to address the role of *POLG1* methylation status in cancer progression. We analyzed mRNA expression of *POLG1* in non-neoplastic human breast epithelial cells MCF-12A and human breast cancer cell lines MCF-7 and MDA-MB-231, and Rho^o (mtDNA depleted) cells. A very low expression of *POLG1* was observed Rho^o and MDA-MB-231 cells compared to non-neoplastic breast cell line MCF-12A. To address whether the *POLG1* expression is epigenetically regulated, we treated MDA-MB-231 cells with 5-aza-2'-deoxycytidinie (5-aza), an inhibitor of DNA methylation. Treatment of MDA-MB-231 cells with increasing dose of 5-aza increased expression of *POLG1* in a dose dependent fashion. 5-aza treatment increased expression of mitochondria-encoded genes of mitochondrial OXPHOS complex as well as mitochondrial respiratory activity. 5-aza treatment-mediated increase in *POLG1* expression increased growth rate and reduced matrigel invasion capacity of MDA-MB-231 cells. To further confirm the hypothesis, we analyzed the effect of *POLG1* mediated mtDNA depletion on nuclear epigenetic changes by employing a dominant negative POLG1 D1135A mutation in a doxycycline inducible Tet-on MCF-7 cells. We observed mtDNA depletion impacts nuclear genome, and the reversal of mtDNA mediated effects on nuclear genome when mtDNA is reversed back to control level after growing POLG1 D1135A MCF-7 cells for further 6 days in absence of doxycycline. We will be further studying the effect of 5-aza in modulating methylation status of *POLG1* promoter region. Our results indicate that POLG1 is epigenetically regulated, and its methylation status alter mitochondrial functions which can potentially affect the tumorigenic properties of the breast epithelial cells and could serve as important signature marker.

813F

In-depth analysis of genomics and epigenomics identifies a novel susceptibility lncRNA *GCLET* for gastric cancer. *M. Du[1,2], J. Lu[1,2], G. Ma[1,2], H. Chu[1,2], Y. Ge[1,2], W. Shao[1,2], S. Li[1,2], R. Zheng[1,2], N. Tong[1,2], G. Jin[3], Z. Hu[3], H. Shen[3], M. Wang[1,2], Z. Zhang[1,2].* 1) Department of Biostatistics, Nanjing Medical University, Nanjing, Jiangsu, China; 2) Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, Jiangsu, China; 3) Department of Epidemiology, Nanjing Medical University, Nanjing, Jiangsu, China.

Background: Genome-wide association studies (GWASs) have identified several loci dramatically associated with gastric cancer risk, but it seldom reported the specific mechanism of susceptibility gene for gastric cancer. **Methods:** We implemented an integrative methodology as a part of the "post-GWAS" strategy for gastric cancer including distributional annotation of genetic variants, gene-set analysis of GWAS data and *in silico* analysis along with functional studies. **Results**: By the validation of three-stage and multi-ethnic populations, we identified a significant association between rs3850997 at a long noncoding RNA (lncRNA) Gastric Cancer Low-Expressed Transcript (GCLET) and gastric cancer risk (P = 5.34×10^{-7}). It is mainly attributed to the allele-specific effect of rs3850997 on transcription factor CTCF binding affinity and thus on the expression levels of GCLET. Additionally, it was demonstrated that GCLET as a biomarker with higher expression could significantly predict better clinical progression and outcome of gastric cancer. Furthermore, GCLET could elevate FOXP2 expression by competing with miR-27a-3p. **Conclusion**: These findings suggest that the in-depth analysis by integrating genomics and epigenomics could further investigate the etiology and pathology of cancers. .

814W

Clinical relevance of non-coding A-to-I RNA editing in multiple human cancers. *T. Gu[1], A. Fu[2], M. Bolt[3], K. White[3].* 1) Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, FL; 2) Department of Statistical Science, Institute for Bioinformatics and Evolutionary Studies, Center for Modeling Complex Interactions, University of Idaho, Moscow, ID; 3) Institute for Genomics and Systems Biology, The Department of Human Genetics, The University of Chicago, Chicago, IL.

RNA editing is a post-transcriptional process that alters the nucleotide sequences of certain transcripts, in vertebrate most often converting adenosines (A) to inosines (I). Recently multiple studies have implicated RNA editing in cancer development and progression. However, most studies have focused on re-coding RNA editing events, which change the amino acid sequence of proteins but constitute only a small percentage of A-to-I RNA editing events. The function and clinical relevance of the majority of non-coding RNA editing events in cancers have not been systematically examined. We identified 3,788 non-coding A-to-I RNA editing events of high confidence from four human cancers (liver hepatocellular carcinoma, LIHC; lung adenocarcinoma, LUAD; kidney renal clear cell carcinoma, KIRC; and thyroid carcinoma, THCA). For each tumor type, hundreds to thousands of the non-coding RNA editing events showed significant differential editing (DE) efficiency between tumor and normal samples.These DE sites have distinct profiles across the four cancer types, and most (2,357; 62%) of the non-coding events we observed are at 3'UTRs and enriched at microRNA (miRNA) seed regions (key element for miRNA to recognize their targets). In kidney cancer, which has the largest uncensored survival data among the four cancer types, ninety four DE sites are significantly associated with patient survival. Using sparse regression and bootstrapping methods, we identified 3'UTR RNA editing sites that can affect gene expression, either independent of or through working with miRNAs. We validated the 3'UTR RNA editing sites in transcripts from the *CWF19L1* and *F11R* genes and the edited sites can result in increased protein levels. Furthermore, we demonstrated the alterations of the two genes' expression affect the proliferation of human embryonic kidney cells (HEK293T). These results indicate that 3'UTR editing sites may affect their host gene expression, and thus impacting cell proliferation.

815T

Epigenetic regulation of the Runx2 gene in lung cancer. *A. Herreno[1], A. Ramirez[1], G. Giraldo[1], O. Moreno[1], A. Cañas[2], M. Montecino[3], A. Rojas[2].* 1) Instituto de Genetica, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Unidad de neumología-Hospital Universitario San Ignacio,Bogota, Colombia; 3) Centro de Investigaciones Biomedicas, Facultad de Ciencias Biologicas y Facultad de Medicina, Universidad Andrés Bello, Santiago☐Chile.

Lung cancer is one of the most common causes of cancer death in the world causing 1.6 million deaths per year. This disease can be developed by genetic, epigenetic and environmental factors, among others, In Colombia for 2015, Globocan has estimated an average of 5433 diagnosed, with a mortality rate of 92.3%. In recent years, the influence of specific lineage transcription factors in pathologies such as cancer has been demonstrated. In this regard, it has been reported that the transcription factor Runx2, master regulator of osteoblastic differentiation, can actively participate in processes of progression, survival and transition mesenchymal epithelium in different types of cancer. The expression of *Runx2* is regulated by the presence of covalent modifications mediated by the WDR5/MLL/UTX complex. We selected 31 samples of lung tissues of patients who attended Hospital Universitario San Ignacio for the study of lung lesions suspected of malignancy with indications for pulmonary biopsy obtained from thoracoscopy and tomography-guided fine needle aspiration from 2016 to 2017. In this work demonstrates an increase in the expression of the *Runx2* gene at the level of mRNA and protein by RT-PCR and WB respectively in subjects diagnosed with lung cancer type NSCLC. This deregulation in the expression of Runx2 is related to changes in the levels of enrrichment of the enzymatic complex WDR5 / MLL/ UTX and is related with the presence of covalent modifications of activating histones in the P1 promoter of Runx2.

816F

Molecular mechanisms underlying serrated polyps: Comprehensive DNA methylation analysis reveals new targets in the serrated neoplasia pathway. *V. Khammad[1], N. Zhuchenko[2], T. Lysytsyna[2,3], I. Abramov[3], I. Sapozhnikova[1], E. Khammad[1], O. Zaitseva[1,2].* 1) Medical Rehabilitation Center, Ministry of Health of Russian Federation, Moscow, Russian Federation; 2) I.M. Sechenov First Moscow Medical University, Moscow, Russian Federation; 3) Engelhardt Institute of Molecular Biology, Moscow, Russian Federation.

   Background: Most cases of colorectal cancer (CRC) originate from identifiable precursor lesions. Fearon and Vogelstein described adenomatous polyps as the principal pre-neoplastic leasions leading to CRC. However, recently recognized «serrated pathway» represents a heterogeneous group of polyps with characteristic serrated morphology some of which exhibit a significant risk of neoplastic progression. Our aim was to investigate DNA methylation in serrated polyps (SPs) and achieve a better understanding of their role in colorectal carcinogenesis.   **Methods:** We performed comprehensive genome-scale DNA methylation profiling in 76 sporadic CRCs with corresponding normal mucosa, 35 sessile serrated polyps (SSPs) and 48 traditional serrated adenomas (TSAs) using the Illumina HumanMethylation450 BeadChip. Gene expression profiling was also performed using the Illumina whole-genome DASL HT Assay.   **Results:** Our genome-wide methylation study showed that differentially methylated loci were highly enriched for CpG islands in gene promoter regions: CRCs and SSPs exhibited higher methylation levels compared to TSAs (p<0.0001). Methylation-mediated transcriptional silencing of various genes belonging to the β-catenin/Wnt pathway (SFRP family genes, DKK3, WIF1, CDX2, MCC, WNT5a) was the most frequent molecular abnormality identified in TSAs (21/48; 44%) and CRCs (33/76; 43%). This finding suggests that aberrant DNA methylation has a huge impact on Wnt-driven tumorigenesis in TSAs. Moreover, SOX-Wnt interactions may also be involved in malignant transformation of TSAs. Three key genes of the SOX family: SOX7, SOX17 and SOX9, also known as suppressors of the hyperactive β-catenin activity, were downregulated through promoter hypermethylation in TSAs (11/48; 23%, 7/48; 15% and 4/48; 8%). The disruption of p53 and RB tumor-suppressor signaling pathways by epigenetic inactivation of IGFBP7 and CDKN2A occurred predominantly in SSPs (13/35; 37%) and was observed in some cases of CRC (10/76; 13%). Finally, we discovered that increased methylation of CpG islands in MLH1 occurred more frequently in SSPs (24/35; 69%) than in tumor specimens (9/76; 12%) and TSAs (3/48; 6%).   **Conclusions:** Our data suggests that microsatellite instability, CpG island methylator phenotype high and inactivation of IGFBP7 and CDKN2A play a key role in neoplastic progression in SSPs, while TSAs are more likely to develop cancer through methylation-mediated deregulation of the β-catenin/Wnt signaling pathway. .

817W

Differential DNA methylation aspect of L1-chimeric transcripts in various cancers. *S. Kim, S. Mun, K. Han.* Nanobiomedical Science, Dankook University, Cheonan, Chungnam, South Korea.

   Transposable elements (TEs) have contributed to the genomic structural variations which were lead to primate evolution. TEs are classified in to LINE, SINE, ERV, and SVA. Among these elements, LINE-1s (L1s) make up ~17% of the human genome, with a total of about 520,000 copies. Approximately 1,800 copies of human-specific L1s are integrated into genome after the divergence of human and chimpanzee lineages. Although the major of L1s are inactive, some human-specific L1s are still active in the human genome. L1s are a *cis*-regulatory elements and contain bidirectional internal promoters within the 5' untranslated region (UTR). L1 promoters contribute the transcriptome diversity and differential expression of nearby genes. The gene expression is regulated by the methylation of CpG islands in the 5'UTR of L1s. In this study, we characterized the L1-chimeric transcripts and analyzed methylation status of L1 promoter. First, we collected 150 L1-chimeric transcripts generated from L1 antisense promoters by referring to previous studies. Through manual inspection, we identified 60 L1-chimeric transcripts derived from 48 L1 antisense promoters that may serve as alternative promoters. We performed reverse transcription-PCR to analyze their expression patterns in matched cancer-normal tissue pair (gastric, colorectal, lung, breast, liver, prostate, breast, ovary, uterus, and kidney). Furthermore, we analyzed relative gene expression using real-time PCR and investigated correlation between methylation status of DNA and its expression. Here, we suggest that DNA methylation of L1 antisense promoter could affect the differential gene expression between cancer and normal tissues.

818T

The effect of expression of glycosylation genes, regulators, and targets on cancer cell line sensitivity to drug treatment. *J. Krushkal[1], Y. Zhao[1], C. Hose[2], A. Monks[2], J.H. Doroshow[1,3], R. Simon[1].* 1) Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, MD 20850; 2) Molecular Pharmacology Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702; 3) Division of Cancer Treatment and Diagnosis and Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892.

Glycosylation pathways are vital for cell functioning. In cancer, many glycosylation processes are profoundly altered, affecting malignant cell growth, functioning, and interaction with surrounding environment. Due to their importance in cancer development, progression, dissemination, and treatment, glycosylation components are increasingly being targeted in cancer therapy and used as cancer biomarkers. We used time course gene expression data for the NCI-60 cancer cell lines treated with eleven antitumor agents (bortezomib, cisplatin, dasatinib, doxorubicin, erlotinib, gemcitabine, paclitaxel, sirolimus, sorafenib, sunitinib, and topotecan) to analyze the transcriptional response of 179 genes involved in biochemical glycosylation pathways or encoding glycosylation targets, regulators, and components of cancer pathways affected by glycosylation. For each cancer drug, concentration, and time after treatment, we identified glycosylation genes with a concerted response, defined as those genes in which the expression of the vast majority of the cell lines changed in the same direction (upregulation or downregulation). Expression of 145 glycosylation genes was affected by at least one antitumor agent in a concerted manner. While many genes responded in a concerted manner to multiple agents, the direction and magnitude of their transcriptional responses were specific to individual drugs. Pretreatment expression levels of several candidate genes, e.g., *EXT1, NRP1, GALNT10, GNPTAB,* and *SIGLEC6,* were significantly correlated with chemosensitivity to dasatinib and erlotinib. Drug treatment induced transcriptional changes in many genes. Some of these changes were correlated with chemosensitivity to antitumor agents with different mechanisms of action. Such genes participate in *N-* and *O-*glycosylation, fucosylation, biosynthesis of poly-*N*-acetyllactosamine, removal of misfolded proteins, binding to hyaluronic acid and other glycans, and cell adhesion. Examples of such genes include *C1GALT1C1* (*COSMC*), *FUCA1, SDC1, MUC1,* and members of the *MGAT, GALNT, B4GALT*, *B3GNT, MAN,* and *EDEM* gene families. We provide a catalog of concerted expression changes of glycosylation genes, ligands, targets, and regulators in response to treatment by different agents, and list the genes associated with chemosensitivity. These genes may be targeted in combination therapy in order to enhance the treatment effects. Funded by NCI Contract No. HHSN261200800001E.

819F

Motif disruption domains lead to cancer gene expression rewiring. *F.C. Lamaze[1], A. Chateigner[1], H.A. Edgington[1], A. Ang Houle[1], M.J. Fave[1], P. Awadalla[1,2], PCAWG-3.* 1) Informatics Program, Ontario Institute for Cancer Research, Toronto, Ontario; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario.

Recent advances in sequencing studies have exposed numerous potentially high-impact and, in some cases, actionable somatic variants in protein-coding regions across several tumor types. Yet, 95% of these somatic mutations lie in the non-coding fraction of the genome. Studies of noncoding variation implicate disease-associated loci are commonly located in noncoding regions, and are recurrent in regulatory regions. There is much evidence to suggest that a proportion of somatic single nucleotide variants (SNV) in cancer may have transcriptional regulatory mechanisms. Yet, the functional characterization of non-coding SNV, with respect to gene regulation during tumorigenesis, is a challenge. Somatic non-coding mutations rarely overlap among patients, which necessitates large sample sizes to detect associations. We analysed somatic mutations called from whole-genome sequencing and RNA sequencing from 2881 tumors across the Pan-Cancer Analysis of Whole Genomes to identify and functionally characterize mutation accumulation and its impact on gene dysregulation in cancer. We identified 1,507,105 million motif disruption domains (MDDs) of median size 1.2kb across 40 cancer types, which we characterized as pan-cancer targets for recurrent mutation accumulation. The top 1% of motif disruption scores capture binding sites of transcription factors (TFs) and cofactors known to bind to promoter, enhancer and insulator regions, for example pioneer TFs (FOXA and GATA), chromatin remodeler binding sites (EP300) and CTCF (CCCTC-binding factor). These MDDs may deregulate gene expression in cancer-specific and pan-cancer patterns by disrupting transcription factor binding sites in regulatory and insulator elements. Disruption is most recurrent across individuals at MDDs in conserved open chromatin, revealing potential functional drivers. This accumulation of somatic variants targeting regulatory and structural elements in MDDs generates gene expression dysregulation during tumorigenesis. We believe this work will improve our ability to classify tumoral and non-tumoral tissue samples by identifying pan-cancer and cancer-specific mutational signatures with substantive functional impact on the regulation of genes and pathways that could reveal new therapeutic targets.

820W

Effect of DNA methylation on expression of drug response genes. *J.M. Oh[1], I.W. Kim[1], S. Kim[2].* 1) College of Pharmacy and Research Institute of Pharmaceutical Sciences, Seoul National University, Seoul, South Korea; 2) Department of Bioinformatics and Life Science, Soongsil University, Seoul, South Korea.

 Backgrounds: Pharmacogenomics focuses on the influence of individual genetic variations on drug response and is helping to pave the way toward safer and more effective treatments for patients. Individual differences in drug response are associated with genetic and epigenetic variability. DNA methylation can affect drug response by altering the expression of genes encoding drug-metabolizing enzymes, transporters, and targets. To identify DNA methylation loci for drug response genes, we investigated DNA methylation and RNA expression profiles of drug response using the Cancer Genome Atlas data. Methods: Important pharmacogenomic-related genes were searched on the Pharmacogenomics Knowledge Base and the US Food and Drug Administration Table of Pharmacogenomic Biomarkers in Drug Labels. DNA methylation and RNA expression profiles of drug response genes were selected from colon adenocarcinoma and kidney renal clear cell, liver hepatocellular, and lung squamous cell carcinomas. We identified DNA methylation sites in each sample to obtain β-value levels for all probes. The M-value was calculated by the following equations: M-value$=\log_2\{\beta\text{-value}/(1-\beta\text{-value})\}$ and this M-value was used for a statistical analysis. A Pearson's correlation analysis was used to identify DNA methylation loci related to differential gene expression after adjusted by tissue types. The distance matrix for DNA methylation variables in normal samples was constructed using the Euclidean distance and was visualized by multidimensional scaling analysis. Results: We found DNA methylation loci in drug response genes contributed to different DNA methylation levels in various normal tissues that correlated with their RNA expression levels. Additionally, the different methylation levels of DNA methylation loci in drug response genes were related to their differential expression in normal and tumorigenic kidney, liver, colon, and lung tissues. Conclusions: DNA methylation profiles of drug response genes in normal and tumor tissues may play a role in determining individual variations in drug response.

821T

Evaluation of circulating cell free DNA in bisulfite sequencing applications. *M. Poulin, J. Xu, R. Drennan, S. Miller, A. Meyer, L. Yan.* EpigenDx, Hopkinton, MA., USA.

 Circulating cell free DNA (cfDNA) has become a highly sought-after analyte in both the research and clinical communities. Its use in detecting and evaluating biomarkers for various diseases and its non-invasive isolation make it extremely versatile. However, the isolation, quantification and limited quantities give rise to many technical issues for use as a routine product in diagnostic medicine. We have been isolating cfDNA from serum and plasma for DNA methylation analysis using both pyrosequencing and next-generation bisulfite sequencing. Because the cfDNA must be bisulfite treated prior to either sequencing method, we have analyzed the quantity and quality of cfDNA from different volumes of serum, following both isolation and bisulfite treatment. Quantities of DNA were examined using real-time PCR quantification assays that can assess both unmodified and bisulfite treated DNA at very low levels. Results will be presented for the quantification of cfDNA isolated from various samples and volumes of serum both before and after bisulfite treatment. The quality and utility of the bisulfite treated DNA was analyzed using a targeted NextGen bisulfite sequencing (tNGBS) panel of 18 cancer related genes containing approximately 200 CpG sites; APC, BRCA1, CDKN2A, CST6, DAPK1, ESR2, ETS1, FOXP3, GSTP1, HIST1H3C, HOXB4, IKZF2, ITGAL, MLH1, RARB, RASGRF2, RUNX3, SOX17, TMEFF2. Results were also validated using pyrosequencing assays and will be presented for the different samples and different serum starting material. These results will help to evaluate isolated cfDNA prior to bisulfite sequencing protocols and establish optimal volumes of serum that need to be processed for these downstream applications.

822F

Genetic polymorphism and gene expression of SHH & PI3K gene in ameloblastoma. *H. Singh, AB. Urs.* Oral Pathology, Maulana Azad Institute of Dental Sciences, BSZ Marg, MAMC Complex, New Delhi, New Delhi, India, India.

   Background & Aim: Ameloblastoma is a benign and local aggressive odontogenic tumor. SHH plays a pivotal role in epithelial mesenchymal interaction to induce many cellular changes while PI3K has an important role i.e. cellular quiescence, proliferation, cancer, and longevity in the pathogenesis of Ameloblastoma through PI3K/AKT/mTOR signalling pathway. The study was designed to evaluate the gene expression and gene polymorphism of SHH and PI3K gene. **Materials & Methods:** The present study was a prospective preliminary study, which was carried out in 30 patients of confirmed ameloblastoma cases. Biopsy was taken with patient's consent. Genomic DNA was extracted for polymorphism of SHH and PI3K gene by polymerase chain reaction to assess genetic alteration and Immunohistochemistry technique respectively. **Results:** Mutations were found in Eight cases for PI3K, of these 5 cases had mutations in exon 9 (62.5%) in E542 and E545 region and 3 in exon 20 (37.4%) in H1047. No SHH gene polymorphism was detected in tumor. No differences were noted in the frequency and type of mutations analyzed by sex, age, or histologic features. The gene expression of SHH & PIK3CA was significantly higher in tumor epithelial cells. **Conclusion:** These results suggest that common genetic variations in these pathways may modulate risk and clinical outcomes of ameloblastoma. Further replication and functional studies are needed to confirm these findings. It will be of benefit to the patient, if we target the mutation or aberrant protein products at the appropriate time by intervention of précised therapy.

823W

MiRNA profiling of pre-cancerous and cancerous condition of stomach by next-generation sequencing. *J. Skieceviciene[1], S. Juzenas[1], V. Salteniene[1], R. Steponaitiene[1], G. Kiudelis[2], L.V. Jonaitis[2], L. Venclauskas[3], A. Ivanauskas[2], A. Tamelis[3], G. Hemmrich-Stanisak[4], M. Hubenthal[4], A. Franke[4], L. Kupcinskas[1,2], J. Kupcinskas[1,2].* 1) Institute for Digestive Research, Lithuanian University of Health Sciences, Kaunas, Lithuania; 2) Dept of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania; 3) Dept of Surgery, Lithuanian University of Health Sciences, Kaunas, Lithuania; 4) Institute of Clinical and Molecular Biology, Christian-Albrecht-University Kiel, Kiel, Germany.

   Gastric cancer (GC) is a leading cause of cancer related death worlwide. Therefore, molecular biomarkers for early non-invasive diagnostics of this malignancy are of major interest. Deregulation of miRNAs has been observed virtually in all major types of cancer. The aim of this study was to determine miRNA expression profile in tissue and plasma of patients with GC and atrophic gastritis (AG). High-throughput miRNA profiling of 20 GC, 18 AG and 26 controls was performed using small RNA-seq approach and differentially expressed miRNAs were defined. Highly significantly deregulated miRNAs were selected (criteria: corrected p-value < 0.01, and fold change > 2, expression level in plasma (read count sum > 5)) for validation by Taq-Man low-density array in replication group of 39 GC, 40 AG and 40 controls. Validated miRNAs (FDR adjusted p-value < 0.01 and fold change > 2.0) were further analysed in plasma samples of the study group (60 GC, 60 AG and 60 controls). The study was approved by the bioethical research committee (No BE-2-10). Small RNA-seq revealed: (i) 110 differentially expressed miRNAs in GC group compared to HC group; (ii) 17 differentially expressed miRNAs in AG compared to HC; (iii) 76 differentially expraessed miRNAs in GC compared to AG. In the validation analysis 21 miRNA present in plasma and meeting other criteria were selected for validation study. The expression levels of 8 out of 21 miRNA in GC and 2 out of 21 miRNA in AG compared to normal tissue showed significant differential expression to the same direction as in profiling study. The expression level of miRNA in GC compared to AG was statistically different 5 out of 21 miRNA. The expression levels of validated miRNAs were further analyzed in plasma samples. The expression levels of hsa-miR-148a-3p and hsa-miR-375 showed significant down-regulation in GC plasma; while hsa-miR-223-3p was significantly up-regulated (FDR adjusted p < 0.05). No miRNA expression differences were observed in plasma of AG patients. In conclusion, miRNA expression profiling revealed: (i) 8 deregulated miRNAs in GC tissue, (ii) two deregulated miRNAs in AG tissue, (iii) five miRNAs were different in GC compared to AG. Evaluation of miRNAs as non-invasive biomarkers in GC plasma revealed differential expression of hsa-miR-148a-3p, hsa-miR-375 and hsa-miR-223-3p.

824T

Epigenetics modification and gene expression studies upon human amniotic fluid stem cells treated with chemotherapeutic drugs. *P. Upadhyaya[1], D.Di Tizio[1], P. Ballerini[1], I. Antonucci[1,2], L. Stuppia[1,2].* 1) Department of Psychological , Health and Territorial Sciences ,Laboratory of Molecular Genetics, School of Medicine and Health Sciences, G. d'Annunzio University, Chieti-Pescara, Italy; 2) Stem Tech Group, Center for Aging Studies (CESI), G. d'Annunzio University, Chieti-Pescara, Italy.

   BACKGROUND: Although testicular cancer (TC) has one of the highest cure rates of all cancers with an average five-year survival rate of 95%, following the treatment with chemotherapeutic reagents, it leads to infertility in some of the patients (30%). In the recent years, it has been reported in the literature that human Amniotic Fluid-derived Stem Cells (hAFSCs) show characteristics similar to pluripotent cells of the epiblast and therefore could be considered ideal candidate to generate in vitro germ cells. The purpose of this particular study was to investigate the epigenetic modifications of hAFSCs during treatment with the chemotherapeutic drugs Bleomycin*,* Etoposide and Cisplatin (BEP regime). **METHODS:** hAFSCs were cultered for up to 7 consecutive passages and at 5th and 7th passages Cisplatin, Bleomycin and Etoposide were added into the culture medium. From the dose-response curve, the optimal time of treatment was found to be 48 hours and the IC_{50} of Cisplatin, Bleomycin and Etoposide was found to be 5 μM, 50 μM and 100 μM, respectively. Then, the hAFSCs were treated with the respective IC_{50} concentration and a concentration 10 times less (i.e. 5 and 0.5 μM for Cisplatin; 50 and 5 μM for Bleomycin; 100 and 10 μM for Etoposide) for epigenetics as well as expression studies. **RESULTS:** Expression analysis by Quantitative PCR showed the increase of expression of pluripotency-associated genes (Oct-4, SOX2,c-Myc, KLF4 and c-Kit) in the cells treated with the chemotherapeutic drugs as compared to the control. Finally, Global DNA methylation experiments have demonstrated a gradual demethylation of DNA in cells treated. **CONCLUSION:** The chemotherapeutic agents used for the treatment of testicular cancer were found to be acting epigenetically on the DNA of hAFSCs. The results obtained suggest that BEP regime remove methyl group from the DNA and the expression studies of pluripotent markers support this epigenetic data . In future, hAFSCs could be used in disease modelling of infertility to study mutagenic and epigenetic effects of the drugs.

825F

Integrating chromatin and expression variation in statistical fine-mapping. *M. Roytman[1], G. Kichaev[1], B. Pasaniuc[1,2,3].* 1) Bioinformatics IDP, University of California Los Angeles, Los Angeles, California, USA; 2) Department of Pathology and Laboratory Medicine, Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA; 3) Department of Human Genetics, Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA.

   Recent efforts to characterize how genetic variation impacts intermediate molecular phenotypes have yielded thousands of SNPs that associate with local and distal histone modifications -- termed histone quantitative trait loci (hQTLs). Chromatin states that are defined by combinations of histone marks are in turn, known to associate with changes in gene regulation. Furthermore, recent studies have identified many eQTLs that co-localize with molecular QTLs, implying there may exist a shared genetic influence on epigenetic traits and gene expression. Therefore, one proposed mechanism where regulatory variants may affect gene expression and thereby cause disease is through changes in chromatin state. Here we introduce new frameworks that models the hierarchical relationships between genome, epigenome, and gene expression to predict both the causal SNP and the causal epigenetic mark within a gene region. Our model naturally allows for multiple causal variants and marks to further improve accuracy. We extend our approach to incorporate interaction effects on the histone-expression level, which will allow us to identify cases where two marks influence expression concordantly rather than independently.

826W

Mismatch repair-associated mutations reprogram the colorectal cancer enhancer epigenome. *S. Hung[1], A. Saiakhova[1], D. Neu[1], Z. Faber[1], C. Bartels[1], G. Dhillon[1], M. Kalady[3], S. Markowitz[1,2], P. Scacheri[1].* 1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Department of Medicine, Case Western Reserve University, Cleveland, OH; 3) Lerner Research Institute, Cleveland Clinic, Cleveland, OH.

Somatic mutations in protein coding genes are well defined in many forms of cancer, but far less is known about mutations in noncoding regulatory elements. Active enhancer elements co-localize with acetylated lysine 27 on histone H3 (H3K27ac). Through direct interrogation of H3K27ac ChIP-seq reads, we searched for somatic indel mutations across the enhancer epigenomes of a cohort of clinically and genetically diverse human colorectal cancer (CRC) specimens. We identified potential "enhancer activating" indels that acquire H3K27ac in CRC compared to healthy normal colon cells. Strikingly, CRC samples with deficiencies in mismatch repair (MMR) proteins show significant enrichment of potential enhancer activating indels, and acquire these indels at a 2 to 100-fold higher rate than MMR-proficient samples (P < 9.2e-5). The enhancer indels show evidence of positive selection and are significantly associated with overexpressed cancer-related genes in CRC (P < 0.05). A subset of the indels are recurrent in primary CRC tumors. The indels arise mostly in short (5-10 bp) stretches of poly (A/T) tracts and generate sequences that closely resemble consensus motifs for the FOX family of pioneer transcription factors. Through ChIP-seq studies, we verified enrichment of FOX factors predicted to bind at the indels (P < 0.05). Moreover, at loci harboring the canonical FOX motif-like sequence (TGTTT*), MMR-deficient CRC samples contain 1,500 to 2,300 more active enhancers than CRC samples with intact MMR machinery. MMR-deficient endometrial cancers also show evidence of enhancer-activating indel mutations, suggesting that the effects are not limited to CRC and may be a general feature of cancers with MMR-deficiencies. Our results suggest that indel mutations in noncoding poly (A/T) sequences, previously presumed benign, frequently activate multiple enhancers in the epigenomes of MMR-deficient cancers and provide a selective advantage for tumor growth.

827T

The genetic diversity affects the cell-fate in genotoxicity test. *C.C. Lin[1], S.M Chuang[2], Z.C Hung[3], Y.C. Li[3].* 1) Lab Chromosome Res, China Medical University & Hospital, Taichung, Taiwan; 2) Institute of Biomedical Sciences, National Chung Hsing University, Taichung, Taiwan; 3) Department of Biomedical Sciences, Chung Shan Medical University, Taichung, Taiwan.

The genetic diversity is a concerned issue in population, disease susceptibility, and drug response. However, it is less noticed about the genetic diversity among cell types in cell-based study. Generally, the preclinical study used one kind of cell line for in vitro genotoxicity test. CHO (Chinese hamster ovary) cell lines are commonly used for *in vitro* chromosomal aberration test suggested by FDA for the guideline of *in vitro* genotoxicity testing of chemicals. In addition to the result of chromosome aberrations, the used dosage, and the cytotoxicity responses are the reference indicators for a new drug development in pharmaceutical industry. In our study, we used two different mammalian cell lines, CHO-K1 (Chinese hamster ovary cell line) and SG (Human gingival epithelioid cell line) to test the genotoxicity and cytotoxicity of arecoline. Both cell lines were considered as non-tumorigenic cell lines but both have aberrant karyotypes. We found that the numbers of chromosome aberrations induced by arecoline in CHO-K1 cells is lower than that in SG cells at the same dosage used. CHO-K1 was more resistant to arecoline induced cytotoxicity than SG. Surprisingly, ROS (reactive oxygen species) was dose dependently induced by arecoline in CHO-K1, but not in SG. Moreover, ROS caused the cell cycle arrest at G2/M phase and then the cell death (pyknotic necrosis) of CHO-K1, while chromosome aberrations caused the cell death (apoptosis) of SG. Arecoline induced chromosome aberrations (a kind of DNA damage) in both cell lines. However, arecoline-induced DNA damage resulted in different cell fate between CHO-K1 and SG. It had been known that CHO-K1 is a p53 mutant cell line and SG has a functional p53. The mutant p53 lacks of G1 checkpoint. This could be one of reasons to cause cell cycle arrest at G2/M phase in CHO-K1, and the functional p53 of SG once detected the damaged DNA and then induced apoptosis. Our findings suggested that the genetic diversity affects the response of genotoxicity (quantitatively or qualitatively) and the mechanism of genotoxic agent induced cytotoxic response. More attention should be paid for the genetic diversity than cell phenotype (such as normal, non-tumorigenic, and carcinogenic) in studying genotoxic and cytotoxic responses in future.

828F

Genome-wide map of APA in lung cancer: A pilot study. *A. Zingone[1], M. Ante[2], KA. Mitchell[1], BM. Ryan[1].* 1) Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, MD; 2) Lexogen GmbH, Vienna, Austria.

Alternative polyadenylation (APA) is emerging as an alternative mechanism for proto-oncogene activation. The process generates shorter mRNA transcripts, which results in the loss of miRNA repression and mRNA stability. The 3'UTR is globally shortened in many types of cancer. In addition, environmental exposures can also mediate usage of a proximal polyadenylation site (PAS) over a distal one. However, the role of 3'UTR shortening in cancer and the contribution of environmental exposure to this process remains understudied. We hypothesized that smoking could induce APA in lung epithelial cells and contribute to their transformation. Previously we found a shorter isoform of the KRAS 3'UTR over-represented in primary lung tumors, expression of which correlated with increased pack-years of smoking. Thus, we sought to systematically map the APA landscape in lung cancer using a novel RNA-seq method termed QuantSeq-Rev 3'UTR. This method allows the identification of distinct APA within a single transcript, as well as quantification of gene expression. A pilot study comparing 8 matched primary lung tumors/non-involved samples was designed to validate the method. To measure the 3'UTR length of transcripts we developed a PAS index. Lung cancer samples had a lower PA index value compared with controls, indicating a significant switch in the use of proximal PAS $(P<0.0001)$. We found over 4000 genes that were significantly more likely to use the proximal PAS in lung cancer tissues (FDR<0.01). Interestingly, these genes were associated with pathways regulating translation initiation and protein synthesis. Conversely, genes more likely to use distal sites associated with pathways present in cells that undergo differentiation or low proliferative status. DeSeq2 analysis identified over 2000 genes differentially expressed in tumor versus normal. Interestingly, the Cleavage Stimulating Factor 2 (CSTF2) expression was 2-fold higher in tumor tissues. Overexpression of CSTF2 has been associated with poor prognosis of patients with non-small cell lung cancer. In parallel, we studied the effect of nicotine on APA *in vitro* and found that nicotine exposure induces preferential usage of the proximal PAS. Currently, we are studying APA in 100 primary tumors and 100 matched normal tissues. We have hypothesized that we will identify distinct APA signatures that can predict clinical outcome and exposure related changes.

829W

A genome-wide association study (GWAS) implicates NR2F2 in lymphangioleiomyomatosis pathogenesis. *K. Giannikou*[1], S. Won*[2,3,4], W. Kim[2], J.R. Dreier[1], X. Qiu[5], M. Tyburczy[1], E.K. Silverman[1,6], E. Radzikowska[7], P. Rao[5], H. Long[5], E.P. Henske[1], G. Hunninghake[1], J. Moss[8], M. Pujana[9], D.J. Kwiatkowski[1], * equal contribution.* 1) Division of Pulmonary and Critical Care Medicine and of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, 02115, MA, USA; 2) Department of Public Health Science, Seoul National University, Seoul 151742, Korea; 3) Interdisciplinary Program of Bioinformatics, Seoul National University, Seoul 151742, Korea; 4) Institute of Health and Environment, Seoul National University, Seoul 151742, Korea; 5) Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston 02215, MA, USA; 6) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA; 7) National Tuberculosis and Lung Diseases Research Institute, Warsaw, Poland; 8) Cardiovascular and Pulmonary Branch, NHLBI, NIH, Bethesda, 20824-0105 MD, USA; 9) Catalan Institute of Oncology, Barcelona 08908, Spain.

Background: Lymphangioleiomyomatosis (LAM) is a rare aggressive low-grade neoplasm which affects almost exclusively women and causes cystic lung destruction. LAM can either be associated with Tuberous Sclerosis Complex (TSC) or be sporadic and it is known to be due to TSC2/TSC1 genetic alterations. Risk factors for development of sporadic LAM (S-LAM) are unknown, and understanding of disease origins remains limited. We hypothesized that DNA sequence variants outside of TSC2/TSC1 are associated with susceptibility/risk for LAM disease development, and performed a Genome-Wide Association Study (GWAS) on sporadic adult female S-LAM patients, the largest LAM cohort ever assembled. **Methods:** Over 600 S-LAM subjects were collected from 14 countries, and saliva DNA from 479 S-LAM and blood DNA from 1261 matched healthy female control subjects from the COPDGene cohort were genotyped on the Infinium *OmniExpress*-24 v1.2 BeadChip (discovery phase). Standard quality controls were employed, and 550,923 SNPs were analyzed in 1690 subjects. Principal components analysis was performed to adjust for genetic ancestry and exclude outliers. Conditional logistic regression for matched cases and controls was done. **Results:** Two non-coding SNPs met genome-wide significance, rs4544201 (minor allele frequency(MAF) 0.143 in cases vs. 0.255 in controls, P=$2.8×10^{-9}$) and rs2006950 (MAF 0.164 in cases vs. 0.274 in controls, P=$2.9×10^{-8}$), which are in the same 34kb LD block on chr15q. These findings were replicated with an independent set of 220 S-LAM cases and 1214 controls, with P=$2.1×10^{-5}$ and $1.5×10^{-6}$, respectively. Another intergenic SNP mapped within this LD block, rs4561414, has a tissue-specific eQTL relationship with *NR2F2* (located 700kb away) expression in thyroid, P=$3.7×10^{-6}$, a member of the steroid thyroid hormone superfamily of nuclear receptors and a key developmental transcription factor in neural crest, candidate cell of origin for LAM. NR2F2 expression was found remarkably high in RNA-seq data for kidney angiomyolipoma, a LAM-related tumor after comparison with all other cancer types (TCGA consortium) and identified as a super-enhancer gene in H3K27ac ChIP-seq data. **Conclusions:** This S-LAM GWAS has identified replicable SNPs on chr15q whose alleles are associated with LAM in patients of European ancestry. The biological mechanism related to this genetic association requires further investigation and functional assessment of NR2F2 is in progress.

830T

In silico reanalysis reveals novel prognostic miRNAs in pancreatic neuroendocrine tumors. *V.K. Grolmusz[1,2], K. Borka[3], A. Szentpéteri[1], Zs. Vásárhelyi[1], P. Igaz[1], K. Rácz[1,4], A. Patócs[2,4,5].* 1) 2nd Department of Medicine, Semmelweis University, Budapest, Hungary; 2) MTA-SE "Lendület" Hereditary Endocrine Tumours Research Group, Hungarian Academy of Sciences – Semmelweis University, Budapest, Hungary; 3) 2nd Department of Pathology, Semmelweis University, Budapest, Hungary; 4) MTA-SE Molecular Medicine Research Group, Hungarian Academy of Sciences – Semmelweis University, Budapest, Hungary; 5) Department of Laboratory Medicine, Semmelweis University, Budapest, Hungary.

Pancreatic neuroendocrine tumors (PNETs) are rare tumors arising from the endocrine islets of the pancreas. The prognosis and the therapeutical strategy of these tumors are determined by Ki-67 proliferation index and mitotic rate. Upon the proliferative phenotype, WHO classifies PNETs as G1 (least proliferative), G2 and G3 (most aggressive) tumors. MiRNAs are small, noncoding RNAs posttranscriptionally regulating gene expression. The expression of hsa-miR-21 has been shown to be associated with aggressive PNETs. Our aim was to detect novel miRNAs which expression can be used as prognostic biomarkers. Reanalysis of miRNA expression profiles of 40 PNETs were performed in datasets downloaded from Gene Expression Omnibus and differently expressed miRNAs upon tumor grade (discovery cohort) were identified. Of these the expression of 5 chosen miRNAs was evaluated in 67 histopatologically diagnosed PNET tissues by quantitative real-time PCR (validation cohort). Kaplan-Meier analysis was performed to analyze the role of tissue miRNA expression upon survival. Expression of 19 miRNAs were significantly different among tumors with different grade. Among 5 chosen miRNAs hsa-miR-21, hsa-miR-106b and has-miR-10a were upregulated in G3 PNETs compared to G1 and G2 PNETs. Survival analysis demonstrated a significant association between the expression of these three miRNAs and prognosis. The expression of hsa-miR-21, hsa-miR-106b and hsa-miR-10a may be useful prognostic biomarkers in PNETs.

831F

Colorectal cancer-upregulated long non-coding RNA lincDUSP regulates DNA damage response genes and promotes resistance to apoptosis. *M.E. Forrest[1], A. Saiakhova[1], L. Beard[3], V. Varadan[4], D. Buchner[1,2], P. Scacheri[1], S. Markowitz[1,3], A.M. Khalil[1,2,3].* 1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Department of Biochemistry, Case Western Reserve University, Cleveland, OH; 3) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH; 4) Department of Pathology, Case Western Reserve University, Cleveland, OH.

Long non-coding RNAs (lncRNAs) are powerful regulatory molecules involved in critical cellular processes that can control gene expression at multiple levels. LncRNAs are frequently dysregulated in many cancer types and may serve as useful therapeutic targets. Given the high tumor complexity and heterogeneity observed in colorectal cancer (CRC), we hypothesized that specific lncRNAs act as "oncogenic" RNAs to drive the colon tumor phenotype and disease progression. To identify candidate oncogenic lncRNAs in CRC, we leveraged RNA sequencing data from 22 colon tumor and matched normal colon samples from The Cancer Genome Atlas (TCGA), yielding ~200 differentially expressed lncRNAs. LincDUSP emerged as a notable candidate oncogenic lncRNA due to its consistently low expression in normal colon and high expression in a subset of colon tumors. Knockdown of lincDUSP in patient-derived colon tumor cell lines by LNA GapmeRs resulted in significantly decreased cell proliferation and clonogenic potential as well as significantly increased induction of apoptosis. Studies are currently underway to characterize the effects of lincDUSP overexpression in early-stage colon tumor cell lines using a dCas9-VP64 CRISPR activation system. In an effort to elucidate the mechanism of action of lincDUSP, we chose a three-part approach: 1) query transcriptome changes associated with altered lincDUSP expression, 2) map genomic occupancy, and 3) identify protein interactions. RNA-Seq was performed in patient-derived colon tumor cell lines with or without lincDUSP knockdown, yielding ~800 differentially expressed genes. NCI pathway enrichment analysis of this gene list showed enrichment of cell cycle control, DNA replication, and DNA repair pathways. Further, identification of genome-wide lincDUSP chromatin occupancy sites by ChIRP-Seq demonstrated association with genes involved in the replication-associated DNA damage response (including *PCNA*, *EXO1*, and *XRCC2*) and cell cycle control (including *MYC* and *CCND2*). Consistent with these findings, preliminary cell cycle phase analysis in colon tumor cell lines showed early S-phase arrest upon lincDUSP knockdown; further, lincDUSP knockdown showed increased γH2AX foci, indicating DNA damage response induction. Future studies to elucidate the mechanism of action of lincDUSP will involve identification of interacting protein effectors by ChIRP-MS and functional analysis of the effect of lincDUSP knockdown on DNA damage response pathways.

832W

Antisense long non-coding RNAs in breast cancer: A transcriptome-wide disruption. *S. Wenric[1,2], S. El Guendi[1], JH. Caberg[3], W. Bezzaou[1], C. Fasquelle[1], B. Charloteaux[4], L. Karim[4], B. Hennuy[4], P. Frères[2], J. Collignon[2], M. Boukerroucha[2], H. Schroeder[2], F. Olivier[2], V. Jossa[5], G. Jerusalem[2], C. Josse[1,2], V. Bours[1,3].* 1) GIGA-Research, Unit of Human genetics, University of Liège, Liege, Belgium; 2) University Hospital (CHU), Department of Medical Oncology, Liège, Belgium; 3) University Hospital (CHU), Center of Genetics, Liège , Belgium; 4) University of Liège, GIGA-Genomics Platform, Liège, Belgium; 5) Clinique Saint-Vincent (CHC), Department of Pathology, Liège, Belgium.

Introduction: Non-coding RNAs (ncRNA) represent at least 1/5 of the mammalian transcript amount, and about 90% of the genome length is actively transcribed. Many ncRNAs have been demonstrated to play a role in cancer. Among them, antisense long non-coding RNAs (lncRNAs), also called natural antisense transcripts (NAT) are RNA sequences which are complementary and overlapping to those of protein-coding transcripts (PCT). NATs were punctually described as regulating gene expression, and are expected to act more frequently in cis than other ncRNAs that commonly function in trans. **Materials and Methods**: In this work, 22 breast cancers expressing estrogen receptors and their paired healthy tissues were analyzed by strand-specific RNA sequencing. Differential expression analysis of all expressed antisense transcripts, differential correlation analysis between each antisense transcript and their paired protein coding genes, and read counts ratios variation analysis between the antisense transcripts and their paired protein coding genes have been performed. Besides providing information regarding the disruption happening both at the transcriptome-wide scale and at the scale of gene pairs, these three methods yielded candidate antisenses lists for which the association with survival of the corresponding protein coding gene has been tested on an independent cohort of 1060 breast cancer patients from the TCGA. **Results**: We observed a significant global over-expression of antisenses in breast cancer tumors. We also reported a disruption of mainly positive correlations between antisenses and protein-coding genes. Finally, all three lists of antisenses obtained with the different methods are enriched in genes significantly associated with survival.

833T

Single cell RNA sequencing identifies novel gene expression signatures in a mouse model of multiple myeloma treated with the cIAP antagonist LCL161. *N.E. Banovich[1], M. Boateng[1], L. Bergsagel[2], M. Chesi[2], J.J. Keats[1].* 1) TGen, Phoenix, AZ; 2) Mayo Clinic Arizona, Scottsdale, AZ.

Multiple myeloma (MM) is the second most common hematological cancer, accounting for roughly 2% of all cancer deaths. Unfortunately, MM remains incurable and nearly all patients experience relapse. Recently, we have identified a novel cIAP1 antagonist, LCL161, with remarkable anti-MM effects. Interestingly, unlike in other tumor types, our prior work found LCL161 does not directly induce apoptosis of MM cells, but rather drives activation and infiltration of macrophages to the tumor bed and phagocytosis of tumor cells. Furthermore, it appears the anit-MM activity of the agent relies on complex interactions between the tumor and the immune microenvironment. While LCL161 has been advanced to phase 2 clinical trials, the molecular mechanisms underlying the immune response remain poorly understood. We believe identifying cell-type specific gene expression changes after LCL161 treatment will aid in elucidating these mechanisms, potentially leading to additional and more targeted therapies. To this end, we collected whole spleens from three control and ten VK*MYC tumor bearing mice – three untreated, two treated with LCL161, two treated with cyclophosphamide, and three treated with bortezomib – for single cell RNA-sequencing (scRNA-seq). All samples were collected and processed in tandem 48 hours after treatment using the 10X genomics chromium platform. Ultimately, we measured mRNA levels in over 52,000 individual cells. We identified groups of cells with shared expression profiles, corresponding to known cell-types, using graph based clustering. Consistent with our previous work we observed a marked reduction in tumor cells and activation of the non-canonical NFkB pathway in mice treated with LCL161. Moreover, we identified immune cell-types with enrichments and depletions of cells treated with LCL161, compared to other tumor bearing mice. Focusing on cells annotated as macrophages, we found a clear separation between those treated with LCL161 and other conditions. Using a binomial model we identified over 500 upreglated genes in LCL161 treated macrophages. Of note, we identified two genes, *Saa3* and *Spp1*, which are expressed exclusively in LCL161 treated macrophages (~0.4% of total cells). Due to this small number cells, the signal is unobservable using traditional bulk RNA-seq; yet we believe it may play a critical role in the efficacy of LCL161. We are currently performing experiments to reveal the role of these genes in LCL161 mediated tumor phagocytosis.

834F

Investigation of a transcription factor network involved in exocrine pancreatic development and homeostasis reveals a putative tumor suppressor role and a novel genetic interaction. *J. Hoskins, E.M. Mobaraki, M. Zhang, I. Collins, A. Jermusyk, L. Amundadottir.* LTG, National Cancer Institute, Gaithersburg, MD.

In contrast to many cancers, early detection and treatment for pancreatic cancer has shown little improvement. Understanding the functional consequences of germline risk altering variants is one approach that may aid both of these aspects of cancer management. However, despite successful linkage and genome wide association (GWA) studies, much of the estimated heritability for pancreatic cancer remains unaccounted for. One suggested source of this missing heritability is interaction effects, either gene by gene or gene by environment, but the false positive rate of genome wide scans for such interaction effects is prohibitive due to the number of tests performed. Reducing the number of analyzed variants can greatly relax this constraint, but requires prior knowledge of candidates likely to interact. GWA studies to date have identified 16 loci with 20 independent signals associated with risk of pancreatic cancer among the European population, and four of these loci lie in or near genes encoding transcription factors (TFs) thought to be important for exocrine pancreas development and homeostasis. These TFs (HNF1A, HNF1B, PDX1 and NR5A2) form a putative network of positive feedback expression regulatory relationships based on previous studies with other tissues (e.g. liver and islet cells). Such functional interactions imply possible genetic interactions between these loci. Indeed, our preliminary pairwise interaction analyses revealed a significant interaction effect between the *HNF1A* and *HNF1B* risk loci ($P_{interaction}$ = 0.017 in 5113 cases and 8846 controls). Functional studies with inducible expression cell lines indicated reduced cell proliferation upon over expression of each of the four TFs, though the effects were most striking for HNF1A and HNF1B. ChIP-seq analysis of binding sites and time-course RNA-seq of downstream expression effects are underway for all four TFs, which should yield a more complete model of the gene regulatory effects of this pancreatic homeostasis TF network.

835W

Disruption of Mi2b and MBD2/3 corepressor functions mediates LINE-1 reactivation andtumorigenicity in human bronchial epithelial cells challenged with benzo(a)pyrene. *P. Bojang[1], K.S. Ramos[1,2,3].* 1) Pulmonary, Allergy and Critical Care, University of Arizona, Tucson, KY; 2) Division of Medical Toxicology and Precision Medicine, Department of Medicine, University of Arizona College of Medicine, Phoenix, Arizona; 3) Center for Applied Genetics and Genomic Medicine, University of Arizona Health Sciences, Tucson and Phoenix, Arizona.

Long Interspersed Nuclear Element-1 (LINE-1 or *L1*), the only autonomous retroelement that remains active in the human genome, is silenced early during embryonic development and reactivated by tobacco carcinogens such as benzo(a)pyrene (BaP). The genetic and molecular mechanisms of *L1* reactivation are of interest because retroelement expression is associated with poor prognosis in Non-Small Cell Lung Carcinoma (NSCLC). In this report, we show that challenge of human bronchial epithelial cells (HBECs) with BaP shifts the *L1* promoter from a heterochromatic to euchromatic state through the disassembly of the Nucleosomal and Remodeling Deacetylase (NuRD) complex followed by shifting of constituent proteins from the nucleus to the cytoplasm. Under these conditions, the loss of NuRD corepression is associated with transcriptional activation and accumulation of *L1* proteins and loss of Mi2β (Chromodomain helicase DNA-binding protein 4) interaction with the *L1* promoter. The repressor functions of Mi2β on *L1* required the DNA and ATPase binding domains, but not the chromo domain, and were confirmed in genetic ablation experiments where either Mi2β and Methyl-CpG Binding Domain proteins 2/3 (MBD2/3) downregulation increased retroelement activity. Inoculation of HBECs expressing constitutively active *L1* in SCID mice induced tumours that failed to grow, suggesting that epigenetic silencing of ectopic *L1* may contain the tumorigenic response. Together, these results identify Mi2β and MBD2/3 as key effectors of *L1* silencing in HBECs and raise important questions about the tumorigenicity of *L1 in vivo*.

836T

Genetic variations in alcohol-metabolizing genes (*GSTM1, GSTT1, CYP2E1, ADH2 and ADH3*) and pancreatitis risk in alcoholics. *V. Aaren[1], L.R.S. Girinadh[2], G. Sudhakar[1].* 1) Department of Human Genetics, Andhra University, Visakhapatnam, Andhra Pradesh, India; 2) Department of Gastroenterology and Hepatology, Andhra Medical College, King George Hospital, Maharanipeta Visakhapatnam- 530002, India.

The pancreas produces hormones such as trypsin, which is important for digestion, and insulin, which controls blood sugar levels. Long-term alcohol use can cause pancreatitis, an inflammation of the pancreas. The mortality rate of patients with alcoholic pancreatitis is about 36 percent higher than that of the general population. Approximately 50 percent of patients with alcoholic pancreatitis die within 20 years of onset of the disease. To determine the genetic susceptibility to alcoholic pancreatitis (AP) in South Indian populations, we investigated the genetic variations of exon 3, Arg47His G/A of *ADH2/ ADH1B*; exon 8, Ile349Val G/A of *ADH3/ADH1C*; intron 6, T7678A of *CYP2E1; GSTM1*; and *GSTT1* genes in 100 cases and an equal number of age, ethnic and sex-matched controls. A detailed history of each case including the age, alcohol consumption (quantity and type of beverage), smoking, tobacco intake and other lifestyle factors were recorded. *PCR*-restriction fragment length polymorphism (RFLP)-based analysis of blood DNA samples were analyzed using gene and locus specific oligos. Gene polymorphisms were compared between alcoholic pancreatitis subjects (AP) vs alcoholic controls (AC) and alcoholic pancreatitis (AP) vs healthy controls (HC). Genotypic frequencies were calculated and odds ratios (OR) were determined using 2 × 2 contingency tables with 95% confidence intervals. Statistical significance for Odds Ratio (OR) was calculated by chi-square test using SPSS Version 16 software. Fisher's exact test was used when expected cell frequency was less than five. The association of the gene polymorphisms in patients and controls was examined using OR and Chi-square tests with CI 65%. $P<0.05$ was considered as significant. A significant association between *ADH3 Ile349Val* polymorphism and AP was observed, when compared to the AC controls ($P=0.0072$) and HC controls ($P=0.024749$). Markers *CYP2E1* T7678A failed to show any significant association with AP compared to controls ($P=0.42736$). *GSTM1* ($P=0.87417$ and $P=0.32052$), and *GSTT1* ($P=0.39718$ and $P=0.305749$) genotypes were not significantly associated with AP when compared to AC controls and HC controls. Similarly, *ADH2/ ADH1B* Arg47His showed a lack of association with AP. Our research on the causes of alcoholic pancreatitis may support more effective disease management, provide hope for a potential cure and aid in developing strategies for the prevention or treatment of alcoholic pancreatitis.

837F

Determinants and consequences of ribosomal poverty and subsistence in *C. elegans*. *E. Cenik, N. Jain, AZ. Fire.* Pathology, Stanford University, Stanford, CA.

Mutations in ribosomal protein (RP) genes can lead to diverse pathologies that include heritable ribosomopathy syndromes and several cancers. For 16 such genes, heterozygous loss-of-function leads to a syndrome, Diamond Blackfan Anemia (DBA), with variable anemia, bone marrow failure, cancer predisposition, and organ malformation. We have applied the power of *C. elegans* genetics to characterize mechanisms responding to RP heterozygosity and to understand the substantial phenotypic variation in human ribosomopathies. With Cas9, we generated heterozygous knockouts for six *C. elegans* DBA homologs. Paralleling the phenotypic variability in human syndromes, these mutant heterozygotes exhibited a variety of phenotypes with different severity, including slow growth, dominant-lethality, and sterility. RNAseq of heterozygote populations showed a global feedback process regulating ribosomal gene expression. Imbalances in large and small subunits of the ribosome exhibited different consequences. In addition, shared changes in gene expression highlight disturbances in various signaling pathways and metabolic processes. Ongoing work on RP heterozygosity is directed toward identifying specific genes/pathways where intervention will mitigate pathogenic consequences. Our RP heterozygote strains combined with this animal's intensively characterized (and transparent) embryogenesis provide an additional window on an extreme (and likely biologically relevant) regulatory situation. To date, few studies have addressed consequences of a complete loss of ability to produce new ribosomal components. Although organisms clearly require new ribosomes for growth and proliferation, certain cell populations (such as platelets) can persist and provide biological contributions without new ribosome synthesis. Ribosomes are a developmentally diverse and dynamic population. Surprisingly, we find for *C. elegans* that RP null homozygotes can produce fully functional first stage larvae. These animals, relying on maternal but not zygotic contribution of specific ribosomal components, are capable of completing embryogenesis and of fully differentiating every tissue type in the complex organism. These observations suggest that all diversity required of the translational machinery during functional tissue diversification can derive from the pool of RP transcripts and protein components present in the oocyte.

838W

Genetic variations in ERCC2 gene and the risk of developing head and neck cancer in an Indian population. *K. Chukka[1,2], Z. Vishnuvardhan[3], U. Radhakrishna[4].* 1) Department of Biotechnology, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India; 2) Department of Biochemistry, University College of Sciences, Osmania University, Hyderabad- 500007, India; 3) Department of Botany and Microbiology, Department of Environmental Sciences, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur, India; 4) Department of Obstetrics & Gynecology, Oakland University William Beaumont School of medicine. MI, USA.

Head and Neck Cancers (HNCs) constitute a difficult challenge for oncologists across the globe, with one person dying every hour of every day. HNCs can disfigure the face, strip away the voice and rob one of the basic ability to eat, drink and swallow. The psychosocial impact is severe. HNCs are a significant problem in India, and constitute approximately one-third of cancer cases in contrast to 4-5% in the developed world. *XRCC1* (X-ray cross-complementing group 1) and *ERCC2* (excision repair cross-complementing group 2) are two major DNA repair proteins. Polymorphisms of these two genes have been associated with altered DNA repair capacity and increased cancer risk. In this study, 60 HNC patients and 60 matched controls were selected from the Department of Radiotherapy, Government General Hospital, Guntur, India. The mean age of participants was 54.23 years of age, and most were 40 to 60 years of age. Most participants were illiterate daily wageworkers and accustomed to hard work. Smoking, alcoholism and high consumption of spices and very hot beverages are predisposing factors in this region. We explored the gene-environment interaction between *ERCC2* (Lys751Gln) polymorphism and smoking exposure, alcohol consumption and dietary factors in the development of HNC by PCR-RFLP. Data showed no significant allelic associations for *ERCC2* (Lys751Gln). 86.66% (N=52) of the HNC cases were heterozygous and only 13.32% (N=8) of the cases were homozygous. All the controls (N=60) were heterozygous. The X^2- value calculated for heterozygotes frequency in controls and cases did not differ significantly showing that there is no direct association between SNPs of *ERCC2* gene and outcome of HNC. The analysis of other genetic factors along with this parameter may provide the answer for personal risk assessment and an opportunity for individualized therapy.

839T

MiR-450a and miR-450b-5p negatively impact the tumorigenic potential of ovarian epithelial cancer cells. *B.R. Muys[1,2], J.F. Sousa[2], L.F. Araújo[1,2], F. Turaça[1], I.S. Rodrigues[3], D.O. Vidal[4], W.A. Silva Júnior[1,2].* 1) Ribeirã Preto Medical School, Ribeirã Preto, Sã Paulo, Brazil; 2) Center for Cell-Based Therapy (CEPID/FAPESP); National Institute of Science, and Technology in Stem Cell and Cell Therapy (INCTC/CNPq), Regional Blood Center of Ribeiroã Preto, Riberã Preto, Brazil; 3) Faculty of Philosophy, Sciences and Literature of Ribeiroã Preto, University of Soã Paulo, Ribeiroã Preto, Brazil; 4) Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, Brazil.

Ovary cancer is the fifth largest cause of cancer-related deaths for women worldwide, with patients presenting extremely low 5-year survival rates. Differently from most types of solid tumors, which disseminate through the bloodstream, ovarian tumor cells frequently metastasize via transcoelomic dissemination, spreading to the peritoneum and many peritoneal organs. MicroRNAs (miRNAs), non-coding RNAs around 22 ribonucleotides long, play a critical role in gene expression regulation in virtually all biological processes. miRNAs induce gene silencing of their targets through translation repression or mRNA cleavage. In a previous work we showed that miR-450a and miR-450b-5p, two functionally underexplored miRNAs in gynecologic tumors, were downregulated in ovarian tumor cell lines in comparison to normal tissue. The aim of this work was to verify the hypothesis that both miR-450a and miR-450b-5p display tumor suppressor functions in ovarian cancer. For this purpose, we utilized in vitro assays and an in vivo xenographic model. Using a lentiviral expression vector, miR-450a and miR-450b-5p were stably overexpressed in SKOV-3 and A2780 ovarian tumor cell lines. First, apoptosis was analyzed in transduced cells submitted to low adherence conditions for 24 hours. We observed that miR-450a overexpression resulted in lower viability in A2780 and SKOV-3 cells in comparison to the control (p<0.001 and p<0.05, respectively). We also measured cell migration and invasion rates using transwell assays. Overexpression of miR-450a or miR-450b resulted in lower migration for both cell lines (p<0.001 and p<0.01, respectively). The same result was verified for invasion after up-regulation of miR-450a (p<0.001) or miR-450b in A2780 and SKOV-3 cell lines (p<0.05 and p<0.01, respectively). Transduced A2780 cells were also injected intraperitoneally in 9-11 weeks immunodeficient mice and the weight of dissected tumors compared after 29 days. In this in vivo xenographic model of ovarian cancer, we verified that both miRNAs impaired tumor growth (p<0.001 for miR-450a and p<0.01 for miR-450b) and therefore their ability to metastasize, which corroborates our in vitro assays. To date, our analysis points to the scenario of miR-450a and miR-450b-5p playing functions proven to be relevant to the tumorigenic process of ovarian epithelial tumors. Financial Support: CAPES and FAPESP..

840F

Integrative genome analysis of somatic p53 mutant osteosarcomas identifies Ets2 dependent regulation of small nucleolar RNAs by mutant p53 protein. *R. Pourebrahimabadi[1], Y. Zhang[1], B. Liu[1], R. Gao[1], PP. Lin[2], MJ. McArthur[3], MC. Ostrowski[4], G. Lozano[1].* 1) Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Orthopedic Oncology, University of Texas MD Anderson Cancer Center, Houston, TX; 3) Department of Veterinary Medicine and Surgery, University of Texas MD Anderson Cancer Center, Houston, TX; 4) Department of Cancer Biology & Genetics, The Ohio State University, Columbus, OH.

   TP53 is the most frequently mutated gene in cancer. Many mutant p53 proteins exert oncogenic Gain-of-Function (GOF) properties that contribute to metastasis, but the mechanisms mediating these functions remain poorly defined *in vivo*. To elucidate how mutant p53 GOF drives metastasis, we developed a traceable somatic osteosarcoma model that starts with either a single p53 mutation (p53R172H), or *p53* loss specifically in osteoblasts. p53 mutant mice developed osteosarcomas with higher metastasis rate as compared to *p53*-null mice. Comprehensive transcriptome analysis of sixteen tumors identified a cluster of small nucleolar RNAs (snoRNAs) being highly upregulated in p53 mutant tumors. Regulatory element analysis of snoRNA genes identified an Ets2 binding site to be the most common transcription factor binding site enriched in those genes. Homozygous deletion of *Ets2* in p53 mutant mice resulted in strong downregulation of snoRNAs and reversed the pro-metastatic phenotype of mutant p53, but had no effect on the rate of tumor formation, and mice continued to produce osteosarcoma with 100% penetrance. These studies identify Ets2 inhibition as a potential therapeutic vulnerability in p53 mutant cancers.

841W

NF1 mutation structure-function analyses using a full-length mouse cDNA. *D. Wallis, K. Li, M-J. Chen, A. Challa, J. Li, H. Liu, J. Kang, B. Korf, R. Kesterson.* Genetics, University of Alabama at Birmingham, Birmingham, AL.

   To determine the functional consequences of NF1 patient mutations and variants of unknown significance, we have established a heterologous cell culture expression system using a full-length mouse *Nf1* cDNA and human cell lines. We demonstrate a full-length mouse *Nf1* cDNA (longest known coding *Nf1* mRNA of 2841 amino acids) placed downstream of the CMV promoter will produce a >250 kDa neurofibromin protein capable of modulating Ras signaling when expressed in wild type or *NF1*-deficient cells. Knockout HEK293 and human iPS cell lines have been established wherein *NF1*-deficiency was induced with CRISPR/Cas9. When the m*Nf1* cDNA is transiently transfected into HEK293 cells (or expressed in stable cell lines), western blot shows overexpression of neurofibromin in both *NF1* replete (*NF1*+/+) and deficient (*NF1*-/-) cells. Moreover, the re-expression of mouse neurofibromin from the cDNA is able to suppress the elevated p-ERK seen in the *NF1*-/- HEK293 cells, thereby demonstrating that the mouse neurofibromin protein is functional in the human cell line. This system provides the ability to assess the functional effect of *NF1* genetic variants found in patients with neurofibromatosis. We have thus far found that a relatively frequent cryptic splicing mutation (c.1466A>G, p.Tyr489Cys reported 34 cases that leads to a 62bp deletion in exon 13) will however produce functional neurofibromin if a full length mRNA is made (i.e. mouse p.Tyr489Cys *Nf1* cDNA is functional). The ability to express mouse *Nf1* cDNAs in human cell lines has also led to the validation of antibodies that recognize mature and truncated versions of neurofibromin proteins with different specificity to either C-terminus (SC-67, rabbit polyclonal) or N-terminus (H-12, mouse monoclonal) regions. Additional studies are underway with other NF1 mutations, including out-of-frame deletions (c.2393_2408del16), nonsense point mutants (c.2041C>T, p.R681*), and missense mutations (c.5425C>T, p.Arg1809Cys and c.3827G>A, p.R1276Q) that are associated with different phenotypes clinically.

842T

Whole exome sequencing of patient cell lines with highpersensitivity to radiation exposure identifies ATIC as a novel target for chemoradiosensitization. *X. Liu[1], D.U. Paila[2], N.S. Teraoka[1], J. Wright[1], X. Huang[1], A. Quinlan[3], R. Gatti[4], P. Concannon[1].* 1) University of Florida Genetics Institute, Gainesville, FL; 2) University of Virginia,Department of Public Health Sciences, Charlottesville, VA; 3) University of Utah, Department of Human Genetics and Department of Biomedical Informatics Salt Lake City, UT; 4) UCLA Department of Pathology and Laboratory Medicine, and Department of Human Genetics, Los Angeles, CA.

Purpose: A frameshift mutation in the gene encoding 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC), a bifunctional enzyme that catalyzes the final two steps of the purine *de novo* biosynthetic pathway, was identified during a screen of cell lines displaying unexplained hypersensitivity to ionizing radiation. Functional studies were carried out to determine if ATIC inhibition was radiosensitizing, and if so, to elucidate the mechanism of this effect and to determine if small molecule inhibitors of ATIC could act as effective radiosensitizing agents. **Methods and Materials:** Both siRNA knockdown and small molecule inhibitors were used to inactivate ATIC in cell culture. Clonogenic survival assays, the neutral comet assay, and γH2AX staining were used to assess the effects of ATIC inhibition on cellular DNA damage responses. **Results:** Depletion of ATIC or inhibition of its enzymatic activity significantly reduced the surviving fraction of cells in clonogenic survival assays in multiple cancer cell lines. In the absence of ionizing radiation exposure, ATIC knockdown or chemical inhibition activated cell cycle checkpoints, shifting cells to the more radiosensitive G2/M phase of the cell cycle, and significantly depleted cellular ATP, but did not result in detectable DNA damage. Cells in which ATIC was knocked down or inhibited and then treated with ionizing radiation displayed increased numbers of DNA double-strand breaks, and a delay in the repair of those breaks relative to irradiated, but otherwise untreated, controls. Supplementation of culture media with exogenous ATP reversed these phenotypes, indicating that the ATP depletion resulting from ATIC inhibition is critical to its radiosensitizing effects. **Conclusions:** These findings implicate ATIC as an effective, and previously unrecognized, target for chemoradiosensitization and more broadly suggest that the energy balance in cells may have an underappreciated role in modulating the efficiency of DNA damage responses that could be exploited in radiosensitizing strategies. .

843F

Transcriptomic changes in NF1 deficient cells. *C. Skefos, M-J. Chen, A. Challa, K. He, H. Liu, BR. Korf, R. Kesterson, D. Wallis.* University of Alabama at Birmingham, Birmingham, AL.

In efforts to define pathways responsible for the initial development of neurofibromas, we have been evaluating transcriptomic changes in a non-malignant *NF1* null cell line. We created *NF1* null HEK293 cells using CRISPR/Cas9 technology and performed transcriptomic analysis in comparison to HEK293 wild type (*NF1* +/+) cells. We have found that there are 352 genes that are up-regulated and 445 genes down-regulated by 2 fold or more (q< 0.05) in the null cells. Ingenuity Pathway analysis has revealed that the top canonical pathways include: GABA Receptor Signaling (p=9.73E-06), Regulation of the Epithelial-Mesenchymal Transition Pathways (p=1.55E-04), and Axon Guidance Signaling (p=1.8E-04). In addition, the top three Diseases and Biological functions associated with these expression changes are "Dermatological Diseases and Conditions" (p<5.43E-06), "Organismal injury and abnormalities" (p<2.64E-05), and "Cancer" (p<2.64E-05). Next, we compared our data set to nine other published studies that involve analysis of cells or tumors that are deficient for NF1 and a KRAS-mutant cross-tumor screen. We found that 125 genes reoccur in our sample and others. Panther Gene List Analysis indicates that over a third are involved in binding (GO: 0005488). Based on our interest, we chose to validate a few genes via q-RT-PCR and see statistically significant differences between genotypes for CACNG2, ETV4, LAMB3, and SLITRK5. Though still in the validation stages, our differential expression data support continued evaluation for known disease targets and therapeutic inhibitors (e.g., FOSL1 and alisertib or BET inhibitors, HNMT and ketotifen or other mast cell stabilizers, RUNX and Ro5-3335, and inhibition of the GABA-ergic system). Some genes we've identified could be used as biomarkers or prognostic indicators (ANXA1, LAMB3, ETV4). Finally, we have implicated new genes and potentially druggable targets in the NF1 phenotype. Our identification of SLITRK5 and SLIT2 is a novel association for this gene family involved in both neurite outgrowth and the development of neoplasia with NF1. LAMB3 also offers a potential therapeutic target as its over-expression has been associated with KRAS-driven cancers and poorer survival while its knock-down was able to selectively initiate cell death. We are working to further investigate some of our hits and complement our RNA-Seq data with proteomics data with the goal of defining new pathways and functions of neurofibromin.

844W

Consequences of miR-122 loss in the development of hepatocellular carcinoma. *P.N. Valdmanis[1,2], K. Chu[2], F. Zhang[2], M.A. Kay[2].* 1) Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Pediatrics and Genetics, Stanford University, Stanford, CA.

MicroRNAs play important roles in fine-tuning gene expression and directing cell fate in various tissues. miR-122 is a highly-expressed liver microRNA that accounts for up to 70% of hepatic microRNAs. miR-122 expression is activated perinatally and aids in regulating cholesterol metabolism, promoting terminal differentiation of hepatocytes and suppressing genes that are active in liver embryonic development. Disrupting the expression of miR-122 can re-activate these embryo-expressed adult-silenced genes and this ultimately leads to the development of hepatocellular carcinoma (HCC) in mice. We performed small RNA sequencing and RNA-seq at various time points after genomic excision of miR-122 to determine the pathways and transcriptional progression to oncogenesis. The loss of miR-122 led to specific and progressive increases in imprinted clusters of microRNAs and mRNA transcripts at the *Igf2* locus and the *Dlk1-Dio3* locus on mouse chromosome *12qF1* that are typically restricted to embryo development. This progression can be curbed by re-introduction of exogenous miR-122. Interestingly, the chromosome *12qF1* locus is also enriched for adeno-associated virus (AAV) genomic integrations in liver tumors after gene therapy approaches using AAV. From a quantitative standpoint, we have utilized this system to elucidate the degradation pathway of miR-122 isoforms. We also demonstrate that mRNA targets of other abundant hepatic microRNAs such as miR-21 and let-7 are functionally repressed leading to global hepatic transcriptional de-regulation though this is alleviated though administration of a various microRNA-like small hairpin RNAs. Together, this reveals a transcriptomic framework for the hepatic response to loss of miR-122 and the outcome on other microRNAs and their cognate gene targets. Recognizing the progression and reversibility from miR-122 loss to HCC can provide insight into disease pathogenesis.

845T

Modeling human cancer syndromes using TALEN and CRISPR/Cas9 mediated genome editing in *Xenopus tropicalis*. *K. Vleminckx[1,2,3], T. Naert[1,3], R. Noelanders[1,3], D. Dimitrakopoulou[1,3], T. Van Nieuwenhuysen[1].* 1) Department of Biomedical Molecular Biology , Ghent University, Ghent, Belgium; 2) Centre for Medical Genetics, Ghent University, Ghent, Belgium; 3) Cancer Research Institute Ghent, Ghent, Belgium.

CRISPR/Cas9 and TALEN mediated genome editing is creating unique and unmatched opportunities in several research fields. For the first time it is now possible to create functional gene knockouts in a number of model organisms. The tetrapod species *Xenopus tropicalis* is extremely well positioned for modeling human disease. It shares with zebrafish the aquatic habitat and easy manipulations associated with its external development. However, it manifests unique features making it a more favorable and powerful organism for modeling human genetic disease. (1) unlike zebrafish (and *Xenopus laevis*), *Xenopus tropicalis* has a true diploid genome. Hence, gene disruption studies are not suffering from redundancy. (2) The *Xenopus tropicalis* genome shows high synteny with humans, which greatly facilitates the identification of orthologs of human disease genes. (3) Gene manipulations can be restricted to specific tissues and organs via targeted injection of individual blastomeres. More than 60 genes have meanwhile been analyzed in our experimental CRISPR/Cas9 pipeline, including genes associated with retinal disease and limb malformations. Furthermore, we have recently generated the first genetic tumor models in *Xenopus tropicalis*. Via mosaic targeting of the tumor suppressor gene *apc* we generated tadpoles that rapidly (< 1.5 months) and efficiently (>90%) developed a range of neoplasia characteristic for Familial Adenomatous Polyposis, including intestinal neoplasia, desmoid tumors and medulloblastomas (Van Nieuwenhuysen *et al*., Oncoscience 2015). Similarly, using CRISPR/Cas9 techniques we found that *rb1/rbl1* double mosaic mutant tadpoles rapidly develop retinoblastoma (Naert *et al*. Sci. Rep. 2016). More recent work also indicates the possibility of modeling T-cell acute lymphoblastic leukemia (T-ALL) and bladder cancer. The rapid kinetics of tumor development in the tadpoles and froglets pave the way for use as pre-clinical models. Additionally, these tumor models provide unique possibilities for fast identification of modifier genes and novel drug targets by multiplexed CRISPR/Cas9 gRNA injections (*tumor suppressor gene + candidate therapeutic target gene)*. We present our first promising results with multiplexed gene targeting in desmoid tumors. We believe that genetic *Xenopus tropicalis* models offer a unique experimental platform that can be easily plugged into the research lines of several groups active in the field of human cancer and medical genetics.   .

846F

Five cases report with maple syrup disease over a period of 16 years: Metabolic screening, detection of inborn errors of metabolism at the Hospital para el Nino Poblano, Mexico. *P. Sánchez Meza[1], J.M. Aparicio-Rodríguez[2,4], M.L. Hurtado-Hernandez[3]. 1) Hematology; 2) Genetics; 3) Cytogenetics, Hospital para el NiÑo Poblano, Puebla; 4) Estomatology Benemerita Universidad Autonoma de Puebla, Mexico.*

The congenital inborn errors of metabolism (IEM) are detecTABLE diseases since 1908. Some of the most frequent IEM reported are fenilketonuria, galatosemy as albinism, cystinuria and porfiria and more rarely maple syrup urine disease reported in this study. It is caused by a deficiency of the branched-chain alpha-keto acid dehydrogenase complex (BCKDC), leading to a buildup of the branched-chain amino acids (leucine, isoleucine, and valine) and their toxic by-products are detected in the blood and urine. Case reported. 102 clinical cases with polymalformed muscle esqueletal and skull and dental dismorfies from this Hospital were already published Ma. De Lourdes Hurtado-Hernández et al, 2009, Figure I. Only 5 patients with maple syrup urine disease have been reported in a 16 years period of time Table 1, where the clinical evolution of one of the patients is reported in this study. Conclusions. Actually, the IEM are defined as monogenic inherited diseases or mendelian, due to a metabolic error for a protein or enzyme absence. It might be incompatible with the patient life and sometimes if the patient lives it will modify its quality of life, especially in a severe metabolic disease as it is maple syrup urine disease. Key Words: Inborn errors of metabolism, protein, enzyme and maple syrup.

847W

The use of livers with metabolic disease for domino transplantation: Can a patient with an inherited metabolic disorder act as a liver transplant donor? *P.A. Levy[1], K. Gallagher[1], N. Ovichinsky[2], RM. Cunningham[2], D. Kogan-Liberman[2], S. Bellemar[3], M. Kinkhabwala[3], D. Jan[3]. 1) Pediatrics, Genetics, Children's Hospital at Montefiore, Bronx, NY; 2) Pediatrics, GI & Nutrition, Children's Hospital at Montefiore, Bronx, NY; 3) Surgery, Montefiore Medical Center, Bronx, NY.*

Domino transplantation involves the transplantation of the native liver from a transplant recipient into another patient. It is generally assumed that patients with inherited metabolic disease that primarily affect the liver are not good candidates to be donors due to fear that the disease would be expressed in the recipient. We recently had a 4 year old patient with propionic aciduria who was receiving a liver transplant from a deceased donor. Another patient, a 14-year-old male with acute liver failure of unknown etiology, had no suitable liver for transplant. It was decided that the 4-year old's explanted liver would be used for the 14 year old as a temporizing measure. Post-op results showed improvement in organic acids for the propionic aciduria patient: 3-OH propionic 1996±699 pre , 172±53 post; propionylglycine 1606±1262 pre, 94±123 post; methylcitric 307±33 pre, 295±247 post). Normal values for 3-OH propionic 2-42 mmol/mol Cr, propionylglycine ≤ 2 mmol/mol Cr, and methylcitric ≤ 6 mmol/mol Cr. The recipient of the liver with propionic acidura, showed only a slight rise in metabolites (3-OH propionic 13.3±11 post; propionylglycine 0±0 post; methylcitric 15±11 post). We had anticipated that medical and nutritional therapy for propionic aciduria would be required post operatively, but this proved not to be necessary. There are reports of inadvertent liver transplants where the donor was found to have an inherited metabolic disorder when the recipient became ill. Ornithine transcarbamylase was fatal in the recipient. There are also reports of domino liver transplants with maple syrup urine disease, and methylmalonic aciduria with good results in the recipient. Our report is the first with propionic aciduria that has been reported. Our patients are now almost 2 years out from transplant and are doing well. With an increasing need for liver grafts to donate, this experience suggests that metabolic patients who receive a liver can have their liver used in another patient with some known disorders. We have shown that with Propionic aciduria, domino transplantation is successful. This could open a pool of donors that in the past might have been summarily rejected. .

848T

CPT1A p.P479L and long QT syndrome in Northern BC: Evidence for an increased risk for symptomatic hypoglycemia and implications for management. *S.A. Collins[1], S. McIntosh[1], C. Ens[2], G. Sinclair[3], H. Vallance[3], S. Stockler[4], C. Greenberg[5], S. Sanatani[2], L. Arbour[1].* 1) Dept of Medical Genetics, University of British Columbia, Victoria BC, Canada; 2) Division of Cardiology, Dept of Pediatrics, BC Children's Hospital, Vancouver BC, Canada; 3) Dept of Pathology & Laboratory Medicine, University of British Columbia, Vancouver BC, Canada; 4) Dept of Pediatrics, University of British Columbia, Vancouver BC, Canada; 5) Dept of Pediatrics and Child Health, University of Manitoba, Winnipeg MB, Canada.

Usually rare, autosomal dominant Long QT syndrome (LQTS) is common in First Nations people of Northern BC (1/125) due to 2 different pathogenic variants in *KCNQ1* (p.V205M and p.R591H) originating in 2 different communities. LQTS is a potentially fatal condition characterized by a prolonged QT interval, corrected for heart rate (QTc) on electrocardiogram. Reflective of prolonged repolarization in the cardiac cycle, there is risk of serious arrhythmia that can deteriorate to ventricular fibrillation and cardiac arrest. Loss-of-function mutations in *KCNQ1* (LQTS1) and *KCNH2* (LQTS2) have recently been shown to also predispose to symptomatic postprandial hyperinsulinemia due to the role of potassium channels in insulin secretion. Beta-blockers are commonly prescribed for children with LQTS; however, hypoglycemia is a potential side-effect of treatment. Recently, it has been shown (Poterucha et al 2015) that children with LQTS on beta-blocker therapy may be at risk for hypoglycemia (3% N=9/322); however, it remains unclear under what circumstances and whether secondary genetic factors can influence the risk. Of note, one case in the above study was subsequently diagnosed with a metabolic disorder. Alarmingly, hypoglycemia is known to prolong the QTc possibly increasing risk for cardiac events. This potential risk for hypoglycemia is a concern in the BC First Nations LQTS population since ~50% of the Northern BC First Nations population, in general, are carriers for the *CPT1A* p.P479L variant (25% homozygous). This variant, common in Northern Indigenous populations, is important in fatty acid oxidation during fasting and may predispose to hypoglycemia. Currently, it is not known how many mutation positive LQTS patients in this population are also carriers of the *CPT1A* p.P479L variant. We present a case report of four First Nations children between the ages of 2.5 and 4 years who presented with documented hypoglycemic seizures. All of the children were carriers of a pathogenic *KCNQ1* variant, were heterozygous or homozygous for the *CPT1A* p.P479L variant, and were taking beta-blockers. All four children presented with seizures in the morning, presumably after fasting through the night. We propose that the *CPT1A* p.P479L variant in combination with a LQTS causing variant in *KCNQ1* increases risk for hypoglycemia in this population. We will describe the multiple mechanisms of hypoglycemia involved and considerations for optimum management strategies to lower risk.

849F

Chart review is insensitive to ascertain pathogenicity of MODY gene variants of unknown significance. *J. Goehringer[1], B. Federman[4], K. Nori[4], U. Mirshahi[1], T. Govina[1], Y. Hu[1], J. Williams[1], T. Pollin[3], A. Shuldiner[2], D. Carey[1], Geisinger-Regeneron DiscovEHR Collaboration.* 1) Geisinger Medical Center, 100 North Academy Ave, Danville PA 17822; 2) Regeneron Genetics Center, 777 Old Saw Mill Road, Tarrytown, NY 10591; 3) University of Maryland, 800 West Baltimore Street, Baltimore MD 21201; 4) Sarah Lawrence College, 1 Mead Way, Bronxville NY 10708.

Maturity onset diabetes of the young (MODY) is an autosomal dominant form of monogenic diabetes that accounts for 1-2% of diabetes. While usually characterized as onset before age 25 in lean individuals, and not requiring insulin, many patients do not fit these criteria. Diagnosis is complicated by phenotypic heterogeneity, overlap with more common forms of diabetes mellitus (DM), and uncertain pathogenicity of MODY variants, leading to mis- and under-diagnosis. In this study we determined the value of systematic chart review of MODY gene variant carriers for variant interpretation. Within the DiscovEHR Cohort exome sequence database of 60,873 exomes, we identified carriers of rare (MAF cutoff 0.002) coding variants in *HNF4A*, *GCK*, and *HNF1A*, the genes that account for ~90% of known MODY cases. The variant data is linkable to longitudinal electronic health records, which contain diagnosis, medication, and laboratory data. We focused on 48 carriers of MODY gene variants of unknown clinical significance, and 48 non-carriers matched for age, sex, and BMI. This clinical population has a high prevalence of obesity (mean and median BMI = 30 kg/m[2]), which is a confounding factor in diagnosing MODY. Chart review was blinded and included 48 carriers of variants in *HNF4A* (14), *GCK* (7), *HNF1A* (26) and one patient with *HNF1A* and *HNF4A* variants. Charts were reviewed for DM-associated factors (fasting blood glucose [FBG] > 125 mg/dL or hemoglobin A1c [HbA1c] ≥6.5%) or pre-DM (FBG 100-125 mg/dL or HbA1C 5.7 - 6.4%), family history of DM, medication, age of diagnosis, and diabetic complications. Chart review yielded no documentation of MODY diagnoses for carriers. Overall, chart review did not show a difference in diabetes prevalence between gene variant carriers and non-carriers. Of the carriers, 24% had documented DM, and 39% were suspected to have pre-DM or DM based on FBG or HBA1c values. Of the non-carriers, 33% had documented DM and 48% were suspected to have pre-DM or DM. There was no difference in the age at diagnosis between carriers and non-carriers. We conclude that chart review of existing medical data does not facilitate classification of MODY variants due to phenotypic overlap of MODY with other forms of DM, inconsistencies in clinical documentation, and missing data. Future studies to determine pathogenicity of MODY variants of unknown significance will require information not systematically captured in the EHR or part of standard medical care.

850W

Genotype-phenotype and structure-phenotype correlations of the insulin receptor gene mutations in patients with severe insulin resistance. *J. Hosoe[1], H. Kadowaki[2], F. Miya[3,4], M. Takakura[1], K. Aizu[5], I. Miyata[6], T. Kawamura[7], K. Satomura[8], T. Ito[9], K. Hara[10], M. Tanaka[11], H. Ishiura[11], S. Tsuji[11], K. Suzuki[1], T. Tsunoda[3,4], N. Shojima[1], T. Yamauchi[1], T. Kadowaki[1].* 1) Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Pediatrics, Sanno Hospital, Tokyo, Japan; 3) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 4) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Division of Endocrinology and Metabolism, Saitama Children's Medical Center, Saitama, Japan; 6) Department of Pediatrics, The Jikei University School of Medicine, Tokyo, Japan; 7) Department of Pediatrics, Osaka City University Graduate School of Medicine, Osaka, Japan; 8) Department of Pediatric Nephrology and Metabolism, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Japan; 9) Department of Pediatrics, Atsugi City Hospital, Kanagawa, Japan; 10) Department of Endocrinology and Metabolism, Saitama Medical Center, Jichi Medical University, Saitama, Japan; 11) Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

Mutations in the insulin receptor (*INSR*) gene cause a clinical spectrum of severe insulin resistance syndromes that range from Donohue syndrome (DS, the most severe) to Rabson-Mendenhall syndrome (RMS, intermediate severity) and type A insulin resistance (type A-IR, the least severe). Here, *INSR* was analyzed and five previously unreported mutations and a deletion that removed exon 2 were identified in four patients with extreme insulin resistance (one patient with DS, two with RMS, and one with type A-IR). The patient with DS had a previously unknown mutation, c.1969G>T (p.V657F), in the second fibronectin type III domain (FnIII-2), which contains a part of the insulin binding site and an α-β cleavage site. Furthermore, one of the two patients with RMS was a compound heterozygote for two novel mutations, c.2504G>T (p.S835I) and c.2525C>T (p.A842V), both of which were also within FnIII-2. A patient with type A-IR had a c.3160G>A (p.V1054M) mutation in the tyrosine kinase domain (TK) of *INSR*. Insulin proreceptor processing and the activation of signaling cascades were impaired by the mutation in FnIII when expressed in Chinese hamster ovary cells. Using mutation data from the NCBI ClinVar database, Human Gene Mutation Database (HGMD), and UniProt database, we analyzed the distribution of 82 *INSR* missense mutations for a correlation between the localization of these mutations and the severity of insulin resistance. We found that missense mutations in the FnIII domains of *INSR* resulting in DS were significantly more frequent than those resulting in type A-IR. In addition, mutations in the TK domain were found at a significantly higher frequency in patients with type A-IR than in patients with DS or RMS. Next, we analyzed missense mutations in the FnIII and TK domains by a structural bioinformatics approach to determine the correlation with phenotypic expression. This analysis revealed that missense mutations predicted to severely affect formation of the hydrophobic core of FnIII and domain stability caused DS, while mutations predicted to result in localized destabilization of FnIII and not affect domain folding led to RMS. Additionally, we showed that a number of mutations situated in structurally important/functional sites in the TK domain caused extreme insulin resistance. These results suggest the importance of the FnIII and TK domains, and thus could facilitate early diagnosis, and may also provide new therapeutic approaches.

851T

Blue genes or red genes: Using large scale sequencing cohorts to reassess the pathogenicity of monogenic diabetes genes. *T.W. Laver, M. Wakeling, O. Knox, K.A. Patel, R.C. Caswell, E. De Franco, S.E. Flanagan, S. Ellard, A.T. Hattersley, M.N. Weedon.* University of Exeter Medical School, University of Exeter, Exeter, Devon, United Kingdom.

Genetic testing relies on accurate information on the pathogenicity of genes. Projects such as the Transforming Genetic Medicine Initiative (TGMI) are attempting to definitively map genes to diseases (red genes) and just as importantly re-classify those without robust evidence for association with the disease (blue genes). There is considerable controversy about whether some genes proposed to cause Maturity Onset Diabetes of the Young (MODY) actually cause young-onset diabetes when mutated. MODY is an autosomal dominant, Mendelian form of diabetes which typically presents before 25 years of age. MODY is an example of where genetics enables precision medicine: a genetic diagnosis can dictate the correct treatment for a patient. Large sequencing cohorts like ExAC (n=60,706) along with extensive sequencing efforts of potential MODY cases offer new opportunities to investigate the pathogenicity of the proposed MODY genes - most were proposed before the availability of large sequencing cohorts, which may have led to incorrect association with the disease. We used a cohort of patients referred for MODY genetic testing and the ExAC cohort to assess the likelihood that variants in these genes are highly penetrant MODY mutations. We performed targeted next generation sequencing of 13 proposed MODY genes on a cohort of 596 MODY patients. For genes reported to cause MODY by haploinsufficiency we would expect to see enrichment of rare loss of function (LoF) and missense variants in a MODY cohort compared to ExAC. We see enrichment for the most common cause of MODY, *HNF1A* (LoF $P=4\times10^{-6}$, missense $P=5\times10^{-9}$), but not for *BLK* (LoF $P=0.4$, missense $P=0.5$) or *PDX1* (LoF $P=1$, missense $P=1$). Additionally, reported variants in *BLK*, *KLF11*, *PAX4*, *PDX1*, and *NEUROD1* are present in ExAC at a frequency not consistent with the frequency of MODY. MODY has an estimated prevalence of 1 in 10,000. Variants in all of these genes, reported in HGMD (Human Gene Mutation Database) as pathogenic, are present at greater than 1 in 2,000 in ExAC. For example, the sole non-synonymous variant reported in the original *BLK* study occurs at a frequency of 1 in 85 in ExAC. Large scale sequencing cohorts are an important tool for assessing pathogenicity in rare Mendelian diseases. Our results suggest *BLK* and *PDX1* should be classified as TGMI blue genes and not screened on diagnostic MODY panels.

852F

Personalized medicine in diabetes mellitus: Lessons from the US Monogenic Diabetes Registry. *M. Sanyoura[1], D. del Gaudio[2], L. Letourneau[1], A. Knight Johnson[2], S. Greeley[1], L. Philipson[1].* 1) Section of Adult and Pediatric Endocrinology, Diabetes, and Metabolism, The University of Chicago, Chicago, USA; 2) Department of Human Genetics, The University of Chicago, Chicago, IL, USA.

Monogenic forms of are often difficult to distinguish from type 1 and type 2 diabetes. As a result, a majority of patients remain misdiagnosed and often inappropriately treated. An accurate diagnosis is important, as the molecular subtype can direct appropriate treatment and define prognosis. The U.S Monogenic Diabetes Registry, created in 2008 at the University of Chicago, facilitates recruitment of appropriate patients and collects genetic, medical and clinical information. As of May 2017, 1430 families have been enrolled in the registry.   Due to the vast number of genes implicated and the phenotypic overlap within these disorders, a gene panel-based approach to genetic testing is preferred, as it offers the ability to analyze multiple genes simultaneously. We developed and validated a next-generation sequencing (NGS) assay to evaluate all genes known to cause monogenic diabetes.   To date, a pathogenic cause was identified in 429 families (774 individuals). Mutations in *KCNJ11* were found to be the most common cause of neonatal diabetes diagnosed under 1 year of age (43.2%), followed by *INS* (16.5%), 6q24 (12.2%) and *ABCC8* (10.6%). 11% of cases has pathogenic variants in syndromic genes, namely, *WFS1*, *FOXP3*, *GATA4*, and *GATA6* that would not have been correctly diagnosed otherwise. Mutations in *GCK* were found to be the most common cause of Maturity Onset Diabetes of the Young (MODY) (62%) followed by *HNF1A* (30%), *HNF4A* (7%), *HNF1B* (1%).   Of all GCK-, HNF1A-, and HNF4A-MODY positive probands, 23% were misdiagnosed at type 1 diabetes, 42% were misdiagnosed as type 2 diabetes, and more than 55% of them were inappropriately treated with either insulin or other glucose-lowering agents. Similarly, 45% of probands positive for mutations in *INS*, *KCNJ11*, and *ABCC8* were misdiagnosed as type 1 diabetes. Those with diabetes due to mutations in *KCNJ11*, *ABCC8*, *HNF1A* and *HNF4A* may be more effectively managed with sulfonylurea therapy instead of insulin injections, while individuals with GCK-related diabetes generally do not require any treatment. Monogenic diabetes should be considered in any diabetic patient who has features inconsistent with their current diagnosis or presents with additional features characteristic to a specific syndromic subtype of monogenic diabetes. Rapid and cost-effective molecular genetic tests are now readily available to help define the diagnosis, alter prognosis and optimize treatment of children, young adults, and their families.

853W

***MAT1A* variants in methionine adenosyltransferase deficiency (MATI/III) suggesting dominant inheritance.** *J. Higgs, D. Wang, R. Salomon Kent, S. Suchy.* GeneDx Inc., Gaithersburg, MD.

Elevated methionine (Met) on newborn screening can detect abnormal Met-homocysteine metabolism that may be associated with cystathionine β-synthase deficiency, tyrosinemia type I, glycine N-methyltransferase deficiency, adenosylhomocysteine hydrolase deficiency, adenosine kinase deficiency or MAT I/III. MAT I/III is inherited as an autosomal recessive (AR) or AD disorder. Most individuals, particularly those with the AD R264H variant, have a benign clinical course, however, patients with AR MATI/III may have more severe findings including cognitive deficits, abnormal brain MRIs and other neurologic abnormalities. In addition to R264H, other AD variants (R249Q, G280R) have been described in association with a benign phenotype. These AD variants are located at the dimer interface of the MAT protein. We report here the results of *MAT1A* analysis in 84 cases suspected of MAT I/III. A definitive diagnosis was made in 43%: 10 homozygous/compound heterozygous for pathogenic (PATH)/likely pathogenic (LPATH) variants, 11 heterozygous for R264H and 15 with no reportable variants. Of the remaining patients, 5 were heterozygous for 2 variants of uncertain significance (VUS) or for a VUS and a PATH/LPATH variant and 43 had a single heterozygous variant not known to be associated with AD inheritance. Further review of these 43 cases revealed that 33 were heterozygous for a PATH/LPATH variant and 10 were heterozygous for a VUS. In 41 of these 43 cases, exon-level deletion/duplication testing of *MAT1A* was also performed and was negative. The 33 cases heterozygous for a single PATH/LPATH variant included 11 different variants with 7 (64%) reported as located in either the dimer interface or the substrate binding pocket or involved in Met binding/positioning (A55D, I322F, I322T, R249W, G253R, P255S, A259V) where other AD variants have been proposed. Of the 11 variants 4 were observed as single het variants in multiple individuals referred for elevated Met (A55D, P255S, A259V, I322F) comprising 17 of the 33 cases (52%). If other causes of hyper-Met have been excluded, the clinical outcome for the 33 patients is expected to be favorable as they are either heterozygous carriers of AR MAT I/III or have AD MAT I/III. Whether any of the 4 recurrent single het variants have a dominant negative effect requires further study. Such information would confirm the cause of the abnormal biochemical results, aid in the prediction of phenotype and enable accurate genetic counseling.

854T

Case report of congenital disorder of glycosylation caused by novel variant on COG6 gene diagnosed in early infancy. *Z. Wei[1], B. Wu[2], K. Meirelles[1], W. Li[2], L. Yang[2], H. Wang[2], W. Zhou[2].* 1) WuXi NextCode Genomics (Shanghai) Co., Ltd., Shanghai, China; 2) The Molecular Genetic Diagnosis Center, Shanghai Key Lab of Birth Defect, Children's Hospital of Fudan University, Shanghai, China.

   Background: Congenital disorders of glycosylation (CDG) comprise a group of autosomal recessive inherited metabolic diseases caused by synthesis defect of glycans and glycoprotein. CDG type III is caused by mutations of the conserved oligomeric Golgi (COG) complex, which is important for the proper transport and positioning of Golgi proteins. Only 10 COG6-deficient patients have been reported in the literature so far, showing a broad spectrum of clinical features. **Cases report:** Here we report a 1.5 month old Chinese male with COG6 deficiency. The patient is the first child of non-consanguineous parents with noncontributory family history. He was conceived *via in vitro* fertilization and he was born prematurely at 34 weeks of gestation *via* c-section due to fetal distress. The major clinical findings include recurrent hyperpyrexia, hypohydrotic ectodermal dysplasia, elevated liver enzymes, bronchopulmonary dysplasia, congenital heart disease (ASD, PFO, pulmonary hypertension, left superior vena cava, bradycardia), prolonged APTT, multiple small crystals in left kidney, micrognathia, no calcification of second molar and eosinophil deficiency. Whole-exome sequencing of the proband identified two compound heterozygous variants in the COG6 gene (c.511C>T, c.540G>A). The nonsense variant c.511C>T (p.Arg171*), maternally inherited, was previously reported as pathogenic, while the novel synonymous variant c.540G>A (p.Glu180Glu) paternally inherited was predicted to affect mRNA splicing. c.540G>A is not reported in any known exome and whole genome database including the 1000 genomes and ExAC database. Sanger sequencing revealed that c.540G>A resulted in a 26bp deletion (c.515_c.540del) of cDNA, leading to the premature truncation of the protein (p.Thr173Phefs*13). **Conclusion:** This is the first case of COG6-CDG reported from China, which suggests hypohidrotic ectodermal dysplasia and abnormal liver enzymes could represent main clinical features for this neonatal presentation. Gene sequencing can help diagnose COG-CDG in the neonatal period and provide the basis for accurate diagnosis, clinical decision-making and outcome prediction of COG-CDGs.

855F

The incidence and mutational spectra of hyperphenylalaninemia in the Xinjiang Uygur population. *Y. Su[1], H. Wang[2], N. Rejiafu[1], B. Wu[2,3], H. Jiang[1], H. Chen[2], X. A[1], Y. Qian[2], M. Li[1], Y. Lu[2], Y. Ren[1], W. Zhou[2,3], L. Li[1].* 1) Department of Neonatology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China; 2) Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defects, The Translational Medicine Center of Children Development and Disease of Fudan University, Shanghai; 3) Key Laboratory of Neonatal Diseases, Ministry of Health, Shanghai.

   Objective: Hyperphenylalaninemia(HPA) is an autosomal recessive disease caused by deficiencies in phenylalanine hydroxylase (*PAH*) activity. Blood phenylalanine (Phe) levels greater than 360μmol/L often cause mental retardation. This is the first study to interrogate the HPA incidence rate within a large Xinjiang Uygur population sample**.** A further aim was to characterize the mutation spectrum of the *PAH* gene within the Uygur population. **Methods :** Cross-sectional data from the 2010 to 2016 National Direct Reporting System database was used to calculate the incidence rate. All HPA patients were diagnosed at birth using tandem mass-spectrometry (TMS) following exclusion of a defect in tetrahydrobiopterin metabolism, and further confirmed by direct Sanger sequencing. A low Phe diet was immediately implemented following Phenylketonuria (PKU) diagnosis. Intelligence tests were performed at the age of 18 months to assess the level of mental development. Eligible patients were followed-up from the date of diagnosis until the present day. **Results** A total of 580,607 Uygur neonates were screened and 110 were diagnosed with HPA; resulting in an incidence of 1:5278. Among these 110 Uygur patients, 105 were followed up in subsequent evaluations, 61.97% of them had normal intelligence scores (DQ > 85); 14.08% had marginal normal intelligence (DQ 75-85) and 23.94% have mental retardation. All 105 Uygur patients were genetically sequenced for *PAH,* as well as 3 children of their relatives and 3 children diagnosed as outpatients. We detected 55 different mutations, *c.1238G>C* (p. R413P) was the most frequent mutation. Furthermore, we identified 6 novel variants: *c.32T>A, c.590T>A, c.1102G>A, c.1109A>G, c.1200-2A>C and c.1304A>T,* which predicted to be pathogenic. Finally, the hotspot mutation of R413P was found in 4 patients from a village in Aksu in the Southern Xinjiang with a carrier rate of approximately 1:2.1. **Conclusions and Relevance:** The HPA incidence rate within the Uygur population is significantly high. Our data further highlights regional differences in *PAH* genotypes and mutation rates, suggesting not only a consanguineous relationship but also distinct differences between Asian and Caucasian populations.

856W

Mucopolysaccharidosis type VI (MPS VI) and molecular analysis: A review of published classified variants in the *ARSB* gene. *M. Bailey[1], R. Tomanin[2], L. Karageorgos[3], M. AlSayed[4], E. Izzo[1], N. Miller[1], H. Sakuraba[5], A. Zanetti[2], J.J. Hopwood[3].* 1) Medical Affairs, BioMarin Pharmaceutical Inc., Novato, CA; 2) Laboratory of Diagnosis and Therapy of Lysosomal Disorders, Department of Women's and Children's Health, University of Padova, Padova, Italy; 3) Lysosomal Diseases Research Unit, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia; 4) King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; 5) Department of Clinical Genetics, Meiji Pharmaceutical University, Tokyo, Japan.

   MPS VI (Maroteaux-Lamy syndrome) is an autosomal recessive lysosomal storage disorder caused by mutations in the *ARSB* gene that result in deficient activity of the lysosomal catabolic enzyme N-acetylgalactosamine-4-sulfatase (arylsulfatase B, ASB). Deficient ASB activity causes lysosomal storage and elevated urinary excretion of the glycosaminoglycan enzyme substrates dermatan sulfate and chondroitin-4-sulfate. MPS VI is diagnosed through clinical findings and deficient ASB enzyme activity with normal activity of control enzymes; detection of pathogenic variants in each *ARSB* allele can independently confirm the diagnosis and facilitate genetic counseling. A previous report (Karageorgos et al. 2007) on a cohort of 105 individuals diagnosed with MPS VI documented that most *ARSB* variants are rare or private. Since 2007, with the increasing use of molecular testing, many new *ARSB* variants have been published. To uniformly summarize all *ARSB* variants, we collected and analyzed from the literature and public databases 822 reports of 193 distinct variants in the *ARSB* gene from individuals diagnosed with MPS VI. In agreement with previous reports (Karageorgos et al. 2007), most variants are missense (60%; 115 of 193); next most common are deletions (18%; 34 of 193), followed by nonsense (12%; 23 of 193), and splice site/intronic variants (6%; 11 of 193). Many reported *ARSB* variants are rare, with 33% (64 of 193) reported only once. Zygosity of individuals with MPS VI (n=411) distributed as: 51% homozygous (209 of 411) and 42% heterozygous (173 of 411); in 7% of cases only one allele was identified (29 of 411). Of the 193 unique *ARSB* variants summarized here, only 18% (34 of 193) are recorded in public databases (Clinvar, EmvClass, Invitae, ARUP) in association with supporting evidence/ clinical significance. This analysis illustrates the heterogeneity of alleles linked to MPS VI and lack of representation of otherwise characterized pathogenic *ARSB* alleles in publicly available variant databases. We emphasize the importance of maintaining high clinical suspicion during MPS VI diagnosis and confirming diagnosis via ASB enzyme testing as many *ARSB* alleles may have yet to be formally classified as MPS VI-associated. Timely submission and classification of *ARSB* variants in public databases in association with biochemical and clinical data will help improve a timely diagnosis of MPS VI.

857T

Teeth loss and ungueal dysplasia as atypical features in Hunter syndrome. *P. Garavito[1], J. Estrada[2], B. Orozco[1], A. Cortés[1], C. Silvera[1], C. Gallón[2].* 1) Universidad del Norte, Barranquilla, Colombia; 2) Clínica San Martín, Barranquilla, Colombia.

   Background: Hunter syndrome (OMIM 309900) or mucopolysaccharidosis type II (MPS II) is a rare disorder caused by deficiency of iduronate-2-sulfatase (I2S) enzyme, which is inherited as an X-linked recessive disorder. **Case report:** The patient is the younger of three boys of nonconsanguineous parents. At 30-month-old he presented with weight loss, short stature, and hepato-splenomegaly. He has a past medical history of fever, spontaneous teeth loss, otitis media, gastroenteritis, and intermittent abdominal pain. On clinical examination, as positive findings: brachycephaly, posterior rotation of the ears, mild icteric skin, loose teeth (upper incisors), prominent abdomen with hepato-splenomegaly, bilateral ear secretion. No coarse face, joint contractures, or hernias were observed. Dermatologic assessment described the typical cutaneous lesions of ivory-white papules on the pectoral region, pebbling of the skin, and ungueal dystrophy in hands. In the diagnostic workup, abdominal US showed hepatosplenomegaly and enlarged bile ducts. Laboratory work-up evidenced increased liver proteins (ALT, AST, GGT), hypercholesterolemia, anemia, and hypothyroidism (TSH 15.96). Serologic I2S was 4,2 umol/L/h (normal values >5,6 umol/L/h), and molecular confirmation revealed a pathogenic hemizygous mutation c.854A>G. This mutation leads to a substitution of a Tyrosine for Cysteine, p.(Tyr285Cys) and has not been described in the literature. Among the differential diagnosis that were considered in this patient were Langerhans cell histiocytosis, hypophosphatasia, and others lysosomal storage diseases. **Discussion:** The dermatologic and oral cavity findings in our patient could be secondary to glycosaminoglycans accumulation in skin and gums, and could correlate with this newly described mutation.

858F

Type 2 Gaucher disease in an infant despite a non-mutated maternal *GBA1* gene. *R. Grey, E. Hagege, G. Lopez, T. Roshan Lal, B. Stubblefield, E. Sidransky, N. Tayebi.* Medical genetics, NHGRI, NIH, Bethesda, MD.

Gaucher disease (GD), a recessively inherited autosomal lysosomal storage disease, results from a deficiency of the enzyme glucocerebrosidase. Over 400 mutations have been identified in *GBA1*. This gene is located in a gene-rich region on chromosome 1q21 and has a highly homologous pseudogene located 16kb downstream, which leads to a large number of recombinant alleles. The disease most frequently results from the inheritance of one *GBA1* mutation from each parent. Here we report an infant with type 2 GD who inherited one mutant allele, Arg257Gln (c.887G>A; p.Arg296Gln) from his father, while his second mutation, Gly202Arg (c.721G>A; p.Gly241Arg) was not identified in three different DNA sources (skin, fibroblasts and buccal swab) from the mother. Quantitative real-time PCR, performed on genomic DNA, failed to demonstrate any deletions, duplications or recombinations in the proband not detected by sequencing. The proband and mother's results were confirmed by sequencing cDNA made from fibroblast RNA. RNA expression and glucocerebrosidase enzyme activity and protein levels in the mother's samples were in the control range, while both were almost undetectable in the proband. These results suggest that the maternal mutation arose from germline mosaicism in the mother's ova or a single *de novo* mutation in one ovum. While there have been other cases implicating *de nova* mutations in patients with type 2 GD previously, this is the first time mutation Gly202Arg has been reported to be inherited non-traditionally. This case is part of a growing literature suggesting that Gaucher disease can be inherited via germline or *de novo* mutations, and emphasizes that it is critical for clinicians to consider such inheritance when making diagnostic decisions or providing genetic counseling. Furthermore, it will be important to evaluate the factors underlying this disease mechanism, and to explore why it seems to occur more frequently in type 2 Gaucher disease. .

859W

Comparative plasma proteomic analysis in Korean patients with Fabry disease pre and post enzyme replacement therapy. *S. Heo[1], M. Kang[1], G. Kim[3], J. Choi[2], S. Jung[4], R. Desnick[5], H. Yoo[2], B. Lee[2].* 1) Genome Res Ctr/Asan Inst, Asan Medical Ctr, Seoul, South Korea; 2) Department of Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, Korea; 3) Medical Genetics Center, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, Korea; 4) Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, Korea; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Fabry disease is a rare X-linked lysosmal storage disease caused by a deficiency in α-galactosidase A (GLA). It is characterized by the progressive accumulation of globotriaosylceramide (Gb3) and related glycosphingolipids in the vascular endothelial, perithelial, and smooth muscle cells. Enzyme replacement therapy clears this accumulation. Herein, we analyzed plasma proteome profiles before and after enzyme replacement therapy in eight patients with Fabry disease to characterize its molecular pathology. After short-term enzyme replacement therapy (4–12 months), the abundances of 15 plasma proteins involved in inflammation, oxidative and ischemic injury, or complement activation were reduced significantly. Among them, ACTB, inactivated C3b (iC3b), and C4b were elevated significantly in pre-enzyme replacement therapy Fabry disease plasma compared with control plasma. After longer-term enzyme replacement therapy (46-96 months), iC3b levels gradually decreased, whereas the levels of other proteins varied. The gradual reduction of iC3b was comparable to that of Gb3 levels. In addition, iC3b increased significantly in pre-enzyme replacement therapy Fabry disease mouse plasma, and C3 deposits were notable in renal tissues of pre-enzyme replacement therapy patients. These results indicated that C3-mediated complement activation might be altered in Fabry disease and enzyme replacement therapy might promote its stabilization.

860T

Evaluation of intracerebroventricular enzyme replacement therapy treatment with rhNAGLU-IGF2 from birth onwards in MPS IIIB mice. *S.-h. Kan[1], S.Q. Le[1], J.D. Cooper[1], M.S. Sands[2], P.I. Dickson[1].* 1) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, 90502; 2) Washington University School of Medicine in St. Louis, St. Louis, MO 63110.

   Mucopolysacccharidosis IIIB (MPS IIIB, Sanfilippo B syndrome) is an inherited lysosomal storage disease caused by the deficiency of the enzyme, alpha-*N*-acetylglucosaminidase (NAGLU) associated with the degradation pathway of heparan sulfate. Patients with MPS IIIB have a progressive neurological disorder and a lifespan of about two decades. Current studies have shown that intracerebroventricular (ICV) enzyme replacement therapy (ERT) with a fusion protein (rhNAGLU-IGF2) of recombinant human NAGLU and an IGF-2 motif is a feasible treatment in MPS IIIB mice. To further evaluate the biochemical efficacy of single dose of rhNAGLU-IGF2 *in vivo*, neonatal MPS IIIB mice (postnatal day 1 or 2) were given rhNAGLU-IGF2 (100 μg) into the left lateral ventricle. Mice were harvested at 2, 4, 8 and 12 weeks old and their brains were homogenized for enzyme activity assays. Heterozygous littermates and untreated MPS IIIB mice were used as controls. The average NAGLU enzyme activity from the whole brain homogenate ranged from 26x (2 weeks, ranged from 0.93-43 fold, n = 3), 12x (4 weeks, ranged from 3.6-21 fold, n = 3), 1.7x (8 weeks, ranged from 0.18-2.90 fold, n = 4), and 0.74x (12 weeks, ranged from 0.29-1.38 fold, n = 7) in rhNAGLU-IGF2 treated mice vs. heterozygous littermates. The estimated half-life of the enzyme *in vivo* is approximately 13.2 days. The secondary elevation of hexosaminidase activity was reduced to the heterozygous levels up to 8 weeks post-treatment with ICV rhNAGLU-IGF2 ($p < 0.001$), and 81% reduced at 12 weeks post-treatment vs untreated MPS IIIB mice ($p < 0.001$).   The overall therapeutic effect of single dose ICV ERT with rhNAGLU-IGF2 in newborn MPS IIIB mice was stable in brain up to 12 weeks. These results, together with ongoing behavioral and pathological analyses will inform the further development of treatment protocols in MPS IIIB.

861F

Genetic modifiers of *NGLY1* deficiency, a rare deglycosylation disorder, identified by exploiting natural variation in *Drosophila*. *C.Y. Chow, K. Owings, N. Russell, J. Lowry, R. Palu.* Department of Human Genetics, University of Utah, Salt Lake City, UT.

   Autosomal recessive loss-of-function mutations in *N-Glycanase 1* (*NGLY1*) cause *NGLY1* deficiency, the only known human disease of deglycosylation. *NGLY1* deficiency is a devastating, rare, neglected disease. Patients with *NGLY1* deficiency present with developmental delay, movement disorder, seizures, liver dysfunction, and alacrima. NGLY1 is a conserved component of the endoplasmic reticulum associated degradation (ERAD) pathway. ERAD degrades misfolded proteins that accumulate in the lumen of the ER. NGLY1 deglycosylates misfolded proteins as they are translocated from the ER lumen to the cytoplasm for degradation. Patients with *NGLY1* deficiency show a wide spectrum of severity. This is especially striking since all identified patients carry two null mutations. This phenotypic heterogeneity poses a great challenge to treating this disorder. To understand what may be underlying inter-individual differences in severity of *NGLY1* deficiency, we undertook a natural genetic variation screen. We developed a *Drosophila* model of *NGYL1* deficiency and crossed it into 200 strains from the *Drosophila* Genetic Reference Panel (DGRP), a collection of wild-derived *Drosophila* strains that harbor polymorphisms present in a natural population. We assessed the effect of natural genetic variation in modulating the impact of loss of *NGLY1* activity by quantifying the number of mutant flies that survived to adulthood. Across the 200 strains, we found a very large phenotypic spectrum in survival in the absence of *NGLY1*. We found that some strains showed nearly 100% survival without *NGLY1* function, while, as expected, a number of other strains were completely lethal without *NGLY1*. We performed an association analysis to identify natural polymorphisms that modify survival in flies lacking *NGLY1*. Modifier genes fell into diverse categories not predicted to be involved in *NGLY1* function, including ER Ca2+ homeostasis (*TMTC1/2*), ER proteostasis (*ERMP1*), ubiquitination (*MYCBP2*), and vesicle trafficking (*RAB26*). Over half of the candidate genes that we tested strongly suppressed or enhanced the original *NGLY1* survival phenotype. The genes identified from this study have excellent biological support and may serve as candidate modifiers of *NGLY1* deficiency. This study represents an important step to understanding the pathogenesis underlying *NGLY1* deficiency and provides a general framework for incorporating phenotypic variability, when developing personalized therapies for rare diseases.

862W

Clinical and molecular variability in Niemann-Pick disease type B. *I. Focsa[1], S. Macovei[2], I. Dobrescu[2], L. Bohiltea[1], M. Budisteanu[2,3,4].* 1) Carol Davila University of Medicine and Pharmacy Bucharest, Romania; 2) Prof. Dr. Alexandru Obregia Clinical Hospital of Psychiatry, Bucharest, Romania; 3) Victor Babes National Institute of Pathology, Bucharest, Romania; 4) Titu Maiorescu University, Bucharest, Romania.

Niemann-Pick disease (NPD) is a lysosomal storage disorder, transmitted in autosomal recessive manner with a wide spectrum of phenotype. Based on clinical and molecular findings, several types have been described to date: while types A and B represents allelic variations in a single gene (*SMPD1*), type C1 is caused by mutation in the gene *NPC1* and in the *NPC2* gene (*HE1*) - type C2. Although type B is known as a "visceral non-neurologic form" with adult-onset disorder, an increased number of patients with highly variable phenotypes have been described, making it difficult to delineate the types in the absence of molecular diagnosis. We report on tree pediatric patients with Niemann-Pick type B disease and variable phenotypes. The first patient is a girl with hepatosplenomegaly (HSM) from the age of 1 year, with psychomotor regression starting at the age of 3 years followed by a slow gradual decline resulting in quadriparesis, absent speech, severe intellectual disability, and sleep disorders (insomnia) at the age of 12 years. At 10 years epileptic seizure occurs. She died at the age of 15 years. The second case is an 8 years girl who was first admitted to the hospital at the age of 1 year because of HSM. At 6 years growth retardation, kyphosis, neurodevelopment regression (mild intellectual disability, poor speech with dyslalia) were noted. The third patient: a 17-year-old boy exhibited HSM from the age of 1,3 year, recurrent respiratory infection, delayed developmental milestones (first steps at 2,3 years, first syllables at 9 months), moderate intellectual disability (QI 47), behavior problems. In 1 and 3 cases lysosphingomyelin-509 marker was elevated. NPC1, NPC2, SMPD1 genes sequencing was performed in all patients. Two heterozygous mutations c.573delT (p.Ser192Alafs*65) and c.1177T>G (p.Trp-393Gly) was found in the first case. In the second case we also identified two heterozygous mutations: c.560_606del in exon 2 leading to a premature stop at codon 208 and c.177T>G (p. Trp393Gly), while a homozygous mutation c.1177T>G (p.Trp393Gly) was revealed in our third case. All the mutations was detected in SMPD1 gene and were previously described as pathogenic. Since the c.1177T>G mutation is common in all patients and seems to be associated with mild phenotype, the other two mutations could play a role in phenotype severity. The c.573delT mutation is related with the most severe phenotype. Our study brings new evidence in a better genotype-phenotype correlation.

863T

Monoallelic mutations in *OXCT1* in clinically and biochemically proven SCOT deficiency: Evidence of deep intronic mutations? *C. Murali, M. Yudkoff, C. Ficicioglu, R. Ganetzky.* Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA.

Ketone body utilization defects are metabolic disorders with impaired ketolysis. Patients typically present in the newborn period or in childhood. Many manifest continual ketosis; others become ketotic in response to metabolic stress. A deficiency of succinyl-CoA 3-oxoacid CoA transferase (SCOT) is a common cause. Inheritance is thought to be autosomal recessive, but here we present 2 cases with severe ketosis and only a single mutation in *OXCT1*. Subject 1 is a full term exclusively breastfed girl who presented with lethargy and tachypnea on day of life 4. She had severe anion gap acidosis (pH 7.01, bicarbonate 3.8 mmol/L) with large ketones and euglycemia. Ammonia was mildly elevated at 141 μmol/L. Ketosis resolved in response to treatment with intravenous dextrose. Urine organic acids showed massive ketonuria (beta-hydroxybutyrate: 4156 mg/g creatinine; acetoacetate: 690 mg/g creatinine). Exome sequencing revealed a maternally-inherited heterozygous variant of uncertain significance in *OXCT1* (c.1376C>G). At 4 months, she has had no further episodes of decompensation, but does have intermittent, unprovoked, trace ketonuria (41 mg/g creatinine). She is managed with frequent feedings and strict fasting precautions. Subject 2 is a boy who presented with ketotic hypoglycemia at age 2 years. He has had recurrent episodes of ketotic hypo- and euglycemia over the subsequent 11 years. He requires chronic bicarbonate supplementation and a high-carbohydrate diet; he previously required a gastrostomy tube to maintain euglycemia. Enzymology showed low SCOT activity in fibroblasts (2.3 nmol/min/mg; normal > 2.6). Exome sequencing revealed a maternally-inherited heterozygous likely pathogenic variant in *OXCT1*, as well as a heterozygous mutation in *SLC22A5*. His medical problems include executive function problems, depression, hand tremor, and migraines. Targeted del/dup analysis of *OXCT1* was negative in both subjects. These cases show the classic, severe ketosis of SCOT deficiency, despite having only monoallelic mutations in *OXCT1*. We speculate that these subjects either have a second deep intronic mutation missed by exome sequencing; or, in the setting of additional metabolic stressors (exclusive breast feeding, *SLC22A5* carrier status), there may be haploinsufficiency. We propose that manifestation of disease in apparent *OXCT1* heterozygotes is an important consideration in the differential diagnosis of severe ketosis.

864F

Newborn screening for Hunter Disease: Is the c.103+56_34 dup allele a pathogenic variant or a pseudodeficiency variant? *Y.H. Huang[1], S.P. Lin[1,2,3], C.C. Chiang[4], M.J. Chan[4]*. 1) Rare Disease Center, MacKay Memorial Hospital, Taipei, Taiwan; 2) Department of Pediatrics, MacKay Memorial Hospital, Taipei, Taiwan; 3) Department of Medical Research, MacKay Memorial Hospital, Taipei, Taiwan; 4) Newborn Screening Center, Chinese Foundation of Health, Taipei, Taiwan.

Mucopolysaccharidosis type II (MPS II, Hunter Disease) is an X-linked lysosomal storage disease caused by a deficiency of iduronate-2-sulfatase (IDS). Since August 2014, we started large-scale newborn screening for Hunter disease by using dried blood spots (DBS) of LC-MS/MS method for measuring IDS activity. In the past 20 months, a total of 103,025 newborns were screened, 66 newborns were referred to the confirmatory center for further evaluation. Leucocyte enzymatic assay was done and IDS gene analysis was performed for those with leukocyte enzyme deficiency. 50 newborns were confirmed with reduction of IDS enzyme activity and a variant was found in IDS gene. According to the current database, only one case has been categorized as Hunter disease patient with a R273W mutant and others were variants of unknown significance (VOUS): c.103+34_56dup (43 cases), R297H (1 case), R493H (3 cases), P197S (1 case), and H342R (1 case). The c.103+34_56dup variant was identified in 86% of the VOUS groups and average IDS activity was 3.67±1.38 (NR: 30-53 nmol/4h/mg. protein). A total of 21 adjunctive maternal family members undertook the genetic testing for this variant, and 5cases were found to have this variant. None of the cases with c.103+34_56dup variant have any feature and characteristic of Hunter disease, including heart and abdomen sonography, X-ray of T-L spine and bilateral hand, even disaccharide units (KS, DS and HS) of urine GAG detected by LC-MS/MS were no specific findings, except only one newborn has mild hepatomegaly (liver span: 8.08 cm at mid-clavicular line). This information collected over 20 months of newborn screening for Hunter disease highlights the importance of genetic counseling and testing among newborns and high-risk family members. The high prevalence of the c.103+34_56dup variant in Taiwan raises an issue that whether this variant is truly a pathogenic mutation or just a pseudodeficiency variant. Further studies are demanded to analyze functional data and to evaluate the frequency of this variant in general population, furthermore, newborns and family members who have this variant also need to long term follow up.

865W

Pharmacokinetics of glycerol phenylbutyrate in pediatric patients 2 months to 2 years of age with urea cycle disorders. *S.A. Berry[1], J. Vockley[2], A.A. Vinks[3], M. Dong[3], R. Perdok[4], B. Robinson[4], R.J. Holt[4], N. Longo[5]*. 1) University of Minnesota, Minneapolis, MN; 2) University of Pittsburgh, Pittsburgh, PA; 3) Cincinnati Children's Hospital Medical Center and University of Cincinnati, Cincinnati, OH; 4) Horizon Pharma USA, Lake Forest, IL; 5) University of Utah, Salt Lake City, UT.

Introduction. Glycerol phenylbutyrate (GPB) is approved in the US and EU for the management of patients ≥2 months of age with urea cycle defects (UCDs) who cannot be managed by dietary protein restriction and/or amino acid supplementation alone. It is hydrolyzed by pancreatic lipases to yield phenylbutyric acid (PBA), converted in the liver to phenylacetic acid (PAA) that binds with glutamine to form phenylacetylglutamine (PAGN) which is excreted in urine (UPAGN). Here we examine the short- and long-term PK of GPB in very young UCD patients whose pancreatic and liver function might be immature. **Methods**. PK was analyzed using WinNonlin®/NONMEM® in UCD patients aged 2 mos to <2 yrs from two open label clinical studies. Study 1 (NCT02246218) analyzed PK for 1-2 days and UPAGN for 1 week in 10 patients with blood/urine sampling completed pre-dose and periodically thereafter. Only 3 patients were dosed with GPB for ≥6 mos. Study 2 (NCT01347073) compared the 24 hour PK in 4 patients switched from sodium PBA to GPB. These patients plus 3 additional patients were analyzed for up to 1 yr on GPB. **Results.** Short Term: Mean PBA/PAA/PAGN C_{max} were 42.4/36.5/62.5 mcg/mL on the first day of GPB-only dosing; UPAGN averaged 7,561 mcg/mL at 12-24 hr on day 1 and 8,859 mcg/mL on Day 7 (Study 1). Body surface area (BSA) correlated with PAA C_{max} and AUC_{0-last}, PAGN C_{max} and to a lesser extent AUC_{0-last} (Study 1). Population modeling found BSA to be the most significant covariate for PK analytes. PAA exposure (AUC_{0-24}) with GBP and Sodium PBA was 3,322 & 4,138 mcg*hr/mL (Study 2). Long Term: Plasma PBA/PAA/PAGN values did not significantly change over time. Overall median plasma PAA was ~1 mcg/ml and median plasma PAA:PAGN ratio <1 (Study 2). One PAA value exceeded 500 mcg/ml but was lower 9 months later at similar GPB doses. Mean targeted ammonia values were maintained at scheduled visits throughout the studies. **Conclusions.** GPB dosing was followed by prompt appearance (first day) of metabolites in blood and PAGN in urine. Elevated PAA levels were infrequent, transient and the median PAA:PAGN ratio of <1 indicate that PAA to PAGN conversion was not saturated. The findings collectively indicate that pediatric UCD patients aged 2 mos–2 yrs effectively hydrolyze GPB to release PBA and convert PAA to UPAGN so as to maintain blood ammonia within normal limits.

866T

Increased expression of *SLC26A9* delays age at onset of diabetes in cystic fibrosis. *A. Lam, M. Atalar, B. Vecchio-Pagan, S. Blackman, G. Cutting.* Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Diabetes is a common complication of cystic fibrosis (CF) with a number of clinical and histological similarities to type 2 diabetes. Genome-wide studies have identified significant association with CF-related diabetes for variants near and within *SLC26A9*, which encodes a chloride/bicarbonate transporter that interacts with CFTR, the protein defective in CF. The *SLC26A9* locus is of particular interest as it has also been identified as a modifier of intestinal obstruction and drug response in CF. A biologic role for *SLC26A9* in exocrine pancreatic function is supported by single cell RNA sequencing indicating that *SLC26A9* is predominantly expressed in pancreatic ductal cells, along with *CFTR*. To assess the genetic architecture of *SLC26A9*, the entire locus was sequenced in 762 individuals with CF (all F508del homozygotes). Two linkage disequilibrium (LD) blocks were identified and defined by a single recombination in intron 8. Two common haplotypes (24.2% and 28.4% frequency) exist within the 5' LD block corresponding to the variants associated with higher-risk (HR) and lower-risk (LR) of diabetes, respectively. No coding variants were in LD with either haplotype, and no individual variant showed greater association with diabetes than the two common haplotypes. To determine if the region containing the diabetes-associated variants encompasses regulatory sequences, 2.3 kb corresponding to the HR and LR haplotypes were cloned into a dual luciferase/renilla reporter system (DLRS). Combined analysis of the normalized data from 3 independent transfections into PANC-1, a surrogate for pancreatic ductal cells, with 4 biological clones per haplotype (technical replicates: transfection well N=78 for HR and N=77 for LR; DLRS N=234 for HR and N=230 for LR) showed that HR had a 13% lower promoter activity compared to LR (p=5.9E-11). Luciferase expression was negligible when the DNA sequences were cloned in the reverse orientation indicating that the 2.3kb region 5' of *SLC26A9* drives expression in an orientation-specific manner. These results are consistent with eQTL predictions from RNA sequencing data suggesting that LR variants in the 2.3 kb region correlate with higher levels of expression than HR variants (rs61814953; p=6.24E-03; beta=0.238). Taken together, these results indicate that increased expression of *SLC26A9* delays onset of diabetes in individuals with CF, possibly by modification of pancreatic ductal function. *Supported by CFF, Gilead Sciences.*.

867F

Genetic causes of hypercholesterolemia in the Emirati population. *H. Daggag[1], A. Rimbert[2], A. Buckley[1], RJ. Sinke[3], MT. Barakat[1], JA. Kuivenhoven[2].* 1) Imperial College London Diabetes Centre, Abu Dhabi , Abu Dhabi, United Arab Emirates; 2) Department of Pediatrics, University of Groningen, University Medical Center Groningen, The Netherlands; 3) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands.

Cardiovascular disease is the leading cause of morbidity and mortality in the UAE. Of the 5434 registered deaths in the UAE in 2000, 1381 were due to cardiovascular disease. Worldwide the two most important risk factors for ischaemic heart disease and myocardial infarction are smoking and dyslipidaemia. Familial hypercholesterolemia (FH) is an autosomal dominant genetic disorder that leads to premature arterial disease and cardiac death if left untreated. FH is widely under diagnosed and the prevalence in the Emirati population has not been established. The molecular diagnosis includes screening for monogenic as well as polygenic causes of this disorder. Of 39 patients screened as potential index cases by our clinical service, only five were found to have mutations in known FH genes (13.1%). Further five patients were carriers of a variant of unknown significance (VUS), further supporting the need to perform a research study and identify FH genetic causes specific to the Emirati population. In this study, we will apply a two-stepped approach using Next-Generation Sequencing (NGS) tools to investigate genetic bases of FH in our population. We will first screen known and candidate FH related genes using a novel dedicated kit which combines: targeted capture of genes and loci related to FH, NGS and an up to date bio-informatics pipeline. We recently showed for the first time that this kit can assess all possible causes of FH known to date. Secondly, for patients without established genetic diagnostics, we will investigate novel genetic causes using whole-exome sequencing. To date, 330 patients have been recruited from clinical areas at Imperial College London Diabetes Centre (ICLDC), an outpatient diabetes and general endocrine clinic based in the United Arab Emirates. 68 patients with LDL-c value higher than 6mmol/L (after correction for lipid reducing agents) will be selected for a molecular diagnosis of FH. This approach will contribute to our understanding of the epidemiology of FH in Emirati people and could potentially identify novel FH-causing genes susceptible to be of therapeutic interest. Proactive genetic screening of individuals who meet clinical criteria for FH and subsequent cascade screening of families may well provide a comprehensive tool for accurate diagnosis, facilitate early preventative clinical intervention and reduce overall healthcare costs.

868W

CEBPA mutation as a potential clinically novel cause of congenital generalized lipodystrophy. *R.M. Mostafavi[1], A. Lahoti[2,3], S.F. Miller[2,3], E.K. Pivnick[2,3].* 1) Le Bonheur Children's Hospital, Memphis, TN; 2) University of Tennessee Health Science Center, Memphis, TN; 3) UT Le Bonheur Pediatric Specialists, Memphis, TN.

Congenital generalized lipodystrophy (CGL) is a rare autosomal recessive disorder characterized by absent adipose tissue at birth, bone cysts and insulin resistance. Mutations in *AGPAT2* and *BSCL2* account for the two most common types of CGL. We present a patient with CGL in whom whole exome sequencing (WES) identified a mutation in *CEBPA*, a gene involved in adipocyte differentiation and a potential candidate for unexplained cases of CGL. An AA female born at 35 weeks gestation had hepatomegaly, lack of subcutaneous fat, muscular hypertrophy, hyperpigmentation and hypertrichosis. Parents were non-consanguineous and healthy with negative family history. Laboratory findings showed elevated direct bilirubin, mildly high triglycerides and transaminitis. At 8 months, she developed growth failure. Skeletal survey demonstrated profound changes of vitamin D deficiency rickets with secondary hyperparathyroidism and hypophosphatemia. Treatment with Vitamin D and calcium supplementation led to improvement but not complete resolution of rachitic changes, improvement of biochemical abnormalities and improved growth. Skeletal survey at 18 months demonstrated mild persistent rickets along with mild osteosclerosis. Subtle intramedullary lucent foci were seen within the femoral and tibial diaphyses, as well as within multiple metatarsals and phalanges. Ultrasound demonstrated moderate hepatomegaly with hypoechoic hepatic parenchyma but no focal lesions. At 2 years age, she has developmental delay, wasted appearance, hepatomegaly, severe generalized acanthosis nigricans and hirsutism. Lab testing showed normal glucose levels and HbA1c with low HDL and mild hypercholesterolemia. Leptin was <0.6 ng/mL. CGL was suspected. Molecular analysis of *AGPAT2* and *BSCL2* were negative. WES revealed a de novo heterozygous missense mutation of *CEBPA* gene (c.874A > C; p.N292H). The CCAAT enhancer binding proteins (CEBPs), which include CEBPB, CEBPD, and CEBPA, are a family of basic-leucine zipper proteins that function as transcription factors in terminal adipocyte differentiation. In the process of adipocyte differentiation, *CEBPB* and *CEBPD* are upregulated in committed preadipocytes and induce expression of *PPARG* and *CEBPA*, which promote terminal differentiation of adipocytes. We propose that *CEBPA* mutations may account for unexplained cases of CGL given the findings in our patient and the role of *CEBPA* in adipocyte differentiation.

869T

Biallelic mutations in *GPD1* gene in a Chinese boy mainly presented with obesity, insulin resistance, fatty liver, and short stature. *N. Li[1], G.Y. Chang[2], Y.F. Xu[1], Y. Ding[2], G.Q. Li[1], T.T. Yu[1], R.E. Yao[1], Y.P. Shen[1,3], X.M. Wang[2], J. Wang[1].* 1) Department of Medical Genetics and Molecular Diagnostic Laboratory, Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China; 2) Department of Endocrinology and Metabolism, Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China; 3) Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA.

Biallelic mutations in the *GPD1* gene cause a rare autosomal recessive inherited disease known as transient infantile hypertriglyceridemia (OMIM #614480). To date, only five pathogenic variants have been reported in 15 patients from 3 studies. The clinical symptoms of the affected individuals present a certain degree of heterogeneity. Here, we describe a 13 years and 8 months old Chinese male patient who mainly presented with obesity, insulin resistance, fatty liver, and short stature. Targeted next-generation sequencing revealed a novel compound heterozygous variant in *GPD1* gene (c.220-2A>G and c.820G>A; p.Ala274Thr). *In vitro* studies demonstrated that the Ala274Thr variant induced a decrease in GPD1 protein expression. Further *in vitro* investigation of the splicing pattern in a minigene construct in HEK293 cells showed that the c.220-2A>G variant generated an altered transcript with one cryptic splice site in exon 3, resulting in the loss of 69 bases in exon 3 (c.220_288del, p.74_96del). This is the first report involving an Asian who harbored *GPD1* mutations. Our work not only expands the mutant spectrum of the *GPD1* gene but also provides new insights on its resulting phenotype. Moreover, we suggest that the term GPD1 deficiency is more appropriate in describing this phenotype in OMIM than the phrase transient infantile hypertriglyceridemia.

870F

Systematic characterization of mutations in familial hypercholesterolemia linked genes in Estonia. *M. Alver[1,2], A. Reigo[1], L. Leitsalu[1], M. Palover[1,2], T. Nikopensius[1], P. Natarajan[7,8], S.M. Zekavat[7,8], A. Irs[3,5], T. Ainla[4], A. Saar[3,4], R. Mägi[1], M. Kals[1], T. Marandi[3,4], N. Tõnisson[1,6], A. Metspalu[1,2], S. Kathiresan[7,8], T. Esko[1,7].* 1) Estonian Genome Center, Tartu, Estonia; 2) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Department of Cardiology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; 4) Cardiology Clinic, North Estonia Regional Hospital, Tallinn, Estonia; 5) Chair of Clinical Pharmacology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia; 6) Department of Genetics, Tartu University Hospital, Tartu, Estonia; 7) Broad Institute, Cambridge, MA, USA; 8) Massachusetts General Hospital, Boston, MA, USA.

While familial hypercholesterolemia (FH) is one of the most common single-gene disorders, it remains significantly under-diagnosed and inadequately treated, underscoring the paramount need to identify individuals affected by FH for timely diagnosis and treatment. Estonian Biobank's population-based collection in combination with comprehensive health data and genomic profiles offers a state-of-the-art platform to reliably determine carriers of FH mutations. Via implementation of a genotype-first approach we determined index cases by screening for rare pathogenic/likely pathogenic variants in 3 FH linked genes (*LDLR, APOB, PCSK9*) in 30X whole genome (n=2420) and 60X whole exome sequence (n=2384) datasets associated with untreated baseline LDL cholesterol (LDL-C) level ≥4.0 mmol/L (≥130 mg/dl). All research participants and their 1st and 2nd degree relatives undergo clinical investigation to determine the presence of overt or subclinical cardiovascular pathology and genetic counselling. We have identified 9 *LDLR*, 3 *APOB* and 2 *PCSK9* gene variants, of which 5 are novel (1 novel substitution) and 8 have previously not been associated with FH, in 30 index cases. So far, 21 individuals with LDL-C values between 2.72 to 9.21 mmol/L (mean=4.91, SD=1.96; not adjusted for statin treatment) and their 26 family members have attended a multi-module phenotyping appointment with special cardiovascular refocus highlighting generally low pre-established treatment adherence (38% were on lipid-lowering therapy) and unawareness of the familial background of the disease (FH suspected in 3 index cases (11%)). Compared to 970 non-FH-mutation carriers, individuals who harbour a disease-causing variant, display on average higher LDL-C level as well as larger variability of LDL-C concentration. Based on genetic feedback survey, all participants have expressed interest of disclosure of genetic finding. Preliminary results indicate that the established genotype-first framework provides a powerful platform to determine the Estonian population specific genetic landscape of FH and that the implementation of cascade screening and deep-phenotyping proves highly valuable for estimating the pathogenicity of discovered variants. Further attempts are being made to infer discovered mutation carriers in the whole cohort of Estonian Biobank (n=52,000; array genotyped only).

871W

Androgens and antioxidants management improve clinical & hematologic response of Fanconi Anemia Egyptian patients to bypass hematopoietic stem cell transplantation unavailability. *A. Attia, G. El-Kamah, M. Eid.* Medical Molecular Genetics, Human Genetics and Genome Research,National Research Centre of Egypt, Giza, MI., Egypt.

Fanconi anemia (FA) is the most common inherited bone marrow failure syndrome characterized by hypersensitivity to clastogenic agents particularly diepoxybutane (DEB) and cancer predisposition. Oxidative stress plays a major role in the pathogenesis of leukemia-prone diseases such as FA. Recently there is controversy concerning hematopoietic stem cell transplantation (HSCT) as they don't always abolish the threat of malignancies, which with progressive bone marrow failure are the cause mortality in FA. Aim: Evaluating clinical & hematological response to androgens and antioxidant therapy as an alternative when HSCT is not available. Subjects and Methods: 45 cases descending from unrelated consanguineous pedigrees were included in the study and confirmed through DEB hypersensitivity. DNA micronuclear damage, among other parameters, was measured as an indicator to FA patients' stress condition. Patients were kept on androgens, prednisone and Vitamins A, E, and C according to the recommended daily allowance. Clinical and hematological follow-up were conducted for all patients over a period of two years with 3 monthly reevaluation schemes. Results: The median age at onset was 7 years (4m–10.5 years). Average Hb (7.6 g/dl), RBC (2.6x10^{12}/L), WBC (3.7 x10^9/L) and platelet count (53x10^3/uL). 20% had low birth weight, 90% café au lait spots, 38% radial reduction, renal disease in 4 cases & other skeletal affection in 78%. DEB showed wide range of break/cell (3.4-12.5), with mean 6.0. Elevated levels of oxidative stress and DNA damage were detected. Hematological parameters improved in response to androgen therapy, mostly WBCs by 40%, followed by hemoglobin (9%) and lastly platelets by only 5%. Oxidative stress parameters and DNA damage improved after antioxidant administration. Conclusion: Androgen therapy improved the hematological condition of patients. Early administration of antioxidants in FA slows disease course, ameliorate symptoms and improve general health.

872T

Very early-onset inflammatory bowel disease in a Mexican patient with an IL10 receptor deficiency due to a novel homozygous *IL10RB* mutation.

D.E. Cervantes-Barragan[1], F.H. Campos-Romero[2], C. Maldonado-Rivera[3], M.J. Gaytan-Garcia[1], J.I. Navarrete-Martinez[1], J. Silva-Estrada[3], C. García-Parra[3], G.H. Wakida-Kusunoki[3], I. Rivera-Salgado[4], J. Reyna-Figueroa[5], A.E. Limón-Rojas[6], M.R. Delgado-Calvillo[7], L. Flores-Lagunes[8], C. Molina[8], M. Olivares[8], K. Carrillo-Sánchez[8], L. Arriaga-Pizano[9], C. López-Macias[9], L. Cervantes-Barragan[10], C. Alaez-Verson[8]. 1) Genetics Department, Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico City, Mexico; 2) Allergology Department, Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico City, Mexico; 3) Pediatrics Department, Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico City, Mexico; 4) Pathology Department, Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico City, Mexico; 5) Education and Research Department, Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico City, Mexico; 6) General Dictatorate, Hospital Central Sur de Alta Especialidad, PEMEX, Mexico City, Mexico City, Mexico; 7) Pediatrics Department, Hospital Regional de Villahermosa, PEMEX, Villahermosa, Tabasco, Mexico; 8) Genomic Diagnosis Laboratory, National Institute of Genomic Medicine, INMEGEN, Mexico City, Mexico; 9) Unit of Immunochemistry, Specialities Hospital, National Medical Centre "Siglo XXI" of the Mexican Institute for Social Security, IMSS,, Mexico City, Mexico; 10) Department of Pathology and Immunology, University of Washington, St. Louis, MO, USA.

Very early-onset inflammatory bowel disease (VEOIBD) refers to IBD presented in children before the 6th year of life, few cases havebeen reported occurring before the first year of life. VEOIBD share the clinical picture of multifactorial Crohn´s and ulcerative colitis clinical and endoscopic characteristics, such as deep intestinal mucosal ulcerations, perianal inflammation with abscess and fistula formation. Linkage studies and candidate gene approach found recently, that monogenic autosomal recessive mutations in IL10 receptor chains, alpha or beta, leads to a severe pancolitis with perianal inflammation syndrome. Since today few patients with alpha or beta (*IL10RA* and *IL10RB*) mutations had been reported. We present a case of a VEOIBD in a mexicna patient with a novel homozygous novel mutation in *IL10RB* gene. The index patient was from Tabasco, Mexico, he is the second child of a young consanguineous parents. The sib present at 3 months with VEOIBD and recurrent pneumonias, at 3 years the patient died secondary to intestinal perforation. The propositus presented at the age of 1 month with perianal bleeding secondary to multiple perianal abscesses. Colonoscopy revealed multiple ulcerations through the entire colon. Colon biopsy reavealed cryptitis and crypt abscesses. At 2 months intestinal present coinfection with CMV that was treated with valganciclovir. Peripheral plasma flow cytometry revealed a low concentration of NKs cells 0.63 (1-6%). The patient was treated initially with mesalizine and then with azathioprine and ciclospirine, with poor response, and nowadays, at 11 month, is in bone marrow HSCT protocol. Clinical picture suggests a monogenic VEOIBD, molecular analysis was performed using DNA peripheral blood, that was extracted using Maxwell 16 DNA purification kit ® and perform a NGS approach using Nextera Rapid Illumina inc ®. The NGS revealed an homozygous variant c.49C>T (p.Ala17Pro), the carrier status was confirmed in the parents. Mutation Taster and Polyphen 2 predicts pathogenicity, the variant also is absent in ClinVar, HGMD, ExAC and 1000 genomes databases. Literature review revealed 18 cases of *IL10RB* mutations reported in series of VEOIBD mainly in caucasian and asian ancestry with severe clinical picture like our patient phenotype. Monogenic etiology has to be ruled out in patients with VEOIBD patients.

873F

PIDDGEN: A multi-disciplinary team providing molecular diagnoses of primary immunodeficiency diseases in South Africa.

C. Kinnear[1], M. Moller[1], B. Glanzmann[1], N. Schlechter[1], E. Banda[1], E. Hoal[1], M. Schoemann[1], R. Nortje[2], M. Urban[1], M. Esser[2]. 1) SAMRC Centre for Tuberculosis Research, Biomedical Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa; 2) Immunological Unit, National Health Laboratory Service, Cape Town, South Africa.

The diagnosis of Primary Immunodeficiency disorders (PIDs) remains challenging. In South Africa, based on global estimates, the projected number of individuals with PIDs should range from 2850 to 45723, however, to date, less than 500 cases have been reported and approximately 300 have been enrolled in the South African national PID registry. To raise awareness for PID in our country and to facilitate the molecular diagnosis of these patients, the Primary Immunodeficiency Diseases Genetics group (PIDDGEN), a multi-disciplinary group of specialists interested in the immunological and genetic diagnosis of PID, was founded in 2013 at Tygerberg Hospital, Stellenbosch University in Cape Town. PIDDGEN pursues the genetic diagnosis of PID diagnoses from across South Africa. Thus far, PIDDGGEN has sequenced exomes of 17 South African PID patients and, where available, unaffected family members. Exome sequencing allowed for the definitive molecular diagnosis for thee of these patients. Furthermore, using targeted sequencing of known PID-causing genes, we have been able to provide molecular diagnosis for five patients with X-linked Chronic Granulomatous Disease (CGD) (three novel *CYBB* mutations), two patients with an interferonopathy (novel *SAMHD1* mutation), and one patient with Mendelian Susceptibility to Mycobacterial Disease (MSMD) (*INFGR1* mutation). These mutations and patients only represent a fraction of the PID cases reported on the South African PID registry. Hence, we believe that PIDDGEN has an important role to play in the improved awareness, definitive diagnosis and targeted management of PID patients in South Africa.

874W

Exome sequencing of extreme phenotypes identifies potential novel genes as modifiers of leg ulcer in sickle cell anemia. *G.Q. Carvalho-Siqueira[1], B.B. Souza[1], M.G. Borges[2], M.T. Ito[1], G. Ananina[1], A.S.C. Streva[3], M.S. Gonçalves[4], I.T.L. Cendes[2], F.F. Costa[5], M.B. Melo[1].* 1) CBMEG, University of Campinas-UNICAMP, Campinas, Brazil; 2) FCM, University of Campinas-UNICAMP,, Campinas, Brazil; 3) Foundation of Hematology and Hemotherapy-HEMOBA, Salvador, Brazil; 4) Institute Gonçalo Moniz-Fiocruz, Salvador, Brazil; 5) Hematology and Hemotherapy Center-UNICAMP, Campinas, Brazil.

   Background: Although sickle cell anemia (SCA) [603903] results from homozygosity of a single mutation at position 6 of the β-globin locus, the clinical aspect of this disease is very heterogeneous. Leg ulcer is one of the worst complications, exerting a highly negative impact in the quality of life of patients with SCA, and is related to the severity of the clinical aspect of the disease. **Objective:** Seeking for novel candidate genes associated with leg ulcer in sickle cell anemia, we proposed this search through exome sequencing of extreme phenotypes in a sample of the Brazilian population. **Methods:** The project was approved by the Research Ethics Committee of the Faculty of Medical Sciences of the State University of Campinas. Our sample consisted of 40 patients with sickle cell anemia from the Hematology and Hemotherapy Center, out of which 20 didn't have leg ulcer, with ages varying from 40 to 61 years (it is unusual the appearance of this disorder at such age) and 20 had chronic leg ulcer. DNA was extracted from peripheral blood leukocytes and submitted to exome sequencing. Capture and enrichment were performed with the Nextera Rapid Capture Exome kit (Illumina) and samples were loaded in the Illumina HiSeq 2500. The quality of the reads was verified through FastQC, the files were then submitted to the GATK Pipeline generating a vcf file for each individual. In the bioinformatics processing, several filters were applied using vcftools. Non-synonymous variants were prioritized according to various conditions, such as score quality, depth and Hardy-Weinberg equilibrium. The Fisher association test was performed with the remaining variants, through PLINK, resulting in 734 variants, selected with a p<0.05. The annotation was made through Annovar. In the annotated file, the synonyms, the non-deleterious variants, as well as the barely conserved variants were removed, generating 54 variants. **Results:** Six genes were selected, according to the literature, involved in angiogenesis, inflammation, coagulation and thrombosis: *AHNAK*2 [608570], *CEACAM*8 [615747], *FLG*2 [616284], *LAMA*1 [150320], *RNF*213 [613768], and *ZNF*540 [613903]. **Conclusion:** Therefore, we suggest that these genes could potentially be related to the development of leg ulcer in SCA. However, they should all be validated in other populations for confirmation.

875T

Genomic characterization of *F8* and *F9* copy number variants in the My Life, Our Future TOPMed hemophilia cohort. *M.M. Wheeler[1], D.J. Aschman[2], G.F. Pierce[3], B. Konkle[4,5], J. Morales[6], J.M. Johnsen[4,5], D.A. Nickerson[1], NHLBI Trans-Omics for Precision Medicine.* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) American Thrombosis and Hemostasis Network, Chicago, IL; 3) National Hemophilia Foundation, New York, NY; 4) Bloodworks NW, Seattle, WA; 5) Division of Hematology, University of Washington, Seattle, WA, USA; 6) Bioverativ, Waltham, MA.

   Hemophilia A and B are X-linked bleeding disorders commonly resulting from genetic alterations of the genes, *F8* and *F9*. These genes encode Factor VIII and Factor IX of the coagulation cascade; deficiencies in one of these two factors define hemophilia A and hemophilia B, respectively. Numerous genetic variants in *F8* and *F9* have been identified. The majority include single nucleotide variants (SNVs), copy number variants (CNVs), and in the case of *F8*, large intra-chromosomal inversions. Most CNVs have been found through analyses of targeted regions or techniques that detect differences in gene dosage. However, these methods cannot fully characterize the extent of each CNV as they are unable to detect breakpoints precisely. In this study, our goal was to characterize *F8* and *F9* CNVs in the My Life, Our Future hemophilia cohort that was recently whole genome sequenced (WGS) by the NHLBI TOPMed program. This cohort mainly comprised samples from males diagnosed with hemophilia A (N=1,959) or hemophilia B (N=227). Comprehensive analysis of WGS data, revealed 64 CNVs in hemophilia A (57 deletions and 7 duplications) and 24 deletions in hemophilia B subjects. In hemophilia A, CNV size ranged from <100bp to > 100Kb and comprised single exon *F8* deletions to multi-genic CNVs. Several *F8* CNVs (N=14) included the first or last exon and exhibited heterogeneous breakpoints that differentially affected neighboring genes. In hemophilia B, CNV size ranged from approximately 1Kb to 6.5Mb and most included deletions of multiple exons or multi-genic deletions with substantial breakpoint heterogeneity. Most CNVs detected are causal for severe hemophilia (Factor VIII and Factor IX levels < 1%) and a subset of CNVs (N=16) are likely novel causal variants as they have not been previously reported. CNVs in two samples co-occurred with likely deleterious variants and a subset of CNVs co-occurred with likely benign missense variants. Together, these analyses demonstrate substantial heterogeneity in CNV size and breakpoint location. The identification of CNVs which impact a hemophilia gene and neighboring gene(s) may have clinical implications and raises the possibility of previously undetected hemophilia-associated syndromes. Moreover, combined analysis of SNVs and fully characterized CNVs will enable increased understanding of hemophilia presentation and complications.

876F

Exome sequencing reveals novel compound heterozygous mutations in FOXN1 in patients with severe immunodeficiency and no alopecia. *S. Khan[1], G. Pardon[2], Q. Du[1], B. Wakeland[1], K. Viswanathan[1], M. Louise Markert[2,3], M. Teresa de la Morena[4,5,6], N. van Oers[1,4,7], E. Wakeland[1].* 1) Department of Immunology, Univeristy of Texas Southwestern Medical Center, Dallas, TX; 2) Department of Pediatrics, Duke University, Durham, NC; 3) Department of Immunology, Duke University, Durham, NC; 4) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; 5) Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX; 6) Children's Health, Dallas, TX; 7) Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, TX.

   Primary immune deficiency diseases (PIDs) are a diverse group of genetic disorders in humans of varied clinical and genetic heterogeneity that result in altered immune function. More than 220 PIDs have been identified thus far, however, many patients with a PID phenotype remain with an unknown genetic etiology. Recent approaches to identify novel mutations in genes resulting in PIDs have included both whole genome and exome sequencing. Here we utilized Exome sequencing in a child that presented with unique clinical presentation: T cell lymphopenia and no alopecia. In order to identify underlying causative genetic mutations, we sequenced proband, parents and all the unaffected siblings. Detailed genetic and phenotypic characterization identified novel compound heterozygous mutations in Forkhead Box N1 (*FOXN1)* gene in the proband. In addition we independently identified compound heterozygous mutations in *FOXN1* gene in two other patients with similar clinical presentation. FOXN1 is an epithelial-specific transcription factor essential for the development of the thymus. Patients with mutations in *FOXN1* (OMIM # 600838) are born with a severe T-cell lymphopenia that presents in combination with alopecia and nail dystrophy. However all three patients in this study had normal hair and skin development. In order to better understand the functional impact of mutations, we used CRISPR/Cas techniques and generated the corresponding mutations in the mouse *FOXN1* sequence. We are in the process of phenotypically characterizing these mice. We hope that our findings will provide novel insight into distinct functional role of FOXN1 in thymus and skin. .

877W

An update on the diagnostics, phenotype and treatment of deficiency of ADA2 (DADA2). *I. Aksentijevich[1], N. Deuitch[1], W. Pei[2], K. Bishop[2], J. McElderry[2], S. Burgess[2], R. Sood[2], D. Stone[1], A.K. Ombrello[1], D.L. Kastner[1], Q. Zhou[1].* 1) Inflammatory Disease Section, National Human Genome Research Institute, Bethesda, MD; 2) Translational and Functional Genomics Branch, National Human Genome Research Institute, Bethesda, MD.

   Background: DADA2 is a recessively inherited disease caused by loss-of-function mutations in the *CECR1* gene, which encodes adenosine deaminase 2 (ADA2). Patients typically present with fevers, livedoid rashes, polyarteritis nodosa (PAN) and recurrent ischemic strokes leading to neurological impairment. Recently, DADA2-associated phenotypes have expanded to include pure red cells aplasia, Diamond-Blackfan anemia (DBA), Sneddon's syndrome, and patients with common variable immunodeficiency (CVID). The precise molecular mechanism of this disease is not clear and it appears that ADA2 plays an important role in the development of myeloid cells. **Methods**: We Sanger sequenced all 9 exons of *CECR1* in 84 new patients with suspected DADA2. Serum ADA2 enzyme activity was tested using a modified kit from Diazyme. Patient samples were analyzed by Nanostring, RT-PCR, and cytokine profiling. CRISPR/Cas-9 KO zebrafish were made to study the mechanism of vascular phenotype. **Results**: We identified 39 patients with pathogenic mutations in *CECR1*. Thirty-three patients carry bi-allelic homozygous or compound heterozygous mutations, while 6 patients carry only a single detectible pathogenic mutation suggesting the presence of cryptic mutations. We report 16 *CECR1* mutations that had not previously been associated with DADA2. The 20 patients tested for ADA2 activity had significantly decreased protein activity compared to healthy controls and unaffected family members. Interferon-mediated gene expression signature correlate with the disease severity. Intracellular cytokine staining in primary mononuclear cells demonstrated increased production of IL-1b, TNF and IL-6 in patients. Zebrafish KO embryos presented with hemorrhagic strokes. This extended cohort of patients has helped identify other previously unappreciated clinical features associated with DADA2. Many of the patients have been misdiagnosed until the adult age and subjected to unnecessary and non-effective therapies. Treatment with TNF inhibitors proved to efficiently reduce inflammation and prevent recurrence of strokes. **Conclusion:** Since the initial description of the disease in 2014, there have been 99 reported cases in the literature. The phenotype variability is significant among family members and even in individuals with the same genotype emphasizing a role for other modifying gene alleles. Therapies with TNF inhibitors have been life changing. .

878T

Factor (F)VIII gene mutation type and type of FVIII therapeutic influence the risk of developing neutralizing anti-FVIII antibodies independent of genetic relatedness, age, race, hemophilia A (HA) severity, therapeutic exposure days, and haplotype in HA patients of the PATH Study. *T.E. Howard[1], V.P. Diego[1], K.R. Viel[2], B.W. Luu[1], A. Ameri[3], A. García-Hernández[1], L.V. Dinh[4], M. Almeida[1], S.A. Cole[5], R. Rajalingam[6], C.K. Kasper[7], J. Blangero[1], L. Almasy[8].* 1) South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley School of Medicine, Brownsville, TX; 2) Histonis, Inc, 200 Cowan St, #304, Pittsburgh, PA, 15211; 3) Department of Pediatrics, Division of Pediatric Hematology, Augusta University, Augusta, GA, USA; 4) Haplomics Biotech, 2651 FJRM AVE, Ste 111, Brownsville, TX, 78520; 5) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA; 6) Immunogenetics and Transplantation Laboratory, Department of Surgery, University of California San Francisco, San Francisco, CA , USA; 7) Orthopaedic Hemophilia Treatment Center, 403 W Adams Blvd, Los Angeles, CA 90007; 8) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

There has been a reinvigorated interest in Mendelian diseases. Hemophilia A (HA), an X-linked bleeding disorder and classical Mendelian disease, is caused by Factor VIII (FVIII) gene (F8) mutations and plasma deficient in FVIII clotting activity. While FVIII replacement is the standard treatment for HA, the development of neutralizing antibodies ("inhibitors") to FVIII therapeutics is the most serious obstacle to effective patient care. Because there are many important variables that influence the development of inhibitors and because the number of patients that may be enrolled in investigative studies tend to be small, compared to numbers available for complex disease studies, it is important to develop powerful approaches that can simultaneously and yet robustly account for multiple variables given limited data on small sample sizes. Toward this end, we have applied the mixed linear model of statistical genetics to data on inhibitor development in the PATH Study. To control for confounding due to genetic relatedness, we used a genetic relationship kernel estimated from the study participants' own genomic data measured with the ImmunoChip. We analyzed three dichotomous inhibitor variables: historical inhibitor status (H-Inh); enrollment inhibitor status (E-Inh); and lifetime inhibitor status (L-Inh). We used backward-selection, Bayesian model averaging and selection methods to evaluate the importance of age, race, baseline HA severity, therapeutic exposure days, FVIII haplotype, a mutation-type ranking score (Mu), and a dichotomous variable contrasting plasma derived (pd) vs. recombinant (r)FVIII (scored 1 if rFVIII and 0 if pdFVIII) denoted by PvR. Mu was constructed by assigning ordered ranks to a patient's mutation-type, where the ranks were determined in accordance with current knowledge about the severity-level associated with each mutation-type. We found that both Mu and PvR were significantly associated with increased risk for H-Inh (Mu: $p = 0.003$; Posterior Probability (PP) = 0.66; PvR: $p = 0.001$; PP = 0.78), for E-Inh (Mu: $p = 0.005$; PP = 0.82; PvR: $p = 0.003$; PP = 0.72), and for L-Inh (Mu: $p = 0.0009$; PP = 1; PvR: $p = 0.002$; PP = 0.88). Our statistical genetic approach to studying HA is both novel and powerful, and the results lead us to conclude that both the type of HA-causing F8 mutation and type of FVIII therapeutic (with rFVIII having greater risk than pdFVIII) have a significant influence on the development of inhibitors.

879F

Immuno-genomic association analysis of Factor VIII immunogenicity in hemophilia patients of the PATH Study using the ImmunoChip Array. *B.W. Luu[1], M. Almeida[1], V.P. Diego[1], L.V. Dinh[2], K.R. Viel[3], K. Haack[4], R. Rajalingam[5], C.K. Kasper[6], A. Ameri[7], A. García-Hernández[1], S.A. Cole[4], L. Almasy[8], J. Blangero[1], T.E. Howard[1] On behalf of the other PATH (Personalized Alternative Therapy for Hemophilia) Study investigators.* 1) South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA; 2) Haplomics Biotechnology, 2651 FJRM Avenue, Ste. 111,Brownsville, TX, 78520; 3) Histonis, Inc, 200 Cowan St, #304, Pittsburgh, PA, 15211; 4) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA; 5) Immunogenetics and Transplantation Laboratory, Department of Surgery, University of California San Francisco, San Francisco, CA , USA; 6) Orthopaedic Hemophilia Treatment Center, 403 W Adams Blvd, Los Angeles, CA 90007; 7) Department of Pediatrics, Division of Pediatric Hematology, Augusta University, Augusta, GA, USA; 8) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

Hemophilia A (HA) is the X-linked recessive bleeding disorder caused by loss-of-function factor (F)VIII gene (F8) mutations leading to deficiencies in plasma FVIII coagulant activity. While FVIII replacement is currently the standard treatment for HA, the development of neutralizing anti-FVIII antibodies ("inhibitors") is the most serious and common obstacle to effective patient care. In this report, we present results from the PATH Study subjects for the identification of population specific immune-response gene variants associated with the development of FVIII inhibitors. We analyzed the relationship between causative F8 mutations, patient race and disease severity to the development of these neutralizing anti-drug antibodies (ADAs). PATH participants were genotyped using the ImmunoChip array, which enabled empirical estimates of genetic relatedness between all pairs of individuals. This information was used in a linear mixed model to test the association of each single-nucleotide polymorphisms (SNP) with a dichotomous lifetime inhibitor status variable. We stratified our overall sample by race into Black and White subsets of patients. We located two SNPs with significant regulatory potential—one each on chromosome (Chr)1 and Chr9—in the White patients only, which remained significant after combining their likelihood ratio test statistics with those from the analysis of the Black patients. The most significant result was for rs2276404 (p-value < 4.27e-08) in IFI16, which encodes γ-interferon-inducible (IFI) protein 16. The only other array-wide significant result was for a variant located at position 4,291,392 on Chr9 which, while not present in dbSNP, is in a highly-conserved region of GLIS3 intron-1. As neither IFI16 nor GLIS3 has ever been associated with the development of inhibitors, our results further support the involvement of immune-response gene variation in FVIII immunogenicity risk and provide the first evidence that their influence may be race-specific. Furthermore, they indicate the need for patient-specific immune- cell-based platforms to elucidate the mechanisms by which these variants influence ADA development and to identify potential therapeutic immune system targets.

880W

A new patient with common variable immunodeficiency (CVID) and autoinflammation due to biallelic mutations in HOIP. *H. Oda[1,2], H. Kuehn[1], P. Hoffmann[1], N. Deuitch[1], J. Stoddard[1], M. Gadina[1], S. Rosenzweig[1], I. Aksen-tijevich[1], D. Kastner[1]. 1) NIH, Bethesda, MD; 2) Kitano Hospital, Japan.*

   [**Background**] HOIP is the catalytic subunit of the linear ubiquitination chain assembly complex (LUBAC), which is important for the activation of the innate and adaptive immune responses. LUBAC depletion leads to attenuation of NF-kB signaling, and LUBAC-deficient patients present with recurrent invasive bacterial infections. In contrast to immunodeficient fibroblasts, monocytes of LUBAC-deficient patients are highly responsive to IL-1 stimulation and produce high levels of proinflammatory cytokines IL-6 and MIP-1α. To date only one patient with the HOIP deficiency has been reported; her symptoms include, multiorgan autoinflammation, immunodeficiency, amylopectinosis, and systemic lymphangiectasia (Boisson, 2015). [**Case**] We investigated a 7 year-old girl with CVID and systemic autoinflammation. She has a clinical history of severe viral and bacterial infections, and while her immunoglobulin levels were normal, impaired antibody response to pneumococcal polysaccharide vaccines was observed. She also experiences episodes of autoinflamma-tion including recurrent fevers, arthritis, rash, and persistent ESR elevation. [**Method and Results**] We performed NGS-based targeted sequencing of 353 immunologically related genes, followed by Sanger sequencing. In the *RNF31* gene, which encodes HOIP, we identified a rare missense variant (NM_017999;c.1336C>T;p.R446W) in exon 8 and a novel intronic variant (c.1737+3A>G) in intron 9, which were confirmed to be biallelic by subcloning. RT-PCR amplification of exons 6 through 10 of the patient's cDNA followed by subcloning demonstrated that both of the variants create alternatively spliced *RFN31* transcripts. siRNA-mediated knockdown of HOIP in THP1, a human monocytic cell line, resulted in higher responsiveness to cytokine stimulations, which phenocopied the cellular phenotype of the previously reported patient with HOIP deficiency. [**Discussion**] We report a second patient with suspected HOIP-deficiency. The phenotypic similarity and the preliminary results suggest that these two HOIP mutations could explain the patient's phenotype. Further molecular analyses are required to elucidate mechanisms of the apparently contradictory observation of immunodeficiency and autoinflammation associat-ed with the HOIP deficiency.

881T

Novel truncating variant in single immunoglobulin Interlekin-1 receptor related (*SIGIRR*) gene in a dominant family with early-onset inflammatory bowel disease. *J.E. Horowitz[1], J. Staples[1], A. Klauer King[1], E. Coonrod[2], A. Rutherford[2], J.G. Reid[1], J.D. Overton[1], J. Singer[2], B. Moore[2], M. Yandell[2], A.R. Shuldiner[1], F.E. Dewey[1], O. Gottesman[1], S. Guthery[2], C. Gonzaga-Jauregui[1]. 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 2) Department of Pediatrics, University of Utah, Salt Lake City, UT.*

   Inflammatory bowel disease (IBD) is a genetically heterogeneous, chronic inflammatory response to commensal microbiota in the GI tract with average age of onset at 30y. Severe, monogenic forms of IBD can present with pediat-ric onset (<18y) and have been attributed to rare, highly-penetrant variants in ~50 'Mendelian' genes. However, the genetic architecture of early onset IBD (EO-IBD) is poorly understood, and the majority of patients remain genetical-ly undiagnosed.   To identify novel candidate EO-IBD genes, we performed whole-exome sequencing and trio-based variant analysis on a 13-year-old EO-IBD patient, his IBD-affected mother, and his unaffected father. Through this, we identified a novel, predicted-deleterious truncating indel variant in the Single Immunoglobulin Interleukin-1 Receptor Related (*SIGIRR: c.557delA*; p.K186fs*31) in the EO-IBD proband inherited from his affected mother.   *SI-GIRR* encodes a negative regulator of interleukin (IL)-1 and toll-like receptor (TLR) signaling expressed in gut epithelial and hematopoietic cells. In mice, *Sigirr* is a critical regulator of gut homeostasis, and humans with variants in the extracellular domain of *SIGIRR* develop necrotizing enterocolitis in infancy. However, a role for *SIGIRR* in EO-IBD has not been established.   To function-ally validate *SIGIRR* as a mediator of EO-IBD, we generated a lymphoblastoid cell line (LCL) from the EO-IBD patient and analyzed it using a pro-inflamma-tory cytokine screen complemented with RNA-sequencing. In unstimulated cells, we observed upregulation of key immune modulators IL-1β, IL-8, and IL-6 in the *SIGIRR* EO-IBD patient LCLs relative to both LCLs generated from healthy controls and from EO-IBD patients not harboring truncating *SIGIRR* variants, suggesting a unique inflammatory signature in IBD patients carrying truncating *SIGIRR* variants. Further, we observed that *SIGIRR* EO-IBD pa-tient-derived LCLs are refractory to either IL-1β or TLR stimulation, suggesting constitutive activation of these inflammatory pathways in the truncation of *SIGIRR*.   Our analysis has identified a cytoplasmic truncation in *SIGIRR* in an EO-IBD patient with a unique, constitutively active pro-inflammatory cyto-kine profile. Future studies will seek to determine how this truncation affects downstream IL-1β and TLR signaling, to elucidate how *SIGIRR* maintains commensal tolerance, and to identify potential therapeutic pathways to amelio-rate inflammation in EO-IBD patients.

882F

Utility of whole genome sequencing for population newborn screening for immunodeficiencies. *D.L. Bodian[1], A.R. Pavey[1,2,3], T. Vilboux[1], A. Khromykh[1], N.S. Hauser[1], K. Huddleston[1], E. Klein[1], A. Black[1], M.S. Kane[1], R.K. Iyer[1], J.E. Niederhuber[1], B.D. Solomon[1,4,5].* 1) Inova Translational Medicine Institute, Falls Church, VA; 2) Department of Pediatrics, Walter Reed National Military Medical Center, Bethesda, MD; 3) Department of Pediatrics, Uniformed Services University of Health Sciences, Bethesda, MD; 4) Inova Children's Hospital, Falls Church, VA; 5) GeneDx, Gaithersburg, MD.

Immunodeficiency screening has been added to many state-directed newborn screening programs but is currently limited to screening for severe T-cell lymphopenia disorders. To evaluate the potential of genomic sequencing to augment current immunodeficiency screening, we analyzed whole genome sequences (WGS) and clinical data from a cohort of 1,349 newborn-parent trios by genotype-first and phenotype-first approaches. For the genotype-first approach, we analyzed the WGS data for predicted protein-impacting variants in 329 immunodeficiency-related genes. As a phenotype-first approach, electronic health records were used to identify children with clinical features suggestive of immunodeficiency. Genomes of these children and their parents were analyzed using a separate pipeline for identification of candidate pathogenic variants for rare Mendelian disorders. The results show that WGS provides adequate coverage for most known immunodeficiency-related genes. Within these genes, 8,502 distinct predicted protein-impacting variants were identified in the cohort, and 5 individuals carried potentially pathogenic variants requiring expert evaluation. One clinically asymptomatic individual was found genomically to have C9 deficiency. Of the symptomatic children, one was molecularly identified as having an immunodeficiency condition and two were found to have other molecular diagnoses. These results suggest that neonatal genomic sequencing can potentially augment newborn screening for immunodeficiency by broadening the range of immunodeficiencies that can be screened, and by identifying likely causative mutations.

883W

Severe EDS III with cell activation syndrome (MCAS) in infancy and young children. *C. Tsai.* pediatrics/genetics, childrens Hospital colorado, Aurora, CO.

Mast cell activation syndrome (MCAS), is a condition in which mast cells inappropriately and excessively release chemical mediators, resulting in a range of chronic symptoms. Recently, it has been noticed that there is a higher-than-normal representation of MCAS among the EDS population and often complicated with flushing, pruritus, dysautonomia, functional gastrointestinal symptoms, chronic pain, and connective tissue abnormalities, POTS, migraines and psychiatric symptoms. I herein present 5 young children with EDS type III with MCAS: two presented as infants with severe feeding difficulty, inconsolable crying and irritability, hypermobile joints and multiple fractures in ribs, without other finding consistent with NAT. One of them has bleeding from a torn frenulum during bottle feeding. Both received molecular work up for 9 plus gene OI/ brittle bone disease panels and were negative. MOC of both infants have history of depression required antidepressant during pregnancy. Both infant required G tube and one required GJ tube and eventually both improved dramatically with H1 and H2 antihistamine. Family three have two siblings sets, both presented with severe feeding difficulty, flushing, vomiting and hypermobile joints, thin transparent skin, and required G tube; exome sequencing was not revealing and both respond well to antihistamine treatment. The 4th patient presented with low tone, poor feeding requiring G tube, rashes, food allergy, periodic fever, severe sensory defensiveness, but respond well with Periactin and fluid therapy. Elevated tryptase levels were not found in any of these patients. Mast cells are found in tissues throughout the body, particularly in association with structures such as blood vessels and nerves. Activation through various receptors leads to distinct signaling pathways and synthesis of chemokines and cytokines. There is no universal recognized explanation to why MCAS is disproportionately found in EDS. I hypothesize that the mast cells in some patients with EDS are more fragile, and subject to mechanical changes such as over stretching of the cell membrane/ receptors and subsequently trigger histamine, cholinergic, inflammatory and serotoninergic pathways. Proper blockage by antihistamine can avoid further complications and improve feeding. I also emphasize the recognition of clinical features of young infants to avoid wrongful accusation of NAT. This should be considered as a differential diagnosis of brittle bone diseases. .

884T

Case report of a patient with uncharacterized IFN-γ mediated autoinflammatory disorder. *O. Schnappauf, A. Ombrello, P. Hoffmann, D. Kastner, I. Aksentijevich.* NHGRI (National Human Genome Research Institute), Inflammatory Disease Section, Bethesda, USA.

Background: We report a 3.5y old female patient from a non-consanguineous marriage who was seen at the NIH in September of 2015. She was born full-term with a normal weight, but soon after birth developed fevers accompanied by an erythema annulare rash on her cheeks that spread to her entire body. She also presented with conjunctivitis and abdominal bloating. During her disease, she had multiple episodes of infection, septicemia and two episodes of aseptic meningitis. Multiple lumbar punctures revealed increased lymphocyte counts while brain CT showed increased CSF and micro calcifications. Her skin biopsy revealed lobular panniculitis with inflammatory infiltration composed mostly of myeloid cells. Previous treatments included; immunosuppressives, steroids, IVIG, and anti TNF and anti IL-1 therapies. She did not respond to either cytokine inhibitor. **Results**: During her visit to the NIH the patient had extremely high serum cytokine levels of APRIL and BAFF, consistent with low or absent B-cells. Additionally, the levels off pronflammatory cytokines IFN-α, IFN-γ, IP-10, IL-12, TNF-R1 and TNF-R2 were significantly increased. In particular, CXCL9/MIG (monokine induced by gamma interferon) was too high to be determined. Functional assays showed increased levels of pSTAT1 and pSTAT3 at baseline and in IFN-α stimulated CD4[+], CD8[+] T-cells and monocytes. Gene expression analysis of whole blood samples showed a strong interferon signature, which was comparable to that seen in CANDLE-patients. Together, these results suggest that she has an IFN-γ mediated disease. Genetic testing of known primary immunodeficiency genes, autoinflammatory genes and four hemophagocytic lymphohistiocytosis (HLH) -associated genes (*PRF1*, *UNC13D*, *STX11*, *STXBP2*) was inconclusive. No obvious candidate gene was identified via the patient's, unaffected parents and an unaffected sibling by WES. Patient had excellent response to anti-IFN-γ therapy, however she later died from the complications of allogeneic bone marrow ttransplantation. **Conclusion:** We describe a patient with an uncharacterized type II (IFN-γ)-mediated disease. Most patients with interefronopathies have type I (IFN-a, IFN-b)-mediated inflammatory diseases. Analysis of the family whole genome sequencing data is in progress.

885F

A novel *GFI1B* mutation at the first zinc finger domain causes congenital macrothrombocytopenia. *Y. Uchiyama[1,2], Y. Ogawa[2], S. Kunishima[3], M. Shiina[4], M. Nakashima[5], K. Yanagisawa[2], A. Yokohama[6], E. Imagawa[1], S. Miyatake[1], T. Mizuguchi[1], A. Takata[1], N. Miyake[1], K. Ogata[4], H. Handa[2], N. Matsumoto[1].* 1) Human Genetics, Yokohama City University, Yokohama, Kanagawa, Japan; 2) Department of Hematology, Gunma University Graduate School of Medicine, Gunma, Japan; 3) Department of Advanced Diagnosis, Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan; 4) Department of Biochemistry, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 5) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan; 6) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Congenital macrothrombocytopenia (CMTP) is characterized by thrombocytopenia and giant platelets with variable bleeding tendency. *GFI1B* encoding growth factor independent 1B transcriptional repressor is one of the mutant genes causing autosomal dominant CMTP. Thrombocytopenia, giant platelets, and reduction in platelet α-granule content were common features in previous seven CMTP cases with *GFI1B* mutations. The residual CD34 expression on the giant platelets and the abnormally shapes of red blood cells were reminiscent of the three reported cases. *CD34*, an optical marker for haematopoietic stem cells, is expressed on the surface of platelets in patients with *GFI1B* macrothrombocytopenia, and abnormal differentiation of megakaryocyte–erythroid progenitor cells is implicated. Functional analysis based on cell lines and null mice showed that mutant *GFI1B* led to aberrant development and differentiation of megakaryocytes and erythroid cells. Previous studies also showed that mutations depleting the fifth to sixth zinc finger domains lose their binding ability to the target genes and *GFI1B*-macrothrombocytopenia was occurred to the dominant negative fashion. Here, we identified a novel missense *GFI1B* mutation c.494G>T:p.(Cys165Phe) located at the first zinc finger domain in five cases of a CMTP family by whole exome sequencing. Current cases also showed that thrombocytopenia and giant platelets. However, number of α-granules in the platelets and red blood cell morphology were normal in our patients. Persistent CD34 expression was recognized in the surface of platelets in our cases. These findings imply that the first and second zinc finger domains of *GFI1B,* which do not directly bind to the target genes are also important for the development and the differentiation of megakaryocytes. Furthermore, the detection of CD34 on platelets might be a useful diagnostic biomarker for *GFI1B* macrothrombocytopenia.

886W

Identification and characterization of adenosine deaminase 2 variants in pediatric vasculitis. *K. Gibson[1,2], D.A Cabral[3,4], B. Drogemoller[2,5], X. Xhan[2,5], F. Miao[2,5], K. Morishita[3,4], E. Gill[6], R.E.W. Hancock[6], C. Ross[5,2], K. Brown[3,2] On behalf of the PedVas Initiative Investigator Network.* 1) Dept of Medical Genetics, The University of British Columbia, Vancouver, British Columbia, Canada; 2) BC Children's Hospital Research Institute, Vancouver, BC; 3) Dept of Pediatrics,The University of British Columbia, Vancouver, British Columbia; 4) BC Children's Hospital, Vancouver, British Columbia; 5) Faculty of Pharmaceutical Sciences, The University of British Columbia, Vancouver, British Columbia; 6) Dept of Microbiology and Immunology, The University of British Columbia, Vancouver, British Columbia.

   Deficiency of adenosine deaminase 2 (DADA2) is a recently recognized, autosomal recessive genetic disease with various systemic vascular and inflammatory manifestations; patients present with early-age onset and often recurrent strokes. The clinical features and histological findings of DADA2 overlap with those of early childhood onset polyarteritis nodosa (PAN), a primary "idiopathic" systemic vasculitis, characterized by necrotizing inflammatory lesions of small and medium-sized vessels. Despite similar clinical presentation, individuals with PAN and DADA2 may benefit from different therapy. While treatment of primary chronic vasculitis is usually with toxic immunosuppressive drugs, there has been clinical indication that DADA2 patients respond better to less toxic IL-6 receptor antagonists and anti-TNFα therapy. We aimed to screen patients with PAN, cutaneous PAN, unclassifiable phenotype, or chronic vasculitis of any type with onset-age less than 5 years for variants in adenosine deaminase 2 (*ADA2*).   Of the 493 pediatric patients included in our international, multi-ethnic cohort, ARChive (A Registry of Childhood Vasculitis), there were 99 patients who provided DNA samples, and 41 of these fulfilled screening criteria. We have sequenced the coding exons of *ADA2* in these 41 patients and identified variants in *ADA2* with known and novel association with DADA2; four patients were found to be homozygous or compound heterozygous for rare (MAF < 0.01) predicted pathogenic variants. The impact of the identified variants on ADA2 expression, secretion, and enzymatic activity were quantified by qPCR, standard ELISA, and colorimetric assays. Whole blood RNA sequencing enabled transcriptomic profiling of patients with DADA2 versus PAN and other types of chronic vasculitis that could inform treatment. At present, the gold standard for direct diagnosis of vasculitis is histopathological examination of biopsy specimens from involved organs. Screening for pathogenic variants in *ADA2* among patients with possible early onset chronic vasculitis or PAN phenotype may identify and diagnose patients, perhaps without requirement for biopsy. Early diagnosis of DADA2 patients may spare them treatment with toxic systemic immunosuppressive drugs, and allow more effective intervention with targeted biologic and/or gene therapy.   This work was supported by a Canadian Institutes of Health Research grant for the PedVas Initiative [TR2-119188 to DAC].

887T

Variants in *FOXP1* cause syndromic genitourinary tract defects. *N. Bekheirnia[1], J.A Rosenfeld[1], M.A Manning[2], M.F Wangler[1], J.S Pluciniczak[3], M. Pietryga[3], N.L Rider[1], Y. Yang[1], C. Eng[1], M.C Braun[1], D.J Lamb[1], M.R. Bekheirnia[1].* 1) Baylor College of Medicine, One Baylor Plaza, Houston, TX.77030; 2) Stanford University School of Medicine, 291 Campus Drive, Stanford, CA 94305; 3) Nationwide Children's Hospital, 700 Children's Drive ,Columbus, OH 43205.

   Background: Forkhead box P1 (FOXP1) is a key transcription factor in the development of major organs including heart, lung and brain. FOXP1 helps regulate cell-type- and tissue-specific gene transcription both during development and in the adult stage. *De novo* variants in *FOXP1* (OMIM # 613670) have been associated with a syndrome consisting of intellectual disability (ID), hydrocephalus, cognitive impairment, cardiac defect, kidney and genitourinary (GU) tract abnormalities by our group in 8 individuals **Methods**: Whole exome sequencing (WES) was performed in a clinical diagnostic setting in ~8,000 individuals with variable indications including ID and other neurological presentations. Patient information was collected as part of clinical laboratory intake form and families were consented accordingly. **Result:** *De novo* variants in *FOXP1* were identified in 5 additional individuals. These variants include p.E463A, p.V445fs, p.R544X, p.N516S, and p.R525X. None of these variants expect for p.R525X have been previously reported in patients or controls. The patients' phenotypes are consistent with the *FOXP1*-associated syndrome, with the additional finding of GU/kidney abnormalities in 4 out of 5 patients, including hydronephrosis, undescended testes, micropenis, duplicated collecting system, and retractile testes. Information about one patient's GU/ kidney phenotype is not available at this time. **Conclusion**: In this study, we now report five additional cases (total of 13) with de novo variants in *FOXP1*. Ten out of thirteen individuals have kidney and GU abnormalities. We conclude that kidney and GU defects can be part of the *FOXP1* syndrome, and therefore these patients require clinical evaluations for such abnormalities. These findings also implicate *FOXP1* in kidney and GU development. This study was supported in part by K12 DK083014 Multidisciplinary K12 Urologic Research Career Development Program.

888F

Comprehensive analysis using targeted sequencing panel for congenital anomalies of the kidney and urinary tract and nephronophthisis in Japan. *N. Morisada[1,2], A. Shono[1], K. Nozu[1], R. Tanaka[3], K. Iijima[1].* 1) Pediatrics, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan; 2) Clinical Genetics, Hyogo Prefectural Kobe Children's Hospital, Kobe, Hyogo, Japan; 3) Nephrology, Hyogo Prefectural Kobe Children's Hospital, Kobe, Hyogo, Japan.

Background Chronic kidney disease (CKD) in children has very important problems for childhood growth and neurocognitive development. Genetic renal disorders such as congenital anomalies of the kidney and urinary tract (CAKUT), chronic hereditary glomerulonephritis (GN), steroid resistant nephrotic syndrome (SRNS) and nephronophthisis (NPHP) are major leading diseases to a pediatric CKD. Urinalysis can indicate chronic GN and SRNS with CKD, CAKUT and NPHP however show only slight abnormalities in urinalysis. In addition, both CAKUT and NPHP are highly genetically heterogeneous disorders, therefore the exact molecular diagnosis for CAKUT and NPHP is much difficult. **Methods** Next-generation sequencing (NGS) analysis was performed on 157 Japanese individuals who were clinically diagnosed as CAKUT, NPHP or renal insufficiency of unknown cause in childhood using three custom designed panels containing 91, 128 and 172 genes associated with CAKUT or NPH. All procedures were reviewed and approved by the Institutional Review Board of Kobe University School of Medicine. **Results** Seventeen responsible genes in 30 individuals were identified (diagnostic rate 19.1%). The detected responsible genes were as follows; *BBS10, GATA3, HNF1B, INF2, NPHP1, NPHP3, OFD1, PAX2, PKD1, PKHD1, REN, RET, SALL1, SDCCAG8, TMEM67, TTC21b* and *WDR19*. It is noted that renal cystic phenotypes including NPHP were most frequently found in cases, of which causative genes were identified (50.0%). Furthermore, we found that five individuals of which clinical diagnosis and genetic diagnosis were considerably different. Two individuals with serum electrolyte abnormalities, who were clinically diagnosed as Bartter syndrome but no associated responsible genes were identified, were both resulted in harboring *PAX2* pathogenic variants and thereby diagnosed as papillorenal syndrome. Another two individuals with multiple renal cysts, who were diagnosed as NPHP or autosomal recessive polycystic kidney disease, were resulted in carrying *HNF1B* pathogenic variants and followed by diagnosis of an autosomal dominant CAKUT; renal cysts and diabetes (RCAD). **Conclusion** It is suggested that NGS analysis using targeted sequencing panels is useful for the definitive diagnosis and genetic counselling for CAKUT and NPHP.

889W

***HES1* gene screening in a cohort of patients with hipopituitarism reveal an allelic variant c.578G→A (p.G193D).** *R. Kertsz, A.P Otto, J.L.O Madeira, L.R.S Carvalho.* Unit of Developmental Endocrinology, University of Sao Paulo Medical School, Sao Paulo, Sao Paulo, Brazil.

Hypopituitarism is defined as deficiency of one or more hormones produced and secreted by the pituitary gland. *HES1* is a gene that encodes a transcription factor that plays a role in the pituitary development and differentiation, which makes it a possible candidate gene for hypopituitarism in humans. In order to look for allelic variants in *HES1*, 193 Brazilian patients (124 male) with congenital hypopituitarism were analyzed with mean age of 11.9 years (± 6.9), -4.4 Z score for height, -0.8 (± 1.5) Z score for BMI, 60 were Isolated growth hormone deficiency (GHD) and 132 combined GHD. The code region was amplified by polymerase chain reaction, followed by the Sanger sequencing method. The allelic variant c.578G→A (p.G193D) was found in heterozygous state in a male patient with growth deficit since childhood, diagnosed with GH and TSH deficiencies at 11 years that evolved with LH/FSH and ACTH deficiencies. The pituitary resonance image reveals an ectopic neurohypophysis, non visualized stalk and hypoplastic adeno-hipophysis. Familial segregation revealed that mother and 3 out 5 siblings did not harbor the variant. Father's DNA was unavailable. The two siblings harboring the allelic variant did not present growth deficit. *In silico* analysis using 4 different tools showed a contradictory result. Polyphen, Sift and SNAP tools suggested being benign and Mutation Taster, disease-causing by altering the chemical physical amino-acid characteristic leading to a possible splicing site modification. It is under rs771445679 in the GenomAD data base with 7.083e-5 allele frequency (AF). Two other allelic variant were found in our cohort c.741C→T (p.Val247Val) in 7 patients (1 homozygous) that it was described in the genomAD data base under rs7629144 with 0,01361 AF. Another allelic variant in a splicing region c. 293-5 C→T was described under rs148500265 in the genomAD presenting 0,003399 AF. This residue is not conserved among different species. The 88 final residues of *HES1* (193-280) interacts with the 60 final residues of *CBFA1* leading to transactivation, what prompted us to test the functional impact of *HES1* c.G578A (p.G193D) in vitro. *Hes1* cDNA WT plasmid was obtained and the mutant generated through mutagenesis. The assay for the in vitro analysis is under development. We conclude that *HES1* in the pedigree presents an incomplete penetrance and in vitro study will collaborate to establish a relationship between phenotype and genotype.

890T

Rapid paediatric sequencing (RaPS) from patient to variant: A step-by-step workflow and case report. *L. Boukhibar[1], H. Wiliams[1], A. Gagunashvili[1], L. Ocaka[1], P. Le Quesne Stabej[1], E. Clement[2], W. Jones[2], L. Jenkins[3], S. Marks[4], C. Bacchelli[5], M. Bitner-Glindzicz[6], P. Beales[1].* 1) GOSgene, Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health. UCL., London, United Kingdom; 2) Clinical Genetics and Genomic Medicine, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom; 3) NE Thames Regional Genetics Laboratory, Great Ormond Street Hospital, London, United Kingdom; 4) Department of Paediatric nephrology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK; 5) Experimental and Personalised Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom; 6) Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom.

Disease in severely ill children is often undifferentiated at birth, clinically heterogeneous and progression is extremely rapid. To address the unmet need to quickly identify a genetic cause, we started the **RaPS** (Rapid Paediatric Sequencing) project in collaboration with Great Ormond Street Hospital and NE Thames Regional Genetics Lab. We developed a workflow template to rapidly sequence, analyse and report likely causative variants from patient/parent trios. We describe our workflow and illustrate its clinical impact with a case report. The RaPS workflow comprises 3 stages: 1) patient referral, evaluation and consent 2) whole genome sequencing and data analysis, and 3) variant selection. **Stage 1** involves the referring clinicians and clinical geneticists to rapidly assess suitability based on RaPS Inclusion Criteria. In **stage 2**; library preps and rapid sequencing (Illumina NextSeq/Hiseq Rapid mode 29 hrs) are followed by rapid data analysis from fastq to vcf (*Genalice*, 1.5 hrs/trio). **Stage 3** variant analysis and selection is performed using *Ingenuity Variant Analysis* and consists of filtering steps based on population allele frequencies, predicted deleterious effect and mode of inheritance. Importantly, Stage 3 is a phased variant selection: each phase is powered by panel-specific filters. Phase I analysis comprises phenotype-specific gene panels and is performed as first line of investigation. If no likely-causative variant is identified in Phase I; Phase II and/or III analyses are performed. Phase II panels consist of broader disease-associated genes (OMIM, DDG2P databases). Phase III analysis is research-oriented and consists of variant sharing with the scientific community through GeneMatcher. Likely-causative variants are discussed with at least two clinical geneticists for a consensus and a diagnostic report is issued. We present the RaPS workflow and a case report side-to-side to depict each stage of the process. We report a three week old neonate admitted to PICU with metabolic acidosis, hyperkalaemia and oligoanuric kidney injury. We identified a pathogenic *WT1* variant thus providing a diagnosis of Denys-Drash syndrome. RaPS allowed an early diagnosis and re-shaped clinical management to minimize future tumour development before kidney transplant. The strength of the RaPS workflow lies in its rapidity (5 days) and phased analysis. Our workflow is adaptable for other clinical settings where rapid and accurate genetic diagnosis is a necessity.

891F

A novel deletion in ABCC9 gene identified through whole-exome sequencing of patient with clinical spectrum of Cantú syndrome. *O. Migita[1,2], M. Mizuno[1], K. Aso[1], H. Yamamoto[1], K. Hata[2].* 1) Department of Pediatrics, St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan; 2) Department of Maternal-Fetal biology, National Center Institute for Child Health and Development.

Cantú syndrome is a clinically characterized by hypertrichosis, osteochondrodysplasia and cardiomegaly. Most of the patients have their unique DNA sequences in the *ABCC9* gene that codes for an sulfonylurea receptor 2 (SUR2) protein, which is a member of the superfamily of ATP-binding cassette (ABC) transporters. We conducted a next generation sequencing analysis of whole exon from a 4 month female with phenotypes of a hypertrichosis and cardiomegaly, which were compatible to Cantú syndrome. We detected a novel heterozygous deletion in *ABCC9* that may impact to *ABCC9* proteins. This is the first report to detect *ABCC9* deletion mutation in a patient with Cantú syndrome.

892W

Population-scale linkage mapping in a healthcare system uncovers novel signal for primary biliary cirrhosis. *G.M. Belbin[1,2], S. Kohli[1], M.C. Yee[3], G.L. Wojcik[4], E. Sorokin[4], C.R. Gignoux[5], J.H. Cho[1,2,6], R.J.F. Loos[1,2,7], G. Nadkarni[1], E.E. Kenny[1,2].* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, NY, USA; 2) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) Carnegie Institution for Science, Department of Plant Biology, Stanford, CA, USA; 4) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 5) Colorado Center for Personalized Medicine, University of Colorado, Aurora, CO, USA 5. Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York; 6) Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 7) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

   Linkage mapping is a classical method for the genetic identification of monogenic disease in the absence of a known molecular mechanism. This approach has been used to uncover loci underlying suspected genetic disorders segregating in pedigrees or probands. Broad-scale adoption of genomic data in health systems offers opportunities for population-scale disease mapping. Here we introduce a method that combines genome-wide association with linkage mapping to discover genomic disorders in the BioMe Biobank in New York City. First we detected tracts of genetic homology shared between pairs of patients in the Puerto Rican founder population in BioMe (N=5213) to infer distant pedigree relationships in the absence of known genealogy. Next, we performed statistical association of regions of shared genetic homology with ICD9 billing codes derived from the Electronic Health Record (billing codes; n=8386). One ICD9 code was significantly associated (FDR<0.05) between a recessive locus and a rare ICD9 code (occurrence 27/5213) encoding for "Primary Biliary Cirrhosis" (PBC, $p<2.9\times10^{-9}$, frequency=0.73%). The significant locus was homozygous in three distantly related individuals (greater than 5th cousins), each of whom had a diagnosis of PBC (11% of total PBC diagnoses in Puerto Ricans in BioMe). The mapping interval spanned 2.8cM at 7q21.12 and included the *ABCB4* gene, which has been extensively implicated in monogenic forms of hepatobiliary disease. Whole genome sequencing ruled out any of the 15 known variants underlying dominant or recessive forms of hepatobiliary disease, but revealed sharing of multiple rare variants in the mapping interval, suggesting that the detected signal may be driven by previously unreported variation at the *ABCB4* locus. Finally, analysis of ICD9 codes in heterozygous carriers (n=70) uncovered significant associations to "biliary calculus with cholecystitis" ($p<4.5\times10^{-5}$, O.R=5, (95% CI=3-9)) and "hydrops of the gallbladder" ($p<0.00016$, O.R=151 (95% CI=16-1690)), which were both ranked within the top three hits. This analysis provides a methodological framework for bridging statistical genetics and clinical genomics, and demonstrates that EHR-embedded, population-level research can elucidate the pathology of rare genetic disorders.

893T

A case of vitamin D-dependent rickets type 1a with a novel pathogenic variant in a Mexican patient. *M. Abreu-González[1], C. Candia-Tenopala[2], Y. Mena[1], S. Martínez-Gutiérrez[1], BA. Rodríguez-Espino[1], V. Flores[1], OC. Azotla-Vilchis[1].* 1) Genos Medica, Centro Especializado en Genética, Mexico City, Mexico; 2) Centro Endocrino, Mexico City, Mexico,.

   Vitamin D-dependent rickets type 1A (VDDR-1A) (MIM #264700), also called pseudovitamin D deficiency, is an inborn error of vitamin D metabolism involving defective or diminished conversion of 25-hydroxyvitamin D3 (25(OH) D3) to the active form 1,25-dihydroxyvitamin D3 (1,25(OH)2D3) in the kidney by the enzyme 25(OH) D-1-hydroxylase. This type of rickets presents early in infancy with hypotonia, weakness, severe skeletal disease, failure to thrive, and growth retardation. The treatment consists of supplementation with calcium and active forms of vitamin D. VDDR-1A is an autosomal recessive disorder due to pathogenic variants in the *CYP27B1* gene (MIM *609506).   In this study we present a male Mexican patient born to the first pregnancy of a non-consanguineous, young, healthy couple. He started at 7 months with bilateral tibial bowing and at 1 year old he had two long bone fractures. He was referred to our clinic with the clinical suspicion of osteogenesis imperfecta. The physical examination at 2 years old revealed that his weight and height were below the 3rd centile, mid facial hypoplasia, rachitic rosary, thickness of the wrists with shortening and bowing of the limbs. Biochemically, he had hypocalcaemia, hypophosphatemia with elevated paratohormone levels, leading the diagnosis to hypophosphatemic rickets. A next generation sequencing targeted panel was performed (TruSight One Sequencing Panel). The bioinformatics pipeline analysis of the 26 genes associated with rickets identified two genetic variants located in *CYP27B1*. The first one is an insertion of 7 bases (NM_000785.3: c.1325_1326insCCCACCC) that have been previously reported as pathogenic and a novel transition (NM_000785.3:c.227G>A) that generates a premature stop codon (NP_000776.1:p.Trp76*). This novel variant has not been described, but *in silico* analyses indicated a loss of function for the coded protein due to either a truncated protein or a nonsense mediated decay of the *CYP27B1* mRNA. No additional changes of clinical significance were found in the rest of the analyzed genes, suggesting that the cause for the observed phenotype was indeed due to variants in *CYP27B1*. This case shows a patient who presented with severe skeletal signs of VDDR-1A at a young age probably due to the type of the variants detected. Using a next generation sequencing strategy we could identify the genetic basis of the disease and start an early appropriate pharmacological treatment.

894F

From genetics to therapy: CD55 loss of function leads to protein-losing enteropathy responsive to eculizumab. *A. Kurolap[1,2], O. Eshach Adiv[2,3], T. Hershkovitz[1], T. Paperna[1], A. Mory[1], D. Oz-Levi[1], Y. Zohar[2,4], H. Mandel[2,5], J. Chezar[6], D. Azoulay[6], S. Peleg[2,7], E.E. Half[8], V. Yahalom[9], L. Finkel[9], O. Weissbrod[10,11], D. Geiger[10], R. Shaoul[2,3], D. Magen[2,12], L. Bonstein[2,13], D. Mevorach[14], H.N. Baris[1,2].* 1) The Genetics Institute, Rambam Health Care Campus, Haifa, Israel; 2) The Ruth and Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel; 3) Pediatric Gastroenterology Unit, Rambam Health Care Campus, Haifa, Israel; 4) Institute of Pathology, Rambam Health Care Campus, Haifa, Israel; 5) Metabolic Unit, Rambam Health Care Campus, Haifa, Israel; 6) Laboratory for Clinical Cell Analysis & Translational Research, Hematology, Western Galilee Hospital, Nahariya, Israel; 7) Pediatric Gastroenterology Unit, HaEmek Medical Center, Afula, Israel; 8) Department of Gastroenterology, Rambam Health Care Campus, Haifa, Israel; 9) National Blood Group Reference Laboratory (NBGRL), Magen David Adom (MDA), National Blood Services, Israel; 10) Computer Science Department, Technion - Israel Institute of Technology, Haifa, Israel; 11) Department of Statistics and Operations Research, School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv, Israel; 12) Pediatric Nephrology Institute, Rambam Health Care Campus, Haifa, Israel; 13) Blood Bank and Platelet & Neutrophil Immunology laboratories, Rambam Health Care Campus, Haifa, Israel; 14) Rheumatology Research Center, Hadassah Medical Center and the Hebrew University, Jerusalem, Israel.

Protein-losing enteropathy (PLE) is caused by disruption of lymphatic or intestinal integrity, manifesting as hypoalbuminemia, edema, hypogammaglobinemia, lymphopenia, and malabsorption. One primary disorder manifesting as PLE is primary intestinal lymphangiectasia (PIL), a rare life-threatening disorder of unknown etiology. We studied an extended consanguineous Muslim-Arab family comprising of six patients diagnosed with PLE; one of the patients had pathological and endoscopic evidence of PIL. Hypercoagulability was observed in three of the patients. Whole exome sequencing was performed on three affected individuals and a healthy mother, revealing a candidate frameshift homozygous variant in the *CD55* gene. The finding was supported by Two-Point LOD score for the variant of 4.43, based on genotype information from 32 tested family members. *CD55* encodes the decay-accelerating factor (DAF), a membrane-bound complement-regulatory protein that inhibits the C3 and C5 convertases, thus preventing C3 and C5 cleavage and activation, and protecting cells against complement-induced self-injury. Previous studies in a mouse model of dextran sulfate sodium (DSS)-induced colitis demonstrated that DAF loss leads to a more severe phenotype compared to wild-type animals, reinforcing the protein's role in intestinal tissue protection. We demonstrate by hemagglutination tests and flow cytometry the absence of DAF, confirming the variant's predicted loss-of-function effect. Furthermore, membrane attack complex (MAC) and iC3b depositions on white blood cells (WBC), tested by flow cytometry, were significantly increased in patients compared to age-matched controls. Evidence of abnormal complement activation in the patients corroborated the genetic findings, and prompted off-label approved compassionate therapy with eculizumab (Soliris), a humanized anti-C5 monoclonal antibody, that serves as a terminal complement inhibitor. We report on 100 days of successful treatment in three PLE patients, with significant clinical improvement, normalization of serum albumin, and 60% reduction in MAC deposition on WBCs. The path from genetic analysis to tailored therapy with anti-C5 has provided hope for amelioration of symptoms in *CD55*-related PLE patients. This proof of principle demonstration of eculizumab treatment benefits in CD55-related PLE patients should prompt further study regarding the role of complement and CD55 in other intestinal disorders that may benefit from similar treatment.

895W

High prevalence of PKD2 R803* mutation in Taiwan. *D. Hwang, C. Yu.* Division of Nephrology, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan.

Purpose To study the disease-causing mutations of *PKD1* and *PKD2* of autosomal dominant polycystic kidney disease (ADPKD) in Taiwan, we enrolled 109 families with the diagnosis of ADPKD which were established by clinical and image studies from nephrology investigators in Taiwan. Most of the families were from southern Taiwan. The study was approved by the institutional review board of the Kaohsiung Medical University Hospital, and all procedures were performed in accordance with the declaration of Helsinki. All subjects provided written informed consent before participating in this study. Methods Long-range PCR primers were designed for *PKD1* and *PKD2* genes to avoid the amplification of *PKD1* pseudogenes. Second PCR primers were designed to cover the coding and intron-exon boundaries. Products from long-ranged PCR were 2500-fold diluted to prevent genomic DNA carryover. Multiplex microfluidic PCR system (Fluidigm Access-Array) was used for the second PCR, followed by barcoding PCR with next-generation re-sequencing on Illumina MiSeq. Fastq files from MiSeq were loaded into CLCbio Genomic Workbench for bioinformatics analysis. Data was compared with dbSNP database, Mayo Clinic ADPKD Mutation Database, and functionally prediction with polyphen2. Five microsatellites markers (D4S1534, D4S1542, D4S1563, D4S1544, D4S414) were used to analyze 17 families of *PKD2* R803* mutation. Results We potentially can identify mutations in 75% (82/109) of the ADPKD families in Taiwan by our high-throughput method. Forty percent of our cohort (34/82) was due to mutation in the PKD2 gene, and half of them (17/34) had *PKD2* R803* mutation. Microsatellite analysis showed only marker D4S1563 differentiated those 17 *PKD2* R803* families into two groups, indicating two different but closely related founders leading to the high prevalence of *PKD2* R803* mutation in Taiwan.

896T

Novel genotype-phenotype correlations in X-linked Alport syndrome: Serum albumin level, age at onset of hematuria and hypertension. *L.I. Shagam[1], A.N. Gerasimov[2], T.A. Kuznetsova[1], M.E. Aksenova[1], V.S. Sukhorukov[1,2], T.V. Akopova[1], V.V. Dlin[1].* 1) Research Institute of Pediatrics, Pirogov Russian Medical University, Moscow, Russian Federation; 2) Sechenov Moscow State Medical University, Moscow, Russian Federation.

Alport syndrome (AS) is a hereditary nephritis with an estimated population frequency of 1:5000 – 1:53000 that varies from region to region. It is characterized by hematuria, proteinuria, progressive loss of kidney function and usually accompanied by hearing loss and/or ocular abnormalities. X-linked AS caused by *COL4A5* gene defects comprises ~85% of all Alport syndrome cases. Males with the X-linked AS harboring missense mutations are known to demonstrate milder phenotype than those with frameshift, splice site, nonsense mutations or deletions covering at least 1 exon including later age at end stage renal failure, hearing loss and eye lesions' manifestation. However, little is known about genotype correlations with other phenotypic characteristics such as hematuria and hypertension manifestation age, biochemical characteristics of blood and urine. We have performed a longitudinal clinical and comprehensive genetic research on patients with AS. A gene panel covering coding exons and adjacent 5bp intronic regions of *COL4A3*, *COL4A4* and *COL4A5* genes has been used with Ion Torrent sequencer for confirmation of the diagnosis. A sample of 43 patients from Russia aged 1 to 17 years old (average 11) at last examination, 26 males and 17 females with X-linked AS and *COL4A5* pathogenic or likely pathogenic variants has been subjected to study of genotype-phenotype correlations. Here we show that patients with missense mutations resulting in glycine substitutions in the *COL4A5* collagenous domain (n=26) are characterized by later age at onset of hematuria and hypertension (p<0,05 according to Mann-Whitney U test) than those with frameshift, splice site, start codon, nonsense mutations or deletions covering at least 1 exon (n=16). The abovementioned associations have been shown for males and for both sexes taken together, whereas they do not reach statistical significance threshold in the female sample. Moreover, among patients of both sexes harboring the glycine substitutions (n=21) we demonstrate a negative correlation of the mutation position with age-adjusted serum albumin level (i.e., C-terminal mutations tend to lead to lower albumin level than N-terminal) with Pearson correlation coefficient of -0,46 and Spearman correlation coefficient of -0,45 (p<0,05). The results herein reported expand the knowledge on phenotypic characteristics of patients with X-linked Alport syndrome caused by different mutation types.

897F

Exome sequencing reveals novel candidate genes and potential oligogenic inheritance in patients with hypergonadotropic hypogonadism. *A. Jolly[1], Y. Bayram[1], S. Turan[2], Z. Aycan[3], T. Tos[4], B. Hacihamdioglu[5], Z.C. Akdemir[1], S. Bas[2], Z. Atay[2], T. Guran[2], S. Abali[2], J.J. White[1], G. Yesil[6], E. Karaca[1], S.N. Jhangiani[7], D.M. Muzny[7], A. Bereket[2], R.A. Gibbs[7], J.E. Posey[1], J.R. Lupski[1,7,8,9].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatric Endocrinology and Diabetes, Mamara University Hospital, Istanbul, Turkey; 3) Department of Pediatric Endocrinology, Sami Ulus Children's Hospital, Ankara, Turkey; 4) Department of Medical Genetics, Sami Ulus Children's Hospital, Ankara, Turkey; 5) Department of Pediatric Endocrinology, Suleymaniye Training and Research Hospital, Istanbul, Turkey; 6) Department of Medical Genetics, Bezmialem University, Istanbul, Turkey; 7) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 8) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 9) Texas Children's Hospital, Houston, TX.

Hypergonadotropic hypogonadism (HH) is characterized by hypogonadism due to an impaired response of the gonads to the gonadotropins (FSH and LH) and in turn a lack of sex steroid production and elevated gonadotropin levels. HH can be caused by environmental factors such as acquired damage to the ovaries, and congenital disorders that affect normal ovarian development and function, as well as syndromic and non-syndromic single gene disorders. However, in most cases of gonadal dysfunction the molecular etiology and molecular diagnosis remain an enigma. To identify novel molecular etiologies in HH we applied whole exome sequencing (WES) to 32 affected female individuals from 29 unrelated families including 21 with reported parental consanguinity. WES revealed likely pathogenic variants in known HH-associated genes in 7/29 (24%) including *AR*, *NOBOX*, *MCM8*, *PADI6*, *PSMC3IP*, and *TG*. In six families, we found likely pathogenic variants in candidate disease genes including *SOHLH1*, *C3*, *IGSF10, MRPSS22,* and *MAST4*. GeneMatcher facilitated the identification of 1 family of 3 affected individuals with homozygous variation in *MRPSS22* and 1 family of 4 affected individuals with homozygous variation in *IGSF10*. Interestingly, in three families we found evidence for a potential mutational burden or oligogenic inheritance model. In one patient with HH and obesity we identified homozygous variants in two known genes associated with hypogonadism (*CHD7* and *MCM9*) and another homozygous variant in *PRKD1* that was reported as a candidate gene for obesity. In a second subject, we identified homozygous variation in *GALT*, associated with Galactosemia, a known cause of HH. In the same subject, we identified a homozygous variant in *DNAH6* that had been previously reported in a patient with ovarian failure. Variation at both *ITGA11* and *TMPRSS9* were found in a third family with potential multilocus inheritance. In this study, we identified both potential novel genes associated with HH, as well as variants in known HH genes in a cohort of 32 affected female individuals. A genome-wide approach enabled both identification of seven potential novel disease genes as well as an oligogenic mutational burden model in three families. We suggest that increased application of WES and WGS in clinical genomics of genetically heterogeneous phenotypes such as HH will provide further molecular etiological insights including the potential contributions of multilocus variation.

898W

Enzyme replacement therapy during dialysis in a patient with Fabry disease in a community hospital in New York. *E. Astiazaran-Symonds, S. Lee, P.D. Lopez, M. Maik, M. Banikazemi.* Internal Medicine, Metropolitan Hospital Center, New York, NY.

A 31 year old Mexican male was admitted to the hospital due to acute presentation of end stage renal disease. He reported experiencing acroparesthesia since early childhood. He also complains of anhidrosis, poor heat tolerance and fatigue. Enzyme and molecular testing for Fabry disease were performed. Patient was found to have low α-galactosidase A activity. DNA sequencing confirmed a *GLA* nonsense mutation (c.748C>T). Patient was placed on hemodialysis three times per week, as well as enzyme replacement therapy (ERT) with r-hαGalA every other week. It has been reported that administration of r-hαGalA during hemodialysis is not associated with a reduced efficacy of r-hαGalA therapy or loss of enzyme into the dialysate. Therefore, we have arranged that his ERT be continued during hemodialysis. Since the start of ERT, patient reports decreased acroparesthesia, better performance at work and increased heat and exercise tolerance. Fabry disease is an X-linked lysosomal storage disorder, caused by reduced activity of the lysosomal enzyme α-galactosidase A. Globotriaosylceramide (GL3) accumulates, leading to renal, cardiac, and cerebrovascular manifestations, and risk of early death from end-stage renal disease, coronary artery disease or stroke. Characteristic features of FD begin early in life and include acroparesthesia, angiokeratomas, anhidrosis, corneal opacities, gastrointestinal symptoms, tinnitus and hearing loss. Specific treatment for Fabry is available by means of ERT with recombinant agalsidase alpha and beta. ERT slows disease progression by reducing levels of GL3. This in turn protects organs from irreversible damage and improves patient's symptoms and quality of life. To the best of our knowledge, this is the first case of a patient with Fabry receiving ERT during dialysis in the United States. This arrangement has increased patient's satisfaction and quality of life.

899T

***CFTR* gene mutations in the São Miguel island (Azores, Portugal): 20 years follow-up study.** *L. Mota-Vieira[1,2], P. Gaspar-da-Silva[3], P. Pacheco[4], C. Silva[4], C.C. Branco[1,2], A. Carreiro[5], M.D. Amaral[2,6], J. Gonçalves[3], J. Rosa[3].* 1) Molecular Genetics and Pathology Unit, Hospital of Divino Espirito of Santo Ponta Delgada, Azores Islands, Portugal; 2) BioISI - Biosystems & Integrative Sciences Institute, Lisboa, Portugal; 3) Pediatrics Dept, Hospital of Divino Espirito of Santo Ponta Delgada, Azores Islands, Portugal; 4) Human Genetics Dept, INSA, Lisboa, Portugal; 5) Pneumology Dept, Hospital of Divino Espirito of Santo Ponta Delgada, Azores Islands, Portugal; 6) Faculty of Sciences, University of Lisbon, Portugal.

The advent of new drugs targeting specific *CFTR* gene mutations is highlighting the importance to characterize, clinical and molecularly, each patient with cystic fibrosis and their populations (CF, MIM#219700; *CFTR*, MIM*602421). Here, we performed a cross-sectional study on CF patients admitted to the Hospital of Divino Espirito Santo of Ponta Delgada, over the last 20 years (1997–2016). This hospital is the only one in the Azorean island of São Miguel (137,699 inhabitants; 55.9% of the total Azores population; Portugal Census, 2011). The diagnosis of the CF was based on clinical symptoms and elevated sweat chloride concentrations (≥60 mmol/L). Two disease-causing *CFTR* mutations were identified in each patient, using a two-step mutation screening that covers the most frequent (98%) mutations found in Portuguese CF population. This strategy allowed us to confirm the disease in 14 CF patients, being 7 male and 7 female, 11 unrelated and 3 related. There was 1 dead patient (18 years-old), and 13 patients are alive with a mean age of 14.8 years in 2016 (range 6.4-22.2). Clinically, all patients have, or had for the deceased patient, exocrine pancreatic insufficiency (fecal elastase-1 activity: <200 µg/g) and respiratory infections. We identified 4 different *CFTR* mutations in 28 CF chromosomes, namely (NM_000492.3): c.1521_1523delCTT (p.Phe508del) [allele frequency: 85.7%], c.-12_10del23 [7.1%], c.11C>A (p.Ser4Ter) [3.6%], and c.3299A>C (p.Gln1100Pro) [3.6%]. In genotypic terms, only 1 patient (a compound-heterozygous for p.Gln1100Pro and p.Ser4Ter) does not carry the p.Phe508del. In other words, the p.Phe508del, which is the most common among Caucasian population, was found more frequently in homozygosity (78.6%: 11 patients of which 3 were born from consanguineous parents); only 2 patients were compound-heterozygous with the c.-12_10del23. Indeed, until now, the c.-12_10del23 is considered a private mutation, since it was reported in few CF patients, but it could be found in patients of Azorean origin living abroad. In terms of population genetics, we calculated the prevalence of the p.Phe508del in 469 individuals from São Miguel Island. We found 7 heterozygous individuals for this mutation, corresponding to an allele frequency of 0.7% and predicting 1 in 17.956 individuals could be homozygous and 1 in 68 should be heterozygous. Now, we are studying the specific genotype-phenotype correlation in three non-classical CF patients carrying other *CFTR* mutations.

900F

TRPV4 alternative splicing transcripts in metatropic dysplasia. *S.M. Kirwin[1], KM. Robbins[1], MB. Bober[2], VL. Funanage[1].* 1) Molecular Diagnostics Laboratory, Nemours/duPont Hospital for Children, Wilmington, DE; 2) Orthopedics Department, Nemours/duPont Hospital for Children Wilmington, DE.

The transient receptor potential vanilloid 4 (TRPV4) cation channel mediates calcium influx in tissues in response to chemical, physical, and hormonal stimuli, and is a member of the Transient Receptor Potential (TRP) super family of cation channels. Deleterious mutations cause a group of skeletal dysplasias including Metatropic dysplasia (MD). Although phenotypes may differ widely, they all share defects in bone ossification. Metatropic dysplasia is a rare skeletal dysplasia (OMIM 156530) characterized by a narrow chest, reduced spinal integrity with platyspondyly and progressive kyphoscoliosis, shortened long bones with metaphyseal widening and often a delay of endochondral ossification. Pathologic *TRPV4* variants are also associated with several neurological disorders, including hereditary motor and sensory neuropathy type IIC ; and scapuloperoneal spinal muscular atrophy. *TRPV4* mutations associated with disease are located throughout the gene, resulting in a gain of function in this heterogeneous group of disorders ranging from mild to lethal. *TRPV4* mRNA is reported with variable expressivity in a wide range of tissues and cell types: cardiac, immune system, CNS, respiratory system, skin, and musculoskeletal system. There are six reported isoforms: full length (TRPV-A), isoform 2 (TRPV-B, Δ7), isoform 3 (partial exon16 deletion), isoform 4 (TRPV-C, Δ5), isoform 5 (TRPV-D, Δ2), isoform 6 (TRPV-E, Δ5Δ7). This study sought to determine which alternative splice variants exist in samples collected from skin fibroblasts, bone or chondrocyte cells, in patients with metatropic dysplasia. Furthermore, we sought to characterize why the same mutation (p.P799L) in patients may present with widely differing phenotypes, and determine whether mutations of the *TRPV4* gene affected RNA splicing in the affected tissues. It is unknown whether differential expression of splice variants based on cell type play a role in pathology of the *TRPV4* channelopathies. Preliminary results show the full-length transcript and the minor splice variant isoforms with skipped exons 5 and 7 as the primary transcripts detected in bone and fibroblasts. Of interest, both the wild-type and p.P799L mutation were detected in an RNA delta 7 transcript isolated from cultured bone cells. Thus, the expression of the *TRPV4* mutation in various RNA transcripts may explain the phenotypic variability in this autosomal dominant disease.

901W

Nora's Lesion or something less "bizarre": Case report of family with benign bone tumors and review of the literature. *E. Carter, J. Davis.* Ctr Skeletal Dysplasias, Hosp Special Surgery, New York City, NY.

Siblings with a history of benign bone tumors were referred to our service for genetic evaluation. Family history was unremarkable. The brother, now 5 yrs old, presented at birth with painless masses between the base of his 2nd/3rd toes on his left foot (surgically excised at 3yrs) and along the radial aspect of his right index finger. Nora's lesion, also known as bizarre parosteal osteochondromatous proliferation (BPOP), was suspected given the localization in hands and feet. His 3-yr-old sister subsequently developed a painless lump on the outside of her right knee. She had been diagnosed prenatally with bilateral clubfeet; amniocentesis revealed a normal female karyotype. Radiographically, no lesion was contiguous with cortical and cancellous bone, as seen in osteochondromas, and none appeared to represent an osteochondroma emanating from the physis. Small foci of mineralization were noted in the boy's left foot 1st MTP joint and right hand 2nd PIP and 1st MCP joints. MRI of the right index finger described an osteochondromatous lesion originating from the distal proximal phalanx with no prominent overlying cartilaginous cap, and a mass along the dorsal aspect of the distal metacarpal physis. His sister's X-ray showed an anterolateral soft tissue mass at the proximal fibula with calcifications suggestive of a chondral lesion; there was no clear abnormality of the underlying bone to indicate osteochondroma. Post-operatively the inter-toe tumor was described as being attached in a stalk-like manner to the metaphysis of the 2nd proximal phalanx. It had sparsely cellular cartilaginous nodules, overlying fibrous periosteal tissue, and extensive endochondral ossification. *EXT1* and *EXT2* analysis is pending. Benign bone tumors can be categorized into 8 types. Although incidence varies by tumor type, overall they most commonly arise in people <30 yrs old and can be triggered by the hormones that stimulate normal growth. BPOP lesions are rare; recently an association with t(1:17)(q32;q21) was reported. Osteochondromas are the most common benign bone tumor. ~85% of individuals with hereditary multiple osteochondromas have dominant mutations of the *EXT1* or *EXT2* genes in the growth plate cartilage. Clinical manifestations vary; individuals may be asymptomatic, may experience pain, or restricted motion. Surgery may be indicated with pain and/or functional limitations. Risk of malignant degeneration is reported in 0.5-20% of cases. .

902T

A novel mutation in the C-Terminal Associated Peptide (TCAP) region of Teneurin 3 found to co-segregate in all affecteds in a multi-generation family with developmental dysplasia of the hip. *G.J. Feldman[1], T. Freeman[1], A. Fertala[1], J. Parvizi[1,2].* 1) Division of Orthopaedic Research, Thomas Jefferson University, Philadelphia, PA; 2) Rothman Institute, Philadelphia, PA.

DDH is a debilitating condition characterized by incomplete formation of the acetabulum leading to dislocation of the hip, suboptimal joint function and accelerated wear of the articular cartilage resulting in early onset crippling arthritis of the hip in 20-40 year olds. Undetected DDH in this age group accounts for 40-60% of hip osteoarthritis. Current diagnostic tests in newborns using physical manipulation of the femur or ultrasound either under or over-diagnose this condition. DDH is a complex disorder having environmental and genetic causes. Evidence for the genetic cause is well established and includes concordance between identical twins and high rates of transmission among first degree relatives. To better understand the biologic pathways involved in acetabular development, DNA from severely affected individuals in a 4 generation family that showed inter-generational transmission of the disorder was isolated and whole exome sequenced. Shared variants were filtered using various parameters including biological impact and relevance to chondrogenesis and osteogenesis. A novel A to C transversion at position 183721398 on human chromosome 4 was found to co-segregate with the affected phenotype in this family. This mutation encodes a glutamine to proline change at position 2665 in the Teneurin 3 (Tenm3) gene and was judged damaging by 4 prediction programs and by computer modeling of the scFv structure. Teneurin 3 is located on human chromosome 4q35, a region previously strongly linked to DDH in an unrelated 4 generation family. This transmembrane protein is thought to function in signal transduction,and is most strongly expressed in the perichondrium and proliferating chondrocytes in the femoral cartilage of 1 day old mice. This mutation resides near the amino terminal, cleavable end of TCAP domain of Teneurin 3. Within the TCAP domain is a motif bearing strong homology to calcitonin- a hormone strongly associated with endochondral bone formation in early life. This variant and others that remain to be found in other families and sporadic individuals may begin to elucidate the pathways responsible for normal hip development. .

903F

Dual genetic diagnoses identified in a large family with brachydactyly type A1 and insulin resistance using whole-exome sequencing. *R. Ho[1,2], A.D. McIntyre[1], B.A. Kennedy[1], R.A. Hegele[1,2,3].* 1) Robarts Research Institute, Western University, London, Ontario, Canada; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 3) Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada.

Brachydactyly is an umbrella term describing shortened fingers and toes, often part of an autosomal dominant complex malformation syndrome. Although various forms of brachydactyly have been characterized and several causative genes have been found, many forms remain genetically undefined. Here, we describe a 15–year–old unsolved medical case of a Canadian family with a history of brachydactyly and insulin resistance that remained genetically undiagnosed. To characterize the genetic basis of the brachydactyly and insulin resistance in this family, DNA was extracted from 18 family members (14 definitely affected) who were previously Sanger sequenced but not found to have mutations in known genes underlying the disorder. One patient was recently whole-exome sequenced at the London Regional Genomics Centre, and a rare variant prioritization approach was used to identify variants of interest. In this patient, a novel candidate variant for brachydactyly (exon 1, c.285_287dupGAA, p.Glu95_Asn96insLys) was identified in the Indian hedgehog (*IHH*) gene, and confirmed through Sanger sequencing to co-segregate with affected status in the pedigree. Mutations in this gene are known to cause autosomal dominant brachydactyly type A1, which correlates with the clinical phenotype of the family. Using statistical analyses, we found that in comparison to an unaffected control population, the *IHH* variant is associated with shortened middle phalange length by 21.1% ($P<0.001$), palm length by 13.8% ($P<0.01$), digit-palm ratio by 6.8% ($P<0.03$) and stature by 9.5% ($P<0.001$), strengthening the association. Current studies are focusing on *ex vivo* and *in vitro* assessment of the *IHH* variant, and identifying the genetic basis of the insulin resistance segregating in this family. These data highlight the power of next-generation sequencing to solve or clarify the genetic basis of "cold cases". They also indicate the need to consider independently segregating etiologies for dual diagnoses when studying complex disorders. Determining the molecular basis of these disorders will allow for further understanding of disease mechanisms, and will help other health care providers caring for families with the same condition caused by the same genes.

904W

Mutated *DMRT2* causes a distinct type of spondylocostal dysostosis (SCD). *Q. Waisfisz[1], A. Bouman[2], J. Admiraal[3], M. van de Loo[3], R.R. van Rijn[4], R.J. Oostra[5], I.B. Mathijssen[2].* 1) Department of Clinical Genetics, VU University Medical Center, Amsterdam, Netherlands; 2) Department of Clinical Genetics, Academic Medical Center, Amsterdam, Netherlands; 3) Department of Neonatology, Emma Children's Hospital, Amsterdam, The Netherlands; 4) Department of Radiology, Academic Medical Center, Amsterdam, The Netherlands; 5) Department of Anatomy, Embryology & Physiology, Academic Medical Center, Amsterdam, The Netherlands.

Spondylocostal dysostosis (SCD) is a rare disorder characterized by vertebral segmentation defects and malformations of the ribs. In addition, SCD-patients have some degree of (kypho)scoliosis, short stature and suffer from respiratory impairment due to the reduced size of their thoracic cage. Mutations in *DLL3, MESP2, LFNG, HES7, TBX6* and *RIPPLY2* are known to cause different subtypes of SCD. Here, we present a male neonate with an apparent severe type of SCD only partly overlapping the previously described SCD-subtypes. Diagnostic radiological imaging demonstrated multiple costovertebral abnormalities. The proband presented with severe costal malformations (missing, fused, bifid, and hypoplastic ribs), vertebral malformations (including intervertebral fusions of the laminae and irregular ossification of the vertebral bodies) and a mild scoliosis. Other dysmorphic features included: cleft palate (soft palate), posteriorly rotated ears, short neck, asymmetry of the thorax, extra nipple, transverse palmar creases, sandal gap (bilateral), and sacrum acutum. Severe respiratory insufficiency was present from birth. Whole exome sequencing (WES) identified a homozygous start-loss variant in *DMRT2* (NM_006557.6: c.1A>T p.(Met1?)). Mutations in *DMRT2* (OMIM#604935) have not been described in relation to SCD before. It has been demonstrated that *DMRT2* knock-out mice exhibit severe rib and vertebral defects overlapping with the radiological phenotype of our patient. Therefore, it seems plausible that mutations in *DMRT2* can cause a severe type of SCD characterized by a striking skeletal phenotype in combination with extra-skeletal features. This is the first report of a human proband with a severe SCD-like phenotype most probably caused by a homozygous mutation in *DMRT2*.

905T

Potential pathogenic variants identified in a Turkish tooth agenesis cohort via whole exome sequencing. *R. Du[1], N. Dinckan[2,10], L.E. Petty[3], Z. Coban-Akdemir[1], S.N. Jhangiani[4], I. Paine[1], E.H. Baugh[5], A.P. Erdem[6], H. Kayserili[7], H. Doddapaneni[4], J. Hu[4], D.M. Muzny[4], E. Boerwinkle[3,4], R.A. Gibbs[1,4], J.E. Below[3], J.R. Lupski[1,4,8,9], Z.O. Uyguner[2], A. Letra[10,11].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Department of Medical Genetics, Istanbul Medical Faculty, Istanbul University, Istanbul, 34093, Turkey; 3) Human Genetics Center, University of Texas Health Science Center at Houston School of Public Health, Houston, TX 77030, USA; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 5) Department of Biology, New York University, New York, NY 10012, USA; 6) Department of Pedodontics, Faculty of Dentistry, Istanbul University, Capa, Istanbul, 34390, Turkey; 7) Department of Medical Genetics, Koc University, School of Medicine (KUSOM), Istanbul, 34010, Turkey; 8) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 9) Texas Children's Hospital, Houston, TX 77030, USA; 10) Department of Diagnostic and Biomedical Sciences, University of Texas Health Science Center at Houston School of Dentistry, Houston, TX 77054, USA; 11) Center for Craniofacial Research, University of Texas Health Science Center at Houston School of Dentistry, Houston, TX 77054, USA.

Tooth agenesis (TA) is a common craniofacial abnormality in humans and represents the failure to develop one or more permanent teeth. Isolated TA has complex underlying genetic etiology and variations in at least ten genes have been reported as contributing to TA. Although comprehensive efforts were conducted to elucidate genetic susceptibility, novel causative variants remain to be identified. We recruited a Turkish TA cohort consisting of 11 multiplex families affected with TA, and applied whole exome sequencing (WES) using HGSC VCRome capture reagent to identify candidate causative variants for TA. Array-based genotyping and linkage analysis were also employed to facilitate the identification of candidate variants. Sanger sequencing confirmation and segregation analysis were also conducted for all variants of interest. Homozygous and heterozygous rare variants in *WNT10A, LRP6, LAMA3, KREMEN1, DKK1, COL17A1,* and *ANTXR1* were identified as known or potentially pathogenic in the studied families. Notably, *WNT10A* variants (c.697G>T, c.682T>A, and c.433G>A) were found in six families with isolated TA, one of which had a dual molecular diagnosis and presented also with a variant in *LAMA3* (c.1097G>A), segregating with tooth agenesis. Different variants in *KREMEN1* (c.146C>G, and c.773_778del) were found in two families with additional ectodermal dysplasia findings. An indel variant in *LRP6* (c.3607+3_6delAAGT) was found to segregate in one family, and functional assay suggested that the variant mRNA was subject to nonsense-mediated decay as predicted. A homozygous rare variant was found in *ANTXR1* (c.1312C>T) in one family. *ANTXR1* has been reported to be associated with GAPO syndrome. However, in our case, the patient was not affected with GAPO syndrome but oligodontia. To the best of our knowledge, this is the first report of association between homozygous *ANTXR1* gene mutation and TA. We also identified one family in which novel heterozygous variants in *DKK1*(c.548-4G>T), *COL17A1*(c.3277+3G>C), and *LAMA3*(c.2798G>T) were segregating with TA, suggesting potential oligogenic inheritance. In conclusion, our results support a major role for genes (*WNT10A, LRP6, KREMEN1*) encoding proteins that function in the WNT signaling pathway in the etiology of TA while revealing new candidate genes. In addition, mutational burden and oligogenic co-segregation was suggestive for complex inheritance contributing to improved understanding of the genetic etiology of familial TA.

906F

Clinicaland molecular heterogeneity in VCP autosomal dominant inclusion body myopathy. *S. Al-Tahan[1], E. Al-Obeidi[1], M. Omizo[2], J. Palmio[3], M. Grafe[4], Y. Harati[5], B. Udd[6,7].* 1) Division of Genetics and Genomic Medicine, Depart, UC Irvine, Irvine, CA; 2) Oregan Osteoporosis Center, Bend, OR; 3) Neuromuscular Research Center, Tampere University and University Hospital, Neurology, Tampere, Finland; 4) Department of Pathology, Oregon Health and Science University, Portland; 5) Department of Neurology, Baylor College of Medicine, Houston, TX, USA; 6) Folkhälsan Institute of Genetics and the Department of Medical Genetics, Haartman Institute, University of Helsinki, Helsinki, Finland; 7) Vaasa Central Hospital, Department of Neurology, Vaasa, Finland.

Introduction: Our lab has focused on Autosomal Dominant Inclusion Body Myopathy (IBM) for almost two decades. Over forty missense mutations in the gene coding for valosin-containing protein (VCP) are associated with a unique autosomal dominant adult onset progressive disease associated with combinations of proximo-distal inclusion-body myopathy (IBM), Paget's disease of bone (PDB), frontotemporal dementia (FTD), and amyotrophic lateral sclerosis (ALS). We report the clinical, and molecular findings in four new patients/ families carrying novel VCP mutations. Interestingly in two families, individuals also had Parkinson's disease (PD). We have recently identified **Results:** The novel VCP mutations included: c.473 G>A, p.M158I associated with IBM and PDB in addition to ALS; c.478 G>C, p.A160Passociated with IBM, ALS and PD; c.383G>C, p.G128A IBM and PD. Interestingly patient with.c.382G>T, p.G128C only had had non familial IBM. There was wide inter and intra-familial variation making genotype-phenotype correlations difficult in these families. **Conclusions:** We report four novel mutations to the >50 previously reported in VCP disease associated with inclusion body myopathy, PDB, FTD and ALS. Furthermore with the addition of these families we have further established VCP in the etiology of PD. Previous cases associated with PD included p.R159C, p.R191Q, p.T262A, and p.P137L VCP mutations. Because of wide spectrum of mutations and heterogeneity establishing genotype-phenotype correlations is challenging. We however recommend routine testing of VCP in familial PD. By understanding the variability of clinical presentations clinicians will be better able to identify and diagnose VCP-related diseases and thus proactively manage associated treatable features more effectively.

907W

Survey of patients with Ollier disease and Maffucci syndrome over Facebook compared to review of clinical literature. *C. Smith[1], S.M. Robbins[1,2,3], N.L.M. Sobreira[1,3].* 1) Johns Hopkins University School of Medicine, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, Baltimore, MD; 3) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Ollier disease (OD) and Maffucci syndrome (MS) are rare disorders characterized by multiple enchondromas. OD is characterized by multiple enchondromas, with onset in early childhood, that are typically asymmetrical in distribution with a predilection for the appendicular skeleton. MS is characterized by multiple enchondromas distributed symmetrically in combination with vascular anomalies; about 25% of cases are noticeable at birth. Both disorders can cause multiple swellings on the extremities, deformity around the joints with limitations in joint mobility, scoliosis, bone shortening, leg-length discrepancy, gait disturbances, pain, and loss of function. The risk of developing a chondrosarcoma in OD and MS is up to 57%, and other malignancies have also been associated with these disorders. This clinical information about the natural history of OD and MS is based on a few cases described in the literature and most likely represent the extreme presentation of these disorders. The prevalence of other malignancies outside of chondrosarcomas is unknown. To address this gap in the natural history, we created a survey that was recently distributed directly to patients through Facebook. The survey has 84 questions and collects self-reported clinical and family history, molecular test results, and demographic information. Here we discuss the results of the survey's responses and compare to a review of the cases described in the literature. To date, 54 patients have answered the survey. Forty-eight patients have OD, and 2 have MS. Twenty-four have more than 10 enchondromas. Twelve have enchondromas affecting their ribs, 1 has the spine affected, 1 has the skull affected, and 2 have their scapula affected. Three patients also have exostoses. Nine had chondrosarcomas. Ten have hemangiomas, and 1 had a spindle cell hemangioma. Two had other kind of malignancies: a brain tumor not specified and a grade II astrocytoma. Seven had developmental delays, and 2 have intellectual disabilities. Two have a family history of enchondromatoses. This information will be compared to the data collected in a literature review of 197 publications found in Pubmed. We expect that the analysis of the survey will give us a better understanding of the natural history, severity and prognosis of these diseases, as well as the prevalence of chondrosarcomas and other malignancies in these patients. This survey may also uncover unexpected diagnoses in these patients.

908T

Dyggve-Melchior-Clausen syndrome, a case report with typical family tree. *L. Mora, F. Suarez, A. Paredes.* Medical Genetics, Pontificia Universidad Javeriana, Bogota, Colombia.

Introduction: Dyggve-Melchior-Clausen (DMC) syndrome belongs to the group of spondyloepimetaphyseal dysplasias. Is caused by mutation of the DYM gene, encodes for Dymeclin, a protein that is involved in skeletal development and brain function, the inheritance pattern could be autosomal recessive or X-linked recessive. Patients presents short trunk, microcephaly, pectus carinatum and intellectual disability, characteristically known as a progressive dwarfism, very similar to the clinical presentation of mucopolysaccharidosis (MPS), therefore genetic testing is necessary to confirm a diagnosis. **Case presentation:** We present a Colombian family of consanguineous parents and five children, four of them referred for study of disproportionate short stature, intellectual disability and facial dysmorphism (coarse facies), with the suspicion of MPS were requested, extension studies: echocardiogram was normal, dorso-lumbar vertebrae radiography which evidences kyphosis, cuneiform vertebral bodies some members with platyspondyly, dysplasic iliac crests, long-bone radiographic with metaphyseal abnormalities and enzymatic tests without alterations. Thinking in differential diagnosis of MPS molecular study was performed confirming homozygous mutation on DYM gene, associated with Dyggve-Melchior-Clausen disease. **Discussion:** Dyggve-Melchior-Clausen syndrome is a progressive spondyloepimetaphyseal dysplasia, result of mutations in DYM gene, located in 18q12-12.1 chromosomal region, codify for dymeclin protein that participated in bone development, necessary for correct organization of Golgi apparatus, perhaps its role within the cell is unknown. Lack of this protein is associated with two pathologies Smith-Mccort Dysplasia and Dyggve Melchior Claussen. The first one differs from our exposed case because it does not have intellectual disability. Other differential diagnosis is morquio disease, but has specific radiological and enzymatic findings, like in this family. Due to the clinical and the low prevalence of DMC syndrome it could be confused with MPS and subdiagnosticated, been necessary the molecular study. **Conclusion:** Over time the correlation phenotype-genotype of patients with DMC syndrome will allow functional studies to understand differences between this disorders or a possible spectrum disease and future treatments.

909F

Case series of individuals with novel syndromic phenotypes characterized by enchondromas and/or exostoses with or without vascular anomalies. *S.M. Robbins[1,2,3], C. Smith[1], B. Cohen[1], E. Lie[1], C. Haldeman-Englert[4], E. Boyd[4], J. Hoover-Fong[1], M. Cernach[5], N.L.M. Sobreira[1,3].* 1) Johns Hopkins University School of Medicine, Baltimore, MD, USA; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins University, Baltimore, MD, USA; 3) Institute of Genetic Medicine, Johns Hopkins Hospital, Baltimore, MD, USA; 4) Fullerton Genetics Center Mission Health, Asheville, NC, USA; 5) Escola Paulista de Medicina, Sao Paulo, Brazil.

Enchondromatoses are a group of 9 rare diseases characterized by multiple enchondromas. Here we describe 4 individuals with novel syndromic phenotypes characterized by enchondromas and/or exostoses with or without vascular anomalies. Case 1 is a 23 yo male with a history of a region of epidermal and dermal atrophy on the left chest, lymphoepithelioma of the nasopharynx with parapharyngeal involvement, multiple enchondromas, an exostosis, and a vascular anomaly affecting his chest, back and left arm. He also had developmental delay and has intellectual disability. Case 2 is a 14 yo with history of developmental delay, hypotonia, swallowing difficulties, vocal cord dysfunction, mild cerebral palsy, immunodeficiency, short stature, low weight, a vascular anomaly in the right arm noticed at 1 year of age, and enchondromas in his hands and feet first diagnosed at 3 years of age. DNA extracted from one of his enchondromas was negative for variants in *IDH1* and *IDH2*. Case 3 is a family where the proband (8 yo), his mother, maternal uncle, and maternal grandfather are affected with short stature, acromesomelia, and rhizomelia. The proband also has attention deficit, epicanthal folds, a broad nasal bridge, pectus excavatum, asymmetric sternum, kyphosis, brachydactyly, increased upper to lower segment, decreased arm span to height ratio, and exostoses. Case 4 is a 4 yo female with disproportionate short stature (short trunk), frontal bossing, strabismus, conical teeth, microdontia, tooth agenesis, short neck, asymmetric chest, fused ribs, pectus carinatum, scoliosis, platyspondyly, irregular ilium, clinodactyly of 5th fingers, dysplastic nails, enchondromas and asymmetric legs size. For each of these cases, we performed whole exome sequencing in the affected proband and family members or in the proband's germline and tumor DNA. Analysis in these cases did not reveal pathogenic variant in a known disease gene associated with similar or overlapping phenotypes, but candidates have been submitted to GeneMatcher, and further functional studies are being performed to investigate their pathogenicity. In summary, here we expand the clinical variability of the phenotypes associated with enchondromatoses. We anticipate that the identification of the molecular basis of these cases will uncover pathways involved in the mechanism by which enchondromas, exostoses and vascular anomalies arise in patients with multiple enchondromas with unknown molecular basis.

910W

Biallelic mutations in FLNB cause a skeletal dysplasia with 46,XY gonadal dysgenesis by increasing β-catenin expression. *K. Upadhyay[1], J. Loke[1], V. Wu[1], B. Taragin[2], H. Ostrer[1].* 1) Albert Einstein College of Medicine, Bronx, NY; 2) Montefiore Medical Center, Bronx, NY.

Filamin B (FLNB) functions as a switch that can affect chrondrocyte development and endochondral bone formation through a series of signaling molecules and transcription factors that also affect Sertoli cell development. Here, we report a subject with a novel skeletal dysplasia and co-existing 46,XY gonadal dysgenesis and biallelic mutations in *FLNB*. The presence of these mutations *FLNB* (NM_001457.3) c.2890T>C p.F964L and c.4730C>T p.A1577) increase binding of FLNB protein to a signal transduction complex that includes the MAP3K1 and RAC1 proteins and activation of β-catenin relative to wildtype. These mutations have different effects on the phosphorylation of intermediates in the MAP kinase pathway and the expression of SOX9; thus, their mechanism of action varies from those reported previously for loss-of-function mutations in *SOX9* and gain-of-function mutations in *MAP3K1*. These findings highlight that subjects with skeletal dysplasias or 46,XY gonadal dysgenesis should be assessed for dual phenotypes. Variants in newly identified candidate chondrogenesis or sex determining genes can be annotated for pathogenic functional effects on the activities of β-catenin, SOX9 and their intermediate regulators.

911T

Multi-center cohorts with animal model and genotype-phenotype analyses: deciphering a new and undefined subtype of congenital scoliosis, TBX6-associated congenital scoliosis (TACS). *N. Wu[1,2,3,4], J. Liu[1,2,3], N. Yang[5], K. Takeda[6,7], S. Liu[1,2,3], W. Li[5], K. Watanabe[7], P. Liu[4], Z. Wu[2,3,8], J. Lupski[4,9], F. Zhang[5], S. Ikegawa[6], G. Qiu[1,2,3].* 1) Department of Orthopedic Surgery, Peking Union Medical College Hospital; 2) Beijing Key Laboratory for Genetic Research of Skeletal Deformity; 3) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences; 4) Department of Molecular and Human Genetics, Baylor College of Medicine; 5) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University; 6) Laboratory for Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences; 7) Japan Early Onset Scoliosis Research Group (JEOSRG); 8) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Science; 9) Human Genome Sequencing Center, Department of Pediatrics and Texas Children's Hospital, Baylor College of Medicine.

Introduction: Congenital scoliosis (CS) is a common type of vertebral malformation. The genetic basis of a significant portion of sporadic CS has been elucidated and explained by a model of compound inheritance. The observation of CS in selected patients with either deletion of the 16p11.2 region or point mutations in *TBX6*, a critical gene mapped to the same chromosomal locus, led to the investigation of the molecular pathogenesis of CS among these individuals (Wu 2015 *NEJM, 372:* 341-350). **Methods**: Although it has been clearly demonstrated that *TBX6*-associated CS (TACS) has unique molecular genetic underpinnings, the clinical characterization remains less well established. We enrolled two cohorts (PUMCH and JSCRG) of CS patients. All patients were divided into two groups according to the genetic characteristics of variant alleles at the *TBX6* locus, and we compared the clinical data based on specific genetic types. Furthermore, the animal model of TACS harboring compound *TBX6* heterozygous mutations was generated by using CRIPSR/Cas9. A systemic detailed phenotypic evaluation of the patients and the mice was also conducted. **Results:** Of the 332 Chinese cases, we identified 34 patients with *TBX6* loss-of-function mutations *in trans* with the risk haplotype at the second allele and classified them as TACS. The remaining patients were classified into the non-TACS (NTACS) group. The patients with TACS were significantly younger at onset (Avg age 7.76±4.70y; p=1x10⁻⁴). For the vertebral defects, failure of formation was more commonly seen in the TACS group. Vertebral anomalies were significantly more extensive in the NTACS group. Remarkably, all 34 TACS patients uniformly exhibited a phenotype with one or more hemivertebrae or butterfly vertebrae. Involvement of the lower part of the spine was more prevalent in TACS. These phenotypic features were confirmed by the replication study in the Japanese cohort of 96 CS cases. Similarly, vertebral hypoplasia was the most common malformation observed in the *Tbx6*^mh/- mice, and most vertebral defects (92.31%) involved the lower part of the spine (T10-S4). **Conclusions:** TACS share the same genetic etiology and have a unique and relatively uniform clinically recognizable presentation. TACS represents a distinct clinical entity, with specific etiologic genetic and molecular underpinnings, that separate it from the other CS for which the mechanisms remain elusive.

912F

Multicentric carpotarsal osteolysis syndrome in mother and daughter misdiagnosed as juvenile rheumatoid arthritis. *K. Chen[1], M.Z. Souza[1], M.F.F. Soares[2], M.I. Melaragno[1], V.A. Meloni[1].* 1) Genetics Division, Department of Morphology and Genetics, Universidade de Sao Paulo, SP, Brazil; 2) Department of Imaging Diagnosis, Universidade de Sao Paulo, SP, Brazil.

We report on two cases of Multicentric Carpotarsal Osteolysis syndrome (MCTO) previously misdiagnosed as Juvenile Rheumatoid Arthritis (JRA). MCTO (MIM #166300) is a rare skeletal disorder that in early childhood mimics JRA, and is characterized by articular pain, progressive osteolysis, predominantly of the carpal and tarsal bones, minor facial anomalies, and oftentimes nephropathy and chronic renal failure, with autosomal dominant inheritance. The disease is caused by heterozygous mutation in the *MAFB* gene (608968), mapped on chromosome 20q12, reported on at least 35 patients. We report on a 13-year-old female patient with skeletal deformities, whose mother presented similar deformities but had been misdiagnosed as JRA. The patient is the first offspring of unrelated parents and has a nine-year-old brother and a seven-year-old maternal half-sister, both without similar symptoms. She presented unremarkable antenatal and postnatal histories, and normal growth and psychomotor development. At the age of 12, she developed pain in the upper limbs and feet, progressive joint restriction, and finger swelling. The first genetic evaluation at the age of 13, she presented mild facial dysmorphic features, flexion contracture of the right elbow, radial deviation of hands, swollen fingers and *pes cavus*. The proband's skeleton X-rays showed osteopenia, carpal and tarsal bones dissolution, proximal and distal metacarpal and metatarsal erosions. The mother's radiological findings indicated progressive bone resorption as the carpal and tarsal bones were completely destroyed. The clinical history and findings and the radiological abnormalities are consistent with MCTO. Similar to other cases of MCTO, this one was misdiagnosed and was followed up as JRA. For a precise assessment, complementary tests for renal function monitoring were requested. In this family, the siblings must be evaluated and monitored since they present 50% risk for the disease. The correct and early diagnosis allows the appropriate follow-up and specific therapeutics for changing the natural history of the disease. Studies indicate that drugs targeting *Mafb* signaling such as anti-RANKL may perform better than bisphosphonates in the treatment of the bone disease of MCTO, thus presenting as an alternative therapeutics. (Financial support of FAPESP).

913W

Recurrence of perinatal lethal osteogenesis imperfecta due to parental mosaicism for a novel dominant mutation in *COL1A1*. *A. Ruiz-Herrera[1,2], M. Abreu-González[3], Y. Mena[3], V. Flores[3].* 1) Medical Genetics, Hospital de Especialidades Pediátrico de León. León, Guanajuato, Mexico; 2) Medical Genetics, Hospital Médica Campestre. León , Guanajuato, Mexico; 3) Genos Médica, Centro Especializado en Genética. Mexico City, Mexico.

Introduction. Perinatal lethal Osteogenesis Imperfecta (PL-OI, MIM #166210) is a skeletal dysplasia that is frequently attributable to heterozygous *de novo* variants in genes encoding type 1 collagens (*COL1A1* and *COL1A2*). However, recurrence can take place in siblings with normal parents. In these cases, the two possibilities are germline mosaicism and autosomal recessive inheritance. **Objective.** We present detailed clinical, radiographic, and molecular findings of a newborn with PL-OI due to parental mosaicism for a novel heterozygous variant in *COL1A1*. **Case Presentation.** The index case is a male newborn patient, born to a healthy young non-consanguineous Mexican couple with medical history of spontaneous abortion and a previous male son with PL-OI. In the proband, PL-OI was suspected prenatally until the third trimester. The patient was born at 35 weeks gestation with dark blue sclerae, absence of cranial ossification, multiple fractures, and bilateral clubfeet; and died shortly after birth. **Materials and Methods.** DNA was isolated from a buccal swab of the proband, and a next-generation sequencing panel for the 18 genes involved in Osteogenesis Imperfecta was performed. Both parents were screened for the *COL1A1* variant in DNA extracted from blood and oral samples in order to detect a possible mosaicism. **Results.** A novel heterozygous variant (NM_000088.3:c.59C>A, NP_000079.2:p.Thr20Lys) in *COL1A1* was detected in the proband and was confirmed by Sanger sequencing; however it was not found in none of the samples from the parents. **Conclusions.** We report a novel variant in PL-OI, with recurrence in a family due to suspected germline mosaicism in one of the parents.

914T

Case report of a mild skeletal phenotype secondary to mutations in *LBR* gene. *M.D.F. Carvalho[1,2,3,4], K.M Carvalho[1,4], I.F Carvalho[2,4], D.P. Cavalcanti[5,6], E.D.F Carvalho[1,2,3,4].* 1) Ceara State University, Fortaleza, CE 04013-000, Brazil;; 2) Christus University Center, Fortaleza, CE, 04013-000, Fortaleza, Ceara, Brazil; 3) APAE, Fortaleza, Ceara, Brazil; 4) Genpharma Consultoria Farmaceutica e Genetica, Ltda, Fortaleza, CE 60160-230, Brazil; 5) Skeletal Dysplasia Group, Department of Medical Genetics, Faculty of Medical Sciences, State University of Campinas, Campinas, SP, Brazil; 6) Perinatal Genetics Program, Department of Medical Genetics, Faculty of Medical Sciences, State University of Campinas, Campinas, SP, Brazil.

Mutations in the The Lamin B receptor (*LBR*) gene were described capables to affect neutrophil and sterol reductase activity. Compound heterozygosity or homozygosity mutations in *LBR* gene can result in different clinical phenotypes: the perinatal lethal autosomal recessive skeletal dysplasia called Greenberg skeletal dysplasia (GSD) and milder skeletal fenotypes whose association with this gene was recently discovered. Heterozygous carriers do not present clinical manifestations but usually have Pelger–Huet anomaly (PHA) that result in a bilobed neutrophil nucleus. We report a child who has mild skeletal dysplasia associated with PH. The proband is an 8-month-old female infant that is the second child of a nonconsanguineous and healthy couple. There was no family history of congenital abnormalities. The prenatal ultrasound exams showed micromelia. The patient was born at term by elective cesarean section, at 36 weeks of gestation, the birth weight was 2880 g, birth height 42 cm and occipitofrontal circumference (OFC) 35,5 cm. Apgar scores were 9 and 9 at 1 and 5 min, respectively. At birth, short limbs were noted and with 1 month and fifteen days old of age the physical examination disclosed: 49 cm of height, normal OFC, eumorphic face and normal hands and feet. At 8-months-old the clinical exam showed discrete decreasing of severity of the rhizomelic shortening of the limbs and a normal neuropsycho-motor development . Radiographs(at 15 days of age) showed: shortening of long bones with possible metaphyseal involvement and mild generalized platy-spondyly. G-banded karyotype and echocardiography were normal. Whole exome sequencing revealed two novel mutations in the *LBR* gene: c.1756C>A (p.Arg586Ser) and c.43C>T (p.Arg15Ter). This and previously reported cases suggest that mutations in LBR gene can result in a spectrum of phenotypes, especially associated with skeletal manifestations.

915F

Loss of inhibition of mTOR signaling in a new form of a metaphyseal chondrodysplasia due to a recessively inherited mutation in salt inducible kinase 3 (*SIK3*). *F. Csukasi[1], I. Duran[1], M. Barad[1], J. Martin[1], D.H. Cohn[1,4,5], D. Krakow[1,2,3,5].* 1) Department of Orthopaedic Surgery, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 3) Department of Obstetrics and Gynecology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 4) Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, Los Angeles, CA; 5) International Skeletal Dysplasia Registry, University of California, Los Angeles, Los Angeles, CA.

Newer genetic technologies have allowed for the identification of causative genes in uncharacterized disorders. Here we describe a recessively inherited unclassified metaphyseal chondrodysplasia characterized by extremely widened/flared metaphysis with irregular ossifications, motheaten long bones, fragmentation of proximal metacarpals, coronal clefts in the cervical spine region, absence of pubic bone ossification and widened spheno-occipital synchondrosis. Both patients (siblings) also manifested immunodeficiency, muscle weakness and developmental delay. Exome analysis identified a homozygous missense mutation in the kinase domain of the gene encoding salt-inducible kinase 3 (*SIK3*). Modeling of the protein showed that the amino acid change (p.Arg129Cys) is likely to compromise the ability of the mutated kinase to phosphorylate its substrate. Patient derived fibroblasts showed diminished amount of SIK3 protein, but not cDNA levels, indicating that the *SIK3* mutation affects the stability of the protein. We found that SIK3 regulates the mechanistic target of rapamycin (mTOR). SIK3 acts as a negative regulator of mTOR complex 1 (mTORC1) and 2 (mTORC2) as evidenced by reduced phosphorylation of their targets S6K1, S6 and AKT when WT SIK3 is overexpressed in chondrocytes. Contrarily, mutated SIK3 is not fully active and cannot negatively regulate. SIK3 is expressed in the cartilage growth plate, particularly in proliferating and hypertrophic chondrocytes where it is co-expressed with mTORC1 and mTORC2 components. Our results reveal a role for SIK3 in producing a previously uncharacterized skeletal disorder and demonstrate that in part, the phenotype results from misregulation of mTOR activity.

916W

Enzyme replacement therapy in perinatal hypophosphatasia: Case report and recommendations for clinical practice. *L. Dupuis[1], G. Costain[1], G. Nimmo[1], A.M. Moore[2], L. Munroe[2], J. Harrington[3], A. Williams[4], R. Zlotnik Shaul[4], C. Rockman-Greenberg[5], M. Offringa[2], P. Kannu[1].* 1) Division of Clinical Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Neonatology, Hospital for Sick Children, Toronto, Ontario, Canada; 3) Division of Endocrinology, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Bioethics, Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Paediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada.

We present the case of a newborn with respiratory distress transferred to the Hospital for Sick Children. She required CPAP in the immediate neonatal period, then was intubated for respiratory failure beginning on day 3. On examination, she was found to have depressible skull bones, large anterior fontanelle, bell-shaped chest, and rhizomelia. Based on X-rays findings, undetectable serum ALP activity and markedly elevated urine phosphoethanolamine, a diagnosis was made of perinatal hypophosphatasia (HPP), a rare metabolic bone disorder. This was later confirmed by the identification of known pathogenic variants (compound heterozygote) in the *ALPL* gene.   Enzyme replacement therapy (ERT) is a newly approved and costly disease-modifying treatment for HPP. Given that ERT is an orphan drug and that HPP is a rare metabolic bone disorder, it is important to share information concerning individuals' response to ERT, as any one centre's familiarity with its use will be limited. Two recently completed clinical trials have reported generally positive outcomes in infants and young children with life-threatening HPP. Here, we describe our patient with perinatal HPP in whom treatment with ERT was not successful. Lessons learned from this case are of importance in making decisions for treatment in patients with new diagnoses of perinatal HPP. We propose a basic protocol for initiating and monitoring treatment of neonates with HPP. We also discuss practical and ethical challenges related to treatment of a rare disease with a costly new drug within the context of a publicly-funded healthcare system.

917T

Spinal manifestations of patients with musculocontractural Ehlers-Danlos syndrome caused by CHST14/D4ST1 deficiency (mcEDS-*CHST14*). *T. Kosho[1,2], M. Uehara[3], H. Kato[3], J. Takahashi[3].* 1) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Orthopedics, Shinshu University School of Medicine, Matsumoto, Japan.

Background: Musculocontractural Ehlers-Danlos syndrome caused by CHST14/D4ST1 deficiency (mcEDS-*CHST14*) is a recently delineated form of EDS and also the first disorder affecting dermatan sulfate (DS) biosynthesis. Clinical characteristics include craniofacial, skeletal, visceral, and ocular malformations; and progressive cutaneous, skeletal, vascular, and viceral fragility-related manifestations. Though spinal deformities are one of the most important complications having a serious impact on the patients' quality of lives and activities of daily living, they have not been systematically investigated to date. **Materials and Methods:** Shinshu University Hospital has the largest cohort of patients with mcEDS-*CHST14*. The cohort includes 10 Japanese patients: five male and five female subjects of an age of 11.8 ± 7.1 (mean ± SD) years at the first visit. We comprehensively reviewed the prevalence of spinal legions, vertebral malformations, changes in severity according to the ages, and atlantoaxial subluxation (AAS), through whole spine radiographs in a standing position and dynamic radiographs focused on the cervical spine. **Results:** Six patients (60%) had a scoliosis with a Cobb angle ≥ 10˚, and one had a severe scoliosis with a Cobb angle ≥ 45˚. Six patients (60%) had a kyphosis at the thoracolumbar junction with a kyphotic angle ≥ 10˚, and three (30%) had a severe thoracolumbar kyphosis with a kyphotic angle ≥ 50˚. All patients with a severe thoracolumbar kyphosis had thoracic lordosis. The transitional level of kyphosis occurred in the thoracolumbar junctional level. Five patients (50%) had a cervical kyphosis, all of whom had more than 20 ˚ kyphosis at the thoracolumbar level. Two patients (20%) had AAS. Eight patients (80%) had vertebral malformations. **Conclusion:** This is the first series describing spinal deformities of patients with mcEDS-*CHST14*, demonstrating scoliosis, kyphosis, cervical kyphosis, and vertebral body malformations as frequent manifestations; also implicating potential importance of dermatan sulfate in the development and maintenance of human vertebral system.

918F

Longitudinal growth curves for OI caused by structural mutations in type I collagen. *J. Marini[1], V. Nakhate[1], C. Abbott[2], L. Barber[1], A. Blissett[1].* 1) Section on Heritable Disorders of Bone and Extracellular Matrix, NICHD/NIH, Bethesda, MD; 2) Office of the Clinical Director, NICHD/NIH, Bethesda, MD.

   Objective: Growth deficiency is a cardinal feature of OI caused by collagen structural mutations, which in practice refers to Types III and IV OI. Growth deficiency is generally marked by the first year of life and progresses thereafter. OI-specific longitudinal growth charts are need to guide caregivers and families. **Methods:** From 100 participants in the NICHD/NIH longitudinal research program on OI, we compiled length, weight and head circumference from ages 2-17. Of the total, 55 are females, 56 have type IV OI and 56 have mutations in *COL1A1*. Data was binned into half year intervals and was analyzed by gender, type and mutated gene using non-linear multilevel modeling (SPSS) with percentile curves constructed using the GAMLSS package from R. **Results:** Sex and OI type had significant effects on stature and weight while the specific collagen gene causing the OI did not. The effect of OI type was greater than that of gender. Interestingly, head circumference was not significantly different by gender, type or mutated gene. OI-specific standard curves for height and weight were generated by gender and type. When OI height curves are superimposed on USA CDC growth curves, types III and IV curves overlap in both girls and boys, with the type III 50th centile curve similar to the type IV 5th centile in both genders. In girls, the type III 95th centile curve exceeded the type IV 50th centile curve, while the reverse occurs in boys. In both genders, the type IV 95th centile curve overlaps the lower general population curves. A pubertal growth spurt is generally not seen in type III OI, and is blunted in type IV. **Conclusions:** OI type is a stronger discriminating factor than gender for OI growth, while curves do not differ for *COL1A1* vs *COL1A2* mutations. OI-specific growth curves will facilitate care.

919W

Gain of function germline mutations in *ABL1* are associated with congenital heart defects, skeletal malformations, and failure to thrive. *Y. Yang[1,2], X. Wang[1,2], C.A. Chen[1], J.A. Rosenfeld[1], A. Al Shamsi[4], L. Al-Gazali[5], M. McGuire[2], N.A. Mew[6], G.L. Arnold[7], C. Qu[2], Y. Ding[3], H. Northrup[10], J.M. Davis[10], D.M. Muzny[3], R.A. Gibbs[1,3], C.M. Eng[1,2], M. Walkiewicz[1,2], F. Xia[1,2], S.E. Plon[1,8,9], J.R. Lupski[1,3,8], C.P. Schaaf[1,9].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Genetics, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Tawam Hospital, Al-Ain, Abu Dhabi, United Arab Emirates; 5) Department of Pediatrics, College of Medicine & Health Sciences, United Arab Emirates University, Al-Ain, Abu Dhabi, United Arab Emirates; 6) Rare Diseases Institute, Children's National Health System, Washington DC; 7) Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA; 8) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 9) Texas Children's Hospital, Houston, TX; 10) Department of Pediatrics, The University of Texas Medical School at Houston, Houston, TX.

   Abelson murine leukemia viral oncogene homolog 1 gene, *ABL1*, is a proto-oncogene which encodes a non-receptor tyrosine kinase. Although the gene is well known as part of the fusion gene BCR-ABL on Philadelphia chromosome in leukemia cancer cells, inherited germline changes in *ABL1* had not been reported in patients until recently. We first reported that *ABL1* germline variants are associated with an AD disorder characterized by congenital heart disease, skeletal abnormalities, and failure to thrive [PMID: 28288113]. Here we provide characterization of eight affected individuals in five unrelated families identified from the clinical exome cohort in our laboratory. The eight individuals shared similar syndromic features including dysmorphic facial features (8/8), skeletal abnormalities (8/8), congenital heart disease (6/7), failure to thrive (6/7), joint laxity or arthritis (5/8), skin anomalies (4/8), GI issues (4/8) and male genital/sexual abnormalities (3/4). The recurrent variant, c.734A>G (p.Tyr245Cys), was seen in seven affected individuals from families 1-4 as either *de novo* or disease co-segregating. A *de novo* c.1066G>A (p.Ala356Thr) variant was identified in the eighth patient (family 5). Of note, the p.Tyr245 residue is one of the two key tyrosine residues required for autophosphorylation-induced activation of ABL1 intrinsic kinase activity, while p.Ala356 plays an important role in myristoylation induced autoinhibition function. We transfected the human embryonic kidney 293T cells with wild-type or mutant constructs and compared the endogenous tyrosine phosphorylation levels of whole cell lysates and of the specific ABL1 substrates. We observed increased phosphorylation suggesting increased ABL1 kinase activity associated with both substitutions. Intriguingly, severe congenital malformations have been reported in fetuses exposed to imatinib, a selective tyrosine kinase inhibitor drug, which inhibits BCR-ABL and is used in the treatment of Philadelphia-chromosome positive CML, raising concern for potential teratogenic effects of the drug. Our findings, together with previously reported teratogenic effects of selective BCR-ABL inhibitors in humans and developmental defects in *Abl1* knock-out mice, suggest *ABL1* plays an important role and needs to be tightly regulated during organismal development. The phenotypic features in our patients suggest *ABL1*-related disorder as a differential diagnosis for patients suspected of connective tissue disorders.

920T

Investigation of the molecular basis of familial and isolated Tarlov cysts.
M. Muriello[1], C. Boehm[1], K. Murphy[2], D. Valle[1], H. Dietz[1], N. Sobreira[1]. 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Radiology, University of Toronto, Toronto, Canada.

Tarlov cysts (TCs) are extrathecal cerebrospinal fluid-filled cavities that develop at the junction of the dorsal ganglion and posterior nerve root between the endoneurium and perineurium. The cause of the formation of TCs is unclear and hypotheses include microcommunication with the subarachnoid space and a ball-valve mechanism allowing CSF influx and restricting efflux, secondary to inflammation or trauma and/or a developmental predisposition. Symptoms are caused by involvement of nerve roots and include pain and neurologic deficits. TC disease is a frequently underdiagnosed condition with a prevalence ranging from 2 to 15%. Cases associating TCs with Marfan and Ehlers-Danlos syndromes have been reported. Familial cases suggest the possibility of autosomal dominant inheritance. Here we report the results of whole exome sequencing (WES) of a cohort of 20 affected individuals with TCs from 14 families. Initial analysis of WES data focused on identifying pathogenic or likely pathogenic variants related to connective tissue disease. Both autosomal dominant and recessive inheritance models were investigated using the PhenoDB Variant Analysis tool to select rare coding or UTR variants with mean allele frequency (MAF) < 1% and < 5% in the 1000 genomes and Exome Variant Server databases. Next, we also selected the genes that were mutated in 3 or more of the 14 probands. Our analysis identified 33 genes, including *PKHD1*, *LAMB3,* and *FAT4*. Pathogenic variants in *PKHD1* are known to cause autosomal recessive (AR) polycystic kidney and liver disease. Pathogenic variants in *LAMB3*, which encodes a widely-expressed laminin protein, cause both AR epidermolysis bullosa and AR amelogenesis imperfecta. *FAT4*, aka protocadherin, is known to play a key role in vertebrate planar cell polarity as well as several other signaling pathways. *FAT4* knockout leads to cystic kidney disease in Drosophila. *FAT4* pathogenic variants cause AR Van Maldergem and Hennekam syndromes. In our cohort, the patients with variants in these genes were heterozygous. However, because of the association of these phenotypes with cyst formation, we prioritized the investigation of the variants in these genes in our patients. It is unknown if the individuals with any of the conditions caused by these genes or their heterozygous relatives have TCs but the investigation of TCs in these patients may lead to the discovery of the molecular basis of the non-syndromic presentations of disease.

921F

***TMEPAI* mutation causes MFS/LDS-like phenotypes in 2 Japanese families.**
H. Morisaki[1,2,3], I. Yamanaka[3], T. Ohata[4], T. Kosho[5], K. Wakui[5], Y. Fukushima[5], M. Masuno[6], K. Takahashi[7], T. Watabe[7], T. Morisaki[2,3]. 1) Sakakibara Heart Institute, Tokyo, Japan; 2) Tokyo University of Technology School of Health Sciences, Tokyo, Japan; 3) National Cerebral and Cardiovascular Center, Osaka, Japan; 4) Okinawa Prefectural Hospitals, Okinawa, Japan; 5) Shinshu University Graduate School of Medicine, Nagano, Japan; 6) Kawasaki Medical School, Okayama, Japan; 7) Tokyo Medical and Dental University, Tokyo, Japan.

To date, several genes have been identified to be responsible for hereditary aortopathies including systemic connective tissue disorders such as Marfan syndrome (MFS) and Loeys-Dietz syndrome (LDS). Some of these genetic mutations were found to cause high transforming growth factor (TGF)-β signaling. To identify new genes responsible for these diseases, 304 patients with MFS/LDS-like features and young aortopathy negative in *FBN1*, *TGFBR1*, *TGFBR2*, *TGFB2*, *TGFB3* or *SMAD3* mutation were subjected to exome sequencing. Three patients with family history of skeletal and/or aortic features resembling MFS/LDS were identified to have a novel mutation in *TMEPAI*. Case 1 was a 43y tall male with funnel chest, arachnodactyly and a history of osteosarcoma with leg amputation in childhood. His mother was also tall and had abdominal aortic aneurysm. His 13y daughter (Case 2) and 11y son (Case 3) were both tall and showed funnel chest, dural ectasia and arachnodactyly without lens dislocation. Case 2 had scoliosis and mitral valve prolapse. Case 3 showed mild aortic root dilatation. Genetic analysis revealed no pathogenic gene variation in known MFS/LDS disorders in these patients. Exome sequencing revealed candidate variants in 12 genes and only 3 of them (*TMEPAI*: c.624_625insC, p.S209Qfs; *AP5M1*: c.A262T, p.I88L; *ESRRB*: c.C156G, p.H52Q) were segregated with similar stature and deformities in this family members. In addition, Case 4, 33y female in unrelated family, showing arachnodactyly with aortic and axillar arterial dissection, revealed to have the same mutation of *TMEPAI*. Since all individuals with *TMEPAI* mutation showed MFS/LDS habitus and some had aortic diseases, *TMEPAI* mutation were segregated with these phenotypes. Then, we studied the expression of fibronectin after TGF-β2 stimulation and the patient's fibroblasts with the *TMEPAI* mutation showed significant upregulation of fibronectin than control fibroblasts, suggesting upregulation of TGF-β signals. It was reported that *TMEPAI* suppresses TGF-β signals and its overexpression promotes progression of prostate cancer. Since *TMEPAI* gene plays an important role in regulation of TGF-β signals, it is thought that *TMEPAI* mutation can cause MFS/LDS-like features. This is the first report of *TMEPAI* mutation found in familial cases with MFS/LDS-like features. Based on these results, we conclude that *TMEPAI* is essential not only for cancer progression but also for systemic connective tissue disorders similar to MFS/LDS.

922W

A distinct cutaneous blistering phenotype with multi-system manifestations caused by a mutation in CD151, the 20th causative gene in epidermolysis bullosa. *H. Vahidnezhad[1,2], L. Youssefian[1,3], A. Saeidian[1], H. Mahmoudi[4], A. Touati[1], A. Kajbafzadeh[5], J. McGrath[6], S. Zeinali[2,7], P. Fortina[8], J. Uitto[1].* 1) Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA;; 2) Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran; 3) Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran; 4) Department of Dermatology, Razi Hospital, Tehran University of Medical Sciences, Tehran, Iran; 5) Paediatric Urology Research Center, Department of Urology, Children's Hospital Medical Center, Tehran University of Medical Sciences, Tehran, Iran; 6) St. John's Institute of Dermatology, King's College London (Guy's Campus), UK; 7) Kawsar Human Genetics Research Center, Tehran, Iran; 8) Department of Cancer Biology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania, USA.

   Epidermolysis bullosa (EB) is caused by mutations in as many as 19 distinct genes. We have developed a next generation sequencing (NGS) panel targeting genes known to be mutated in EB. This panel includes *CD151*, encoding a transmembrane tetraspanin CD151 expressed in keratinocytes. The NGS panel was applied to a cohort of 92 consanguineous families, consisting of 134 individuals of unknown subtype of EB. To direct the bioinformatics analysis to appropriate candidate genes, we applied a genome-wide homozygosity mapping with a SNP-based array consisting of 550,000 markers. In one family, a homozygous *CD151* mutation (NM_139029; c.351+2T>C) at the canonical splice junction at intron 4/exon 5 border was identified, predicting aberrant splicing. RT-PCR confirmed deletion of the entire exon 5 consisting of 75 bp and encoding 25 amino acids. The in-frame deletion resides within the third transmembrane domain of CD151, suggesting aberrant membrane targeting of the protein. Immunofluorescence of skin and Western analysis with a monoclonal antibody revealed the absence of the CD151 epitope. Transmission electron microscopy showed intracellular disruption and cell-cell dysadhesion of keratinocytes in the lower epidermis. Clinical examination of the 35-year old proband revealed widespread blistering, particularly on pretibial areas, nail dystrophy and early onset alopecia, as well as esophageal webbing and strictures. The patient also had history of nephropathy with proteinuria. Collectively, the results provide evidence of a distinct form of EB with systemic involvement. The results suggest that *CD151* should be considered as the 20th causative, EB-associated gene.

923T

Early oro-dental manifestations as a clue for the clinical diagnosis of infantile systemic hyalinosis. *I.S.M. Sayed[1], M.I Mehrez[1], G.Y. El-Kamah[2], M.S. Abdel-Hamid[3], G.M.H. Abdel-Salam[2].* 1) Oro-Dental Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 2) Clinical Human Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 3) Medical Molecular Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt.

   Biallelic mutations of *ANTXR2* gene cause hyaline fibromatosis syndrome (OMIM #228600) which is a very rare autosomal recessive disorder. It has two forms; the juvenile and the infantile types. Infantile systemic hyalinosis has an earlier onset and displays a more severe presentation when compared with the juvenile type. Interestingly, infantile systemic hyalinosis shows considerable clinical overlap with Farber syndrome. Both show severe, progressive and painful contractures, papular skin rashes, hyperpigmented plaques over the joints, nodular eruptions, gingival enlargement and early morbidity. We describe five patients from four unrelated families with pathogenic mutations of the *ANTXR2* gene. All patients presented with severe irritability and generalized joint contractures while two of them displayed hyperpigmented skin plaques over the knuckles and papular skin. The segmented alveolar ridges and fibrous nodules at the corners of the mouth were consistent features that manifested as early as the age of 2 months. Gingival enlargement obscuring the erupting teeth and stiff long philtrum were additional distinctive findings observed in this series. Thus, suggesting the specificity of these oral features of infantile systemic hyalinosis, they could be taken to differentiate it from Farber syndrome. Three different novel homozygous mutations in exons 9 and 13 were identified, c.1069_1069delG (p.A357Pfs*52), c.1052A>C (p.D351A) and c.700_700delG (p.E234Nfs*6). The c.700_700delG (p.E234Nfs*6) mutation was recurrent in two families who shared similar haplotype suggesting a founder effect. This study highlights the need for thorough oro-dental examination as part of the initial evaluation. Further, we report the distinctive appearance of segmented alveolar ridge and gingiva as new additional features that could be taken as a clue in the clinical diagnosis of patients with infantile systemic hyalinosis.   .

924F

Novel compound heterozygous variants in the gene *CHUK* associated with AEC syndrome-like phenotype and immune system involvement.
M. Cadieux-Dion[1,2], N. Safina[3,4,5], K. Engleman[3,4], L. Zellmer[1,2], E. Repnikova[1,2], C. Saunders[1,2], I. Thiffault[1,2]. 1) Pathology and Laboratory Medicine, Children's Mercy Hospital, Kansas City, MO; 2) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO; 3) Division of Clinical Genetics, Children's Mercy Hospital, Kansas City, MO; 4) Department of Pediatrics, Children's Mercy Hospitals, Kansas City, MO; 5) University of Missouri, Kansas City School of Medicine, Kansas City, MO.

Ectodermal dysplasias (ED) are a group of disease that affects the development or function of the teeth, hair, nails and exocrine and sebaceous glands. Among ED, ankyloblepharon-ectodermal defects-cleft lip/palate syndrome (AEC or Hay-Wells syndrome), is caused by heterozygous pathogenic variants in the *TP63* gene. A classical feature of AEC syndrome is the presence of skin erosions affecting the palms, soles and scalp. Other clinical manifestations include ankyloblepharon filiforme adnatum, cleft lip, cleft palate, craniofacial abnormalities and ectodermal defects such as sparse wiry hair, nail changes, dental changes, and subjective hypohydrosis. Here, we present a patient with clinical features reminiscent of AEC syndrome with recurrent infections suggestive of immune deficiency. Genetic testing for *TP63*, *IRF6* and *RIPK4* were negative. Using clinical exome sequencing, we uncovered two novel heterozygous variants *in trans* in the *CHUK* gene: a maternally-inherited frameshift (c.1365del, p.Arg457AspfsTer6) and a *de novo* missense (c.1388C>A, p.Thr463Lys). These variants are absent from population databases and paternity was confirmed. To our knowledge, pathogenic variants in the *CHUK* gene have been reported in 3 families. In 2010, a homozygous loss of function variant was found in a Finnish family with Cocoon syndrome, a lethal condition characterized by severe fetal malformations. In 2015, a homozygous splice acceptor site variant was identified in a family with Bartsocas-Papas syndrome (BPS). Recently, a *de novo* missense variant was reported in one patient with ED, orofacial clefting, limb anomalies and hypogammaglobulinemia. This patient had more severe skeletal and immune defect compared to our patient. In summary, we have identified the first case with compound heterozygous variants in the *CHUK* gene with a milder AEC syndrome-like phenotype along with immune system involvement. This shows that pathogenic variants in *CHUK* can cause a range of severity along the AEC-BPS spectrum, and expand the phenotypic spectrum of disorders caused by *CHUK* deficiency. Therefore, it is unclear if *de novo* heterozygous *CHUK* variant are sufficient to cause the phenotype in the patient presented here. Even if the majority of pathogenic *de novo* variants are involved in dominant genetic disorders, there are growing examples of recessive disorders that can be caused by the combination of an inherited variant on one allele and a *de novo* variant on the other.

925W

Understanding the impact of a novel homozygous nonsense *CAST* gene mutation in a PLACK family.
S.G. Temel[1], B. Karakas[2], B. Turkgenc[3], O. Zorlu[4], O. Ozturk[5], U. Kiran[4], B. Erguner[6], M. Sagiroglu[6], C. Yakicier[7], H. Saricaoglu[4]. 1) Dept Histol & Embryol, Fac Medicine, Univ Uludag, Bursa, Turkey; 2) Sabancı University, Molecular Biology, Genetics and Bioengineering Program, Istanbul, Turkey; 3) University of Acibadem, Acibadem Genetic Diagnostic Center, Istanbul, Turkey; 4) University of Uludag, Faculty of Medicine, Department of Dermatology, Bursa, Turkey; 5) Baskent University, Faculty of Medicine, Deparment of Medical Genetics, Adana, Turkey; 6) Advanced Genomics and Bioinformatics Research Group TÜBİTAK BİLGEM UEKAE,Gebze/Kocaeli, Turkey; 7) University of Acibadem, Faculty of Arts and Sciences, Department of Molecular Biology and Genetics, Istanbul, Turkey.

Peeling skin syndrome is an autosomal recessive disease characterized by continuous shedding of the stratum corneum of the epidermis with onset from birth or infancy and lasting throughout life. It can be accompanied by erythema, vesicular lesions, or other ectodermal features like fragile hair and nail abnormalities. We report a 5.5 year old boy who initially presented with skin fragility. He was the first-born of IVF twins from a consanguineous marriage at 33+5 week gestational age with hypotonia at birth. The main clinical findings included fragile skin, woolly hair, sparse eyelashes and brows, palmoplantar punctate keratoderma, follicular hyperkeratosis, knuckle pads and cheilitis. Exome sequencing revealed a homozygous c.544G>T (p.Glu182*) nonsense mutation in the *CAST* gene. Segregation analysis was confirmed by Sanger sequencing. This novel stop-gain E182X variant produces a truncated protein lacking inhibitory domains II-IV. Notably, the variant is absent in all public databases, also the region is highly conserved among species. H&E stainings of the proband's skin biopsy showed hyperkeratosis, mild spongiotic changes, Toluidine Blue stainings of semithin sections showed elongated epidermal rete pegs, dermal papilla, thick stratum corneum, extensive karyolysis and loss of chromatine pattern in all epithelial cells. Electronmicrographs of the lesioned skin showed widened intercellular space, proapoptotic membrane invaginations, nuclear chromatine accumulations. Immunohistochemistry results showed absent calpastatin staining in the proband compared to normal staining throughout the epidermis in the control. Calpastatin activity assay revealed reduced calpain proteolysis in affected ones. Immunoblot results showed tissue-spesific expression of calpastatin, similar to RT-qPCR results. Confocal microscopy results confirmed the expression pattern shown in immunoblot results. CAST is an endogenous specific inhibitor of calpain, a calcium-dependent cysteine protease. Recently, autosomal recessive loss of function mutations in *CAST* were described in PLACK syndrome characterized by peeling skin, leukonychia, acral punctate keratoses, cheilitis, and knuckle pads. This case is the fifth case of PLACK syndrome without leukonychia but with some additional previously unreported features. Treatment with calpain inhibitors could be used to reduce the unwanted symptomps. Our findings might pave the way to explore new routes in proteolytic pathways in skin. .

926T

Delineation of musculocontractural Ehlers-Danlos syndrome caused by dermatan sulfate epimerase deficiency (mcEDS-*DSE*): Report of additional patients and comprehensive review of reported cases. *A. Unzaki[1,2,3], CK. Lautrup[4], K. Wee Teik[5], S. Mizumoto[6], H. Hock Sin[7], IK. Nielsen[4], S. Markholt[8], S. Yamda[6], N. Matsumoto[9], N. Miyake[9], T. Kosho[1,2].* 1) Center for Medical Genetics, Shinshu University Hospital , Matsumoto, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Problem-Solving Oriented Training Program for Advanced Medical Personnel: NGSD (Next Generation Super Doctor) Project, Matsumoto, Japan; 4) Department of Clinical Genetics, Aalborg University Hospital, Aalborg, Denmark; 5) Department of Genetics, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; 6) Department of Pathobiochemistry, Faculty of Pharmacy, Meijo University, Nagoya, Japan; 7) Department of Pediatrics, Sabah Women and Children's Hospital, Kota Kinabalu Sabah, Malaysia; 8) Department of Clinical Genetics, Aarhus University Hospital, Aarhus, Denmark; 9) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Musculocontractural Ehlers-Danlos Syndrome (mcEDS) is a recently identified connective tissue disorder, classified into a subtype of EDS in the 2017 International Classification of the EDS (Malfait et al., Am J Med Genet Semin Med Genet 175C: 8−26, 2017). It is caused by loss-of-function mutations in *CHST14* encoding dermatan 4-*O*-sulfotransferase-1 (D4ST1) (mcEDS-*CHST14*) or in *DSE* encoding dermatan sulfate epimerase (mcEDS-*DSE*), both resulting in defective biosynthesis of dermatan sulfate (DS). Whereas 40 patients with mcEDS-*CHST14* from 27 families have been published to date, only three patients from two families with mcEDS-*DSE* have been published (Müller et al., Hum Mol Genet 22: 3761−3772, 2013; Syx et al., Hum Mutat 36: 535−547, 2015). In this report, we present three additional patients from two consanguineous families. Patient 1, a boy, had developmental delay, hypertelorism, a large anterior fontanel, bilateral adducted thumbs, bilateral hydronephrosis, cryptorchidism, and constipation in infancy. Later, he showed skin hyperextensibility, delayed wound healing, myopia, strabismus, mild hearing loss, joint hypermobility, scoliosis, recurrent shoulder dislocations, large subcutaneous hematomas, and hemoptysis. At age 18 years, he had a normal stature with mild pectus excavatum and bilateral pes cavus, and no intellectual disability. Patient 2, a boy, had normal developmental milestones, myopia, large subcutaneous hematomas, fractures, talipes equinovarus, and severe scoliosis. At age 21 years, he had a short stature with an asymmetric chest, hallux valgus, and multiple contractures. He had mild hypertelorism, blue sclerae, and atrophic scars. Patient 3, a younger sister of Patient 2, had normal developmental milestones, joint pain and swelling, and astigmatism. At age 14 years, she had a short stature with pectus carinatum and kyphoscoliosis. She had hypertelorism and blue sclerae. She had no joint hypermobility, skin hyperextensibility, or fragility. Pathogenic *DSE* mutations were found in all the three patients and no DS was detected in urine sample of Patient 1, indicating generalized depletion of DS. Review of previously published patients and the current series would delineate the disorder as milder phenotypes with less characteristic facial features and less severe skin/joint/vascular manifestations than mcEDS-*CHST14*.

927F

***ANKRD26* loss of function somatic mutation in a female case with Tree Man Syndrome.** *K.M.F. Uddin[1,4], M.R. Amin[2], S.N. Majumder[2], M.A. Aleem[3], N.J. Dity[4], A. Rahman[4], H. Akter[4], M.A. Baqui[1,4], M. Woodbury-Smith[5], S.W. Scherer[6,7,8], M. Uddin[9].* 1) Biochemistry Dept, Holy Family Red Crescent Medical College, Dhaka, Bangladesh; 2) Medicine Dept, Dhaka Medical College & Hospital, Department of Medicine, Dhaka, Bangladesh; 3) Infectious Disease Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr'b), Dhaka, Bangladesh; 4) Genetics and Genome Biology Dept, NeuroGen Technologies Ltd, Dhaka, Bangladesh; 5) Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK; 6) The Centre for Applied Genomics, Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 8) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 9) College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE.

Epidermodysplasia verruciformis (EV), referred to colloquially as 'tree man syndrome' is a lifelong exceedingly rare hereditary skin disease. It is characterized by abnormal susceptibility to the human papilloma virus (HPV) infection causing defect in cell-mediated immunity. Its dermatological manifestations vary from flat wart like lesion, flat-topped papule, plaques to verrucous or seborrhoeic keratosis like lesions and even cutaneous horn like lesion mainly seen in UV- exposed regions. Affected individuals have a lifelong increased risk for the development of cutaneous malignancy, especially squamous cell carcinoma. Both non-surgical and surgical options are implicated for the treatment of this condition. Loss of function mutation within *EVER1* and *EVER2* genes has been reported for 25% of the male cases. Up until now the severe form of this syndrome was described exclusively among males. Here we report the first female case, a 10-year-old girl who presented to Dhaka Medical College & Hospital in Bangladesh with a severe dermatological condition, histopathologically confirmed in resected tissue as EV. Local examination revealed multiple tree bark like lesions 2-3 cm in size and conical in shape projecting from various tissues, including the cheek, chin, nose, and ear lobules. Although surgical intervention initially removed all bark like lesions, aggressive regrowth of lesionswas observed after one week. From whole genome sequencing data we have identified a rare stop gain somatic mutation in *ANKRD26* and subsequently confirmed the mutation in the tree bark like lesion. *ANKRD26* is strongly associated with thrombocytopenia and myeloid neoplasm. Our findings identify a new pleiotropic gene for EV.

928W

Mutations in *SULT2B1* cause autosomal recessive congenital ichthyosis in humans. *L. Heinz[1,2], G.-J. Kim[1,3], S. Marrakchi[4], J. Christiansen[5], H. Turki[4], M.-A. Rauschendorf[1], M. Lathrop[6], I. Hausser[7], A.D. Zimmer[1], J. Fischer[1].* 1) Institute for Human Genetics, University Medical Center Freiburg, Freiburg, Baden-Württemberg, Germany; 2) Faculty of Biology, University of Freiburg, 79104, Freiburg, Germany; 3) Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences, Albert-Ludwigs University, 79104 Freiburg, Germany; 4) Department of Dermatology, Hédi Chaker University Hospital, 3029 Sfax, Tunisia; 5) Department of Dermatology and Venereology, Skanes University Hospital, 22185 Lund, Sweden; 6) McGill University and Genome Quebec Innovation Centre, H3A 0G1 Montréal, QC, Canada; 7) Institute of Pathology, University Hospital Heidelberg, 69120 Heidelberg, Germany.

   Ichthyoses are a clinically and genetically heterogeneous group of genodermatoses associated with abnormal scaling of the skin over the whole body. Mutations in nine genes are known to cause non-syndromic forms of autosomal recessive congenital ichthyosis (ARCI). However, not all genetic causes for ARCI have been discovered to date. Using whole exome sequencing (WES) and multigene panel screening we identified 6 ARCI individuals from three unrelated families with mutations in Sulfotransferase family 2B member 1 (*SULT2B1*), showing their causative association with ARCI. Cytosolic sulfotransferases form a large family of enzymes that are involved in the synthesis and metabolism of several steroids in humans. We identified four distinct mutations including missense, nonsense and splice site mutations. We demonstrated the loss of SULT2B1 expression at RNA and protein levels in keratinocytes from individuals with ARCI by functional analyses. Furthermore, we succeeded in reconstructing the morphologic skin alterations in a 3D organotypic tissue culture model with *SULT2B1*-deficient keratinocytes and fibroblasts. By thin layer chromatography (TLC) of extracts from these organotypic cultures we could show the absence of cholesterol sulfate, the metabolite of SULT2B1, and an increased level of cholesterol indicating a disturbed cholesterol metabolism of the skin upon loss of function mutation in *SULT2B1*. In conclusion, our study reveals an essential role for SULT2B1 in the proper development of healthy human skin. Mutation in *SULT2B1* leads to an ARCI phenotype via increased proliferation of human keratinocytes, thickening of epithelial layers, and altered epidermal cholesterol metabolism.

929T

Cutaneous neurofibromas in neurofibromatosis type 1: A quantitative natural history study. *A. Cannon[1], M-J. Chen[1], P. Li[3], KP. Boyd[2], A. Theos[2], DT. Redden[3], B. Korf[1].* 1) Genetics, University of Alabama at Birmingham, Birmingham, AL; 2) Dermatology, University of Alabama at Birmingham, AL; 3) Biostatistics, University of Alabama at Birmingham, AL.

   Background: Neurofibromatosis type 1 (NF1) is a disorder characterized by a predisposition to develop multiple benign tumors, including localized cutaneous neurofibromas (cNFs) that affects >99% of adults with NF1. Previous reports have correlated increased burden of cNFs with age and pregnancy, but longitudinal data are not available to establish a quantitative natural history. The purpose of this study was to quantify cNF number and size to establish a quantitative natural history of these lesions. **Methods:** A prospective cohort analysis of 22 adults with NF1 was conducted over an 8-year period. The primary outcome measures were number and volume of cutaneous neurofibromas in three body sites: back, abdomen, and upper arm or thigh. **Results:** The average monthly increase in volume for cutaneous neurofibromas was 0.37 mm^3 in the back region (95% CI (0.23, 051), p<0.0001), 0.28 mm^3 in the abdominal region (95% CI (016, 0.41), p<0.0001), and 0.21 mm^3 in the arm/leg region (95% CI (0.08, 0.34), p=0.0022). The number of cutaneous neurofibromas significantly increased in the back (slope=0.032, p=0.011) and abdominal (slope=0.018, p=0.026) regions, while the leg/arm regions retained a positive trend (slope=0.004, p=0.055). **Conclusion:** The number and volume of cNFs significantly increase over an 8-year timespan; however, the rate of increase is variable by body region. These findings may provide insight into cNF development and benefit researchers considering clinical trials targeting cNFs.

930F

Abnormal splicing in a case of epidermolysis bullosa with a novel synonymous mutation in the *ITGB4* gene. *E. Tan, H.W. Lim, E.C.P. Lim, D. Lian, M. Koh.* KK Women s & Children s Hospital, Singapore.

   Introduction: Integrins are protein heterodimers involved in the maintenance of cell-cell adhesion, signal transduction, and regulation of cell growth. We report the findings from a female infant who presented with feed intolerance, vomiting and scattered skin blisters and erosions shortly after birth. Immunofluorescence mapping showed fraying of the integrin alpha-6 / beta-4 complex, consistent with a clinical diagnosis of epidermolysis bullosa with pyloric atresia (EB-PA). There was no family history and parents were non-consanguineous. **Methodology:** Genomic DNA from the child was extracted from venous blood and sequenced using a targeted panel on the MiSeq System. Identified variants were confirmed by Sanger sequencing followed by targeted testing of parental samples. **Results and discussion:** For the *ITGB4* (integrin beta-4 protein) gene, there was a heterozygous frameshift variant (p.S265fs) which has been reported as disease-causing in HGMD and a heterozygous synonymous variant (p.Cys536Cys) which has not been reported previously. Testing of parental samples showed that the frameshift variant is from the mother while the synonymous variant is from the father. In silico analysis predicted that the synonymous variant might cause abnormal splicing. RNA work was thus carried out on the patient's blood sample and intestinal biopsy tissue, PCR products from cDNA of both tissues were of a smaller product size compared to those obtained using cDNA from HeLa cells and a normal stoma tissue sample as templates. Sanger sequencing indicated that the last 51 bp from exon 13 was missing from the shorter PCR product. The absence of a normal RNA species from the patient's tissues is consistent with the immunostaining result. **Conclusion:** We describe a novel synonymous variant and show that the encoded RNA in the patient's tissues is smaller due to abnormal splicing. This variant is inherited from a healthy parent. Together with a known pathogenic frameshift variant inherited from the other healthy parent, the patient presented with EB-PA due to compound heterozygous mutations in *ITGB4*. **Acknowledgment:** Supported by NMRC/CG/006/2013. .

931W

A disease-associated *REEP1* variant affects splicing of the gene's 3'UTR. *C. Beetz[1], S.L. Rydning[2], A. Jahic[1], L. Goldberg[1], C. Tallaksen[2].* 1) Department of Clinical Chemistry, Jena University Hospital, Jena, Germany; 2) Department of Neurology, Oslo University Hospital, Oslo, Norway.

   Heterozygous loss-of-function mutations in *REEP1* cause autosomal dominant hereditary spastic paraplegia (HSP). We identified the previously reported c.*43G>T substitution in the gene's 3'UTR in an apparently sporadic patient. Age at onset was comparatively late, and the patient's father and daughter are unaffected carriers. The variant is found in control populations, but shows a much higher frequency amongst HSP patient. The mode of action of this apparently only mildly pathogenic variant is unclear. We applied MiniGene assays to test for a potential effect on splicing. We found the GT dinucleotide at c.*58_59 to constitute a functional splice donor, the activity of which is decreased in the presence of the c.*43G>T variant. Screening candidate downstream motifs identified a corresponding splice acceptor with the core AG-motif at c.*1645_1646. *In silico* analysis of the spliced out ~1,6 kb sequence revealed a cluster of high scoring AU-rich elements, i.e. regions that mediate RNA degradation. The decrease in 3'UTR splicing as observed for c.*43G>T is predicted to result in an increase of the fraction of molecules carrying degradation motifs and, thereby, in reduced *REEP1* mRNA levels. Our findings are consistent with a novel mechanism for how 3'UTR variants may confer a partial loss of gene function.

932T

Synonymous variant in *KCTD7* causes alternative splicing in siblings with progressive epilepsy. *D.B. Zastrow[1,2], C. Prybol[1,2], J.M. Davidson[1,2], A.M. Dries[1,2], L. Fernandez[1,2], J.N. Kohler[1,2], Z. Zappala[3], L. Fresard[4], K.S. Smith[4], D. Bonner[1,2], C. McCormack[1,2], M. Majcherska[1,2], D. Waggott[1,2], S. Marwaha[1,2], N. Friedman[1,2], P. Ward[5], Y. Yang[5], C. Eng[5], U.D.N Members[6], S.B. Montgomery[3,4], P.G. Fisher[1,7,8], E.A. Ashley[1,2,3], J.A. Bernstein[1,7,9], M. Wheeler[1,2].* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 3) Department of Genetics, Stanford School of Medicine, Stanford, CA; 4) Department of Pathology, Stanford School of Medicine, Stanford, CA; 5) Baylor Miraca Genetics Laboratories, Houston, TX; 6) NIH Undiagnosed Diseases Network, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 7) Department of Pediatrics, Stanford School of Medicine, Stanford, CA; 8) Department of Neurology, Stanford School of Medicine, Stanford, CA; 9) Lucille Packard Children's Hospital Stanford, Palo Alto, CA.

A 13-year-old Hispanic female presented at 18 months with regression, loss of milestones, head control, and speech. Tremors began at 21 months, and seizures at 22 months (occasional myoclonus). She also has scoliosis, dysphagia, encephalopathy, and corticovisual impairment. A diagnostic odyssey began which encompassed 32 laboratory studies (including 17 genetic tests and 9 imaging studies) over 11 years. A genetic etiology could not be uncovered. Her 5-year-old brother had a different onset noted at 13 months of ataxia, autism, developmental delay and recurrent febrile seizures. *KCTD7* sequencing was done in 2014 for the proband which found 2 variants: c.280C>T (R94W) and c.456G>A (V152V). The R94W variant is classified as pathogenic in ClinVar (VarID 37010). The c.456G>A variant was interpreted as likely benign (VarID 206004). Her brother had *KCTD7* deletion/duplication analysis via high-density array CGH, which was normal. As part of her evaluation in the Undiagnosed Diseases Network (UDN) at Stanford, quad whole exome sequencing was performed. The *KCTD7* variants in the proband and her brother were appreciated and further investigated. The R94W variant is paternally inherited, and the c.456G>A variant is maternally inherited consistent with autosomal recessive inheritance. *KCTD7* mutations are associated with Epilepsy, progressive myoclonic 3 (EPM3 [MIM 611726]). Features of EPM3 (myoclonic seizures, neurologic regression, truncal ataxia, limited expressive language, and visual loss) are seen in the proband and her younger brother. During curation of the c.456G>A variant, MutationTaster predicted it to be disease causing based on potential splicing effect (donor increased). We then set out to confirm this prediction. We performed Sanger sequencing of the major *KCTD7* isoform on the siblings' cDNA and confirmed a splicing gain. RNAseq performed on patient blood samples and induced pluripotent stem cells also detected a splicing gain. Therefore, this synonymous variant appears to create a splice donor site that truncates *KCTD7* exon 3. Additional validation is in progress including family segregation and model organism studies. This adds to the small number of reported EPM3 cases (~19) and informs the interpretation of this potential pathogenic synonymous variant. The UDN clinical research network provides a means for the efficient and effective assessement of variants leading to improved care.

933F

NeuroChip genotyping of the Johns Hopkins brain bank reveals common and rare genetic associations. *C. Blauwendraat[1], J.T. Geiger[1], O. Pletnikova[2], F. Faghri[3], L.S. Rosenthal[2], N.A. Murphy[3], B. Crain[2], Y. Abramzon[3], G. Rudow[2], C. Ruth[3], J. Ding[3], T.M. Dawson[2], A. Pantelyat[2], M.S. Albert[2], A.E. Hillis[2], J.C. Troncoso[2], S.W. Scholz[1,3,4].* 1) Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA; 2) Department of Pathology (Neuropathology), Johns Hopkins University School of Medicine, Baltimore, MD, USA; 3) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 4) Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD, USA.

Genetics has proven to be a powerful tool in research and diagnostics, resulting in the identification of causal and risk variants in numerous diseases including neurodegenerative diseases like Alzheimer's disease (AD) and Parkinson's disease (PD). Array based genotyping and sequencing are the primary techniques used to identify these variants. The downsides of these techniques are typically the lack of rare/pathogenic variants included (genotyping) and the high costs (sequencing). To overcome this, we previously developed the NeuroX genotyping array and we report here its updated successor, NeuroChip. The NeuroChip is a low cost (~$40 per sample) custom-designed array containing 179,467 custom content variants alongside a genome-wide tagging backbone of 306,670 variants (Infinium HumanCore-24 v1.0). These custom content variants have been identified in large sequencing and genotyping studies in several diseases including AD, PD, Lewy body dementia (LBD), amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD) and progressive supranuclear palsy (PSP). Additionally, there are over 1,000 known pathogenic variants included e.g. *PSEN1*-AD (n=175), *LRRK2*-PD (n=80), *SOD1*-ALS (n=70) and *MAPT*-FTD (n=39). As a proof of principle, we genotyped the Johns Hopkins brain bank consisting of 1,405 brain tissues. The most common pathological diagnosis was AD (n=621), followed by complex dementia, which constitutes as having two or more distinct pathologies (n=302), control or no significant pathology present (n=118), Lewy body, including both PD and LBD (n=99), ALS-FTD (n=94), atypical PD (n=55) and other or unclassified pathologies (n=116). In this cohort, we identified multiple *PSEN1* variants and *APP* variants (both missense and duplications), but also *LRRK2* variants in brains that were classified as complex dementia. Additionally, mini GWAS and genetic risk score (GRS) calculations were performed showing specific trends for disease groups including *APOE* association in AD, *MAPT* association in PSP and overall GRS association with PD. In summary, we report here the usage of the NeuroChip array, and show its capability of detecting both common and rare (pathogenic) variants in a large cohort. The NeuroChip has a comprehensive and improved custom content which makes it a high-throughput and cost-effective screening genotyping array for future research in neurodegenerative diseases.

934W

A *PSEN2* frameshift variant associated with early onset AD in two families. *S. Jayadev[1], S.A. Bucks[1], C.L. Smith[1], L. Osnis[1], B. Sopher[1], C. Cross[5], K. Scherpelz[2], C.S. Latimer[2], L.F. Gonzalez-Cuyar[2], C.D. Keene[2], M.O. Dorschner[3,4], G.A. Garden[1,2], J.E. Young[2], D. Tsuang[4,5], T.D. Bird[1,4].* 1) Neurology, University of Washington, Seattle , WA; 2) Pathology, University of Washington, Seattle, WA; 3) UW Medicine Center for Precision Diagnostics, University of Washington, Seattle, WA; 4) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 5) Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA.

Heterozygous mutations in Amyloid Precursor Protein, Presenilin 1 and Presenilin 2 (*PSEN2*) cause autosomal dominant Alzheimer Disease (AD). The mechanism by which Presenilin mutations lead to AD has been debated; there is evidence to support both toxic gain of function as well as loss of normal function mechanisms. We previously reported a 56 year-old woman with early onset AD found to have a novel heterozygous *PSEN2* 2-base pair deletion leading to a frame shift and premature termination codon in exon 5 (PSEN2 p.K115Efs*11). The PSEN2 K115Efs variant is predicted to lead to nonsense mediated decay or result in a stable transcript translated into an early truncated PS2 peptide. The variant, if truly causing familial AD, significantly informs our understanding of *PSEN* gene mutation function because heretofore only missense *PSEN2* mutations have been considered associated with AD. Here we describe a second early onset AD case heterozygous for the PSEN2 K115Efs variant discovered during a targeted exome screen of early-onset dementia cases. The subject had a maternal history of AD and was diagnosed with dementia at age 51. Her autopsy confirmed neuropathological AD diagnosis with neurofibrillary tangles (Braak Stage VI), and neuritic plaques. To investigate pathogenicity *in vitro*, we measured the expression and function of PSEN2 K115Efs. Sanger sequencing of cDNA confirmed the presence of mutant transcript in fibroblasts and cortical tissue isolated from two unrelated variant carriers. Western blot analysis of primary cells isolated from control subjects and PSEN2 variant carriers showed decreased PS2 c-terminal fragment in PSEN2 K115Efs suggesting insufficient compensation for PS2 levels by the wildtype allele. Furthermore, similar to what has been previously shown with PSEN2 N141I (familial AD Volga German mutation), we found decreased Ab_{40} released into conditioned medium from PSEN2 K115Efs and PSEN2 N141I primary fibroblasts compared to controls, suggesting that like the PSEN2 N141I mutation, PSEN2 K115Efs leads to impaired APP processing, a biochemical loss of function. In summary, we report evidence demonstrating loss of normal PS2 levels and function associated with the PSEN2 K115Efs*11 variant which has now been identified in two individuals with familial early onset AD. Further, the biological and genetic features associated with this variant type underscore the complexity of *PSEN2* mutation mechanisms of action leading to AD.

935T

Comparison of the mutations that cause Alzheimer's disease on secondary protein misfolding on transgenic mouse models. *G. Xu, M. Pace, S. Fromholt, J. Triplett, Y. Levites, P. Chakrabarty, T. Golde, J. Lewis, D.R. Borchelt.* Neuroscience, University of Florida, Gainesville, FL.

Mutations in amyloid precursor protein (APP), presenilins (PS1, PS2) or microtubule-associated protein tau (MAPT or tau) can cause familial Alzheimer's disease (AD). Multiple transgenic mouse models developed by expressing these mutations reproduce amyloidosis or tau pathology of human disease. In our prior studies, we used proteomic analyses to determine that mice overexpressing mutant human APP and PS1 genes that develop severe amyloidosis show secondary protein misfolding in the brain. In the current study, using the same strategy of bottom-up LC-MS/MS, we identified the proteins in the detergent-insoluble fraction of the brain homogenates of many different mouse AD models, we found the similar secondary misfolding in both amyloidosis mice and tauopathy mice. Additionally, we sought to comparatively analyze the effects that different initiating proteinopathies have upon secondary protein misfolding. We compared the mice expressing mutant human APP with or without co-expressing of PS1, transgenic mice expressing mouse APP with or without PS1, mice expressing Abeta42 fragment, and transgenic mice overexpressing mutant human tau. Although some of the misfolding proteins are the same across all the models, some proteins are unique to each model. We found the amount of secondary misfolding is not simply corresponded to amyloid deposition level. The mice that develop cored amyloid plaques seem to have higher levels of secondary misfolding than the mice develop diffuse plaques. In both Abeta mice and tau mice, as soon as the disease-driving transgene's expression is switched off in the tetracycline-inducing system, the level of secondary misfolding declines even though the primary pathology persisted. Our data demonstrate that secondarily misfolded proteins are not simply co-aggregated with the main protein deposition. Overall, our findings fit with a model in wich an independent dysfunction of protein homeostasis mechanism causes widespread protein misfolding in neurodegenerative disease.

936F

Regulatory role of RNA chaperone TDP-43 for RNA misfolding and repeat-associated translation in SCA31. *K. Ishikawa[1], T. Ishiguro[2], N. Sato[2], M. Ueyama[3], N. Fujikake[4], C. Sellier[5], E. Tokuda[6], B. Zamirini[7], T. Gall-Duncan[7], M. Mirceta[7], Y. Furukawa[6], T. Yokota[2], K. Wada[4], P.J. Taylor[8], C.E. Pearson[7], N. Charlet-Berguerand[5], H. Mizusawa[2], Y. Nagai[3].* 1) Center for Personalized Medicine for Healthy Aging, Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 2) Dept of Neurology and Neurological Science, Graduate School, and The Center for Brain Integration Research (CBIR), Tokyo Med & Dental Univ, Tokyo, Tokyo, Japan; 3) Dept of Neurotherapeutics, Osaka University Graduate School of Medicine, Suita, Osaka, Japan; 4) Dept of Degenerative Neurological Diseases, Natl Inst of Neurosci, National Center for Neurology and Psychiatry, Kodaira, Tokyo, Japan; 5) Dept of Translational medicine and neurogenetics, IGBMC, CNRS UMR7104, INSERM U964, Université de Strasbourg, Illkirch, France; 6) Dept of Chemistry, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan; 7) Dept of Genetics & Genome Biology, The Hospital for Sick Children,Toronto, ON, Canada; 8) Dept of Cell and Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN.

Microsatellite expansion disorders are pathologically characterized by RNA foci formation and repeat-associated non-AUG (RAN) translation. However, their underlying pathomechanisms and regulation of RAN translation remain unknown. We report that expression of expanded UGGAA ($UGGAA_{exp}$) repeats, responsible for spinocerebellar ataxia type 31 (SCA31) in *Drosophila*, causes neurodegeneration accompanied by accumulation of $UGGAA_{exp}$ RNA foci and translation of repeat-associated pentapeptide repeat (PPR) proteins, consistent with observations in SCA31 patient brains. We revealed that motor-neuron disease (MND)-linked RNA-binding proteins (RBPs), TDP-43, FUS and hnRNPA2B1, bind to and induce structural alteration of $UGGAA_{exp}$. These RBPs suppress $UGGAA_{exp}$-mediated toxicity in *Drosophila* by functioning as RNA chaperones for proper $UGGAA_{exp}$ folding and regulation of PPR translation. Furthermore, nontoxic short UGGAA repeat RNA suppressed mutated RBP aggregation and toxicity in MND *Drosophila* models. Thus, functional crosstalk of the RNA/RBP network regulates their own quality and balance, suggesting convergence of pathomechanisms in microsatellite expansion disorders and RBP proteinopathies.

937W

Genetic analysis in pediatric patients with ataxia. *J. Lee[1], B. Lim[2], K. Kim[2], M. Seong[3], J. Chae[2].* 1) Gachon University Gil Medical Center, Incheon, South Korea; 2) Pediatric Clinical Neuroscience Center, Seoul National University Children's Hospital, Seoul National University College of Medicine, Seoul Korea; 3) Seoul National University College of Medicine, Seoul Korea.

Objectives Childhood-onset cerebellar ataxia is usually genetic conditions rather than disorders with acquired etiologies. However, there are few studies about genetic analysis for ataxia. The aim of this study is to investigate the genetic profiles of pediatric patients with ataxia in Korea. **Methods** Fifteen patients with cerebellar ataxia of suspicious genetic origin at the Seoul National University Children's Hospital were included. Some trinucleotide repeat disorders were excluded. Whole-exome sequencing was performed and candidate variants were validated with Sanger sequencing. Their medical records, including neuroimaging studies, were reviewed. **Results** The mean age at onset was 6.2 years (range, 1.5 to 16 years). All cases except one were sporadic. Most patients had cerebellar atrophy, although some had normal or nonspecific findings in cerebellum. Pathogenic variants in genes previously shown to cause ataxia were identified in 5 patients (5/15=33.3%). The results showed the genetic complexity: *SLC2A1*, *SYNE1*, *ITPR1*, *CYP7B1*, and *GFAP*. All the pathogenic variants identified in this study have never been reported. **Conclusion** To our knowledge, this is the first study to identify the genetic causes in pediatric patients with ataxia in Korea, which shows the genetic heterogeneity of child-onset cerebellar ataxia. To establish a molecular diagnosis of ataxia is very important due to genetic counseling and potential therapeutic interventions.

938T

A novel *PRRT2* pathogenic variant in a family with paroxysmal kinesigenic dyskinesia and benign familial infantile seizures. *J.G. Lu[1], J. Bishop[4], S. Cheyette[3], S. Guo[1], N. Sobreira[4], S.E. Brenner[1,2].* 1) University of California, San Francisco, CA; 2) University of California, Berkeley, CA; 3) Palo Alto Medical Foundation, Palo Alto, CA; 4) Johns Hopkins University, Baltimore, MD.

Paroxysmal Kinesigenic Dyskinesia (PKD) is a rare neurological disorder characterized by recurrent and brief attacks of dystonic movements without alteration of consciousness, which are often triggered by the initiation of voluntary movements. The molecular basis of PKD has recently been identified. Whole exome sequencing has revealed a cluster of pathogenic variants in *PRRT2*, a gene with a function in synaptic regulation that remains poorly understood. Further studies have confirmed its role in the disease. Here, we report the discovery of a novel *PRRT2* pathogenic variant inherited in an autosomal dominant pattern in a family with PKD and Benign Familial Infantile Seizures (BFIS). After targeted Sanger sequencing failed to uncover the existence of previously described *PRRT2* pathogenic variants, we carried out whole exome sequencing in the proband and her affected paternal grandfather. This led to the discovery of a novel *PRRT2* variant, p.A320V, altering an evolutionarily conserved Alanine at the amino acid position 320, located in the M2 trans-membrane domain. Sanger sequencing further confirmed the presence of this variant in four affected family members (paternal grandfather, father, brother, and proband) and its absence in two unaffected ones (paternal grandmother and mother). This newly found variant further reinforces the importance of *PRRT2* in PKD, BFIS, and possibly other movement disorders. Future functional studies using animal models and human pluripotent stem cell models will provide new insights into the role of PRRT2 and the significance of this variant in regulating neural development and/or function.

939F

Expanded-$(TGGAA)_n$-associated unconventional translation in spinocerebellar ataxia type 31. *N. Sato[1,2], T. Ishiguro[2], Y. Niimi[2], M. Higashi[2], T. Yokota[2], H. Mizusawa[2,3], K. Ishikawa[1,2].* 1) Center for Personalized Medicine for Healthy Aging, Tokyo Medical & Dental University. Tokyo, Japan; 2) Department of Neurology and Neurological Science, Graduate School, Tokyo Medical and Dental University. Tokyo, Japan; 3) National Center of Neurology and Psychiatry. Tokyo, Japan.

Background: Spinocerebellar ataxia type 31 (SCA31) is a dominantly inherited neurodegenerative disorder that manifests late-onset, slowly progressive cerebellar ataxia. The causative mutation for SCA31 is complex penta-nucleotide repeats including $(TGGAA)_n$ in an intronic region shared by two genes, *BEAN1* and *TK2*, which are transcribed in opposite directions. Like SCA31, some neurodegenerative disorders are caused by intronic expanded variable numbers of tandem repeats (VNTRs), such as myotonic dystrophy type 1 and 2, and amyotrophic lateral sclerosis caused by C9orf72 mutation. Recently, evidence has been accumulating that these disease-related expanded VNTRs are unconventionally translated into polypeptides, and this repeat-associated non-ATG (RAN) translation has been suspected to be part of the pathogenesis leading to neurodegeneration. The SCA31 mutation in the *BEAN1* direction is somewhat unique in this context as it has numerous ATG start condons, and followed by $(TAGAA)_n$ and $(TAAAA)_n$, both of which could serve as stop codons. **Purpose:** To elucidate if unconventional, expanded-VNTR-associated translation takes place in SCA31. **Materials and Methods:** Polyclonal antibodies against $(MEWNG)_3$ were developed by immunizing rabbits with the oligopeptide, part of polypeptides potentially translated from $(AUGGA)_n$ in the direction of *BEAN1*. These antibodies were first tested by western blot with recombinant GST-$(MEWNG)_n$. Then, immunohistochemistry was carried out on cerebellar specimens of two SCA31 patients and two control subjects. **Results:** The obtained in-house antibodies were proved to recognize $(MEWNG)_n$. Immunohistochemistry on the cerebellar sections showed aggregation-like staining in the cytoplasm and in the primary shafts of the dendrites of Purkinje cells in SCA31, but not in controls. The observed staining manifested various morphological appearances, from granular to rod-like shapes. **Conclusions:** Our immunohistochemistry analysis suggests that unconventional translation originating from seemingly "intronic" expanded $(TGGAA)_n$ may take place in SCA31. Further investigations are needed to clarify how and to what extent these polypeptides contribute to the pathogenesis of SCA31, and on potential translation from $(TTCCA)_n$ in the direction of *TK2*, which lacks the ATG codon.

940W

A novel mutation in eukaryotic elongation factor 2 kinase (eEF2K) decreases phosphorylation of eEF2 in a patient with degenerative ataxia.
K.L. Sund[1], J.E. Jones[3], J. Briggs[3], C.A. Prows[1], J. Glass[1], M. Sweeney[4], D. Gilbert[2], M. Sun[1], C. Prada[1], T.A. Burrow[5], J.D. Dinman[3]. 1) Department of Pediatrics, Division of Human Genetics, Cincinnati Children's Hospital Medical Center , Cincinnati, OH; 2) Department of Pediatrics, Division of Neurology, Cincinnati Children's Hospital Medical Center , Cincinnati, OH; 3) Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD; 4) Division of Neurology, University of Louisville, Louisville, KY; 5) Department of Pediatrics, Section of Genetics and Metabolism, University of Arkansas for Medical Sciences, Little Rock, AR.

Eukaryotic elongation factor 2 kinase (eEF2K) has been extensively studied as a potential therapeutic target for treatment of tumors and epilepsy. It is expressed in cerebral cortex ventricular zone during mouse brain development, transiently upregulated during synaptogenesis, and maintained in the adult brain. We performed whole exome sequencing (WES) on a 27 year old female with adult-onset ataxia, tremor, spasticity, declining cognition, and balance problems. Within 2 years, she was nonambulatory, incontinent, and unable to feed herself. MRI showed innumerable white matter loss lesions consistent with progressive multiple sclerosis. Trio WES revealed a de novo missense mutation in a highly conserved nucleotide of *eEF2K*. The mutation fell within a kinase domain, was uniformly called deleterious by computational algorithms, and was not found in control populations. In addition to autophosphorylation, the only known substrate for eEF2K phosphorylation is eEF2, which is mutated in Spinocerebellar Ataxia, type 26 (SCA 26). Functional studies of the equivalent *eEF2* mutation demonstrated an increased rate of -1 frameshifting indicating dysregulation of protein translation as a mechanism of disease in SCA 26. To determine the impact of the identified mutation in *eEF2K*, we transiently transfected HA-tagged WT eEF2K and eEF2K-Asp[244] into HEK293 cells. Despite persistent levels of eEF2 protein in both cultures, western blot analysis revealed decreased phosphorylation of eEF2 in the presence of eEF2K-Asp[244] compared to WT eEF2K, indicating a dominant negative inhibition of kinase function. Typically, eEF2K is believed to confer cell survival under acute nutrient depletion by blocking translation elongation. Tumor cells can exploit this pathway by inhibiting eEF2 and decreasing the demand for protein synthesis, thereby decreasing cell death. Inhibition of eEF2K in cancer cells leads to increased cell death. A common endpoint for SCAs is increased cell death in the Purkinje cells. Further research will determine whether loss of eEF2K function similarly results in increased cell death in Purkinje cells. Ongoing in vitro studies, more patients and animal models are needed to further implicate eEF2K in causation of this neurodegenerative ataxia that may be misdiagnosed as multiple sclerosis. Conclusive studies may also aid treatment for these patients, especially since eEF2K is inhibited by mTOR and mTOR inhibitors such as Sirolimus are currently used in clinic.

941T

Japan Consortium of Ataxias (J-CAT): A cloud -based national registry for degenerative ataxias providing comprehensive genetic diagnosis and prospective natural history researches.
Y. Takahashi[1], K. Ishikawa[2], Y. Ugawa[3], O. Onodera[4], J. Kira[5], S. Kuwabara[6], H. Sasaki[7], G. Sobue[8], H. Takashima[9], Y. Takiyama[10], A. Takeda[11], S. Tsuji[12], K. Nakashima[13], I. Miyai[14], K. Yoshida[15], H. Mizusawa[16]. 1) Dept Neurology, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan; 2) Department of Neurology and Neurological Science, and Predictive and Preventive Medicine, Tokyo Medical and Dental University; 3) Department of Neurology, School of Medicine, Fukushima Medical University; 4) Department of Neurology, Brain Research Institute, Niigata University; 5) Department of Neurology, Neurological Institute, Graduate School of Medical Sciences, Kyushu University; 6) Department of Neurology, Graduate School of Medicine, Chiba University; 7) Department of Neurology, Hokkaido University Graduate School of Medicine; 8) Department of Neurology, Nagoya University Graduate School of Medicine; 9) Department of Neurology and Geriatrics, Kagoshima University, Graduate School of Medical and Dental Sciences; 10) Department of Neurology, University of Yamanashi; 11) Department of Neurology, Sendai Nishitaga National Hospital; 12) Department of Neurology, Graduate School of Medicine, The University of Tokyo; 13) Division of Neurology, Department of Brain and Neurosciences, Faculty of Medicine, Tottori University; 14) Neurorehabilitation Research Institute, Morinomiya Hospital; 15) Division of Neurogenetics, Department of Brain Disease Research, Shinshu University School of Medicine; 16) National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan.

[Background] Spinocerebellar degeneration (SCD) is a group of heterogeneous disorders composed of hereditary and sporadic ataxias: the majority of hereditary ataxias are autosomal-dominant SCD and the remaining are autosomal-recessive SCD or X-linked SCD. Clinical manifestations including natural history vary greatly among SCDs with different molecular backgrounds, necessitating the accurate genetic diagnosis to establish disease-specific natural history and then disease modifying therapy. [Objective] We aim to develop a national registry for SCDs providing framework for comprehensive genetic diagnosis and prospective study of natural history. [Subjects and Methods] We developed a cloud-based registry system for ataxias, Japan Consortium of ATaxias (J-CAT), supported by Ministry of Health, Labor and Welfare, Japan and Japan Agency for Medical Research and Development (AMED). The system implemented a searchable encryption technology to ensure the security. We also established a multi-institutional network for comprehensive genetic diagnosis including PCR fragment analysis, direct nucleotide sequence analysis and whole exome sequencing analysis. The system allows us to conduct disease-specific prospective natural history researches according to established genetic causes. [Results] J-CAT allows patients themselves to register with informed consent and input clinical data with the assistance of physicians online, providing a platform for prospective natural history research. The registered patients are routinely subjected to mutational analysis including PCR-fragment analysis for triplet repeat diseases and SCA31, followed by systematic diagnostic procedure including whole exome analysis. The concept, registration procedure and contact information of J-CAT are readily available on the website (http://jcat.umin.ne.jp/). Eighty-two patients have been registered until June 6th . Genetic analysis has been conducted in 50 patients in whom the molecular diagnosis was established in 26 patients (52%) including 9 patients with MJD/SCA3, 8 with SCA31, 5 with SCA6, 2 with DRPLA, 1 with SCA2 and 1 with SCAR8. [Conclusion] J-CAT has been launched to facilitate genetic diagnosis, establish disease-specific natural history and identify genetic causes of SCD.

942F

Two patients with *PNKP* mutations presenting microcephaly, seizure, and oculomotor apraxia. *M. Taniguchi-Ikeda[1], N. Morisada[2], H. Inagaki[3], N. Okamoto[4], T. Toda[5], I. Morioka[1], H. Kurahashi[3], K. Iijima[1].* 1) Pediatrics, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan; 2) Clinical Genetics, Hyogo prefectural Childrens Hospital, Kobe, Hyogo, Japan; 3) Molecular Genetics, Fujita health University, Toyoake, Aichi, Japan; 4) Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Izumi, Osaka, Japan; 5) Neurology/Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe, Hyogo, Japan.

Microcephaly with early-onset, intractable seizures and developmental delay (MCSZ, OMIM #613402) is an autosomal recessive group of heterogeneous disorders, which can also be associated with other severe neurological defects. Mutations in polynucleotide kinase 3'-phosphatase (*PNKP)* have been suggested to cause MCSZ. Additionally, *PNKP* mutations also cause ataxia-oculomotor apraxia type 4 (AOA4) without symptoms of epilepsy or microcephaly. The reported AOA4 cases carried mutations in the kinase domain of *PNKP*. Here we identified three new mutations in two Japanese MCSZ patients who also exhibited AOA4 symptoms. All mutations resided within the kinase domain and both patients showed severe epilepsy and microcephaly with congenital anomalies. A 38-year old MCSZ patient showed apparent AOA symptoms with spinocerebellar degeneration. Another patient with MCSZ showed lissencephaly and frequent horizontal headshaking from age 1, suggestive of oculomotor apraxia. Kinase domain mutations in *PNKP* may manifest as a wide spectrum of overlapping phenotypes of MCSZ and AOA4. Therefore, we suggest that each MCSZ and AOA is not a distinct phenotype, but sequential phenotypes of *PNKP* mutations in kinase domain.

943W

Identification of novel *de novo* CHD8 variants associated with autism, language disability and overgrowth. *Y. An[1,2], L.N. Zhang[3], W.W. Liu[3], J. Wang[5], J.F. Gusella[2], Y.S. Du[3], Y.P. Shen[2,4,5].* 1) Institute of Biomedical Sciences and MOE Key Laboratory of Contemporary Anthropology, Fudan University, Shanghai, China; 2) Molecular Neurogenetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA., USA; 3) Shanghai Mental Health Center, Shanghai Jiaotong University School of Medicine, Shanghai, China; 4) Boston Children's Hospital, Harvard Medical School, Boston, MA., USA; 5) Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China.

Autism spectrum disorders (ASD) are childhood neurodevelopmental disorders that have a strong genetic basis. *CHD8,* a gene encoding a chromodomain helicase DNA-binding protein that plays an important role on chromatin remodeling has been implicated as a strong effect cause of ASD inactivated by translocation, deletion or point mutation. *Chd8* deficient mice display increased brain weight, craniofacial abnormalities and autistic-like behaviors, as well as defects in synaptic function, resembling the human phenotype. Recently several functional studies of reduced *CHD8* expression by RNAi suppression or CRISPR/Case9 mediated mutation in iPSC-derived neural progenitor cells (NPCs) have showed that *CHD8* regulates, directly or indirectly, a number of other ASD risk genes. To explore the role of *CHD8* in the Chinese population, we used target capture and next generation sequencing of *CHD8* in a total of 96 ASD subjects. We identified three patients with novel *de novo* variants: a nonsense change (p.Arg1188*), a splice site mutation (c.4818-1G>A) and a missense mutation (p.Tyr1168Asn). We also identified a second missense change (p.Lys2286Arg) in a fourth subject where inheritance could not be tested. Retrospective follow-up of the phenotypic manifestations of these ASD patients revealed macrocephaly detectable from the prenatal ultrasound test, dysmorphic features, overgrowth, gastrointestinal complaints and intellectual delay. More specific examination of published *CHD8*-associated phenotypes revealed repetitive behaviors involving the fingers, evident interest in communicating hindered by a significant expressive language defect, large testes and a tendency to overgrow during puberty. To assess potential genotype-phenotype correlations, we compiled all ASD-associated *CHD8* variants from literature and databases. Among 30 variants from patients, 20 are clear loss-of-function variants; the majority of these (15 of 20) are located in highly conserved functional domains of *CHD8*. In particular, individuals who carried *CHD8* mutations located in the helicase domain tended to have a more prominent and consistent overgrowth phenotype suggesting that disruption of this functional domain in *CHD8* might lead to a specific phenotype different from other *CHD8* mutations. Our study has expanded the phenotype spectrum of subjects with ASD due to *CHD8* mutation and added to the list of ASD-associated loss-of-function mutations in this gene. .

944T

Tmlhe and Bbox1 null mouse models of carnitine deficiency. *A. Ye[1], F. Jimenez-Rondan[1], Q. Sun[1], J. Ge[1], J. Heaney[1], D. Lanza[1], C. Burger[2], D. Jiang[2], J. Seavitt[1], M. Samuel[2], A. Beaudet[1]. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Huffington Center on Aging and Department of Molecular Physiology & Biophysics, Baylor College of Medicine, Houston, TX.*

Background: TMLHE deficiency is an X-linked, very common inborn error of carnitine biosynthesis affecting about 1 in 350 healthy males. It is reported to be a risk factor for autism with a penetrance of approximately 3%. We have hypothesized that brain carnitine deficiency can cause up to 10-20% of all autism apart from *TMLHE* deficiency in infants with a normal physical exam and absence of high penetrance Mendelian mutations (*BioEssays* in press). We wish to test whether brain carnitine deficiency in mice can cause any phenotypic features that might have parallels with autism. **Methods:** In both humans and mice, *TMLHE* encodes the first enzyme for carnitine biosynthesis, and *BBOX1* encodes the last enzyme in the pathway. *Tmlhe* null mice were generated using CRISPR methods to delete 10 bp of a coding exon while *Bbox1* null mice were generated through the Knockout Mouse Project (KOMP) with ES cell methods to delete an essential exon. *Bbox1* null mice were phenotyped as part of the KOMP project, and more limited phenotyping was performed on *Tmlhe* null mice. Metabolomic analysis was performed by Metabolon, Inc. **Results:** *Tmlhe* null mice are viable and fertile. *Bbox1* null mice born to -/+ x -/+ matings show 50% survival to weaning. In contrast, *Bbox1* null mice born to -/- x -/- matings all die shortly after birth with the difference likely related to placental transfer of carnitine and the carnitine content of breast milk in -/+ compared to -/- mothers. Plasma carnitine levels are reduced to 2% and 8% of wild-type levels in adult *Bbox1* and *Tmlhe* null mice respectively. Breast milk for null mothers showed 15% and 10.6% of wild-type levels of carnitine for *Bbox1* and *Tmlhe* respectively. Metabolomic analysis of *Bbox1* null mice revealed severely reduced levels of carnitine and many acyl-carnitines in plasma, liver, and brain and marked elevation of γ-butyrobetaine. We first evaluated the well-ordered retinal system. Immunohistochemical staining for microglia revealed increased distances between microglia cell somata and decreased total microglia numbers in *Bbox1* null mice. Remaining microglia were larger in size and exhibited increased process length. Other studies are in progress. **Conclusion:** These *Tmlhe* and *Bbox1* mouse models demonstrate severe carnitine deficiency in brain, liver, and plasma and should be helpful in testing whether brain carnitine deficiency can cause developmental or functional abnormalities that might parallel human autism.

945F

Severity of GABBR2 mutations determines neurological phenotypes ranging from Rett-like syndrome to epileptic encephalopathy. *Y. Yoo[1], J. Jung[2], Y. Lee[3], K. Miller[4], N. Shur[4], C. Clow[4], R. Ebel[5], S. DeBrosse[5], J.Y. Seong[3], H. Jung[2], J.-H. Chae[6], M. Choi[1,6]. 1) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, 03080, Republic of Korea; 2) Department of Anatomy, Brain Research Institute, and Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, 03722, Republic of Korea; 3) Graduate School of Medicine, Korea University, Seoul, 02841, Republic of Korea; 4) Albany Medical Center, Albany, NY 12208, USA; 5) UH Cleveland Medical Center, Center for Human Genetics, Cleveland, OH 44106, USA; 6) Department of Pediatrics, Seoul National University College of Medicine, Seoul National University Children's Hospital, Seoul, 03080, Republic of Korea.*

Rett syndrome (RTT) and epileptic encephalopathy (EE) are devastating neurodevelopmental disorders with distinct diagnostic criteria. However, highly heterogeneous and overlapping clinical features often allocate patients into the boundary of the two conditions, complicating accurate diagnosis and proper medical interventions. Although previous studies revealed genetic causes that lead to the two diseases, the specific molecular mechanism that allows an understanding of the pathogenesis and relationship of the two conditions remains to be uncovered. Here we show GABAB receptor R2 (*GABBR2*) as a genetic factor that determines RTT- or EE-like phenotype expression depending on the variant positions. By screening genetic factors from RTT-like patients without *MECP2* mutations, which account for ~90% of RTT patients, we identified a recurring *de novo* variant in *GABBR2* that reduces the receptor function from four patients (two Korean and two European origins), while different *GABBR2* variants from EE patients possess a more profound effect in reducing receptor activity when tested on a cell culture system and is more responsive to agonist rescue in an animal model. Finally, transcriptome-wide analysis of developing human brain tissues demonstrated *GABBR2* as a bridging factor that provides a link between RTT- and EE-associated genes, further supporting the notion that *GABBR2*-mediated GABA signaling is a crucial factor in determining the severity and nature of neurodevelopmental phenotypes.

946W

Characterization of a complex translocation causing 3q28ter duplication and 10q26.2ter deletion in a child with self-injurious behavior. *I.M. Adeshina[1,2], A.T. Joynt[1], A.L. Norris[2], J. Thorpe[2], J. Schmidt[2], A. Fatemi[2], L. Hagopian[2], J. Pevsner[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Neurology, Kennedy Krieger Institute, Baltimore, MD.

The identification of genetic factors that drive self-injurious behavior (SIB) is complicated by the heterogeneity of patient phenotypes. Here we report a nonverbal female adolescent with severe stereotypic movement disorder with SIB, and complex cytogenetics. The proband's full phenotype included global developmental delay, congenital scoliosis, dysthymia, and anhedonia. Previous cytogenetics revealed a complex translocation that was also present in the mother: 46,XX,t(3;14;10)(q27;q22;q26.1).ish t(3;14;10)(D3S4560-,D10S2490+;D14S308-,D3S4560+;D10S2490-,D14S308+). To clarify the nature of the abnormality and to identify genes disrupted by the translocations, we subjected DNA from the proband and her parents to whole genome sequencing (WGS, Illumina HiSeq X), and whole genome optical mapping with *de novo* assembly (Bionano Saphyr). Translocations were mapped and identified from structural variant (SV) and copy number variant (CNV) analyses of WGS and Bionano data, and confirmed by Sanger sequencing. Finally, to correlate phenotype and genotype in the proband, we systematically reviewed related case reports in the literature. Genomic analysis revealed that the complex translocations were balanced in the mother, but unbalanced in the proband. While the proband did harbor t(10;14) and t(14;3), instead of inheriting a maternal t(3;10), she inherited the normal maternal copy of chromosome 3, resulting in a 10q26.2ter deletion and 3q28ter duplication. Genes in one or both of these regions likely underlie the proband's phenotype, spanning neurodevelopmental, musculoskeletal, and psychiatric domains. Several dozen cases of 10q26 deletion syndrome (MIM #609625) have been reported. However, there are only sporadic reports of duplications involving 3q28ter. Phenotypes reported in 10q26.2 syndrome patients closely mirrored those in the proband, notably aggressive behavior with limited attention span and violent mood swings, alternating between being provocative or destructive and affectionate. Using complementary methods of WGS and Bionano, we characterized complex germline translocations present in an apparently normal female, and the incomplete transmission of these translocations to her child. Consequently, the child harbored 10q28ter deletion and 3q26.2ter duplication. Review of the literature implicates 10q26.2 genes, while there is insufficient evidence to suggest or refute contribution from 3q28 genes.

947T

The novel aldehyde trap ADX-102 reduces accumulations of GHB and GABA in brain tissue from succinic semialdehyde dehydrogenase-deficient mice. *S.G. Macdonald[1], V. Cullen[1], A. Halilovic[1], L. Case[2], T. Brady[1], S.L. Young[1].* 1) Aldeyra Therapeutics, Inc., Lexington, MA; 2) Jackson Laboratory, Bar Harbor, ME.

ADX-102 is a novel aldehyde-sequestering agent that forms irreversible adducts with aldehydes, including succinic semialdehyde (SSA). Succinic semialdehyde dehydrogenase deficiency (SSADHD; MIM 271980) is an inherited disorder associated with intellectual disability, seizures, ataxia, sleep disorders and anxiety. A loss-of-function mutation in succinic semialdehyde dehydrogenase (SSADH; ALDH5A1), which metabolizes SSA to succinate, results in accumulation of gamma-hydroxybutyrate (GHB) and 4-amino-butyrate (GABA). ADX-102, which covalently sequesters aldehydes on a 1:1 stoichiometric basis, may deplete SSA and result in reduced levels of GHB and GABA in SSADHD. Thus, we tested ADX-102 in a mouse model of SSADHD. Wild type (wt) and B6.129-Aldh5a1[tm1Kmg]/J (null) mice were given 50 mg/kg ADX-102 or vehicle intraperitoneally once or twice daily from birth. After 3.5, 7, or 15 days, plasma and brain were analyzed for GHB and GABA. At birth, brain and plasma GHB levels (~400 μM) were 173- and 81-fold higher, respectively, in null mice than wt mice; brain and plasma GABA levels were 2.8- and 3.8-fold higher, respectively, in null mice than wt mice. Brain GHB and GABA accumulated through day 15 of life, with a sharp increase in GHB seen between days 7 and 15. The phenotype was severe, with some null mice dying by Day 11. ADX-102 had little effect on accumulations of GHB and GABA in plasma or brain. The results are consistent with a parallel pharmacokinetic study in wt mice indicating that there was ~10-fold more GHB in plasma and 12-fold more GHB in brain of null mice at birth, relative to ADX-102 levels after one day of dosing; GABA in plasma and brain of null mice at birth was 1.2- and 122-fold higher, respectively, than ADX-102 levels. The data suggest that ADX-102:SSA adduct formation in plasma may result in too little ADX-102 reaching the brain to affect GHB and GABA levels in null mice. To maximize drug exposure, brain slices from null and wt mice treated for 3.5 days *in vivo*, were incubated *ex vivo* for 24 hours, in the presence or absence of 422 μM ADX-102. ADX-102 statistically reduced GHB and GABA levels in brain tissue from null mice and in the incubation medium. Reductions ranged from 32% to 40% for GHB and 24% to 63% for GABA ($p < 0.0001$ to $p < 0.05$). The data show that ADX-102 can significantly reduce the formation of GHB and GABA in null mouse brain. ADX-102, and aldehyde-trapping in general, represent a novel approach to the treatment of SSADHD.

948F

Identification of mutations in patients from southern Italy with amyotrophic lateral sclerosis using multigene panel testing. *G. Annesi[1], M. Gagliardi[1], G. Iannello[1], R. Procopio[1], A. Quattrone[1,2].* 1) National Research Council, Catanzaro, catanzaro, Italy; 2) 2Institute of Neurology, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy.

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that primarily affects motor neurons, resulting in progressive paralysis and death . The majority of ALS cases are sporadic, meaning that they occur with no family history of the disease (sALS). The remaining 5–10% of cases are familial (fALS), where the disease is inherited in a Mendelian, generally dominant, fashion within a family.in recent years several genes have been linked to both fALS and sALS. In this study, we performed multigene panel testing to identify mutations in ALS-related genes We rolled 25 consecutive patients with sporadic ALS.Exome sequencing was performed using the Ion AmpliSeq Exome kit on Ion Torrent Proton Sequencer. In total, 21 genes were ultimately selected for the targeted sequencing panel. Several bioinformatics analyses were performed to identify the functional and structural significance of missense mutations or splice-site variants observed in patients. Based on a comparison with the dbSNP and HGMD databases, we identified known pathogenic variants of the ALS-related genes in 25 ALS patients. We searched dbSNP and HGMD to identify known pathogenic variants that had been previously reported to cause ALS. We also identified novel nonsynonymous variants that were classified as "Probably Damaging" by Poly- Phen2 and Mutation Taster which yielded a list of novel potentially pathogenic variants. All variants screened were validated by Sanger sequencing Among the potentially pathogenic variants, we focused on nonsynonymous variants that did not appear in genome databases and were predicted to be pathogenic by an in silico analysis.

949W

Evaluation of pathogenic non-coding variants within whole genome data using encephalopathies as a model. *D. Misceo[1], D. Sanjeewani[1], A. Holmgren[1], M.D. Vigeland[1], P. Strømme[2,3], M. van de Vorst[4], C. Gilissen[4], S. Rayner[1], E. Frengen[1].* 1) Oslo University Hospital and University of Oslo, Oslo, Norway, Norway; 2) Faculty of Medicine, University of Oslo, Oslo, Norway; 3) Division of Pediatric and Adolescent Medicine, Oslo University Hospital, Oslo, Norway; 4) Department of Human Genetics Radboud University Medical Center, Nijmegen, The Netherlands.

The roles of protein coding DNA variants in human diseases have been extensively studied, while variants in non-coding (nc) DNA have received less attention. In this work we report our progress on the development of a pipeline to investigate nc genomic regions and identifying their association with disease. The project focuses on a patient cohort with progressive encephalopathies (PE). PE encompasses a broad group of diseases with childhood onset that are mostly metabolic or neurodegenerative in nature. The cohort comprises 71 families selected by a single child neurologist among patients that he has seen in the past 25 years. The patients in these families were negative for common lysosomal, mitochondrial, peroxisomal, and relevant metabolic and degenerative diseases, with no history of infectious diseases or perinatal hypoxia. Trio WES and subsequently WGS were performed (average 10X depth 97% and 20X depth 92% in WGS data). Analysis of protein coding DNA revealed a pathogenic variant for 38 families, but not in the remaining 33. We are therefore initiating the investigation of the nc regions of the genome in these 33 families. Because of the magnitude of the nc regions, we focus on regions related to regulation of gene expression (for examples promoters and enhancers) and miRNA. Our primary aim is to develop a pipeline to systematically and comprehensively annotate and prioritize putative pathogenic nc DNA variants in Mendelian diseases using WGS data. A key component of this work is our effort to standardize and integrate annotation to generate a comprehensive reference source for nc features. This can be used to interpret WGS data, to identify signature variants, which can then be used as input to (for example) statistical learning packages to predict associations with phenotype. The nc DNA variants prioritized as putatively pathogenic can then be studied functionally. In this this step we use fibroblasts from patients and/or commercial cell lines. Variants with documented *in vitro* effect will be studied *in vivo*, providing novel knowledge on the function of nc DNA in human biology and pathology. We present our initial efforts on this work, related to both the pipeline (methodology, annotation) and some preliminary finding based on investigation of these regions.

950T

Novel mutations in *CLN6* cause late-infantile neuronal ceroid lipofuscinosis in two unrelated patients. *B. Behnam[1], J. Chin[1], M. Davids[1], P. Sharma[1], C. Wang[1], W. Zein[3], G.A. Gretchen[1,2], C. Toro[1,2], X. Chepa Lotrea[1], D.R. Adams[1,2], C.J. Tifft[1,2], W.A. Gahl[1,2], M.C. Malicdan[1,2].* 1) UDP Translational Lab, NHGRI, National Institute of Health (NIH), Bethesda, MD; 2) 1.Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) National Eye Institute (NEI), National Institute of Health (NIH), Bethesda, MD.

Ceroid-lipofuscinosis neuronal 6 (CLN6) is a transmembrane protein located in the endoplasmic reticulum and involved in lysosomal lumen acidification. Mutations in CLN6 are implicated in late-infantile neuronal ceroid lipofuscinosis (LINCL), or Batten disease, and in adult-onset Kufs disease type A. Here, we report two patients from unrelated families evaluated by the NIH Undiagnosed Diseases Program at age 9 (Patient 1) and 11 (Patient 2), and diagnosed with LINCL. Both patients demonstrated some of the typical phenotypic features, namely motor impairment, seizures and developmental regression, though ophthalmologic examination did not reveal any visual impairment. Whole-exome and Sanger sequencing identified novel compound heterozygous mutations in *CLN6*. Patient 1 has compound heterozygous c.218-220dupG-GT; p.Trp73dup and c.296A>G; p.Lys99Arg mutations, while Patient 2 has a homozygous c.723G>T; p. Met241Ile mutation. Skin biopsy electron micrographs show large numbers of small membrane-bound vesicles, in addition to lipofuscin deposits. Expression analysis in dermal fibroblasts showed no change in *CLN6* mRNA levels from both patients when compared to control. Immunofluorescence studies of patients' cells showed enlarged lysosomes and increase in LysoTracker Red staining suggesting that the acidification of the lysosomes is increased. Western blot analysis showed increased amounts of lipidated autophagosome-specific protein (LC3-II), indicating impaired basal levels of autophagy. To date, among >200 reported cases of *CLN6*-related Batten disease, nine patients, including those in this report, do not manifest or developed delayed onset of visual impairment. Although the precise function of CLN6 is not well understood, its role in regulation of lysosomal acidity may be important in enzymatic degradation of post-translationally modified proteins. Lysosomal accumulation and aggregation of undegraded proteins is cytotoxic, particularly to neurons. The absence of visual impairment in our patients may indicate some residual expression of functional protein. This study supports the role of CLN6 in lysosomal function, and highlights the importance of considering a *CLN6* diagnosis even in patients with seemingly normal vision.

951F

17p13.3 microdeletions between *YWHAE* and *LIS1* (*PAFAH1B1*) cause a unique leukoencephalopathy. *L.T. Emrick[1], J.A. Rosenfeld[1], L.C. Burrage[1], M. Jain[1], J.W. Belmont[1], C.A. Bacino[1], L. Immken[2], B. Lee[1], S.R. Lalani[1], Members of the UDN.* 1) Baylor College of Medicine, Houston, TX; 2) Dell Childrens Hospital, Austin, TX.

17p13.3 deletions cause Miller-Dieker syndrome, when the deletion includes the *YWHAE* and *LIS1* (*PAFAH1B1*) genes. Smaller 17p13.3 deletions may cause isolated lissencephaly sequence when only involving *LIS1*. In larger deletions, it is thought that haploinsufficiency of *YWHAE* and *CRK* contributes to the more severe Miller-Dieker phenotypes, and smaller deletions including *YWHAE* and *CRK* but sparing *LIS1* have been described in individuals with growth restriction, cognitive impairment, dysmorphic features, and various brain abnormalities. We have identified four individuals from three families with 17p13.3 deletions between and not including *LIS1* and *YWHAE/CRK*, who all have multi-focal white matter lesions and hypermobile joints. Cognition is normal, although some behavioral concerns are present. A fifth individual with a deletion of the distal portion of this region was identified with a normal MRI, suggesting that the proximal genes in the region may be critical for the brain findings in these individuals, although reduced penetrance cannot be ruled out. A combination of our patients and those in the literature with white matter changes with deletions in this chromosomal region narrows the overlapping region for this phenotype to ~300kb, including 10 RefSeq genes. Previous literature has hypothesized some dysmorphic features and white matter changes related to *YWHAE*, however our patients do not have this gene deleted. Our cohort contributes evidence to the presence of multiple genes within 17p13.3 required for proper brain formation.

952W

Large-scale systematic analysis of recessive neurodevelopmental disorders in consanguineous families. *A. Gregor[1], R.D. George[2], B. Saunders[2], J.G. Gleeson[1,2].* 1) Laboratory of Pediatric Brain Disease, The Rockefeller University, New York, NY; 2) Howard Hughes Medical Institute, Rady Children's Institute of Genomic Medicine, University of California, San Diego, CA.

Consanguineous marriages are common in about one-fifth of the world's population. In children from consanguineous marriages, homozygous mutations are common and lead to a doubling of birth defects. To analyze the contribution of recessive mutations to developmental brain diseases, we assembled a cohort of more than 2,200 consanguineous families from the Greater Middle East with various neurodevelopmental brain disorders with suspected autosomal recessive inheritance. Each case was phenotyped in detail, which included brain MRI, metabolic testing, electrophysiology and detailed history/physical examination. We performed whole-exome sequencing across these families to systematically discover putatively causal variants in both known and novel disease genes. All samples were processed jointly to identify rare, potentially damaging variants that followed expected recessive inheritance patterns. We identified likely deleterious variants in previously published disease genes in about half of the families. These mutations affected more than 400 genes, highlighting the extreme genetic heterogeneity of neurodevelopmental disorders. Roughly 60 of these genes have previously been implicated in only single cases. These previously published cases had overlapping phenotypes with our patients, but our patients extended the phenotype in specific ways. About half of the identified mutations were predicted protein truncations, and half were predicted missense variants affecting highly conserved amino acids. In line with a recessive mode of inheritance, mutated genes were enriched for genes affecting metabolic processes ($P = 9.8 \times 10^{-6}$). Newborn metabolic screening is not routinely offered at our sites of recruitment, and thus our genetic diagnosis led to changes in diet or treatment for many of these patients. Work is ongoing to identify recurrently mutated novel genes within our cohort and to identify further factors that may contribute to the likelihood of solving cases from this cohort. .

953T

Genetic, clinical, and imaging study of Pelizaeus-Merzbacher disease using the Integrative Brain Imaging Support System (IBISS). *K. Inoue[1], K. Sumida[2], J. Takanashi[3], H. Matsuda[4], M. Sasaki[5], N. Sato[2].* 1) Dept MR & BD Res, Natl Inst Neurosci, NCNP, Kodaira, Tokyo, Japan; 2) Department of Radiology, National Center Hospital of Neurology and Psychiatry, Kodaira, Tokyo, Japan; 3) Pediatrics, Tokyo Wemen's Medical University, Yachiyo Medical Center, Yachiyo, Chiba, Japan; 4) Integrative Brain Imaging Center, NCNP, Kodaira, Tokyo, Japan; 5) Pediatrics, National Center Hospital of Neurology and Psychiatry, Kodaira, Tokyo, Japan.

Background: Pelizaeus-Merzbacher disease (PMD) represents the most common disorder of hypomyelinating leukodystrophies. PMD is caused by mutations in *PLP1*, which encodes a major myelin membrane protein in the CNS. Amino acid substitutions, genomic duplications, and null mutations of *PLP1* cause different severity of the disease through distinct molecular mechanisms. Magnetic resonance image (MRI) findings are the most critical diagnostic tool for hypomyelinating leukodystrophies including PMD. However, correlation among genotype, clinical severity and MRI findings has not been thoroughly studied in a large cohort of PMD. Methods: We established an MR imaging database for hypomyelinating leukodystrophies, termed Integrative Brain Imaging Support System (IBISS). We selected 19 PMD patients from IBISS and analyzed their genetic, clinical, and imaging data. Results: We identified 10 patients with *PLP1* duplication and 9 patients with point mutations (including 1 with in-frame deletion and 1 with splicing mutation). The clinical stage tended to be more severe when the whole brainstem, or corticospinal tract in the internal capsule showed abnormally high intensity on T2WI. Diffuse T2-high signal of brainstem was observed only in the patients with *PLP1* point mutation. Myelination age "before birth" on T1WI was correlated with the clinically severe phenotypes. On the other hand, eight patients whose myelination ages were > 4 months on T1WI were associated with mild clinical phenotypes. Advanced myelination was observed in 3 of the 7 followed-up patients. Four patients had atrophy of the cerebellum, and 17 patients had atrophy of the corpus callosum. Discussion: We have demonstrated a wide variety of imaging findings of PMD using a imaging DB, IBISS. Signal intensity of brainstem and corticospinal tract of internal capsule would be the points to presume clinical severity in PMD patients. The spectrum of MRI findings should be kept in mind to diagnose PMD and to differentiate from other demyelinating leukodystrophies.

954F

Rare *SOX30* variants in juvenile myoclonic epilepsy. *S. Jaishankar[1], R. Ratnapriya[1], S. Nayak[1], R. Reddy[1], M. Sharma[1], R. Ghugari[1], A. Kapoor[1], S. Shankar[3], S. Sinha[4], P. Satishchandra[4], A. Anand[1,2].* 1) Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore 560 064, India; 2) Neuroscience Unit, Jawaharlal Nehru Centre for Advanced Scientific Research (JNCASR), Bangalore 560 064, India; 3) Department of Neuropathology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore 560029, India; 4) Department of Neurology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore 560029, India.

   Juvenile myoclonic epilepsy (JME) is the most common subtype of genetic generalized epilepsy syndrome characterized by frequent early morning seizures beginning during early adolescence. The genetics of JME is complex and not completely understood. To date, 29 genetic loci are known for the disorder. Here, we present our work aimed at identifying the causative gene in a family segregating JME in an autosomal dominant manner. Genome wide linkage analysis helped us to identify a critical region of about 15 megabases between markers D5S2012 and D5S2075 (two-point lod score 3.66 at theta = 0 for D5S415 at chromosome *5q34*). We carried out whole genome sequencing of two members (unaffected father and affected daughter) of the family and examined genomic sequences in the critical interval. The *5q34* JME locus consists of a cluster of GABA genes (*GABRG2*, *GABRA1*, *GABRA6* and *GABRB2*) which are strong candidates for JME. We analyzed exonic, intronic and promoter sequences of these four genes by Sanger sequencing and found no potentially functional variant. In the analysis of whole genome sequence data for the region, 5673 sequence variants were obtained. These variants were scrutinized based on their zygosity, allele frequencies among controls and segregation with the clinical phenotype in the family. We detected four variants: c.-57 G>A in the 5' UTR of *SOX30*; and c.*1437 T>C, c.*1058 C>T and c.*322 G>A in the 3' UTRs of *EBF1*, *CNOT8* and *ZBED8*, respectively. We carried out mammalian cell based luciferase reporter assays for the 5' UTR and 3' UTR variants. The c.-57 G>A *SOX30* variant exhibited significantly altered reporter activity while the other three variants did not produce a significant difference. Further, we examined complete transcript structure of *SOX30* in 480 JME patients and 480 control samples. In this search, 15 additional rare coding variants present almost exclusively in the patients were obtained. Functional correlates of these variants were studied in cultured mammalian cells. Immunocytochemistry indicated no changes in the localization of the proteins carrying the variants, while reporter luciferase assay indicated that a sub-set of these variants had substantially altered their ability to regulate the reporter gene. Our observations provide prima facie evidence of the role of *SOX30* for JME.

955W

AOH-mediated recessive mutation burden can result in blended phenotypes. *E. Karaca[1], J. Punetha[1], J.E. Posey[1], Z. Coban Akdemir[1], X. Song[1], M.M. Khayat[1], T. Harel[1,2], S.N. Jhangiani[3], O. Ozalp Yuregir[4], S. Bozdogan[5], G. Yesil[6], S. Isikay[7], Y. Bayram[1], D. Pehlivan[1], J.M. Fatih[1], D.M. Muzny[3], R.A. Gibbs[3], J.R. Lupski[1,8,9].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 4) Genetic Diagnosis Center, University of Health Sciences, Numune Training and Research Hospital, Adana 01370, Turkey; 5) Department of Medical Genetics, Cukurova University Faculty of Medicine, Adana 01330, Turkey; 6) Department of Medical Genetics, Bezmialem University, Istanbul 34093, Turkey; 7) Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep 27410, Turkey; 8) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 9) Texas Children's Hospital, Houston, TX, USA.

   Phenotypic expression can be impacted by variation at more than one locus, complicating the molecular diagnosis. In some cases, phenotypic expansion observed in association with a known Mendelian disease gene may also be a hallmark of a blended phenotype resulting from multilocus variation. We hypothesized that such multi-locus disorders would be enhanced in situations that favored a high burden of 'absence of heterozygosity' (AOH), as observed in families with high levels of consanguinity. Whole exome sequencing (WES) was applied in 197 families with neurodevelopmental phenotypes, of which the results of initial exome variant analyses have been reported for 128. Approximately 80% of families reported consanguinity, suggesting AOH may contribute to molecular diagnoses. We calculated B-allele frequency from exome variant data to detect regions of AOH for each case. Analysis of the first 128 families resulted in identification of molecular diagnoses for 85.2% (109/128) cases, of which multilocus variation contributing to the observed clinical phenotype was identified in 11% (12/109) of cases. AOH contributed to the multilocus variation in 7 of these 12 families, including one family for which 3 distinct loci contributing to the phenotype were located within the same 44.8 Mb region of AOH on chromosome 1. Analyses of the additional 69 cases are ongoing and have yielded at least 4 more families for which multilocus variation within regions of AOH potentially contributes to a blended phenotype. Our preliminary data suggest that the portion of these diagnosed cases that have apparent multilocus effects is both enriched for AOH, and is potentially greater than the approximately 5% observed in mixed, unascertained populations with no suspectedly high levels of consanguinity (Posey et al., 2017). The observation of AOH may lead to a prediction of more likely involvement of additional loci, and variability of the size and location of AOH regions between affected siblings can lead to differences in recessive disease burden, offering a rich opportunity to dissect genotype-phenotype relationships. Comprehensive implementation of WES has thus enabled genomic medicine to move from individual study of single gene Mendelian disorders to a more comprehensive understanding of the interplay of genetic variation at more than one locus and resulting phenotypic expression; i.e. from Mendelian genetics to Mendelian genomics.

956T

Characterizing genetic causes of neurodevelopmental disorders with brain malformations in a predominantly Turkish cohort. *J. Punetha[1], E. Karaca[1], J.E. Posey[1], Z. Coban Akdemir[1], X. Song[1], M.M. Khayat[1], T. Harel[1,2], S.N. Jhangiani[3], Y. Bayram[1], D. Pehlivan[1], J.M. Fatih[1], D. Muzny[3], R.A. Gibbs[3], J.R. Lupski[1,3,4,5], Baylor-Hopkins Center for Mendelian Genomics.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 5) Texas Children's Hospital, Houston, TX, USA.

Neurodevelopmental disorders (NDDs) show a wide range of genetic and clinical heterogeneity making their genetic diagnoses challenging and highlighting the complexity of the human nervous system developmental processes. Clinical heterogeneity observed in NDDs is generally categorized into either structural or functional brain deficits, requiring a thorough analysis of phenotype, brain imaging, and molecular data. In this study, we have performed whole exome sequencing (WES) for molecular diagnoses of approximately 197 families with NDDs (mostly consanguineous) associated with structural brain malformations. Out of these, 128 families have been previously reported (Karaca *et al*., Neuron 2015) with a molecular diagnosis rate of ~85% which included known single nucleotide variants (SNVs) and copy number variants (CNVs) in known disease genes, novel disease genes, and potential candidate genes. Potential disease-causing variants were identified in 41 candidate disease genes in 46 families. Here, we discuss the fate of candidate genes that have since been confirmed as pathogenic through ascertainment of additional families with similar phenotypes, supporting functional data, and/or reports in the literature: *PRUNE, VARS, SMARCA1, SLC18A2, OGDHL, CPLX1,* and *ASTN1*. We include follow-up studies of these potential candidate genes and an update on their pathogenicity status. In addition, we present new findings from WES analyses of an additional 69 families with brain malformations that have revealed novel candidate genes as well as novel variants in known disease genes. Further studies including segregation and functional studies are being carried out in order to determine pathogenicity for these candidates. As we delve deeper into the molecular basis of neurodevelopmental disorders in a large cohort with structural brain malformations, we aim to maximize our understanding of different genetic mechanisms that underlie the complexity of development in the human central nervous system.

957F

c.105C>A [p.(Tyr35Ter)] variant in *AIMP2* causes microcephaly, intellectual disability, seizures and spastic quadriparesis. *A. Shukla[1], A.D. Bhowmik[2], M. Hebbar[1], K.V. Rajagopal[3], K.M. Girisha[1], N. Gupta[4], A. Dalal[2].* 1) Medical Genetics, Kasturba Medical College, Manipal, Udupi, Karnataka, India; 2) Diagnostics Division, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 3) Department of Radiodiagnosis and Imaging, Kasturba Medical College, Manipal University, India; 4) Division of Genetics, Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India.

We ascertained two consanguineous families with two affected children each with microcephaly, refractory seizures, intellectual disability and spastic quadriparesis. Brain MR imaging showed atrophy of cerebrum, cerebellum and spinal cord, prominent cisterna magna, symmetric T2 hypointensities in the bilateral basal ganglia and thinning of corpus callosum. Whole exome sequencing of three affected individuals revealed c.105C>A [p.(Tyr35Ter)] variant in *AIMP2* as the likely cause of their condition. In human and other mammalian cells, a unique large tRNA multi-synthetase complex organizes 9 cytoplasmic aminoacyl-tRNA synthetases consisting of bifunctional, as well as the monospecific tRNA synthetases and 3 non-enzyme factors, namely p43 (*AIMP1*), p38 (*AIMP2*) and p18 (*AIMP3*). The p38 protein is a key component of the multi-ARS complex and is crucial for the assembly of the complex. *In vitro* and animal studies have shown that *AIMP2* may be a tumor suppressor, however in human cancers the expression status at the tissue level is not well understood. Various components of the ARS complex are now known to cause encephalopathies, peripheral neuropathies and other sensorineural disorders. Among them, mutations in 10 cytosolic ARSs are associated with Charcot–Marie–Tooth and related neuropathies and 14 mitochondrial ARSs are associated with severe leukoencephalopathies. *AIMP1* has been associated with hypomyelinating leukodystrophy-3 characterized by progressive neurodegeneration, microcephaly, spasticity, coarse facies, progressive contractures, generalized brain atrophy and early death. The phenotype of our patients shares marked similarity with that of mutations in closely related gene, *AIMP1*. We hereby report the first human disease associated with deleterious mutations in *AIMP2*.

958W

Expanding the genetic spectrum in myoclonic astatic epilepsy. *S. Tang[1,2], L. Addis[1], A. Smith[1], S. Topp[1], D. Mei[3], A. Parker[4], S. Agrawal[5], E. Hughes[2], K. Lascelles[2], R. Williams[2], P. Fallon[6], R. Robinson[7], J.H. Cross[7], T. Hedderly[2], C. Eltze[7], C. Ferrie[8], T. Kerr[9], A. Desurkar[9], N. Hussain[10], M. Kinali[11], G. Vassallo[12], W. Whitehouse[13], S. Goyal[2], M. Absoud[2], C. Marini[3], R. Guerrini[3], M.A. Simpson[1], D.K. Pal[1,2], Euroepinomics RES consortium.* 1) King's College London, London, London, United Kingdom; 2) Kings Health Partners, London, United Kingdom; 3) Anna Meyer Children's Hospital, Florence, Italy; 4) Addenbrookes Hospital, Cambridge, United Kingdom; 5) Birmingham Childrens Hospital, United Kingom; 6) St Georges Hospital, London, United Kingdom; 7) Great Ormond Street Hospital, London, United Kingdom; 8) Leeds University Hospital, United Kingdom; 9) Sheffield University Hospital, United Kingdom; 10) Leicester University Hospital, United Kingdom; 11) Chelsea and Westminister Hospital, United Kingdom; 12) Manchester Children's Hospital, United Kingdom; 13) Nottingham University Hospital, United Kingdom.

Myoclonic astatic epilepsy (MAE) is a rare childhood onset epilepsy in which the spectrum of neurodevelopmental impairment and genetic aetiology are largely unknown. We therefore deeply phenotyped MAE patients and performed exome analysis filtered on curated epilepsy, intellectual disability, autism and developmental disorder gene sets to identify genetic aetiologies. Variants were considered pathogenic if they were novel in Gnomad, *de novo*, correlated with gene function and correlated with previously associated phenotypes.　We assembled the largest MAE cohort (n=109, 67% male) to date. A family history of epilepsy was reported in 38% and febrile seizures in 9%. The epilepsy phenotype is similar to previously published cohorts.　We identified likely pathogenic or candidate variants in 18 (16.5%) of 109 cases. This compromised known genes associated with MAE: *CHD2* n=2, *KCNA2* n=1, *KIAA2022* n=2, *SCN1A* n=1, *SCN2A* n=1, *SLC6A1* n=1, *STX1B* n=1, *SYNGAP1* n=3; epilepsy associated genes novel for MAE: *KCNB1* n=1, *MECP2* n=1, *KCNH5* n=1, and three new candidate genes: *ASH1L* n=1, *CHD4* n=1 and *SMARCA2* n=1. The major genetic pathways involved are ion channels, synaptic transmission, transcription regulation and chromatin modification. We describe phenotype-genotype correlations and recognised that patients with identified monogenic aetiology tended to have a severe neurodevelopmental phenotype with intellectual disability present in 17/18, autism in 8/18 and attention deficit hyperactive disorder in 6/18 . These findings reveal the shared genetic aetiology and pathways of epilepsy and neurodevelopmental disorders and the extent of genetic heterogeneity of MAE and highlight the promise of precision medicine for these patients. .

959T

Exome sequencing identifies a novel *FBXO38* variant inherited from a mosaic mother to cause distal hereditary motor neuronopathy Type IID with distinct features. *S.A. Ugur Iseri[1], N.H. Akcakaya[1,2], B. Salman[1], Z. Yapici[3].* 1) Istanbul University, Aziz Sancar Institute of Experimental Medicine, Department of Genetics, Istanbul, Turkey; 2) Turkey Spastic Children Foundation, Istanbul, Turkey; 3) Istanbul University, Istanbul Medical Faculty, Department of Neurology, Division of Child Neu-rology, Istanbul, Turkey.

Introduction: The *FBXO38* gene encodes a large protein that contains an F-box domain, which acts as a transcriptional co-activator of the Krueppel-like factor 7. FBXO38 may also be participated in protein ubiquitination. Heterozygous *FBXO38* variations have been implicated in distal hereditary motor neuronopathy type IID (HMN2D) with autosomal dominant inheritance only in two unrelated families. The disease starts with the onset of walking difficulty in the second to fourth decade of life. **Methods and Results:** Two siblings; a 20-year-old male and an 11-year-old female have been referred to our clinic due to walking difficulty. The major neurological symptoms at the age of evaluation were severe atrophy and spasticity predominantly in the lower extremities. EMG studies showed sensory motor polyneuropathy with axonal involvement. Sural nerve biopsy was indicative of hypo/demyelination and axonal loss. Corticol-subcortical atrophy was detected at late-term brain MRI. The parents mentioned breath-holding spells in infancy for the affected sibling pair. There was fatigue and tachycardia with effort starting from the age of 4 years in both children. After the age of 6, the siblings developed progressive weakness, scoliosis and loss of vision. Painful muscle spasms and rapid aggravation were observed only in the male patient. DNA samples of both patients were subjected to whole exome sequencing (WES) to reveal the genetic defect underlying these complex neurologic findings. The data was filtered for both recessive and dominant inheritance patterns, which revealed a shared and novel heterozygous *FBXO38* variation (NM_030793.4: c.173G>A; p.Arg58Gln) inherited form the asymptomatic mosaic mother. Various *in silico* tools modeled a damaging effect for the amino acid substitution. **Discussion:** WES in two affected sib-pairs followed by data analysis with various inheritance models and segregation studies have led us to diagnose the condition as HMN2D in two affected siblings. We have thus accordingly expanded the clinical spectrum associated with *FBXO38* variations and presented a family in which there is juvenile onset unlike the previously reported families in which the age of onset varies between 13 to 48 years of age. We have also managed to detect the mosaic pattern in the unaffected mother, which in turn demonstrates that the etiology of many unresolved neurological disorders might be explained by pathogenic mosaic mutations.

960F

Aspartate supplementation for aspartate-glutamate carrier isoform 1 deficiency. *S. Yano[1], A. Partikian[2], S. Bluml[3], K. Moseley[1], Y. Watanabe[4], H. Saitsu[5], N. Matsumoto[5].* 1) Pediatrics/Gen Div, 1G24, LAC+USC Med Ctr, USC, Los Angeles, CA; 2) Pediatrics/Neurology, LAC+USC Med Ctr, USC. Los Angeles, CA; 3) Radiology, Children's Hospital Los Ageles, USC. Los Angeles, CA; 4) Pediatrics, Kurume University, Kurume, Japan; 5) Human Genetics, Yokohama City University, Yokohama, Japan.

   Introduction: The mitochondrial aspartate-glutamate carrier isoform 1 (AGC1) is a component of the malate-aspartate shuttle which transfers reducing equivalents from NADH from the cytosol to the mitochondria. AGC1 deficiency is a rare autosomal recessive condition causing seizures, intellectual disability, and the characteristic brain MRI finding with hypomyelination as well as the MR spectroscopy (MRS) finding with a low N-acetylaspartic acid (NAA) signal. NAA is synthesized from aspartic acid and acetyl-CoA by aspartate N-acetyltransferase which is localized in the mitochondrion and the cytosol. Aspartic acid supplementation may improve NAA synthesis through the cytosolic pathway in patients with AGC1 deficiency. **Case Study**: Two Hispanic male siblings (9 y and 13 y) with AGC1 deficiency due to a homozygous p.Gly398Val mutation in *SLC25A12* presenting with intellectual disability, seizures, and the characteristic abnormal brain MRI/MRS findings were treated with aspartic acid at 100 mg/kg/d for 5 months and subsequently at an increased dose of 200 mg/kg/d for 7 months. No specific clinical improvement was observed including in their reported high protein diet preference. Brain MRI/MRS studies at base line and the follow up at 12 months after aspartic acid supplementation did not show significant changes: mild generalized cerebral volume loss and increased lactate and decreased NAA signals were noted. Plasma aspartic acid concentrations as well as urine organic acid analyses did not show significant changes after aspartic acid supplementation. **Discussion:** AGC1 deficiency results in decreasing the cytosolic NAD/NADH ratio leading to inhibition of glycolytic activity. In neuronal cells, NAA is synthesized from aspartic acid and acetyl-CoA primarily in the mitochondria and to some extent in the endoplasmic reticulum. Aspartate supplementation did not improve clinical symptoms nor MRI/MRS study findings in these two cases. Our results indicate that aspartate supplementation (1) did not improve the NAD/NADH ratio and glycolytic activity in the neuronal cells and/or (2) did not result in increased aspartic acid concentrations at least in the dose given to increase NAA synthesis in the neuronal cytosol. Based on these negative results, aspartic acid monotherapy appears to have no therapeutic utility in AGC1 deficiency.

961W

De novo missense variants in *GNAI1* gene are associated with epileptic encephalopathy. *M. Liao[1], L. Rhodes[1], H. Dubbs[2], E. Goldberg[2], L. Zitano[3], C. Bupp[3], J. Chen[4], K.J. Wierenga[4], G. Purcarin[5], G.N. Wilson[6], J. Martinez[7], A. Crunk[1], I.M. Wentzensen[1], R.E. Schnur[1], K.G. Monaghan[1], J. Juusola[1].* 1) GeneDx, Gaithersburg, MD; 2) Department of Pediatrics, Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Spectrum Health Medical Genetics, Grand Rapids, MI; 4) Department of Pediatrics, Section of Genetics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 5) Department of Neurology, Section of Child Neurology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 6) Texas Tech Health Science Center, Lubbock and KinderGenome Medical Genetics, Dallas, TX; 7) Department of Pediatrics and Adolescent Medicine, Division of Genetics, University of South Alabama, Mobile, AL.

   The Gi protein belongs to the family of inhibitory heterotrimeric G proteins which, upon activation, inhibit the adenylyl cyclase activity and reduce cAMP levels. Gi activation is promoted by receptor-induced exchange of GDP for GTP on its α subunit, which has an intrinsic GTPase to deactivate Gi through hydrolysis of bound GTP to GDP. The widely expressed *GNAI1* gene, located at 7q21.11, encodes the α-subunit of subtype 1 Gi (Giα1). Disruption of *GNAI1* was reported to cause long term memory defects in mice, suggesting that it may play a role in neurologic function. Here we show that germline variants in *GNAI1* may associate with a neurodevelopmental disorder characterized by developmental delay, seizure, and hypotonia. Exome sequencing (ES) was performed on 9,659 patients with developmental delay or intellectual disability (ID). Four *de novo* heterozygous missense variants in *GNAI1* were identified in five individuals from unrelated proband-parent trios. The five probands were males between 3-11 years of age with developmental delay and ID. Among them, absent or minimal speech, significant gross motor delay, hypotonia, and seizures were common. Two of the four individuals with brain imaging studies had cerebral atrophy. Of all five patients, two had microcephaly and two had macrocephaly. Short stature was also seen in two of those patients. All detected *GNAI1* variants were absent from the ExAC database as well as our internal ES database of unaffected individuals, and occurred at highly conserved codons 40 or 48. Each alteration was predicted to damage protein function by in *silico* algorithms. Two of the variants (p.Thr48Lys and p.Thr48Ile) are located within the GTP binding domain, while the other two (p.Gly40Cys and p.Gly40Arg) lie adjacent to this domain, indicating that changes of those amino acids might disrupt binding of Giα1 to GTP. In addition, the variants p.Gly40Cys (c.118G>T) and p.Gly40Arg (c.118G>C) are both located in the last nucleotide of exon 1 and were found to have potential impact on splicing. We propose that *de novo* heterozygous missense *GNAI1* variants impacting the GTP binding domain may lead to a novel neurodevelopmental disorder. The findings presented in this case series may facilitate the identification of additional affected individuals and increase understanding of the phenotype associated with pathogenic variants in *GNAI1*.

962T

BICD2-related arthrogryposis with unexplained cardiomyopathy. *R.D. Kastury[1], M. MacNeal[2], H. Seiden[3], H. McLaughlin[4], L. Mehta[5].* 1) Medical Genetics and Genomics, Icahn School of Medicine at Mount Sinai Medical Center, New York, NY; 2) Medical Genetics and Genomics Cardiovascular Genetics Program Icahn School of Medicine at Mount Sinai Medical Center, New York, NY; 3) Associate Professor Pediatric Cardiology Icahn School of Medicine at Mount Sinai Medical Center, New York, NY; 4) GeneDx, 207 Perry Parkway, Gaithersburg, MD 20877; 5) Professor Director, Clinical Genetics Department of Medical Genetics and Genomics, Dept. of Pediatrics Icahn School of Medicine at Mount Sinai Medical Center, New York, NY.

Mutations in *BICD2* are a cause of autosomal dominant spinal muscular atrophy (SMALED2) (OMIM#615290), characterized by congenital or early onset lower limb-predominant weakness with slow progression, muscle wasting and contractures. In-utero onset has been associated with arthrogryposis, though most cases present in the first decade. We evaluated a 12 year old girl with history of decreased fetal movement, bilateral femur fractures and contractures of the ankles, digits and wrists at birth. She required gastrostomy feeding and a tracheostomy, which was removed at age 8 yrs. She had severe kyphoscoliosis and underwent multiple orthopedic procedures and spinal fusion surgery. She was non-verbal and non-ambulatory with history of possible seizures. Brain MRI showed volume loss with thinning of the cortex. At age 11, a heart murmur was heard and an echocardiogram showed concentric left ventricular hypertrophy with moderate LVOT obstruction and moderate to severe mitral valve regurgitation. Parents had normal echocardiograms. SNP microarray, chromosome analysis, carbohydrate deficient transferrin and plasma amino acids were normal. Whole exome sequencing (WES) revealed a *de novo*, previously unreported, heterozygous, likely pathogenic variant, c.1636_1638delAAT in exon 5 in the *BICD2* gene. The in-frame 3 bp deletion is predicted to cause loss of an evolutionarily conserved Asparagine residue as position 546 in the protein (p.Asn546del). A heterozygous *de novo*, VUS (p.P957L), was present in the *SBF1* gene which is a cause of autosomal recessive CMT4B3; no other *SBF1* variants were identified. Mitochondrial genome sequencing with deletion testing was normal. The *BICD2* gene encodes a dynein adaptor protein that appears to be critical in axonal transport and neuron development and maintenance. The phenotype has been expanded to include cortical malformations, intellectual disability, seizures and spastic paraplegia, but no cardiac muscle involvement or non-neurological phenotype is described. No pathogenic mutations related to cardiomyopathy were identified in our patient. Therefore, it is more likely that her recent onset cardiomyopathy is a secondary phenotype due to other factors, and not directly related to *BICD2* mutations. Cardiac work up of patients diagnosed with *BICD2* mutations and further follow up of our patient's cardiac findings will be helpful in determining the natural history and progression of cardiac muscle involvement in BICD2-related disorders.

963F

A tuberous sclerosis positive case without cortical tubers and subsequent diagnosis of "unaffected" family members. *R. Caylor[1], L. Grote[3,7], I. Thiffault[1,2,8], E. Farrow[2,7,8], L. Willig[2,4,7,8], S. Amudhavalli[3,7,8], A. Nopper[6,7], A. Abdelmoitey[5,7,8], C. Saunders[1,2,8].* 1) Department of Pathology and Laboratory Medicine; 2) Center for Pediatric Genomic Medicine; 3) Division of Clinical Genetics; 4) Division of Nephrology; 5) Division of Neurology; 6) Division of Dermatology; 7) Department of Pediatrics, Children's Mercy Hospitals; 8) University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA.

Tuberous sclerosis complex (TSC) is an autosomal dominant neurocutaneous disorder that results from variants in *TSC1* or *TSC2*. TSC is characterized by neurocutaneous lesions and benign tumors in multiple organ systems. Common neurological findings include intellectual disability, seizures and cortical tubers. Common cutaneous findings include hypomelanotic macules and angiofibromas. Renal, cardiovascular, ophthalmologic and pulmonary abnormalities are also frequently observed in TSC. Diagnosis is based on well-established clinical criteria or molecular genetic testing; penetrance is thought to be 100%. We describe a family exhibiting extreme variability in clinical presentation and disease severity, including the incidental diagnosis of individuals thought to be unaffected. The proband is a 20-year-old Pakistani woman with intractable epilepsy, autism, absent speech, and normal MRI. Family history was positive for a brother with well-controlled epilepsy, normal intelligence and MRI. Whole exome sequencing of the proband revealed a likely pathogenic variant, c.196G>T (p.Glu66*), in *TSC1*, a surprising diagnosis given a previous negative MRI and no reported skin findings. The patient's mother and all three of her siblings, two of whom had no medical complaints, also tested positive for the variant. Assessment of the family history revealed a diagnosis of renal cell carcinoma in the mother. Follow-up examinations of her siblings revealed renal cysts, seizures and hypomelanotic macules for two siblings. The third sibling's brain MRI showed lesions that likely represent subcortical tubers, and an abdominal MRI revealed a solitary renal mass, which will require further surveillance. The family represented a difficult genetic counseling situation for several reasons. Initially, the variable presentation was challenging for the family to accept and led to distrust and withholding phenotypic information. Cultural practices inhibited full participation from the mother and children. This lead to additional limitations in phenotypic characterization and inability to explore the guilt the mother was experiencing due to these results. This case illustrates how an unbiased molecular approach may help reveal underlying defects in patients harboring pathogenic variants in genes not entirely expected due to the observed clinical manifestations, which can lead to diagnoses that otherwise would not be made or not made until more harmful symptoms of disease have manifested.

964W

Characterizing the rare X-linked dominant variant in *ALG13*: A case report. *J. Kohler[1,2], D.B. Zastrow[1,2], D. Bonner[1,2], J. Davidson[1,2], L. Fernandez[1,2], A.M. Dries[1,2], C. McCormack[1,2], M. Majcherska[1,2], S. Marwaha[1,2], N. Friedman[1,2], D. Waggot[1,2], U.D.N. Members of[3], P. Fisher[1,4,5], E.A. Ashley[1,2,7], J.A. Bernstein[1,2,6], M. Wheeler[1,2,7].* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, CA; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, CA; 3) NIH Undiagnosed Diseases Network, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Department of Pediatrics, Stanford School of Medicine, Stanford, CA; 5) Department of Neurology, Stanford School of Medicine, Stanford, CA; 6) Lucille Packard Children's Hospital Stanford, Palo Alto, CA; 7) Department of Genetics, Stanford School of Medicine, Stanford, CA.

   Introduction: *ALG13* encodes a subunit of the UDP-GlcNAc glycosyltransferase involved in N-linked glycosylation. *ALG13* was first reported in a male patient with congenital disorder of glycosylation type 1 (CDG1) who was hemizygous for a pathogenic variant inherited from his unaffected mother. In recent years, a different recurrent *de novo* variant (*ALG13* c.320A>G) has been identified in at least 9 unrelated females with early infantile epileptic encephalopathy, type 36 (EIEE36; OMIM:300884). All of these cases have been female. Despite the role of *ALG13* in N-linked glycosylation, glycosylation studies performed on a subset of these cases have so far been normal.   **Case report**: A 2yo female was evaluated at the Stanford Center for Undiagnosed Diseases (CUD) for refractory infantile spasms of unknown etiology. She first presented to neurology clinic at 9 months with global developmental delay (GDD), severe hypotonia, and choreoathetoid movements. At 12 months she developed infantile spasms demonstrated by hypsarrhythmia on EEG. Brain MRI showed generalized white matter volume loss. Poor visual tracking and food intolerance were noted. Metabolic work up showed elevated pyridoxal 5'-phosphate and low alkaline phosphatase compatible with hypophosphatasia. During evaluation at the CUD, whole exome sequencing (WES) was sent on the patient, her parents, and her unaffected twin brother. Sequence analysis identified the recurrent *de novo ALG13* c.320A>G variant. Utilizing the resources of the Undiagnosed Diseases Network, CDG studies, metabolomics, RNA sequencing, and fibroblast studies were performed to better understand the effect of the c.320A>G variant at a molecular level.   **Discussion**: This patient's phenotype overlaps with aspects of reported *ALG13* c.320A>G cases in the literature. Similarities include GDD, infantile spasms, hypotonia, movement disorder, visual impairment, abnormal MRI, and poor feeding. Our patient, however, lacks reported characteristics including dysmorphic features, joint contractures, and sleep disturbance. Hypophosphatasia has not been reported and therefore its association with the c.320A>G variant remains unclear.   **Conclusions**: This case contributes to the emerging phenotype associated with the *ALG13* c.320A>G variant. Functional studies will clarify the disease mechanism of the c.320A>G variant.

965T

Clinical presentation and genotype-phenotype correlation of a complex neurodevelopmental disorder caused by mutations in *ADNP*. *F. Kooy[1], A. Van Dijck[1], B.B.A. De Vries[2], E. Cappuyns[1], I.M. van der Werf[1], R. Bernier[3], I. Gozes[4], E.E. Eichler[5], C. Romano[6], M. Meuwissen[1], G. Vandeweyer[1], ADNP consortium.* 1) Department of Medical Genetics, University of Antwerp, Antwerp, Antwerp, Select a Country; 2) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 3) Department of Psychiatry of the University of Washington, USA; 4) Lily and Avraham Gildor Chair for the Investigation of Growth Factors, Elton Laboratory for Molecular Neuroendocrinology, Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Sagol School of Neuroscience and Adams Super Center for Brain Studies, Tel Aviv University,; 5) Department of Genome Sciences, University of Washington, Seattle, Washington, USA; Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA) 6) Unit of Pediatrics and Medical Genetics, I.R.C.C.S. Associazione Oasi Maria Santissima, Troina, Italy.

   Background: In genome-wide screening studies for *de novo* mutations as a cause of autism, intellectual disability and related developmental disorders, mutations in the *ADNP* gene are consistently reported amongst the most frequent causes. Mutations in this gene have been identified in children with autism spectrum disorder co-morbid with intellectual disability, facial features and deficits in multiple organ systems. However, a comprehensive description of the clinical features of the Helsmoortel - Van der Aa syndrome is lacking.   **Methods**: We identified a cohort of 78 individuals with likely disruptive mutations in *ADNP* dispersed across the globe from January 2014 to October 2016 through a systematic literature search, by contacting collaborators, and through direct interaction with parents on social media. Clinicians were asked to fill in a structured questionnaire on the genetic and clinical findings to enable genotype-phenotype correlations. Clinical photographs and medical specialist reports were gathered and parents of several individuals were interviewed to complement the written questionnaires.   **Findings**: We report on the detailed clinical characterization of a large cohort of patients and demonstrate a distinctive combination of clinical features, including mild to severe intellectual disability, autistic features, severe speech and motor delay and common facial characteristics. Hypotonia, short stature, gastrointestinal and visual problems, congenital heart defects, brain abnormalities, hormonal deficiencies, behavioral problems, sleep disturbance and epilepsy are common co-morbidities. Strikingly, individuals with the recurrent p.Tyr719* mutation were more severely affected for a restricted set of clinical features.   **Interpretation**: This overview defines the full clinical spectrum of individuals with *ADNP* mutations, a specific subtype of autism. We show that individuals with mutations in *ADNP* have many overlapping clinical features, distinctive from other known autism/ID syndromes. In addition, our data show evidence for a genotype-phenotype correlation. .

966F

A novel homozygous mutation in two sisters diagnosed with Joubert syndrome. *A. Sen[1], H.M. Serin[2], S. Gungor[3], D. Sen[1], S. Ceylaner[4].* 1) Medical Genetics Department, Firat University Hospital, Elazig, Turkey; 2) Pediatric Neurology Department, Firat University Hospital, Elazig, Turkey; 3) Obtetrics and Gynecology Department, Elazig Medical Park Hospital, Elazig, Turkey; 4) Intergen Genetic Centre, Ankara, Turkey.

Two sisters admited to our medical genetics polyclinic. They were diagnosed with Joubert Syndrome clinically when they were 3 months old and 2,5 years old. Follow-up procedure has done by pediatric neurologist. Their mom had an abort and gynecologist referred her to us to evaluate the complete family for an exact diagnose approved by molecular genetic techniques. Because, fetus was also suspected to have Joubert Syndrome. After evaluating them at clinical genetics department, we also suggested the same syndrome. As known, Joubert Syndrome has phenotypic series and many genes are responsible for different types of this syndrome. Both of them has been underwent to DNA sequencing panel. The genes responsible for Joubert Syndrome sequenced in this panel. Sequencing analysis revealed a novel mutation in *TMEM*237 gene. The mutation detected was homozygous c.394C>T (p.Q132*). Therefore, both of these patients are diagnosed as Joubert Syndrome 14. Genetic counseling is provided. Parental carrier analysis planned.

967W

Parent-of-origin effects in 15q11.2 BP1-BP2 deletion syndrome. *K. Davis, C. Robinson, M. Serrano.* Lineagen, Inc., Salt Lake City, UT.

The 15q11.2 deletion from breakpoints 1 to 2 cause a neurodevelopmental syndrome that is highly variable with low penetrance and includes four core genes: *CYFIP1*, *NIPA1*, *NIPA2*, and *TUBGCP5*. The TUBGCP5 gene is involved in cell division, is ubiquitously expressed, and is associated with obsessive-compulsive disorder and attention-deficit hyperactivity disorder. A recent study has reported that the promoter of *TUBGCP5* is maternally methylated. Deletions of 15q11.2 may therefore show a parent-of-origin effect (POE) like other conditions in this region. We hypothesized that paternal deletions would be more deleterious, as probands would theoretically lack a functional copy of *TUBGCP5*. To test for a POE, we collected 56 total reports with known inheritance from our internal database (N=7), online databases (N=14), and primary literature (N=35). Sibling pairs were omitted from analysis, as were cases with known genetic conditions or *de novo* inheritance from an unknown parental allele. Cases were analyzed by parental inheritance and the proband's clinical features with chi-squared and Mann-Whitney U tests. Maternally inherited deletions (N=30; 53.6%) outnumbered paternally inherited deletions (N=26; 46.4%). Probands with maternal deletions showed a higher, but non-significant average number of total features (3.3 vs. 3.1; standard deviation: 2.3 for both). Mann-Whitney U tests did not find significant differences between parental inheritance and: total number of clinical features, non-dysmorphic features, physical and dysmorphic features, or behavioral features (all *p*>0.1). Chi-squared analyses showed probands with maternal deletions associated with more diagnoses of epilepsy (*p*=0.02) and autism spectrum disorder (ASD; *p*=0.005), while paternal deletions were associated with developmental delays (*p*=0.03). A sex-based difference was also observed, with more females than males having epilepsy (*p*=0.01). Unexpectedly, paternal deletions were not associated with more specific clinical features or more total features than maternal deletions. This evidence indicates a possible POE, although the mechanism causing parental differences in clinical features may not be due to maternal imprinting at *TUBGCP5*. Limitations included: possible ascertainment bias, variable quality in phenotypic information, and that this population was not evaluated by us or the same observer(s). Further studies analyzing POEs in this condition are warranted.

968T

Case report of a patient with a *TANGO2* deletion that provides additional phenotype information. *R. Godshalk, D. Adams.* Jacobs Levy Genomic Medicine and Research Program, Atlantic Health System, Morristown, NJ.

We report the case of a 17 year old male that had history of delayed walking but no other issues at the time of an illness that hospitalized him in a coma for 2 days at the age of 18 months. In his early childhood he was evaluated for a possible mitochondrial disorder. TW communicates with an IPad and his own sign language. Testing of the mitochondrial DNA was performed at the age of 15 and was found to be most likely negative. Chromosome microarray performed at the age of 17 identified a 17kb deletion of chromosome 22q11.21 that includes the TANGO2 gene. This deletion includes exons 4-9 of the TANGO2 gene. There have been other individuals with homozygous and compound heterozygous mutations that have episodic muscle weakness with recurrent rhabdomyolosis, hypoglycemia, hyperammonia, hyperlactacidemia, intellectual disability, seizures, gait disturbances, dysarthria, myopathic facies, cerebral atrophy, and susceptibility to cardiac tachyarrhythmias. Our patient's clinical picture may be an additional case. Sequencing of the other copy of TANGO2 is being pursued. One patient began to develop developmental regression following a febrile seizure at a year of age. Physical exam showed microcephaly and myopathic facies. The second patient had similar physical features after initial cardiac presentation of a long QT. Both patients were also thought to have a mitochondrial disorder when they first presented. Both patients were found to have abnormalities in the TANGO2 gene. Lelani et al identified various homozygous and compound heterozygous alterations in the area of the TANGO2 gene including copy number variants that range from 9 to 34 KB. The families may be an etiology for a progressive neurologic disorder characterized by intellectual disability. There are cases of patient dying during an acute crisis therefore this underscores the importance of testing of a chromosome microarray analysis in patients with a clinical diagnosis of "mitochondrial" disorders. Lelani et al also stated that their study showed that deletion of exons3-9 are common in the European population. Kremer et al also stated that deficiency of TANGO2 should be considered in cases of episodic encephalopathy and/or myopathy in infancy and childhood of unknown cause, especially in individuals with laboratory findings suggestive of impaired mitochondrial fatty acid oxidation.

969F

Periventricular nodular heterotopia as hallmark of a new ciliopathy related to *CRB2* mutation. *G.M.S. Mancini, R. Schot, M. Wilke, M. van Slegtenhorst, F.W. Verheijen.* Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.

Biallelic mutations in this Crumbs homolog-2 gene, *CRB2* (MIM 609720) have been described in 2015 as cause of isolated steroid-resistant nephrotic syndrome and histopathological findings of Finnish type congenital nephrosis, but also as cause of nephrosis with severe congenital ventriculomegaly and periventricular nodular heterotopia, already detected at prenatal ultrasound. Although periventricular nodular heterotopia (PNH) has been documented in 20% of the cases (Slavotinek et al 2016), this abnormality has not been considered a hallmark of the disorder. Two brothers, aged 36 and 20 years, were referred because of detection of bilateral diffuse PNH and mega cisterna magna at MRI, with overall normal CSF spaces. One presented with ID, autism, motor delay, anxiety. The other had autism, psychosis and unexplained persistent proteinuria from childhood. The head circumference was normal. DNA and RNA sequences of *FLNA* were normal. Whole genome sequencing showed in both brothers the same novel compound heterozygote mutation in *CRB2*, a missense change leading to an amino acid substitution and a 16-bp deletion leading to a premature stop. An unrelated infant was referred with dextrocardia, hydrocephalus, diffuse bilateral PNH. A renal biopsy because of persistent proteinuria showed at the age of 23 months prominent glomerular podocytes and accentuated mesangial layer. A ciliopathy was suspected; skin fibroblasts were immunostained for cilia after serum starvation and repeatedly showed strongly reduced number of ciliated cells (av. 25%, normal range 50-80%, n=100). Targeted analysis of 43 genes involved in ciliopathies, including *CRB2,* showed in this gene a novel compound heterozygote amino acid substitution and a base-pair deletion leading to a premature frame shift and, in addition, a heterozygote missense in *WDPCP*. These cases illustrate that PNH is an important hallmark of *CRB2* mutations, which underlie a novel group of ciliopathies with heterogeneous manifestations, possibly modified by additional variants.

970W

Clinical management of patients with GLUT1 deficiency syndrome (De Vivo disease). *T. Kozhanova[1,2], S. Zhilina[1,2], T. Mescheryakova[1], S. Aivazyan[1], K. Osipova[1], L. Suchko[1], E. Luk`yanova[1], A. Prityko[1].* 1) St.Luka Scientific and Practical Center of Specialized Medical Care for Children, Moscow, Russia, Aviatorov Street, 38; 2) Pirogov Russian National Research Medical University, Moscow, Russia. Ostrivityanova street, 1.

GLUT1 (glucose transporter type 1) deficiency syndrome (GLUT1DS, De Vivo disease) is characterized by the infantile-onset encephalopathy (resistant to traditional seizure medications), acquired microcephaly, developmental delay with spasticity, dysarthria and a complex movement disorder consisting of ataxia, dystonia, alternating hemiplegia. In most patients, the level of glucose in the cerebrospinal fluid (CSF) is low to normal. We observed six clinical cases of GLUT1DS in children admitted to the neuropsychiatric department of Scientific and Practical Center of children medical care. The drug-resistant epilepsy, different movement disorders, psychomotor and intellectual disabilities, low cerebrospinal fluid glucose were diagnosed in all children. One patient had microcephaly. The different types of mutations in the *SLC2A1* gene (c.1305-1306insTGAAGA, c.115-2A>G, c.101A>G, c.400G>A) that are responsible for the development of GLUT1DS were detected by targeted sequencing in only 4 patients. The diagnosis in two patients was established according to clinical criteria, as there was no informed consent for genetic testing from parents. At present all patients are taking ketogenic diet (KD). A ketogenic diet is highly effective method of pathogenesis therapy, which can reduce the clinical manifestations: controlling the seizures, improving the movement disorder and speech. Thus, lumbar puncture should be considered the first diagnostic step for diagnosis of GLUT1DS. Mutational analysis of *SLC2A1* gene should be performed in patients with highly suggestive clinical findings and low cerebrospinal fluid glucose. Early diagnosis is critical because it allows prompt initiation of treatment with KD. Childhood is the critical period for treatment of GLUT1DS and for an effective etiological therapy.

971T

***CAD* mutations and uridine-responsive epileptic encephalopathy.** *I. Bader[1], J. Mayr[1], B. Alhaddad[2], C. Rauscher[1], R. Kovacs-Nagy[2], M. Holzhacker[1], H. Prokisch[2,3], F. Distelmaier[4], T. Polster[5], S. Leiz[6], C. Betzler[7], T. Strom[2,3], W. Sperl[1], T. Meitinger[2,3,8], S. Wortmann[1,2,3], T. Haack[2,3].* 1) Department of Pediatrics, Salzburger Landeskliniken (SALK) and Paracelsus Medical University (PMU), Salzburg, AustriaClinical Genetics, Childrens University Hospital, Salzburg, Austria; 2) Institute of Human Genetics, Technische Universität München, Munich, Germany; 3) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Munich, Germany; 4) Department of General Pediatrics, Neonatology and Pediatric Cardiology, University Children's Hospital, Heinrich-Heine University Dusseldorf, Dusseldorf, Germany; 5) Department of Child Neurology, Bethel Epilepsy Center, Bielefeld, Germany; 6) Department of Pediatric Neurology, Childrens Hospital Dritter Orden, Munich, Germany; 7) Clinic for Neuropediatrics and Neurorehabilitation, Epilepsy Center for Children and Adolescents, Vogtareuth, Germany; 8) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.

We report biallelic mutations in *CAD*, encoding an enzyme of *de novo* pyrimidine biosynthesis, in four patients with developmental disability, epileptic encephalopathy, anaemia, and anisopoikilocytosis. Two children died after a neurodegenerative disease course. Treatment of two surviving children with oral uridine led to immediate cessation of seizures in both. A four-year-old girl, who was previously in minimal conscious state, started to communicate and walk with assistance after nine weeks of treatment. A three-year-old girl likewise showed developmental progress. Blood smears normalised and anaemia resolved. Our findings support the efficacy of uridine supplementation rendering CAD deficiency a treatable neurometabolic disorder.

972F

Impact of a targeted next generation sequencing (TNGS) strategy for the genetic diagnosis of early onset epileptic encephalopathies. *S. Gobin[1], E. Azouguene[1], S. Hanein[2], N. Chemaly[4], M. Hully[4], L. Routier[3], C. Barnerias[4], N. Bahi Buisson[4], M. Rio[5,2], I. Desguerre[4], A. Kaminska[6], JP. Bonnefont[1], J. Steffann[1], R. Nabbout[4], G. Barcia[1,2].* 1) Laboratory of Molecular Biology, Department of Medical Genetics, Necker Enfants, Paris, France; 2) INSERM U1663, Imagine Institute for Rare Diseases, Paris Descartes University, Paris, France; 3) Department of Pediatric Neurology, Amiens-Picardie University Hospital, Amiens, France; 4) Reference Centre for Rare Epilepsies, Department of Pediatric Neurology, Necker Enfants-Malades Hospital, Paris Descartes University, Paris, France; 5) Department of Medical Genetics, Necker Enfants-Malades Hospital, APHP, Paris Descartes University, Paris, France; 6) Department of Clinical Neurophysiology, Necker Enfants-Malades Hospital, APHP, Paris, France.

   Objective: The identification of the genetic bases of epileptic encephalopathies (EE) has rapidly increased in the last years. Genetic diagnosis of EE can orient patient management and clinical outcome. We developed a custom targeted next generation sequencing (TNGS) strategy for the molecular diagnosis of EE. We evaluated the efficacy of this approach. Methods: A custom next generation sequencing panel targeting 151 epilepsy genes was applied to a cohort of 92 consecutive patients (51 boys and 41 girls) affected by early onset EE, classified as known syndromes in 58% and unclassified conditions in 41%. Acquired pathologies, metabolic diseases, brain malformations were excluded for all patients. Results: A disease-causing genetic variant was identified XX of patients. Variants lied on XX different genes including *KCNT1, WWOX, SCN2A, SCN1A, SCN8A,* and *CDKL5.* Patients with epilepsy of infancy with migrating focal seizures (EIMFS) and Dravet syndrome (DS) had the highest rate of positive findings, respectively 75% and 60%. TNGS allowed the detection of a *KCNT1* mutation previously undetected by Sanger sequencing in one patient, of a *SCN1A* variant at a low rate mosaicism and of two deletions of the exons 6-7 of *WWOX.* Conclusions: With a diagnostic yield of 38%, TNGS is an efficient strategy for the first-step genetic screening of EE, avoiding incidental findings, and at a lower cost compared to WES. Even if targeted approaches are restricted to a selected group of genes, interpreting variants generated by TNGS is challenging. A dynamic interaction between physicians and molecular biologists is necessary to interpret TNGS results and to draw genotype-phenotype correlations. TNGS, enabling a comprehensive screening of all known genes involved in EE, offers a rapid, cost effective and relevant selection of patients available for the identification of novel genes by WES and WGS in the research setting.

973W

Linkage and haplotype analyses of families with benign adult familial myoclonic epilepsy (BAFME). *H. Ishiura[1], M.K. Matsukawa[1], M. Tanaka[1], M. Higashihara[2], Y. Ichikawa[3], Y. Takahashi[4], K. Abe[5], Y. Sakiyama[6], M. Otsuka[7], A. Ueki[8], K. Kaida[9], J. Mitsui[10], Y. Suzuki[11], S. Sugano[12], S. Morishita[11], J. Goto[13], S. Tsuji[10,14].* 1) Dept Neurology, Univ Tokyo, Tokyo, Japan; 2) Dept Neurology, Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan; 3) Dept. of Neurology, Kyorin University, Tokyo, Japan; 4) Dept. of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 5) Dept. of Neurology, Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama University, Okayama, Japan; 6) Dept. of Neurology, Jichi Medical University, Saitama Medical Center, Saitama, Japan; 7) Dept. of Neurology, International University of Health and Welfare, Tochigi, Japan; 8) Nozomi Research Institute for Higher Cerebral Function, Saitama, Japan; 9) Third Dept. of Internal Medicine, National Defense Medical College, Saitama, Japan; 10) Dept. of Molecular Neurology, Graduate School of Medicine, The University of Tokyo; 11) Dept. of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 12) Dept. of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 13) Dept. of Neurology, International University of Health and Welfare Mita Hospital, Tokyo, Japan; 14) International University of Health and Welfare Graduate School, Tokyo, Japan.

   [Background] Benign adult familial myoclonic epilepsy (BAFME) is an autosomal dominant disorder characterized by myoclonic tremor and infrequent epilepsy with a benign clinical course. Several previous studies demonstrated a linkage to 8q24 in Japanese families, whereas families of Italian, French, and Thai origins with similar clinical presentations have been shown linkages to 2p, 5q, and 3q, respectively, indicating locus heterogeneity. To refine the candidate region of BAFME in Japanese, we performed linkage and detailed haplotype analyses. [Methods] Nineteen patients and three unaffected individuals from six families with a clinical diagnosis of BAFME were included in the study after written informed consent was obtained. Parametric linkage analysis was performed using high-density SNP data (Genome-Wide Human SNP array 6.0, Affymetrix) assuming autosomal dominant inheritance with complete penetrance. Haplotype was reconstructed inferred by manual inspection to minimize recombination events. In one family in which genotypes of parents or offspring were unavailable, droplet digital PCR and 10X GemCode technology were used for haplotype reconstruction. [Results] Linkage analysis revealed a single peak of cumulative multipoint parametric LOD scores at 8q22.1-8q24.13 (a maximal LOD score of 3.1) with the candidate region spanning 30 Mb, overlapping with a critical region of Japanese BAFME families that have been determined by previous studies. Detailed haplotype analysis revealed a core haplotype shared among the families, further narrowing the candidate region to a 134 kb region. [Discussion] The result strongly suggest that Japanese BAFME families share a common founder chromosome. We narrowed down the candidate region of BAFME by linkage and haplotype analyses. Further sequence analysis will be needed to identify the causative gene for BAFME.

974T

New epilepsy genes and variants discovered utilizing patients referred for clinical genetic testing. *K. McCarty[1], O.A. Moody[2], C. da Silva[3], J.J. Alexander[1,3], M. Hegde[1], A. Jenkins[2], A. Escayg[1].* 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Pharmacology, Emory University, Atlanta, GA; 3) EGL Genetics, Tucker, GA.

Epilepsy is a neurological disorder characterized by recurrent, unprovoked seizures caused by neuronal synchrony and hyperexcitability. To date, mutations in over one hundred genes have been identified in different types of epilepsy. Despite this progress, mutations in these genes only explain a fraction of the estimated genetic contribution to epilepsy. The goal of our research is to elucidate the mechanisms of seizure generation through a better understanding of the genetic causes of epilepsy. We are collaborating with EGL Genetics, a CLIA-certified and CAP-accredited laboratory, to examine available sequence data from clinically-referred epilepsy patients. Upon referral, patient DNA samples are screened using the Epilepsy and Seizure Disorders (ESD) panel, a sequencing panel of 110 known epilepsy genes. The ESD panel is derived from a larger Mendeliome library of approximately 5000 evidence-based disease genes, making this a valuable resource for research-based identification of putative disease-causing alleles and novel disease genes. Approximately 25% of referred patients have a positive finding from the ESD panel. We are currently examining variants from the Mendeliome library for the patients that are negative for mutations from the ESD panel. To date, we have identified at least 65 potentially causal variants in 40 genes. Several variants occur in genes not currently associated with epilepsy, including a novel variant in *GABRA2* that was identified in an individual with epileptic encephalopathy. Functional studies of the single point mutation in the TM2 segment of the receptor subunit encoded by *GABRA2* revealed a constitutively open ion channel that was unresponsive to GABA but was blocked by picrotoxin.

975F

An integrated whole-genome, whole-transcriptome approach to genetic diagnosis in developmental and epileptic encephalopathies. *A.M. Muir[1], K. Boysen[2], G. Hollingsworth[2], C. King[3], A.L. Schneider[2], U.W. Center for Mendelian Genomics[4], L.G. Sadleir[3], I.E. Scheffer[2,5,6], H.C. Mefford[1].* 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, Washington 98195, USA; 2) Epilepsy Research Centre, Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Victoria, 3084, Australia; 3) Department of Paediatrics and Child Health, University of Otago, Wellington, New Zealand; 4) University of Washington, Seattle, Washington 98195, USA; 5) Florey Institute of Neuroscience and Mental Health, The University of Melbourne, VIC 3010, Australia; 6) Department of Paediatrics, Royal Children's Hospital, The University of Melbourne, Parkville, Victoria, 3050, Australia.

Objective: Developmental and epileptic encephalopathies (DEE) are a large heterogeneous group of severe, early onset epilepsies characterized by refractory seizures, developmental delay and/or regression, and often a poor prognosis. Next generation sequencing approaches have revealed *de novo* variants in coding regions of genes as the most common cause of DEE, with clinical diagnostic labs reporting a molecular diagnosis in 24-40% of cases. Despite this significant progress, the majority of DEE cases remain unexplained, even after whole exome sequencing. We sought to investigate potential disease-causing variants in noncoding regions of the genome. **Methods:** We are performing whole genome sequencing (WGS) of 20 exome-negative proband-parent trios. However, a major hurdle for genome-based diagnostics is our ability to interpret the functional and clinical impact of variants in non-coding sequence. In order to determine whether noncoding variants of interest impact gene expression, we will use RNA-seq to detect changes in transcript abundance, evidence of allele specific expression, and the presence of aberrant splicing. Recent large-scale gene expression studies have highlighted the importance of using disease-relevant tissue in transcriptome sequencing; analysis of available data confirms that the majority of established DEE genes are poorly expressed in skin and whole blood, limiting the use of these easily accessible tissues. To overcome this limitation, we are approximating brain tissue by reprogramming primary skin fibroblasts into neurons. We will use genome-wide expression data from the reprogrammed neurons to interpret the clinical relevance of noncoding *de novo* mutations identified by WGS in patients with exome-negative DEE. **Results:** To date, we have performed whole genome sequencing on 5 of 20 exome-negative cases and identified 37-68 *de novo* variants per trio. We have collected fibroblasts from three cases and established a protocol to directly reprogram primary skin fibroblasts to neuronal cells that will be used to aid in the functional interpretation of these sequence variants. **Conclusions:** This work establishes a framework for the investigation of non-coding variants in DEE and will facilitate improved diagnostic testing for DEE.

976W

The genetic landscape of the epilepsy-aphasia spectrum disorders. *C. Myers[1], M. Mehaffey[1], A. Schneider[2], G. Hollingsworth[2], A. Rosen[1], A. LaCroix[1], U.W. Center of Mendelian Genomics[3], L. Sadleir[4], I. Scheffer[2,5,6], H. Mefford[1].* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Epilepsy Research Centre, Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Victoria, 3084, Australia; 3) University of Washington, Seattle, WA, USA; 4) Department of Paediatrics and Child Health, University of Otago, Wellington, New Zealand; 5) Florey Institute of Neuroscience and Mental Health, The University of Melbourne, VIC 3010, Australia; 6) Department of Paediatrics, Royal Children's Hospital, The University of Melbourne, Parkville, Victoria, 3050, Australia.

Objective: The epilepsy-aphasia spectrum (EAS) refers to a complex group of epilepsy syndromes associated with language impairment and variable outcome. On the severe end of the spectrum is Landau-Kleffner syndrome (LKS) and epileptic encephalopathy with continuous spike-and-wave during sleep syndrome (ECSWS) which often carry a poor prognosis including impaired cognitive function with speech and language difficulties. Related to the EAS from an electroencephalographic perspective but without major cognitive problems is the mild epilepsy syndrome of childhood epilepsy with centrotemporal spikes (CECTS). CECTS is the most common focal epilepsy in children and carries an excellent prognosis. It is unknown whether there is a common underlying mechanism to syndromes lying along the EAS. A genetic basis was not suspected until recently with the discovery that copy number variations (CNVs) and mutations in *GRIN2A* cause 5-20% of cases of dominant familial and sporadic EAS. We sought to identify novel genetic causes of sporadic EAS through whole-exome sequencing. **Methods:** Whole exome sequencing was performed in 16 proband-parent trios and a single proband-mother duo. We prioritized the following genotypes for review of their relevance to EAS: *de novo* mutations and/or parental mosaic variants, compound heterozygous, newly homozygous, and X-linked hemizygous genotypes. **Results:** We identified pathogenic or likely pathogenic *de novo* variants in 3/17 probands providing a potential molecular diagnosis for 18% of the cohort. The *de novo* variants in these three individuals were in *FOXP1, KCNMA1,* and *SETD1B*. No single gene was responsible for multiple cases, nor did the genes converge on a single molecular pathway. We also identified potentially pathogenic variants in 5 candidate genes (*CSNK1D, ATP1A2, PPFIA3, DEPTOR, NCAPG2*). Targeted sequencing of these 8 genes in a cohort of 525 probands with diverse developmental and epileptic encephalopathies did not yield any additional pathogenic variants. **Conclusion**: We expand our understanding of the genes associated with EAS and highlight the genetic heterogeneity of these severe disorders. We provide important phenotype expansion of the recently discovered neurodevelopmental genes *FOXP1, KCNMA1,* and *SETD1B*.

977T

***STXBP1* encephalopathy with epilepsy: 6-year-old girl with de novo missense variant in *STXBP1* expands the phenotype.** *J. Pappas[1], J. Borsuk[1], S.S. Nadkarni[3], D.A. McKnight[2].* 1) Dept Pediatrics, Clin Genetic Scvs, New York Univ, Sch Med, New York, NY; 2) GeneDx, Neurogenetics, Gaithersburg, MD; 3) Dept Neurology, Dept Psychiatry, Epilepsy program, New York Univ, Sch Med, New York, NY.

The association of epilepsy and encephalopathy with variants in *STXBP1* has been established and reviewed by Stamberger H et al, 2016 and, Khaikin Y and Mercimek-Mahmutoglu S, 2016. Publications in the medical literature report that about 1/3 of the patients with variants in *STXBP1* and seizures in infancy become seizure-free between the ages of 1 and 4 (Stamberger H et al, 2016; Di Meglio C et al, 2015 and Milh et al, 2011). Stamberger H et al, 2016 suggests that the encephalopathy in children with variants in *STXBP1* is unrelated to epilepsy. We present a 6-year-old girl with a novel missense variant in *STXBP1*. She was born full-term after uncomplicated pregnancy and delivery. She had two episodes of generalized tonic-clonic seizures at 4 months. The EEG was abnormal and the brain MRI normal. She was treated with Keppra that was discontinued because she had no more seizures and her EEG normalized. She was hypotonic and started walking and talking at the end of her second year with the help of therapies. Attention deficit, hand flapping, eye rolling without associated epileptiform activity and constant nail biting were her behavioral characteristics. Her development was progressing during her third year of life with attention deficit, hyperactivity (ADHD) and many sleep interruptions being her main problems. She started having seizures again at age 4 controlled with three medications. She developed aggressive behaviors and her ADHD necessitated management with medications. Chromosome analysis and SNP microarray analysis using the Affymetrix Cytoscan HD platform were reported normal. Blood testing of a panel of genes associated with epilepsy revealed the c.1643 A>T p.N548I variant in *STXBP1*. This variant has not been reported before in the medical literature, ClinVar and ExAC. Parental testing revealed that the variant is de novo. The variant is a non-conservative amino acid substitution and it is conserved in mammals. Missense mutations in nearby residues (including R551H, R551L) have been reported in the Human Gene Mutation Database in association with *STXBP1*-related disorders (Stenson et al, 2014). The variant was classified as likely pathogenic. Our case shares onset of epilepsy in infancy and prolonged seizure free interval with many cases in the medical literature. This seizure-free interval may be characteristic of the *STXBP1*-related infantile epilepsy. The regression of the neurodevelopment seen in our patient at age 4 is unique to our case.

978F

High-depth multi-gene panel analysis with integrated sequence and copy number detection is a useful first-tier test with a high diagnostic yield and broad mutation spectrum detection in childhood epilepsy. *N. Patil, R. Truty, D. Riethmaier, L. Elias, A. Entezam, A. Fuller, T. Hambuch, M. Hogue, A. Khouzam, Y. Kobayashi, R. Lewis, K. Nykamp, M. Rabideau, P. Taylor, J. Westbrook, M. Zeman, S. Aradhya.* Invitae Corporation, 1400 16th Avenue, San Francisco, CA 94103.

   Epilepsy is a common disorder that is increasingly known to result from genetic causes, particularly when it occurs in early childhood. Understanding the genetic etiology can inform treatment strategies, prognosis, and recurrence risk. The objective of this study was to use a multi-gene panel with simultaneous sequence and exonic copy number variant detection to evaluate the clinical yield and assess its utility relative to exome sequencing. Using high depth next-generation sequencing, we investigated subsets of 186 genes in 1131 unrelated individuals diagnosed with epilepsy. We observed a total of 1840 clinically reportable variants in 885 individuals. 14% of patients received a definitive molecular diagnosis by panel testing, and another 29% had results that in some cases would likely reach clinical significance with additional evidence. 14% of positive results were chromosomal or exonic copy number changes or large indels. Most pathogenic variants were in genes for early-onset epilepsy and 22% of molecular diagnoses had treatment implications. Mosaic variants were seen in three genes and one individual had diagnostic results in two genes. Half of the variants of unknown significance identified were likely not clinically significant since they were found alone in an autosomal recessive gene or in the presence of a diagnostic result in another gene. A meta-analysis of published exome data showed that >70% of positive results for individuals with seizures were in genes present in available gene panels. We also evaluated sequence data for epilepsy genes in exome data derived from three different exome kits and bioinformatics pipelines and found a risk of up to 1.6% for false negative results, and that does not consider novel deleterious variants. These observations emphasize that a multi-gene panel for epilepsy is a useful first test prior to exome sequencing because of a high diagnostic yield and reliable detection of a broad range of mutations, including both sequence and copy number variants.

979W

Novel biallelic *SZT2* mutations in three cases of early-onset epileptic encephalopathy. *N. Tsuchida[1,2], M. Nakashima[3], A. Miyauchi[4], S. Yoshitomi[5], T. Kimizu[5], G. Vigneswari[6], WT. Keng[7], C. Gaik-Siew[7], M. Kato[8,9], T. Mizuguchi[1], A. Takata[1], S. Miyatake[1,10], N. Miyake[1], H. Osaka[4], T. Yamagata[4], H. Nakajima[2], H. Saitsu[3], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Department of Stem Cell and Immune Regulation, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 3) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan; 4) Department of Pediatrics, Jichi Medical University, Tochigi, Japan; 5) Department of Pediatrics, Shizuoka Institute of Epilepsy and Neurological Disorders, Shizuoka, Japan; 6) Department of Pediatrics, Penang Hospital, Pulau Pinang, Malaysia; 7) Genetic Department, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; 8) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 9) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 10) Clinical Genetics Department, Yokohama City University Hospital, Yokohama, Japan.

   The seizure threshold 2 (*SZT2*) gene encodes a large, highly-conserved protein related to mTORC1 signaling. In mice, *Szt2* is abundantly expressed in the central nervous system and associated with epileptogenesis. Recently, biallelic *SZT2* mutations were found in seven patients from five families presenting with epileptic encephalopathy with dysmorphic features and/or non-syndromic intellectual disabilities. By whole-exome sequencing, we found compound heterozygous *SZT2* mutations in three unrelated patients with early-onset epileptic encephalopathies. Six novel *SZT2* mutations were found, including three truncating, one splice site and two missense mutations. The splice site mutation resulted in skipping of exon 20 and was associated with a premature stop codon. All individuals presented with seizures, severe developmental delay and intellectual disabilities with high variability. Brain MRIs revealing a characteristic thick and short corpus callosum or a persistent cavum septum pellucidum in each of two cases. Interestingly, in the third case, born to consanguineous parents, had unexpected compound heterozygous missense mutations. She showed microcephaly despite the other case and previous ones presenting with macrocephaly, suggesting that *SZT2* mutations might affect head size. These observations indicate that *SZT2* abnormalities are phenotypically heterogeneous.

980T

Whole exome sequencing reveals potential oligogenic inheritance and candidate novel genes in patients with arthrogryposis. *Y. Bayram[1], D. Pehlivan[1,2], D. Alkaya Uludag[3], Z. Coban Akdemir[1], A. Gezdirici[4], E. Yilmaz Gulec[4], Z. Ocak[4], H. Mutlu Albayrak[5], Y. Sahin[6], H.B. Erdem[7], I. Sahin[7], M. Elmas[8], Z. Yuksel[9], O. Ozdemir[10], F. Silan[10], O. Yildiz[10], E. Karaca[1], S.N. Jhangiani[11], D.M. Muzny[11], R.A. Gibbs[11], B. Tuysuz[3], N. Elcioglu[12], J.R. Lupski[1,11,13,14], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 2) Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 3) Department of Pediatric Genetics, Istanbul University Cerrahpasa Medical Faculty, Istanbul, Turkey; 4) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey; 5) Department of Pediatrics, Division of Pediatric Genetics, Faculty of Medicine, Ondokuz Mayıs University, Samsun, Turkey; 6) Department of Medical Genetics, Necip Fazıl City Hospital, Kahramanmaras, Turkey; 7) Medical Genetics Unit, University of Health Sciences, Diskapi Yildirim Beyazit Training and Research Hospital, Ankara, Turkey; 8) Department of Medical Genetics, Afyon Kocatepe University, Afyon, Turkey; 9) Medical Genetics Clinic, Mersin Women and Children Hospital, Mersin, Turkey; 10) Department of Medical Genetics, Faculty of Medicine, Onsekiz Mart University, Canakkale, Turkey; 11) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 12) Department of Pediatric Genetics, Marmara University School of Medicine, Istanbul, Turkey; 13) Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 14) Texas Children's Hospital, Houston, Texas.

   Arthrogryposis is defined as congenital joint contractures in two or more body areas. It is a clinical sign rather than a specific disease diagnosis and is present in more than 400 different disorders with a majority involving the neuromuscular system, skeletal system and connective tissue. Although variants of more than 220 genes have been associated with arthrogryposis, the underlying molecular etiology remains unknown in a majority of cases. We applied whole exome sequencing (WES) to 117 affected individuals from 105 families with the clinical sign of arthrogryposis to identify potential molecular etiologies. We have previously published the results of 48 families with a molecular diagnostic rate of 58.3% by identifying variants in known genes in 17 families (then most variants were identified in *CHRNG* in 5 families and *ECEL1* in 4 families), variants in candidate novel disease genes (*FBN3*, *MYO9A*, and *PSD3*) in 3 families, and potential oligogenic inheritance in 8 families. The family-based and cohort analysis of the WES results in subsequently collected samples from 57 families with arthrogryposis revealed variants in known genes in 20 families with the most common known genes being *CHRNG* (4 families) and *ECEL1* (2 families) which is consistent with our previous findings. Likely pathogenic variants in candidate novel genes including *DRG1*, *HS6ST2*, *GBP5*, and *NUGGC* were found in 4 families. Additionally, as we found in first batch analysis, there was evidence of mutational burden or multi-locus pathogenic variation contributing to the phenotype, i.e. either rare homozygous variants in two known genes (i.e. *FBN2+COL6A3*, *KLHL7+HOXA11*, and *NEB+SCN9A*) or a homozygous and likely pathogenic variant in a novel gene in addition to a homozygous variant in a known gene (*MID1IP1+NEB*) in 4 families. In total, the arthrogryposis manifestation could be explained by a molecular diagnosis in 53.3% of families (known genes in 37 families [35.2%], novel genes in 7 families [6.7%], and oligogenic inheritance in 12 families [11.4%]) using WES. Our approach using genome-wide analysis allowed the identification of novel candidate disease genes as well as the exploration of potential oligogenic and mutational burden models in the arthrogryposis trait. We suggest that increased implementation of genomic sequencing methods in genetically heterogeneous phenotypes such as arthrogryposis provides a comprehensive understanding of the potential contributions of multilocus variation.

981F

A novel *MTMR2* gene mutation (c.1168 G>T, p.E390*) in a patient with Charcot-Marie-Tooth Disease Type 4B1. *M. Erdogan[1], B. Balta[1], A. Kiraz[1], H. Gumus[2].* 1) Medical Genetics, Kayseri Research And Educational Hospital, Kayseri, Turkey; 2) Erciyes University Medical Faculty, Departmant Of Paediatric Neurology, Kayseri, Turkey.

   Charcot-Marie-Tooth Disease (CMT) is a heterogeneous group of inherited disorder associated with over 80 disease causing genes. CMT hereditary neuropathy syndrome can be inherited in an autosomal dominant, autosomal recessive, or X-linked manner. *MTMR2* encodes a protein possesses phosphatase activity towards phosphatidylinositol-3-phosphate and phosphatidylinositol-3,5-bisphosphate. Mutations in this gene are a cause of Charcot-Marie-Tooth disease type 4B1 (OMIM:#601382), an autosomal recessive demyelinating neuropathy. By whole exome sequencing we detected a *MTMR2* homozygous nonsense mutation (c.1168 G>T, p.E390*) in a 6 years old patient with gait disturbance, muscle weakness, mild ptosis, claw hand, bilaterally pes planus and demyelinating polyneuropathy. The aforementioned mutation has not been previously reported.

982W

Exome sequencing in Italian FTD patients reveals probable novel mutations in neurodegeneration associated genes. *M. Hammer[1], R. Capozzo[2], C. Sassi[1], C. Zecca[2], M. Federoff[1], C. Blauwendraat[1], N. Bernstein[1], J. Ding[1], J.R. Gibbs[1], T. Price[1], A. Singleton[1], G. Logroscino[2,3].* 1) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 2) Department of Clinical Neurology and Research, University of Bari, "Pia Fondazione Cardinale G. Panico", Tricase (LE), Italy; 3) Neurodegenerative Diseases Unit, Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari, Italy.

   Background Frontotemporal dementia (FTD) is a group of familial and sporadic neurodegenerative disorders affecting the frontal and temporal lobes. It is the second most common cause of dementia following Alzheimer's disease. **Objective** We aim at identifying known causative mutations and novel variants in genes associated with FTD and other related neurodegenerative disorders. **Methods** Ninety-two unrelated clinically diagnosed FTD patients (9.8% familial, 90.2% sporadic) were selected from the Apulia-FTD Registry in Italy. Whole exome sequencing was performed to identify the genetic origin of the disease. All subjects were screened for a pathological hexanucleotide repeat expansion in *C9ORF72* based on a repeat primed PCR. **Results** By focusing on pathogenic/likely pathogenic variants in established FTD genes, we identified three variants in *GRN*; two missense variants that have been previously published (c.G314A:p.C105Y and c.T415C:p.C139R) and one truncating frameshift deletion (c.1165delT:p.C389fs). Pathogenic repeat expansions in the *C9ORF72* gene were identified in two sporadic patients. In contrast to previous reports on genetic FTD, we did not find any variants in *MAPT*. Despite being rare causes of dementia, we detected variants that are likely pathogenic in *CHMP2B* (c.C49A:p.Q17K), *CSF1R* (c.G1646A:p.R549H and c.G1717A:p.E573K), *FUS* (c.G151A:p.G51S), *TBK1* (c.G1214T:p.R405L) and *CCNF* (c.A1240G:p.K414E and c.C326G:p.A109G). Furthermore, pathogenic and likely pathogenic variants in genes not commonly associated with clinical FTD were found (Alzheimer's disease genes: *PSEN2*, *SORL1*; ALS genes: *FIG4*, *MATR3*; other dementia/ neurodegeneration genes: *VPS13C*, *NPC1*, *ITM2B*). **Conclusion** Our study is of relevance for FTD genetics as it shows that a wide panel of neurodegenerative disorders causing genes should be considered when testing FTD patients.   .

983T

Clinical and molecular insights into developmental abnormalities of corpus callosum. *M. Hebbar[1], C. Galada[1], A. Kanthi[1], R. Kadavigere[1], A. Shrikiran[1], L. Lewis[1], S. Nampoothiri[2], V. Bhat[3], M. Muranjan[4], S. Bajaj[4], SL. Bielas[5], KM. Girisha[1], A. Shukla[1].* 1) Kasturba Medical College, Manipal - 576104, Karnataka, India; 2) Amrita Institute of Medical Sciences and Research Centre, Ponekkara, Cochin, Kerala, India; 3) Department of Neonatology, JIPMER, Pondicherry, India; 4) Seth GSMC and KEM Hospital, Mumbai, India; 5) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA.

   The development of corpus callosum is a composite process likely involving interplay of various genes and regulatory developmental mechanisms. The developmental abnormalities of corpus callosum is the most common brain malformation and genetically heterogeneous. We carried out an observational study in all patients with brain imaging studies consistent with agenesis or dysgenesis of corpus callosum and their families. We recruited 55 affected individuals from 51 families. Parental consanguinity was reported in 55% (28/51) of families. The syndromic form of corpus callosum abnormality was observed in 42% (23/55) of the cases. Non-syndromic form of corpus callosum abnormalities was seen in 58% (32/55) of the subjects, out of which majority were of complex type (90%, 29/32). Hypoplasia of corpus callosum observed in 51% (28/55) of the cases was the most common malformation, followed by partial agenesis of corpus callosum (15%, 8/55), complete agenesis of corpus callosum (27%, 15/55), hyperplasia of corpus callosum (2%, 1/55) and malformed (others) corpus callosum (5%, 3/55). The associated malformations noted in complex non-syndromic type were white matter disorders (52%) and posterior fossa anomalies (15%). Eight of them (23%) patients had neuronal migration defects. Enlarged subarachnoid spaces, hippocampus flattening, hyperintensity in the basal ganglia were other brain anomalies noted in five individuals. Genetic testing was completed in 43 families. Chromosomal microarray was performed in 16 patients, and mosaicism of chromosome 8, neurofibromatosis microdeletion syndrome, *6q21-q22.31* microdeletion syndrome, *13q13.3-21.1* microdeletion and *17q25.1* microduplication syndrome were detected. Whole-exome sequencing analysis in thirty-three families led to a definitive molecular diagnosis in 60% families. They were mostly autosomal recessive disorders constituting a spectrum of monogenic neurodevelopmental disorders including neurometabolic conditions, leukoencephalopathies and complex brain malformations. In addition, we also identified two novel candidate genes associated with these disorders, *ISCA1* (mitochondrial dysfunctions syndrome) in two unrelated families and *AIMP2* (neurodegenerative disease) in one family. .

984F

Heterozygous missense variant in *TDRKH* encoding tudor and KH domain-containing protein associated with autosomal dominant motor neuropathy. *K. Kosaka[1], S. Miura[2], T. Shimojo[1], S. Nagata[3], T. Morikawa[1], R. Fujioka[4], T. Nomura[5], T. Taniwaki[2], H. Shibata[1].* 1) Genomics, Medical Institute of Bioregulation Kyushu University, Fukuoka, Fukuoka, Japan; 2) Division of Respirology, Neurology and Rheumatology, Department of Medicine, Kurume University School of Medicine, Kurume 830-0011, Japan; 3) Department of Radiology, Kurume University School of Medicine, Kurume 830-0011, Japan; 4) Department of Food and Nutrition, Beppu University Junior College, Beppu 874-8501, Japan; 5) Dr. Nomura's Clinic of Internal Medicine and Neurology, Fukuoka 812-0012, Japan.

 Hereditary motor neuropathies (HMN) comprise a group of clinically and genetically heterogeneous inherited disorders mainly characterized by slowly progressive motor neuropathies. We ascertained a four-generation pedigree of a Japanese family with a hereditary motor neuropathy with the autosomal dominant mode of inheritance, consisting of six family members including two patients. The cardinal clinical features were as follows: i) slow progression; ii) adult onset (37 and 51 years old, respectively); iii) drop foot; iv) facial muscle impairment; v) sternocleidomastoid muscle impairment; vi) areflexia; vii) no sensory impairment. Nerve conduction studies demonstrated axonal change in peroneal nerve. To elucidate the genetic basis of HMN in the pedigree, we performed exome sequencing of the two patients with the average depth of 11.1x. From the total of 16,427 single nucleotide variants (SNVs) shared with the patients, we remained eight functional heterozygous variants to be extremely rare (< 0.2 %) in our unrelated in-house exome data (n = 9) and in public databases such as 1000G, ExAC, ToMMo and HGVD. By Sanger validation of the all six of family members, we confirmed cosegregation with the disease in only two SNVs. One is a single nucleotide substitution, c.851G>A (p.(R284H)) in *TDRKH* (NM_001083964) encoding tudor and KH domain containing protein. The other is a 10-bp deletion, c.749-758delATGACATATT (p.(Y250fs*33)) in *LILRB5* (NM_001081442) encoding leukocyte immunoglobulin-like receptor, subfamily B protein. By additional Sanger sequencing of 500 unrelated Japanese control samples, we excluded the deletion variant, Y250fs*33 in *LILRB5* to be relatively common (0.29 %) and unlikely to be pathogenic. On the other hand, we confirmed ;R284H in *TDRKH* is absent in the 500 controls. The SNV is located in a highly conserved region in the tudor domain in *TDRKH*. The tudor domain is also located in survival of motor neuron protein 1, encoded by *SMN1* and most of the responsible variants for another motor neuron disease, spinal muscular atrophy is known to be located in the tudor domain of *SMN1*. Therefore, it is reasonable that the SNV in the tudor domain may critically affect the *TDRKH* protein function. In conclusion, R284H in *TDRKH* is responsible for the hereditary motor neuropathy in our pedigree through functional alteration in the tudor domain. This is the first report of a human disease associated with *TDRKH*.

985W

Familial choreoathetosis: A novel heterozygous mutation in *PDE10A*. *D. Narayanan[1], D. Deshpande[2,3], A. Bhowmik[2], D. Ravi Varma[4], A. Dalal[2].* 1) Medical Genetics, Nizam's Institute of Medical Sciences, Hyderabad, India; 2) Diagnostics Division, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 3) Graduate Studies, Manipal University, Manipal, India; 4) Citi Neurocentre, Hyderabad, India.

 Chorea is a hyperkinetic movement disorder characterized by brief involuntary repetitive movements and is a symptom of many inherited neurological disorders. Recently Mencacci et al described three unrelated European descent individuals with childhood onset chorea and bilateral striatal abnormalities on cerebral MRI with de novo mutations in *PDE10A*. Diggle et al described eight individuals from two unrelated families with biallelic mutations in *PDE10A* and striatal abnormalities and hyperkinetic movement disorder with onset in infancy.We describe a family of Indian origin with multiple affected individuals harboring a novel heterozygous probable pathogenic variant in *PDE10A* identified by exome sequencing. A 24 year old female had progressively increasing involuntary movements of hands, face and head since 5 years of age. She had increasing difficulty in articulation of speech and difficulty in writing. She never had any seizures or worsening of intellect. On examination, she had choreiform movements involving upper limbs, facial grimace and head nodding. Higher mental status, cranial nerves, bulk, tone, power and reflexes, sensory system and cerebellar examination were normal. Her father was 60 years old and had a similar history with onset of symptoms at three years of age. Multiple members in father's family had similar symptoms. The husband of the proband was 35 years male born to consanguineous parents and had biallelic pathogenic variants in *SERPINF1* gene, causing autosomal recessive osteogenesis imperfecta type VI. MRI head of the patient showed volume loss in putamina and caudate nucleus and hypointense signals on T1 weighted images and hyperintense signals on T2 weighted and FLAIR sequences.A novel nonsynonymous pathogenic variant was identified (NM_001130690.2:c.1001T>G,p.F334C) in *PDE10A* by exome sequencing. Sanger sequencing confirmed the presence of variant in the patient and her father. In silico protein modeling showed that the mutation affected the binding affinity of cAMP to the GAF-B domain due to plausible change in the structure of binding pocket. Genetic counseling was offered to the family. Childhood onset of chorea with normal cognition and striatal hyperintensities should prompt further evaluation beyond benign hereditary chorea and heterozygous or homozygous mutations in *PDE10A* can cause such a phenotype. Making an exact molecular diagnosis is important in providing prenatal diagnosis and genetic counseling to families. .

986T

A novel *de novo* alteration in *SLC12A6* in a patient with early-onset severe progressive sensorimotor polyneuropathy and abnormal EEG. *M. Rossi[1], K. Scherer[2], M. Galindo[2], K. Radtke[1], Z. Powis[1], D. Shinde[1].* 1) Clinical Genomics, Ambry Genetics, Aliso Viejo, CA; 2) Children's Clinics, Tucson, AZ.

The *SLC12A6* gene encodes a potassium-chloride cotransporter (KCC3) belonging to a family of transmembrane proteins that regulate cell volume and control neuronal activity by transporting K+ and Cl− ions across the plasma membrane. Biallelic truncating and missense mutations in *SLC12A6* cause autosomal recessive agenesis of the corpus callosum with peripheral neuropathy associated with hydrocephalus, developmental delay, intellectual disability and seizures. However, only one patient with a *de novo* heterozygous missense alteration, c.2971A>G (p.T991A), in the *SLC12A6* gene has been reported till date. This patient presented with progressive, and early-onset axonal motor neuropathy with normal corpus callosum, cognition, and no epilepsy (Kahle *et al.* 2016). Functional studies demonstrated that this alteration results in constitutive KCC3 activity. We report an additional patient with a previously unpublished *de novo* missense alteration c.620G>A (p.R207H) in *SLC12A6* and severe sensorimotor polyneuropathy with distal and proximal weakness and abnormal EEG. The alteration was identified through diagnostic exome sequencing (DES) and reported as a candidate genetic etiology due to limited clinical validity of the association of monoallelic *SLC12A6* alterations with human genetic disease. A different alteration at the same codon, p.R207C, has been previously reported in a patient with developmental delay and mild intellectual disability, significant hypotonia with areflexia, pyramidal signs, complete agenesis of the corpus callosum, and an axonal and demyelinating neuropathy of the motor and sensory nerves with reduced motor nerve conduction velocity. The patient was found to be homozygous for the alteration and the carrier parents were asymptomatic. *In vitro* functional studies of the p.R207C alteration demonstrated decreased transport activity of the mutant KCC3 when expressed in Xenopus oocytes (Salin-Cantegrel *et al.* 2011). Functional studies of the effect of the p.R207H alteration need to be performed in order to determine whether this alteration has a loss- or gain-of-function effect and hence has been classified as a variant of uncertain clinical significance. There is also the possibility of another alteration on the second allele that could be intronic or located in the regulatory region, and therefore not detected by exome sequencing. RNA-sequencing studies could be pursued to confirm the state of the second allele in this patient.

987F

N98S mutation in *NEFL* gene causing a mild form of Charcot-Marie Tooth disease. *A. Sanchez[1], J.C. Prieto[1,2].* 1) Instituto de Genética Humana, Pontificia Universidad Javeriana Bogotá, Bogota, Colombia; 2) Hospital la Victoria, Bogotá Colombia.

Introduction: Charcot-Marie Tooth disease (CMT) is an heterogeneous group of inherited peripheral neuropathies affecting motor and sensory nerves of peripheral nervous system. Prevalence is 1 in 2500 individuals. CMT is classified into demyelinating or axonal forms depending on nerve conduction velocities (CMT1 - CMT2 respectively), however there is a dominant CMT intermediate form with clinical and electrophysiological characteristics that do not fit into either CMT1 or CMT2. Approx 1000 mutations in 80 genes have been identified as a cause. Neurofilament lightchain polypeptide (*NEFL*) gene codes for a protein of neurofilaments involved in axonal transport processes of mature neurons. Mutations distributed throughout *NEFL* may cause either dominant axonal CMT2E or dominant demyelinating CMT1F. **Case Report**: 23 y/o male patient with bilateral pes-cavus since age 13y. He was born from non-consanguineous parents. Family history was negative for neuromuscular or CNS diseases. At age 20, he developed ataxia. EMG showed demyelinating pattern in lower extremities. Brain MRI was normal. On examination, he was able to stand with support and walk few steps with bilateral steppage and marked postural instability. Bilateral forefoot pes-cavus was evident and had scar sequelae from orthopedic surgical procedure. Muscle power of lower extremities was diminished. Upper extremities were normal. Whole exome sequence showed a heterozygous missense pathogenic mutation N98S and a heterozygous likely benign A195V mutation in *NEFL* gene, which is associated with CMT2E and CMT1F. **Discussion**: *NEFL* related-CMT is clinically and genetically heterogeneous, and accounts only for <1% of all CMT cases. Additional features besides neuropathy are seen. Patients with N98S have hearing loss, ataxia and cerebellar atrophy. Here we report a patient with two variants on NEFL, a benign mutation inherited from father (A195V) and a *de novo* mutation (N98S). The N98S mutation has been reported as a cause of severe and very early onset disease, however this patient manifests a unique mild phenotype. Patient has a mild late onset demyelinating form of CMT with ataxia, without cerebellar atrophy. **Conclusion**: this report highlights the phenotypic variability in CMT patients with *NEFL* mutations and expands the clinical phenotype of dominant mutation N98S. There is an increased necessity of more genotype-phenotype correlations to clearly elucidate the pathogenesis of this wide clinical spectrum disease.

988W

Missense mutations and multiplications of alpha-synuclein in familial Parkinson's disease: Genotype-phenotype correlation. *K. Nishioka[1], H. Yoshino[1], Y. Li[1], M. Funayama[1], T. Candido[2], M. Farrer[2], N. Hattori[1].* 1) Department of Neurology, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan; 2) Department of Medical Genetics, Brain Research Centre, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada.

Alpha-synuclein is a major component of Lewy bodies and neuritis which are pathotlogical hallmark of Parkinson's disease (PD). We have assessed the frequency of alpha-synuclein (*SNCA*) mutations in Japanese patients with familial or sporadic Parkinson's disease (PD), and surveyed their associated clinical manifestations. We screened *SNCA* exon 3 in 988 patients without *SNCA* multiplications (430 with autosomal dominant PD and 558 with sporadic PD), and multiplications in 1620 with familial PD, sequenced by Sanger method using BigDye Terminators v1.1 Cycle Sequencing Kit and 3130 Genetic Analyzer (Life Technologies, Foster City, CA, USA) and MLPA methods (Multiplex ligation-dependent probe amplification, MRC_Holland, Netherlands). We detected one patient harboring a homozygous *SNCA* p.A53V substitution (frequency 2/860 = 0.2%), 13 probands with duplication, and one patient with triplication in the regions of *SNCA* (14/1620). According to inheritance pattern, there were autosomal dominant in 13 families, autosomal recessive-like in one family, and sporadic in one family. The average age at onset was 43.1 ±13.7 (±SD) years of duplication, 28 years of triplication, and 55 years of homozygous p.A53V. Triplication presented most severe prognosis of rapid and progressive parkinsonism and cognitive decline with high concentration of psychosis in the pedigree. Following, duplication showed various types of clinical course such as slow or rapid progression of parkinsonism and cognitive decline. The prevalence of cognitive decline was 5/13 in *SNCA* duplication. However, most of them had segregation of cognitive decline, presented rigidity, akinesia, and gait disturbance, of which manifestations resembled with rigid-akinesia parkinsonism. Levodopa response mostly seems to be good at early stage. A patient with homozygous p.A53V, she manifested slow and progressive parkinsonism at 55 years. Cognitive decline was prominent at 74 years. MIBG myocardial scintigraphy indicated decreasing of heart/mediastinum ration in all patients. In Japan, the frequency of multiplications were higher than missense mutation in *SNCA*. Clinical phenotype reflected on gene dosage effect, related to over-expression mechanisms of alpha-synuclein. High dosage of alpha-synuclein likely induces severe form of familial parkinsonism.

989T

Expanding the clinical spectrum of ARL6IP1-associated hereditary spastic paraplegia. *S. Majid[1], A. Alazami[1], H. Al Dossari[1], S. Al-Hissi[1], A. Al Qahtani[1], R. Al Humaidy[1], B. Meyer[1], S. Bohlega[2], A. Al Hashem[3].* 1) KFSHRC, RIYADH, India; 2) Department of Neurosciences, King Faisal Specialist Hospital & Research Center, Riyadh; 3) Prince Sultan Riyadh Military Medical City, Riyadh, KSA.

Mutations in ARL6IP1 (ADP ribosylation factor like GTPase 6 interacting protein 1) encoding a tetra-span membrane protein localized to the Endoplasmic Reticulum, have been recently described in an extended family with a complicated form of Hereditary Spastic Paraplegia (HSP). ARL6IP1 was first identified as an interacting molecule of ARL6, a member of the ARL subfamily of small GTPases which mainly regulates intracellular trafficking pathways in the ER membrane. In our quest to study HSP in the Saudi population, we uncovered a family with a severe and distinct phenotype that resulted in the spontaneous neonatal death of both affected. Clinical features encompassed not only complicated HSP but also developmental delay, microcephaly, periventricular leukomalacia and respiratory distress. Molecular Karyotyping was found to be normal. Homozygosity mapping localized the phenotype to a region on chromosome 12, and whole exome sequencing of the index case identified a homoallelic truncating mutation in ARL6IP1 that was predicted to result in complete loss of the protein. The mutation was confirmed by Sanger sequencing and segregated with the disease state with an autosomal recessive pattern of inheritance and complete penetrance. Our findings expand the phenotypic and mutational spectrum of ARL6IP1-associated neurodegenerative disorders.

990F

Genomic analysis identifies new loci associated with motor complications in Parkinson's disease. *S. Chung[1], J. Kim[2], M.J. Kim[1], M. Kim[3], Y.J. Kim[1], K.J. Kim[1], H.S. Ryu[1]. 1) Asan Medical Center, Seoul, South Korea; 2) Metro hospital, Anyang; 3) Korean University Medical Center Guro.*

   Background: Over the past 20 years, genetic causes and risk factors for Parkinson's disease (PD) have been identified by high-throughput genotyping technologies. However, previous genetic studies have typically focused on disease risk rather than disease-specific symptomatology. **Objectives:** We aimed to identify the genomic variants that are associated with the occurrence of motor complications in patients with PD. **Methods:** Genomic data was produced using the Korean Chip (K-CHIP), Affymetrix Axiom™ KORV1.1, which contains imputation genome-wide association study (GWAS) grid and other GWAS loci, functional variants of nonsynonymous exome, pharmacogenetics variants, variants in genes involved in absorption, distribution, metabolism and excretion (ADME) of drugs, and expression quantitative trait loci (eQTL), in 1,070 PD patients. The association between genomic data and clinical features of motor complications of PD patients was analyzed using Jonckheere-Terpstra test with codominant coding scheme as a primary analysis. **Results**: The intergenic variant (to be presented) in chromosome 2 had the most significant association with the occurrence of levodopa-induced dyskinesia within 5 years after PD onset ($P = 3.57 \times 10^{-9}$). This association was also significant after Bonferroni correction. There are several other genomic variants that showed nominally significant association with the occurrence of levodopa-induced dyskinesia within 5 years after PD onset, which warrant further evaluation. There was no significant association between genomic variants and the occurrence of wearing-off within 5 years after PD onset, although several genomic variants showed a nominal significance. **Conclusions:** This genomic study identified a new locus associated with the occurrence of levodopa-induced dyskinesia within 5 years after PD onset. These findings need to be validated by replication studies and functional studies.

991W

Genome-wide association study identifies potential genetic modifiers in Charcot-Marie-Tooth disease type 1A. *F. Tao[1], G. Beecham[1], S. Blanton[1], L. Abreu[1], F. Baas[2], B. Choi[3], D. Pareyson[4], M. Reilly[5], M. Shy[6], S. Zuchner[1], Inherited Neuropathy Consortium. 1) Dr. J.T. MacDonald Department for Human Genetics, Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Genome Analysis, Academic Medical Centre, Amsterdam, The Netherlands; 3) Department of Neurology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; 4) Department of Clinical Neurosciences, C. Besta Neurological Institute, Milan, Italy; 5) MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, London, UK; 6) Department of Neurology, University of Iowa, Iowa City, IA.*

   Charcot-Marie-Tooth disease type 1A (CMT1A), caused by a canonical *PMP22* duplication on chromosome 17p11.2, is the most common subtype of inherited peripheral neuropathies and affects 1 in 10,000 individuals worldwide. While sharing the same genetic cause, CMT1A patients often present great variability in their phenotypic presentation and disease severity. The cause of the phenotypic variability is largely unclear. In this study, we performed a case only genome-wide association study (GWAS) to identify novel genetic modifiers of various phenotypes in CMT1A. DNA samples from 971 CMT1A patients were genotyped on Illumina OmniExpress platform. After standard quality control, the dataset includes 600k markers in 857 individuals (644 individuals from European ancestry, and 213 individuals from Asian ancestry). We focused our analyses on the European population. Logistic regression in PLINK was used to analyze the association between clinical outcomes and patient genotypes in an additive model. For CMT neuropathy score (CMTNS), the analysis was performed using linear regression in PLINK, adjusting for age of patients. The analyses yielded several suggestive association signals. An association peak on chromosome 6 was identified in difficulty with eating utensils (lead SNP rs12192704, chr6:30792270, P=1.15E-06, odds ratio=3.25). The peak is located within a non-coding gene *LINC00243*. Hearing loss showed an association peak on chromosome 5 (lead SNP rs7720606, chr5:126551732, P=2.22E-07, odds ratio=3.457), located in an intergenic region near the *MEGF10* gene. In foot plantar flexion, an association signal was identified in the *DSCAM* gene on chromosome 21 (lead SNP rs2249498, chr21:41431874, P=5.13E-07, odds ratio=2.437). CMTNS showed an association signal on chromosome 1 (lead SNP rs12137595, chr1:4094068, P=1.14E-07, beta=3.014), located within an intergenic region close to *DFFB*, *C1orf174*, and *LINC01134*. While these suggestive signals require further validation, our study provides novel insights into the genetic architecture of CMT1A. Novel genetic modifiers may serve as potential targets for therapeutic interventions in the future.

992T

Integrative omics analysis of a cohort of 198 singletons with cerebral palsy. *J. Gecz[1,2,3], CL. vanEyk[1,2], JL. Broadbent[1], K. Harper[1], A. Gardner[1,2], BW. Van Bon[4], MA. Corbett[1,2], A. MacLennan[1,2].* 1) Australian Collaborative Cerebral Palsy Research Group, Robinson Research Institute, Adelaide, SA 5006; 2) Adelaide Medical School, The University of Adelaide, Adelaide, SA 5005, Australia; 3) South Australian Health and Medical Research Institute, Adelaide SA 5000, Australia; 4) 4Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands.

Cerebral palsies (CP) describe a heterogeneous group of permanent, non-progressive disorders of movement and posture diagnosed between birth and age 3-5 years. Our pioneering research into genetics of CP shows that, similar to other neurodevelopmental disorders, CP is likely to be genetically heterogeneous, with 100s of genes involved. Many of the CP genes & mutations will be 'shared' with intellectual disabilities, epilepsies or autisms. Our conservative diagnostic rate from the first 198 CP whole exome (WES) cohort, 98 of which were trios, was 14%. This has now been replicated by WES on our new collaborative cohort of 95 CP trios (M Kruer, Phoenix, USA; unpublished), reaching 35% or more genetic contribution to CP. Subsequent analysis of our CP cohort (n=198) for copy number variants (CNVs) resolved additional 7% of cases (7 dels; 6 dups), with some well-known neurodevelopmental loci involved, i.e. 16p11.2-12.2 or 2p25.3 deletions or 1q21.1 microduplication. Currently we are whole genome sequencing part of this CP cohort (25 trios) as well as investigating genome-wide methylation. RNA sequencing of patient cell lines of 182 cases and 120 controls revealed significant upregulation (n=132) of genes involved in the immune response and downregulation (n=320) of genes involved in cell communication and signalling. Application of Weighted Gene Co-expression Network Analysis identified a coordinated downregulation of expression of a module of 376 genes. This module is enriched for genes from the brain-derived neurotrophic factor (BDNF) signalling pathway and microglial expression. Overall the RNA sequencing data suggests a possible interplay between genetic and environmental factors in CP aetiology. Our ultimate goal is to integrate the 'omics' and clinical data to comprehensively address the aetiology of CP as a multifactorial neurodevelopmental disorder. The long term clinical applications include clinical and medico-legal assertion of causation, better genetic counselling before subsequent pregnancies, early pregnancy diagnosis and precision medicine delivery informed by the expanding knowledge of CP neurobiology. This research was supported by grants from the Australian National Health and Medical Research Council, Cerebral Palsy Alliance and Channel 7 Children's Research Foundation.

993F

Novel *TSC1/TSC2* pathogenic variants in Hungarian cohort with tuberous sclerosis complex: Clinical and molecular genetic aspects. *E. Kovesdi[1,2], K. Hadzsiev[1,2], K. Komlosi[1,3], E. Horvath[4], A. Horvath[5], M. Farkas[6], V. Farkas[6], L. Timar[7], A. Csaba[8], K. Sumegi[1,2], B. Melegh[1,2].* 1) Medical Genetics, University of Pecs, Pecs, Hungary; 2) Human Genetic And Pharmacogenomic Research Group, Szentagothai Research Centre, Pecs, Hungary; 3) Department of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, Germany; 4) Department of Medical Genetics, Faculty of Medicine, University of Szeged, Hungary; 5) EEG Diagnostic and Epilepsy Center, Markusovosky University Hospital, Szombathely, Hungary; 6) Children's Neurology, 1st Department of Pediatrics, Semmelweis University, Budapest, Hungary; 7) National Institute of Child Health, Budapest, Hungary; 8) 1st Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary.

Background: Tuberous sclerosis complex (TSC) is an autosomal dominant disorder, affecting approximately 1,600 individuals in Hungary. It is characterized by the appearance of non-malignant brain tumors, skin, eye, heart and kidney abnormalities. The most common and disabling neurological manifestation of TSC are epilepsy, developmental delay, mental retardation, behavioral disorders, and autism. **Purpose:** This study describes the broad spectrum of clinical phenotype -including disease causing mutations- of 14 Hungarian TSC patients, who showed pathogenic *de novo* genetic variants of *TSC1/TSC2* genes. **Methods:** Blood samples of patients with clinically definite or suspected TSC have been collected since 2011 at the Department of Medical Genetics, University of Pecs (Hungary). Before molecular genetic testing, all patients underwent detailed clinical examinations, including dermatology, internal medicine and neurological assessments. *TSC1/TSC2* genes of 83 Hungarian patients with clinical signs of TSC were analyzed by Sanger-sequencing. **Results:** We identified 14 cases (5 male and 9 female) with previously unpublished *de novo* pathogenic mutations in *TSC/TSC2* genes. Most of the affected organs were the central nervous system (n=13) and the skin (n=8), while only 1 patient had renal abnormalities. From the 14 cases, epilepsy (n=10), hypopigmented spots (n=8) and cortical tubers (n=7) were the most frequented clinical manifestation of the disease. Regarding the genetic testing we found 3 cases with mutations in the *TSC1* and 11 in the *TSC2*. In the case of *TSC1*, 3 different exons were affected in 3 cases. One patient had splicing, the other a nonsense mutation, and 1 was subject to an intronic deletion affecting 8 bases. In *TSC2* gene we found 1 patient with splicing, 1 with missense and 1 with nonsense mutations. Three had small, three had gross deletion, and 2 subjects with small insertion. In *TSC2* gene, exon 16 was the mostly effected region (n=3), however with completely different numbers and severity of clinical signs. **Conclusion:** Based on the severity and the number of the clinical signs of TSC, it is not possible to predict which gene and type of the mutation is involved in the clinical manifestation of the disease. However, through molecular genetic analysis, followed by detailed genotype-phenotype comparison of patients showing symptoms of TSC, one can gain significant new information about the disease, and patients can receive the right diagnosis and care.

994W

Structural and sequence characterization of SMN1 and SMN2 genes in SMA patient collection. *C. Sun, M. Liu, J. Hurt, J. Staropoli, J. Carulli, M. Penny.* Biogen Inc. 115 Broadway, Cambridge, MA 02142.

Spinal muscular atrophy (SMA) is a neuromuscular degenerate disease caused by deletion of or mutations in SMN1 gene (Survival of motor neuron 1) on chromosome 5. SMN2, a partially functional centromeric copy of SMN1 gene in the SMA locus, is the major known modifier of this disease, due to the fact each copy of SMN2 can generate 10-20% of functional SMN transcript, and copy number of SMN2 varies in human population. A greater *SMN2* copy number is usually correlated with a milder presentation of SMA. However, SMN2 copy number and SMA subtype are not perfectly correlated, thus SMN2 copy number is not a diagnostic marker**.** To investigate whether structural or nucleotide sequence variation in SMN2 contribute to this imperfect correlation, we characterized SMN1 and SMN2 genes in 39 SMA patients with known disease type from I to III, and 66 of their first degree relatives, by ddPCR and NGS approach covering all coding exons and flanking intronic sequences. Our results lend support to a recently reported modifier in intron 6 of SMN2 (A-44G). A patient with this allele has a milder disease type than SMN2 copy number alone would suggest. More importantly, our data suggests a previously unreported SNP in SMN2 may be another disease modifier, and with higher frequency than that of A-44 G allele.

995T

Heterozygous *COG4* variant causes a non-lethal type of COG4-CDG (formerly CDG-IIj). *R. Hamid[1], A. Clement[2], B. Blanco-Sanchéz[2], A. Robertson[1], L. Wolfe[3], M. Davids[3], M. Malicdan[2], J. Philips[1], J. Newman[1], J. Cogan[1], J. Phillips[2], J. Wegner[2], M. Westerfield[2], UDN collaborators.* 1) Dept Pediatrics, Div Med Gen, Vanderbilt Univ Sch Med, Nashville, TN; 2) Institute of Neuroscience, University of Oregon; 3) NIH/NHGRI, Bethesda, MD.

Congenital disorder of glycosylation COG4-CDG is a rare autosomal recessive (AR) disorder caused by mutations in the *component of oligomeric golgi complex 4* (*COG4*) gene. This disorder has been reported previously in only two individuals and is associated with a severe phenotype with dysmorphism, delays, hypotonia, liver dysfunction, CNS abnormalities including cerebral atrophy, and early death. Here we report a 22-month-old male who presented to the Undiagnosed Diseases Network (UDN) with mild developmental delay, prominent forehead, triangular face, ptosis, nystagmus, blue sclera, bilateral clubfoot, and history of IUGR with proportionate slow growth and growth hormone deficiency. He also had mild-moderate (2000-4000Hz) bilateral SNHL. MRI showed abnormally dilated semicircular canals, chiari malformation and hydrocephalus. A trio exome did not identify an obvious AR or AD candidate that could explain the patient's phenotype. A de novo novel c.1546G>A (p.Gly516Arg) missense variant was identified and Sanger validated in one allele of the *COG4* gene (NM_015386.2). This variant was rare and absent from 1KGP, EVS, ExAC and GnomAD databases. Deletion duplication and cDNA analyses were negative. Given that the only two patients reported previously had homozygous mutations and this patient had a rare novel variant in only one allele, we generated homozygous and heterozygous zebrafish models of the *COG4* deficiency by targeting CRISPRs to the region encoding the variant amino acid and also in an early exon. These two alleles shift the open reading frame, truncate the protein, and fail to complement. By day 5 post fertilization, homozygous mutants present with a shorter body axis and smaller inner ears. Morphogenesis of the semicircular canals is severely disrupted, and they have a hearing and balance defects. Craniofacial cartilage and fin malformations are also observed. Alcian blue staining shows that chondrocytes fail to stack and that extracellular matrix is affected, consistent with a defect in secretion and possibly glycosylation. Heterozygous mutants showed less severe but significant defects, consistent with haploinsufficiency. In support of this, Golgi morphology changes were observed in the patient's fibroblasts. These data suggest that this rare heterozygous *COG4* variant was likely pathogenic and expands the phenotype associated with COG4-CDG and suggests that *COG4* heterozygous mutations can also cause a phenotype, albeit milder and non-lethal.

996F

Spectrum of *TTN* variants in a patient cohort of neuromuscular disorders. *P.S. Lai, Z. Yong, S. Tomar, R. Sethi, P.S. Low, S.K.H. Tay.* Paediatrics, National University of Singapore, Singapore, Singapore, Singapore.

Titin is the largest known human protein encoded by the *TTN* gene spanning 364 exons and has been known to generate numerous isoforms through alternative splicing. Mutations in *TTN* have been associated with a wide range of disorders from cardiomyopathy to muscular dystrophy. Next gene sequencing approaches (NGS) have uncovered large numbers of missense and truncating variants in *TTN*, many of which are novel. However, there is still a lack of complete understanding in the distribution of these variants across the gene and their association with disease phenotypes. Identification of a titin causing muscle disease or cardiac phenotype is important for clinical management and/or early interventions. In this study, we analyzed the spectrum of *TTN* variations in a cohort of 130 neuromuscular disorder cases that have been sequenced using standard NGS pipelines. Identified variants from whole genomes were analyzed using known population databases such as 1000 Genomes Project, Exome Sequencing Project (ESP) and Exome Aggregation Consortium (ExAC), all known *TTN* mutation databases, ClinVar repository and *in silico* prediction tools. A total of 25,663 non-synonymous variants were identified with an average of 197.41 variants found per individual with 6.79 variants being rare (MAF<1%) and 7.62 variants being novel (MAF=0%). Missense mutations accounted for more than 59.39% of the variants. More than half of the 1373 unique variants (57.17%) identified in our cohort were either novel or rare in the general population. Most of the variants (90.60%) were non-synonymous occurring at a hotspot in the Z-disk/I-band junction, further up the I-band (exon 77-93), A-band and the A-band/M-band junction. Heterozygous potentially pathogenic *TTN* truncating variants (TTNtv) were found in 14 cases (10.77%) with 11 being novel and not predicted to generate a recessive neuromuscular condition. The high incidence of novel *TTN* variants points to the challenges in interpreting their significance in a clinical setting with many ending up being classified as VUS, and thus leading to an inconclusive diagnosis. Reporting of such variants in databases and future functional studies can facilitate further understanding on the classification of these mutations for pathogenicity effect. .

997W

Whole exome sequencing data analysis in hereditary spastic paraplegia patients from Turkey. *B. Ozes[1], O. Ari[1], Y. Parman[2], E. Battaloglu[1].* 1) Molecular Biology and Genetics, Bogazici University, Istanbul, Turkey; 2) Istanbul University, Istanbul Medical School, Department of Neurology, Istanbul, Turkey.

Hereditary Spastic Paraplegia (HSP) is defined as a rare group of clinically and genetically heterogeneous neurodegenerative disorders. Main symptoms are lower limb spasticity and progressive weakness manifested in pure type of the disease. In complicated HSP cases, neurological and non-neurological symptoms can be observed, additionally. Disease might be inherited in autosomal dominant, autosomal recessive, or X-linked manner. Fifty-one loci and forty-three genes are associated with the autosomal recessive form of HSP (ARHSP). In this study, we have investigated twenty families with autosomal recessive HSP inheritance. With the aim of identifying causative genes and variants in these families, whole exome sequencing (WES) was performed for one affected individual from each family. Known HSP genes and genes related to other neurological diseases (app. 1800 genes) were analyzed initially in the WES data. Missense, non-synonymous and splice site variants, predicted as 'effective' by *in silico* tools and with MAF value less than 1% were chosen as candidate variants. Sanger sequencing was performed for the family members to understand inheritance pattern of the variants in the family. The causative variants were identified in seven of the families. These were c.1235C>G, c.6215_6219dupAGAT and c.3036C>A in *SPG11* gene, c.825T>A in *CYP7B1* gene, c.2104-2A>G in *SPG7* gene, c.437A>T in *CCT5* gene, and c.637G>A in *SACS* gene. The segregation analyses for the identified candidate variants are currently being performed for other thirteen families. At this stage of the study, genetic diagnosis rate is 35%. This study shows that WES is an effective technique for genetic diagnosis of HSP since causative variants in both HSP and other neurological disease-associated genes can be identified. Besides, it helps identification of overlapping genes in neurological diseases that is an important contribution to fulfill the overall picture of mechanism underlying neurodegeneration.

998T

Mutations of the *ZNF292* gene are a novel cause of neurodevelopmental disability, behavioral problems, and autism spectrum disorders (ASD). *G. Mirzaa[1,2], A. Piton[3], K. Aldinger[1], K. Gripp[4], L. Walsh[5], K. Payne[5], N. Zeid[6], A. Yang[6], B. Popp[7], M. Cho[8], M. Bamshad[2], University of Washington Center for Mendelian Genomics.* 1) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 2) Department of Pediatrics, University of Washington, Seattle, Washington; 3) Laboratoire de diagnostic génétique, Institut de Génétique Médicale d'Alsace, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; 4) Department of Pediatrics, AI duPont Hospital, Wilmington, Delaware, USA; 5) Riley Hospital for Children, Indianapolis, Indiana, USA; 6) Department of Genetics and Genomics Sciences, Icahn School of Medicine, Mount Sinai, New York, NY, USA; 7) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany; 8) GeneDx, Gaithersburg, Maryland, USA.

   Intellectual disability (ID) and Autism Spectrum Disorders (ASD) are highly genetically heterogeneous neurodevelopmental phenotypes, and the genetic landscape of these disorders continues to rapidly expand with the wide use of Next Generation Sequencing (NGS) methods. By far, most of the mutations underlying ID and ASD are *de novo*. However, dominant, recessive and maternally-inherited X-linked etiologies have been identified as well. Using whole exome sequencing, we identified *de novo* and dominant mutations in Zinc Finger Protein 292 (*ZNF292),* in 16 unrelated families ascertained on ID in both clinical and research settings. Affected individuals also had behavioral issues including attention-deficit-hyperactivity disorder (ADHD), and ASD, among others. Fourteen mutation-positive families were sporadic (i.e. no sibling recurrence), whereas sibling recurrence was seen in two families suggesting either parental germline mosaicism or autosomal dominant inheritance. All identified *ZNF292* mutations in this series were truncating (frameshift, nonsense) suggesting loss of gene function (LoF). Mutations were distributed among several critical domains of the gene, with several occurring within critical DNA binding domains.   *ZNF292* encodes a transcription factor that has possible roles in tumor development as a tumor suppressor. Somatic frameshift mutations in *ZNF292* have been identified in tumors of the liver, colon and bone marrow. *ZNF292* is highly expressed in the human brain, with the highest expression identified during the prenatal period. In summary, we report discovery of a novel gene underlying ID, ASD and neurobehavioral issues. We anticipate that mutations of *ZNF292* will be identified in additional children and may in fact be a relatively common single-gene cause of ID.

999F

Correction of *NAGLU* mutation p.R297X using CRISPR/Cas9 gene editing in mucopolysaccharidosis IIIB patient-derived iPSCs. *C.L. Christensen[1,2], S.S. Chu[1,2], F.Y.M. Choy[1,2].* 1) Department of Biology, University of Victoria, Victoria, B.C., Canada; 2) Centre for Biomedical Research, University of Victoria, Victoria, B.C., Canada.

   Mucopolysaccharidosis IIIB (MPS IIIB) is a rare autosomal recessive lysosomal storage disease (LSD), wherein affected individuals experience progressive neurodegeneration and succumb to the disease within the first three decades of life. Mutations in *NAGLU* result in the production of non-functional N-acetyl glucosaminidase (Naglu), an enzyme that is responsible for breaking down heparan sulfate. In MPS IIIB, heparan sulfate accumulates within lysosomes, leading to disease manifestation. There are currently no treatment options for MPS IIIB due to the inability of recombinant *NAGLU* to cross the blood-brain barrier, in order to reach the central nervous system and alleviate the aforementioned neurodegeneration.   CRISPR/Cas9, a recent tool in biotechnology with myriad applications, allows for precise genome editing driven by homology directed repair following double stranded cleavage. This technique has been used for targeting and correction of patient-specific mutations in a number of cell types. CRISPR/Cas9 genome editing, in conjunction with reprogrammed patient-derived somatic cells to induced pluripotent stem cells (iPSCs) has been explored as a regenerative medicine for genetic disease. Edited iPSCs can be differentiated to neural progenitor cells and intracerebrally transplanted into the patient. This process circumvents the blood-brain barrier to deliver continuous functional Naglu via mannose-6-phosphate receptors to neighbouring neurons.   As proof-of-concept, we have generated iPSCs from MPS IIIB patient skin fibroblasts using transcription factors c-Myc, Sox2, Klf4, and Oct3/4 and identified candidate guide RNAs (gRNA) for targeting and correction of a common *NAGLU* mutation, p.R297X, using CRISPR/Cas9. Preliminary Cas9:gRNA lipofection of iPSCs, along with a single-stranded oligonucleotide (ssODN) correction template, resulted in a correction efficiency of 0.3%, as confirmed by restriction enzyme digest and Sanger sequencing. By using second generation CRISPR-Cas9 (Alt-R™ two-part system, IDT) in combination with an asymmetrical ssODN (*Nature Biotechnology* **34**, 339–344 (2016)), we are aiming to improve the correction efficiency of p.R297X, characterize corrected iPSCs, and target three other *NAGLU* mutations in similar fashion. This work will allow for a better understanding of the applications and limitations of this technique as a potential tool in regenerative medicine for MPS IIIB. .

1000W

Progressive abnormal myelination and cerebrospinal fluid volume in canine mucopolysaccharidosis type I: A neuroimaging and neuropathological study. *P. Dickson[1], S. Chen[2], I. Nestrasil[3], S.Q. Le[1], S.-h. Kan[1], J. Jens[4], E. Snella[4], N.M. Ellinwood[4], M. Guzman[5], L.E. White[6], J.D. Cooper[1], M. Egeland[1], C. Vite[7], J. Provenzale[2,8].* 1) Dept Pediatrics, LA BioMed Harbor-UCLA Med Ctr, Torrance, CA; 2) Department of Radiology, Duke University, Durham, NC; 3) Department of Pediatrics, University of Minnesota, Minneapolis, MN; 4) Department of Animal Science, Iowa State University, Ames, IA, USA; 5) Department of Pathology, Saint Louis University School of Medicine, Saint Louis, MO; 6) Department of Orthopedic Surgery, Duke University, Durham, NC; 7) Department of Clinical Studies, University of Pennsylvania School of Veterinary Medicine, Philadelphia, PA; 8) Department of Radiology, Oncology & Biomedical Engineering, Emory University, Atlanta, GA.

We performed neuroimaging studies in a canine model of the lysosomal storage disease, mucopolysaccharidosis type I (MPS I). Brain imaging findings in MPS I patients and dogs include hydrocephalus, white matter hyperintense lesions, enlarged Virchow-Robin spaces, and cortical atrophy. Previously, we found abnormal myelination in the corpus callosum of adult MPS I dogs compared to controls, with reduced fractional anisotropy measured by diffusion tensor imaging and abnormal myelin composition (Provenzale et al, 2015). In the present study, we performed serial neuroimaging of a cohort of 5 MPS I and 6 heterozygous (normal) control dogs using magnetic resonance imaging and DTI prospectively. Imaging was performed on a 3Tesla Siemans Prisma scanner with a 16-channel knee coil at age 3, 6, 12, 18, and 24 months. T1-weighted, T2-weighted, and diffusion weighted images were acquired. For T1 weighting we applied Three Dimensional Magnetization Prepared Rapid Acquisition Gradient-Echo sequence with parameters TR/TE/TI 2530/3.48/1100 ms, flip angle 7 degrees, 1 average, isotropic voxel with slice thickness 0.7mm, 144 slices, FoV phase 100%; scan duration was 10 min. 48 sec. For T2 weighting we applied spin-echo sequence with following parameters TR/TE 3200/563 ms, 1 average, isotropic voxel with slice thickness 0.7mm, 144 slices, FoV phase 100%, acceleration factor GRAPPA of 2; scan duration was 6 min. 22 sec. For diffusion weighting we acquired echo-planar imaging sequence with TR/TE 4900/78 ms, GRAPPA acceleration factor of 2, 1.4 mm in plane resolution, FoV phase 100%, 62 diffusion directions, 40 1.5mm-thick slices. Five dogs died during the study, one under anesthesia (due to dislodged endotracheal tube) and four from progressive MPS I disease. We found that fractional anisotropy in the genu and splenium of the corpus callosum increased in control dogs over time as expected, but decreased over time in MPS I dogs. We performed compartment segmentation to measure volume of cerebrospinal fluid. MPS I dogs show larger volumes of cerebrospinal fluid versus controls at all ages, and these also increased over time. The expression of myelin-related genes was reduced in MPS I dogs versus controls at 6 weeks, 4 months, 12 months, and 18-24 months of age. We conclude that myelination abnormalities in the corpus callosum of MPS I dogs may be progressive, and that cerebrospinal fluid volume may be increased from an early age.

1001T

Magnetic resonance spectroscopy and lipid profiling of myelin composition in corpus callosum of mucopolysaccharidosis I mice. *S. Le[1], S-H. Kan[1], J. Yee[1], J. Tolar[2], I. Tkac[3], I. Nestrasil[2], P. Dickson[1].* 1) Dept Pediatrics, LA Biomed Harbor-UCLA, Torrance, CA; 2) Dept Pediatrics, University of Minnesota, Minneapolis, MN; 3) Department of Radiology, University of Minnesota, Minneapolis, MN.

Brain MRI frequently shows hyperintense lesions of white matter in patients with the lysosomal disease, mucopolysaccharidosis (MPS) type I that are thought to be due to demyelination. Previously, we reported abnormal myelin composition and reduced expression of myelin-related genes in a canine model of MPS I. Here, we studied myelin composition in the corpus callosum of MPS I mice. We performed a cross-sectional study of male MPS I mice and controls at 52 weeks of age. In vivo ^{1}H MR spectroscopy data were acquired from a 3.6 μL volume centered in the corpus callosum using ultra-short echo-time STEAM (TE = 2 ms) at 9.4T. Seventeen brain metabolites were reliably quantified in vivo using LCModel analysis. Mice were sacrificed after scanning and corpus callosum tissue was dissected from whole brains harvested after imaging. Comparison of neurochemical profiles revealed significantly decreased concentrations of N-acetylaspartate (marker of neuronal and axonal integrity, -0.5 μmol/g, p = 0.004) and of phosphoethanolamine (precursor of myelin, -0.6 μmol/g, p = 0.018) in MPS I mice relative to controls (heterozygote littermates). In addition, a trend for increased level of ascorbate (antioxidant, +0.5 μmol/g, p = 0.086) was observed. Total polar lipid content and ceramide/hexosylceramide content was analyzed by LC-MS/MS. Twenty-two compounds showed significant differences between MPS I and control mice at p and q < 0.05. The groups separated by principal components analysis. Phosphoethanolamine levels measured by ^{1}H MRS showed a linear relationship to tissue phosphatidylethanolamine levels (r^2 = 0.17, p = 0.04). There were no intergroup differences in corpus callosum myelin gene expression, MBP, cholesterol, or protein. While intergroup differences in MPS I mice compared to controls were subtle, noninvasive, quantitative measurements of the corpus callosum integrity represent a tempting target for monitoring therapeutic efficacy in MPS I patients. .

1002F

Evaluation of sibling pairs with Gaucher disease discordant for Parkinsonism. *G. Lopez[1], A. Steward[1], G. Monestime[2], C. Groden[1], E. Wiggs[1], N. Tayebi[1], T. Roshan Lal[1], E. Sidransky[1].* 1) NHGRI/NIH, Bethesda, MD; 2) University of Massachusetts Medical School, Worcester, MA.

The association between mutations in the gene encoding glucocerebrosidase (*GBA1*) and the development of parkinsonism has been well established. However, most *GBA1* mutation carriers never develop parkinsonism, implicating contribution of other genetic modifiers in *GBA1*-associated Parkinson disease (GBA-PD). Studies of sibling pairs where both have Gaucher disease (GD) but are discordant for PD (DSP) could help to elucidate some of these modulating factors. In this prospective longitudinal study, nine DSP were evaluated through the Genetics Clinic at the National Institutes of Health. Participants provided relevant medical history, including family history and pedigrees, in addition to undergoing clinical and research evaluations. Clinical assessments included neurological, neuropsychological, olfactory, motor, and non-motor testing, including validated questionnaires to determine the presence and/or severity of PD motor or prodromal symptoms. Several of the DSP were followed longitudinally for over a decade. Multiple genotypes were detected among the DSP. There was no relationship between treatment with enzyme replacement therapy, substrate reduction therapy, genotype, or splenectomy and development of PD symptoms. General severity of Gaucher disease, determined by the degree of skeletal involvement, hematological abnormalities, and organomegaly, varied widely across the sample. However, in the majority of pairs, the younger sibling had PD and milder Gaucher symptoms. No parkinsonian features were observed in any of the non-PD siblings. Continued longitudinal evaluation, next generation sequencing, and ongoing recruitment of additional DSP will be useful in identifying genetic, environmental, and epigenetic factors contributing to phenotypic expression that may diminish or augment risk for PD.

1003W

Phenotypic profiles of *GBA1* mutation carriers with and without Parkinson disease: A data-driven approach. *A.M. Steward[1], L.M. Koehly[1], N.S. Trivedi[1], T.M. Lindstrom[1], G. Monestime[2], N. Tayebi[1], E. Sidransky[1], G.J. Lopez[1].* 1) National Institutes of Health, Bethesda, MD; 2) University of Massachusetts Medical School, Worcester, MA.

Mutations in the glucocerebrosidase gene, *GBA1,* represent the number one genetic risk factor for the development of Parkinson disease (PD). However, not every individual who carries a *GBA1* mutation goes on to develop PD. Careful phenotyping has revealed that there are specific prodromal symptoms of PD, including a decreased sense of smell, depression, and rapid eye movement (REM) sleep behavior disorder (RSBD). Among *GBA1* carriers, a population with a higher-risk for PD, we aimed to classify individuals by symptom profiles using a data-driven approach. Latent class cluster analysis is a computational method which uses Bayesian probabilities to determine the likelihood that an individual within a larger sample belongs to a certain underlying distribution. Clustering was performed on a sample of individuals carrying at least one *GBA1* mutation with and without PD, and individuals without a *GBA1* mutation with PD. The Fatigue Severity Scale (FSS), Geriatric Depression Scale (GDS), University of Pennsylvania Smell Identification Test (UPSIT), State Trait Anxiety Inventory (STAI), and demographic variables were used as predictors for the model. Number of clusters and goodness-of-fit were determined and assessed using the BIC score. Cluster profiles were compared with *GBA1* mutation and PD status to determine whether cluster assignment aligned with clinical and molecular diagnoses. Discriminant analyses were also performed to determine whether the predictors used in the cluster model could independently predict *GBA1* mutation and PD status. This study represents the first use of latent class cluster analysis to analyze phenotypic profiles of *GBA1* mutation carriers in an objective and unbiased manner.

1004T

The Lysosomal Disease Network. *C.B. Whitley, B. Diethelm-Okita, J.R. Jarnes, J. Cloyd, the Lysosomal Disease Network.* University of Minnesota, Minneapolis, MN.

   The combined, integrated efforts of researchers involved in the Lysosomal Disease Network (LDN) continue to focus on using limited resources to create a network of clinical research centers focused on solving the major challenges in diagnosis, disease management, and therapy for these complex, rare disorders. Finding solutions to both the clinical and social problems associated with lysosomal disease will have direct impact on patients and inform medical practice. Toward that goal, the LDN facilitates clinical research in rare diseases through a program that supports (1) collaborative research via nine longitudinal and five pilot project studies that each strive to find answers to the fundamental issues of a given disease, (2) supportive cores providing integral services to investigators across the network, (3) train clinical investigators in lysosomal disease research, (4) provide access to information related to lysosomal conditions by submission of data to the Data Management and Coordinating Center (DMCC), (5) involvement of patient advocacy groups (PAG) in the recruitment, scientific advisement, and transmission of research results to the community at large. Because central nervous system (CNS) involvement is prevalent in lysosomal conditions—while being difficult to measure and treat—we include a major emphasis on qualitative analysis of CNS structure, function, and the discovery of biomarkers for relevant conditions including: mucopolysaccharidosis (MPS), mucolipidosis IV, Batten disease, and the gangliosidoses (including Tay-Sachs disease). Systematically-collected data on brain effects will help us develop outcome measures that are specific and sensitive to change in disease status. To drive cutting-edge research in what are multi-system diseases, we also have projects focused on immune modulation in Pompe disease, kidney structure and function in Fabry disease, the search for undiagnosed Fabry patients in high-risk populations, and the role of oxidative stress in Gaucher disease. We also continue to emphasize the collection of natural history data for those lysosomal conditions for which there are no current therapies: Krabbe disease, mucolipidosis type IV, and Sanfilippo syndromes type C and D. The Lysosomal Disease Network (U54NS065768) is part of Rare Diseases Clinical Research Network funded through collaboration between NCATS, the National Institute of Neurological Disorders and Stroke (NINDS), and the National Institute of Diabetes.

1005F

A novel pathogenic variant of *PURA* in a patient with severe developmental delay, delayed myelination and empty sella. *K. Hosoki[1], T. Kimura[2], K. Yanagi[1], M. Iso[1], Y. Kuroki[1], H. Ogata[3], K. Nakabayashi[3], K. Okamura[4], K. Hata[3], T. Shinozaki[2], Y. Matsubara[5], T. Kaname[1].* 1) Genome Medicine, National Center for Child Health and Development (NCCHD), Tokyo, Japan; 2) Nihonkai General Hospital, Yamagata, Japan; 3) Department of Maternal-Fetal Biology, NCCHD, Tokyo, Japan; 4) Department of Systems Biomedicine, NCCHD, Tokyo, Japan; 5) National Center for Child Health and Development, Tokyo, Japan.

   Introduction: Whole exome sequencing (WES) of case-parent trios is a powerful method to identify genetic causes in undiagnosed patients. Several reports employing such method revealed that patients with severe neurodevelopmental delay, learning disability, neonatal hypotonia, feeding difficulties, abnormal movements and epilepsy were caused by pathogenic variants in *PURA*. Here we repot a patient with severe developmental delay, neurological and endocrinological abnormalities associated with a novel pathogenic variant in *PURA* identified by WES analysis. **Case Presentation**: The patient was a 9-year-old boy. The boy was born at 39 weeks of gestation as the first child of healthy parents. His birthweight and Apgar score was 2210 g and 8/9, respectively. The boy showed nystagmus from infancy with delayed myelination in the brain. He was suspected to Pelizaeus-Merzbacger syndrome at 7 months of age, but it was excluded. He presented with severe developmental delay, intellectual impairment with no meaningful words, short stature, empty sella and pseudobulbar palsy. **Methods**: Written informed consent was obtained from his parents. Whole exome sequencing analysis was performed using Human All Exon V6 kit (Agilent) and HiSeq 2500 (Illumina). The detected variants were confirmed by Sanger sequences. **Results**: A novel heterozygous 8-bp insertion in the *PURA* gene was found in the patient. The variant was *de novo* and it causes frameshift and premature stop codon. **Discussion**: We found the novel frameshift variant at 5'-side of the PUR repeat region I related to severe developmental delay. *PURA* is one of the primary responsible genes for 5q31.3 microdeletion syndrome. Twenty-one patients with neurodevelopmental disorder and brain abnormality caused by *de novo* variants in *PURA* have been reported. Our finding and those reports suggest that loss-of-function of PURA causes the disease.

1006W

Novel *NTRK1* gene mutation and clinical report of HSAN-IV phenotype in a Mexican patient. *L. Patron[1], BL. Camacho[1], VG. Patron[2], HE. Almanza[1].* 1) Universidad Autónoma de Baja California, Calzada University 14418, UABC, Parque Internacional Industrial Tijuana, 22390 Tijuana, B.C; 2) Rush University Medical Center, Psychiatry Department, 1645 W Jackson Blvd Suite 600, Chicago, IL 60612, EE. UU.

The hereditary sensory and autonomic neuropathy Type IV (HSAN-IV; MIM #256800), also known as congenital insensitivity to pain and anhidrosis (CIPA) is a rare condition; its frequency is estimated to be 1 in 125 millions, with an autosomal recessive patter. It is caused by mutations in the neurotrophic tyrosinase kinase type 1 *NTRK1* gene (MIM*191315). The clinical diagnosis criteria include: (1) Lack of perception of painful stimulus from birth onward, with unimpaired touch and pressure sensitivity, (2) anhidrosis, and (3) moderate to severe mental retardation. We present the case of a six year-old Mexican male. The patient is the only child of first degree cousins. He was born by C-section of full term pregnancy. Neonatal screening was reported normal. His mother noticed symptom within 48 hours after birth, when he developed fever every time he covered with clothes, but no sweeting. Also, the patient did not cry when vaccinated. The patient was referred to our center because of chronic arthritis, episodes of fever, self-mutilating behavior and mental retardation. Because of the symptoms and the suspicion of Lesh-Nyhan Syndrome, uric acid was levels were measured, but levels were within normal values. Physical findings included absence of pain response, warm and dry skin. Facial asymmetry secondary to a previous jaw fracture. He showed normal tactile sensation, lacrimation, corneal reflex and anisocoria. There was absence of teeth due self-extraction, absence of left nasal alae due amputation, severe tropic changes in the skin of distal limbs, hyperkeratosis, fissuring and absence of distal phalange in first and second digit of the right hand. He was able to walk but running was limited because of a right tibial fracture and edema secondary to chronic arthropathy. Basic laboratories evaluations were normal. Skin biopsy demonstrated atrophic changes and absence of sweat glands. The molecular genetic analysis detected a homozygous pathogenic variant in the *NTRK1* gene. DNA Sequencing results showed a novel transversion of a guanine to a thymine (NM_001007792.1:c.334G>T), this substitution resulted in a nonsense mutation due substitution a glutamine for a stop codon (NP_001007793.1:p.Glu112*). This pathogenic variant has not yet been described in the international database. The treatment of HSAN-IV is focused on the prevention of injuries and arthropathy, making early recognition is important. Currently there are no guidelines available to treat this disease.

1007T

Further clinical and molecular characterization of the novel autosomal recessive neurodegenerative disorder related to the *ATP8A2* gene. *A. Telegrafi[1], A. Singleton[1], M. Cho[1], C. Mignot[2], D. Doummar[3], C. Nava[2], B. Keren[2], J. Picker[4], S. Lincoln[4], A. Schreiber[5], H. McMillan[6], G. Yoon[7], L. Velsher[8], I. Thompson[8], J. Griffin[9], A. Asamoah[9], K. Retterer[1], C. Forster[1], D. McKnight[1], L. Vincent[1], G. Douglas[1], R. Person[1], L. Henderson[1], M.J. Guillen Sacoto[1], H. McLaughlin[1], J. Juusola[1].* 1) GeneDx, Gaithersburg, MD, United States; 2) Département de Génétique (Pr Soubrier), Hôpital Pitié-Salpêtrière, Paris, France; 3) Service de Neuropédiatrie, Hôpital Armand-Trousseau, Paris, France; 4) Boston Children's Hospital, Boston, MA, United States; 5) Cleveland Clinic, Cleveland, OH, United States; 6) Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 7) Hospital for Sick Children, Toronto, ON, Canada; 8) North York General Hospital, Toronto, ON, Canada; 9) University of Louisville, Louisville, KY, United States.

The *ATP8A2* gene (MIM# 605870) is highly expressed in the retina, spinal cord, and cerebellum with knockout mice showing central and peripheral nervous system abnormalities such as progressive ataxia and neurodegeneration. Previously reported cases with *ATP8A2* variants include compound heterozygous and homozygous variants in patients with intellectual disability, movement disorders, optic atrophy, and hypotonia. Here we describe six unrelated patients ranging in age from one to nine years with clinical features consistent with those previously described.We looked at 11,749 exomes with a neurodevelopmental phenotype, and identified five unrelated affected individuals to be either compound heterozygous or homozygous for variants in the *ATP8A2* gene. A sixth patient was identified through GeneMatcher. Movement disorders in our patients included: chorea in two patients, dystonia in two patients, ataxia in two patients, and tremor in one patient. Additional features reported in our patients include brain abnormalities in two patients, one of whom had macrocephaly, and dysmorphic features, hypermobility, sleep abnormalities, seizures, reported in a single patient each. The variants identified ranged from missense, splicing, intragenic deletions, and frameshift, including: p.W702R, p.K429M, p.D49Y, p.D765Y, p.D428A, c.2211+1G>A, c.1185+5G>A, intragenic deletion of exons 28-33, c.1787delA, c.321+3_321+8delAATGGT. All variants were seen with insignificant frequency in the Exome Aggregation Consortium (ExAC) public database and the GeneDx exome database of unaffected individuals. All variants occurred within highly conserved domains, and were predicted to be damaging or possibly damaging from multiple in silico models. The findings presented in this case series add to the already described *ATP8A2*-related disorder by further characterizing this autosomal recessive neurodegenerative disorder molecularly and phenotypically.

1008F

Evaluating clinical and biochemical endpoints for therapy in Pex7 deficient mouse models. *W. Fallatah[1,2], W. Cui[1], E. DiPietro[1], N. Braverman[1].* 1) Human Genetics, McGill University - Research Institute of the MUHC, Montreal, Quebec, Canada; 2) Department of Medical Genetics, King AbdulAziz University, Jeddah, Saudi Arabia.

Background: Rhizomelic chondrodysplasia punctata type 1 (RCDP1) is a peroxisome biogenesis disorder clinically distinguished by skeletal dysplasia with significant growth deficiency, severe neurological impairment and congenital cataracts. RCDP1 is caused by mutations in *PEX7*, which encodes the cytosolic receptor for peroxisomal matrix proteins carrying the Peroxisomal Targeting Signal 2 (PTS2), resulting in reduced biosynthesis of ether phospholipid compounds, plasmalogens (PL). PL deficiency is the primary cause of RCDP pathology and the severity of the disease directly correlates to residual PL levels. Currently, there is no definitive treatment. Our lab generated and reported a hypomorphic Pex7 mouse model that resembles mild RCDP. To fully understand the RCDP spectrum, we have now generated two additional substrains that represent the intermediate and severe forms of the disease due to a differential decrease in Pex7 transcript and protein levels. **Aim:** To characterize the behavioral, biochemical and neurological phenotypes of our Pex7 mouse models to get ready for future therapies. **Results:** We analyzed PL species in erythrocytes, cerebral cortex and cerebellum of Pex7 mice using LC/MSMS technology and found a significant decrease in total PL levels that correlated to the clinical severity of the Pex7 substrain. In contrast to what is known in human RCDP, very long chain fatty acids (VLCFA) were mildly elevated in some tissues and correlated inversely with the severity of PL deficiency. We evaluated the general locomotor activity of Pex7 mice in the open field environment and found that our Pex7 mouse models - in comparison to their littermate controls - exhibit a hyperactive behavior measured by their increased total distance traveled (P<0.001) and mobility time (P<0.0001), again correlating with the levels of Pex7. Lastly, through a histological evaluation of the brain, we found a reduction in Purkinje cell number in Pex7 mice compared to littermate controls. **In conclusion**, we have established biochemical and neurobehavioral characterizations of our Pex7 mouse models that are useful as valuable preclinical endpoints to determine efficacy in future clinical trials. The Pex7 mouse models also allow us to determine the minimal amount of Pex7 transcript needed to improve the disease.

1009W

Mitochondrial accumulation and increased lipid metabolism in a *Dhtkd1[Tyr486*]* knock-in mouse model of the CMT2Q neuropathy. *M. Gu[1], C. Luan[1], W. Guo[1,4], L. Chen[1], X. Wei[1], Y. Kuang[2], S. Dang[1], L. Van Den Bosch[4], Z. Wang[1,2,3].* 1) Department of Medical Genetics, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 2) Shanghai Research Centre for Model Organisms, Shanghai, China; 3) Research Center for Experimental Medicine, Rui-Jin Hospital at SJTUSM, Shanghai, China; 4) KU Leuven, Laboratory of Neurobiology, Experimental Neurology, Department of Neurosciences and VIB Center for Brain and Disease Research, 3000 Leuven, Belgium.

Charcot-Marie-Tooth disease (CMT) is also known as 'hereditary motor and sensory neuropathy' and is one of the most common inherited disorders of the peripheral nervous system with a prevalence of at least 1 in 2500. Clinically, CMT is characterized by progressive muscular and sensory defects starting at the distal extremities with chronic atrophy and weakness. CMT is divided into two main types: a demyelinating form called CMT1 and CMT2, which is the axonal form. Genetically, CMT is a highly heterogeneous disease with more than 80 disease-associated genes identified to date. Nearly 30 genes have been identified as a cause of CMT2. One of these is the 'dehydrogenase E1 and transketolase domain containing 1' (DHTKD1) gene. We previously demonstrated that a nonsense mutation [c.1455T>G (p.Tyr485*)] in exon 8 of *DHTKD1* is one of the disease-causing mutations in CMT2Q (MIM 615025). The aim of this study was to recapitulate the CMT2Q phenotype and to investigate the physiological function and pathogenic mechanism of the Dhtkd1 gene *in vivo*. Therefore, we generated a knock-in mouse model with the *Dhtkd1[Tyr486*]* mutation. We observed that the Dhtkd1 expression level in liver and sciatic nerve of homozygous (mt/mt) mice was significantly lower than in the wild-type (wt) mice. Moreover, the ATP production in liver of mt/mt mice was significantly lower than in wt mice. We also observed abnormal peripheral nerve phenotypes, including reduced axon diameter and excessive myelination. The mt/mt mice also displayed clear sensory defects, while no abnormalities in the motor performance were observed. Apparently, the lack of energy production is compensated by an accumulation of mitochondria, an enhanced lipid metabolism in the muscle of knock-in mice. Taken together, we show that the *Dhtkd1[Tyr486*]* knock-in mouse model partially recapitulated phenotypes of CMT2Q, especially related to the sensory abnormalities. We hypothesize that the lack of a more pronounced CMT phenotype might be due to a mechanism that compensates for the defects in energy metabolism. This work is partially supported by National Natural Science Foundation of China (Proj. No. 30470951, 31071107).

1010T

Myotonia congenita with a novel missense mutation in *CLCN1* gene (*c.680T>A*, p.Ile227Asn). *A. Kiraz[1], M. Erdogan[1], B. Balta[1], M. Tuna[2], G. Akinci Gönen[1].* 1) Department of Medical Genetics, Health Sciences University, Education and Research Hospital, Kayseri, Turkey; 2) Department of Neurology, Health Sciences University, Education and Research Hospital, Kayseri, Turkey.

   Myotonia congenita (MC) is a group of genetically and clinically heterogeneous congenital neuromuscular disease characterized by muscle stiffness and an inability of the muscle to relax after voluntary contraction, hypertrophy, transient weakness and cramping. It is caused by a mutation in the gene encoding skeletal muscle chloride channel-1 (*CLCN1*). More than 120 mutations have been found in this gene. Here, we report a case of 19-year-old female who exhibited myotonia with periodic paralysis. She has complained of congenital myotonia since she was a child. We analyzed *CLCN1* sequence in this patient and we found a new homozygous missense mutation in *CLCN1* gene (*c.680T>A*, p.Ile227Asn). *In silico* analyses (Mutation Taster, PolyPhen-2, SIFT) this nucleotide change is reporting as a mutation that highly likely affecting the disease. We plan to analyse *CLCN1* sequence on the other members of her family. This new mutation may help identify genetic determinants of the disease as well as clarify genotype-phenotype relationships. Also this variation can be used for prenatal diagnosis in this family.

1011F

Is the association of heterozygous variations in *MORC2*, *MFN2* and *AARS* genes responsible for a severe axonal form of Charcot-Marie-Tooth disease? *A. Lia[1,2], F. Miressi[1], H. Dzugan[1,2], P. Cintas[3], C. Magdelaine[1,2], F. Sturtz[1,2].* 1) Limoges University, Limoges, France; 2) Limoges Hospital, Limoges, France; 3) Toulouse Hospital, Toulouse, France.

   Charcot-Marie-Tooth (CMT) disease is one of the most frequent inherited peripheral neuropathies (1/2500). So far, mutations in more than 50 genes have been identified causing either the demyelinating form (CMT1) or the axonal form (CMT2). While causative mutation is found in 70% of CMT1 patients, only 30% of the CMT2 patients obtain a positive diagnosis. We present here the complex genetic analysis of a family presenting two CMT2 cases. Patient A (mother), a 56-years-old woman of European origin, presents axonal impairment and patient B (daughter), a 23-years-old woman, is more severely affected. Patients C, D and E (respectively the maternal grandmother and the maternal aunts of Patient B) are unaffected. Targeted NGS sequencing screening 89 genes involved in CMT and associated neuropathies was first performed on Patient B to detect pathogenic SNVs. Using Cov'Cop, CNVs were then looked for and confirmed by real time PCR. Last step consisted in a trio exome sequencing (Patients A, B and D). In Patient B, targeted sequencing allowed to detect the heterozygous c.1403G>A (p.Arg468His) mutation in *MFN2* gene. This mutation already described is usually associated to discreet or mild CMT phenotype. Considering the severe symptomatology of Patient B, and the absence of p.Arg468His in her affected mother, we enriched our study with CNVs research in Patient B and found complete duplication of *AARS* gene, whose mutations had been previously associated to CMT2. This duplication is present in the affected mother, absent in the unaffected grandmother, but present in the unaffected aunts. Whole exome sequencing showed then the heterozygous missense mutation c.568C>T (p.Arg190Trp) in *MORC2* gene, which was present in the affected individuals (A and B) and absent in the healthy ones (D, but also C and E). This mutation, already described, is responsible for classical CMT2. In conclusion, the complete *AARS* duplication, which has never been described so far, seems not to be involved in the pathological mechanisms of the disease, since it is also present in two unaffected individuals. According to our findings and previous results, we can assume that the heterozygous *MORC2* mutation is the main cause for the axonal neuropathy in Patient A. The newly described association of *MORC2* and *MFN2* heterozygous missense mutations seems to be responsible for the more complex and severe phenotype observed in Patient B.

1012W

Development of new strategies for the treatment of hereditary cystatin C amyloid angiopathy (HCCAA). *A. Gutierrez-Uzquiza, M. March, L. Matsuoka, K. Nguyen, H. Hakonarson.* The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Hereditary cystatin C amyloid angiopathy (HCCAA) is a dominantly inherited disease caused by a leucine 68 to glutamine variant of cystatin C (L68Q-cystatin C). This cystatin C variant forms amyloid deposits in brain arteries and arterioles that are associated with changes in the arterial wall structure and deposition of extracellular matrix proteins. Most carriers of the mutation suffer micro-infarcts and brain haemorrhages, ultimately leading to paralysis, dementia and death in young adults. Post-mortem studies of patients have shown that mutant cystatin C is deposited in all brain areas. Although the mechanism of action of this disease is known, no treatment to avoid early death by brain hemorrhage is available. The cystatin C amyloid fibrils are generated by propagated domain swapping between cystatin C monomers in which N- and C-terminal parts of the monomers are exchanged, forming long protein chains. The purpose of this study is to develop a cellular model that allows testing the ability of drugs to reduce the formation of cystatin C dimers/oligomers and to identify substances potentially useful for the treatment of patients with HCCAA.   Much of the previous work describing aggregation of the L68Q-cystatin C and identification of compounds to reduce it have been performed using non-physiological systems. We aimed at developing a system that would allow evaluation of the ability of a molecule to interfere with aggregation of mutant cystatin C while also given some information about toxicity to cells or organisms. To this end, we have generated genetically engineered 293T cells with expression of human wild type (wt) and L68Q cystatin C. These cells produce and secrete detectable levels of cystatin C (wt or L68Q) capable of oligomerizing under non-reducing conditions. They constitute a perfect model to test the ability of different compounds to decrease the level of oligomers and fibrils of L68Q Cystatin C. Interestingly, we have found that short incubation with reducing agents breaks oligomers into monomers of both intracellular and secreted cystatin C L68Q.   In summary, this study establishes proof-of-principle that mutant cystatin C is a pharmacological target for reducing redox agent, and most importantly, provide evidence that redox agents could potentially useful to treat this devastating disease and related conditions.

1013T

Search for target genes of transcriptional regulation by dentatorubral-pallidoluysian atrophy protein (DRPLAp) that acts as transcriptional co-regulator. *K. Hatano[1], H. Ishiura[1], H. Date[1], J. Goto[3], M. Tanaka[1], J. Mitsui[1], J. Yoshimura[2], K. Doi[2], S. Morishita[2], S. Tsuji[1].* 1) Department of Neurology, University of Tokyo, Graduate School of Medicine, Tokyo, Japan; 2) Department of Neurology, International University of Health and Welfare Mita Hospital; 3) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Chiba.

Backgrounds: Dentatorubral-pallidoluysian atrophy (DRPLA [MIM 125370]) is an autosomal dominant ataxia caused by unstable expansion of CAG repeats in the coding sequence of *DRPLA* gene (*ATN1*). Previous studies have shown that DRPLA protein (DRPLAp) functions as a transcriptional co-regulator. Its target genes, however, remain unknown. **Purposes:** To determine the target genes regulated by wild type and mutant DRPLAp based on expression profiling in cultured cells expressing full-length wild-type or mutant DRPLAp. **Materials:** Flp-In T-REx 293 cells stably expressing *GFP*-fused full-length human *DRPLA* cDNA (Q19 or Q88) under the control by Dox (doxycycline) -On system were used. **Methods:** Total RNA was extracted from the cells (Q19 or Q88) 24 hours after incubation with Dox. Expression profiling analyses were performed in four independent experiments. Strand-specific cDNA libraries were generated from ribosomal RNA-depleted RNA using random primers. Paired-end RNA-seq was performed using HiSeq2500 and analyzed by Cufflinks2 and edgeR. Differentially expressed genes (DEGs) by expressing wild-type, full-length DRPLAp (Q19) and those by expressing mutant, full-length DRPLAp (Q88) were extracted from each experiment. ;Transcriptional target genes were defined as DEGs obtained by both softwares with reproducibility confirmed by all independent experiments. All the transcriptional target genes were validated by qRT-PCR. **Results:** As transcriptional target genes validated by qRT-PCR, 8 up-regulated genes and 7 down-regulated genes were revealed by expressing full-length DRPLAp (Q19), and 8 up-regulated genes and 11 down-regulated genes were revealed by expressing full-length DRPLAp (Q88). Four up-regulated genes and another 4 down-regulated genes were common when Q19 or Q88 was expressed. A causative gene of epilepsy which is implied to cause epilepsy by loss-of-function mutation was demonstrated to be down-regulated when DRPLAp (Q88) was expressed. **Conclusions:** We found 15 target genes whose expression levels were controlled by full-length wild type DRPLAp and 19 target genes whose expression levels were controlled by full-length mutant DRPLAp. Four up-regulated genes and another 4 down-regulated genes were common when full-length wild type DRPLAp or full-length mutant DRPLAp was expressed. The finding that a causative gene of epilepsy was down-regulated in cells expressing full-length mutant DRPLAp might be related to epilepsy presented in patients with DRPLA.

1014F

New *DNAJC5* mutation initially missed by Sanger sequencing and whole-exome sequencing identified in a familial case of adult-onset neuronal ceroid lipofuscinosis (ANCL). *I. Jedlickova[1], M. Cadieux-Dion[2], A. Pristoupilova[1], V. Stranecky[1], H. Hartmannova[1], K. Hodanova[1], H. Hulkova[1], J. Sikora[1], V. Baresova[1], L. Noskova[1], P. Vyletal[1], M. Zivna[1], P. Cossette[2], E. Andermann[3], F. Andermann[3], S. Kmoch[1], The Adult NCL Gene Discovery Consortium.* 1) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University, Prague, Czech Republic; 2) Centre Hospitalier de L´Universite de Montréal, Montréal, Canada; 3) Montreal Neurological Hospital & Institute, McGill University, Montreal, Québec, Canada.

   Adult-onset neuronal ceroid lipofuscinosis (ANCL) are a group of rare genetic diseases characterised clinically by the progressive deterioration of mental and motor functions and histopathologically by the intracellular and ultrastructurally distinctive accumulation of autofluorescent lipopigment – ceroid – in the brain and other tissues. Clinical heterogeneity of ANCLs is in line with diverse inheritance patterns, increasing number of identified causal genes *DNAJC5*, *CLN6*, *CTSF*, *GRN*, *CLN1*, *CLN5*, *ATP13A2* and various types of disease causing mutations. Diagnosis of ANCLs is challenging from clinical, histopathological as well as laboratory perspectives. Even with advancements in technology, many ANCL families and cases still remain without established genetic diagnosis. Aware of this, we recently established **The Adult NCL Gene Discovery Consortium** involving research groups from Australia, UK, Europe, USA and Canada. Within the Consortium we review clinical and histopathological data and classify recruited cases into categories as definite, probable, possible and not ANCL. ANCL cases are then subjected to candidate gene and whole-exome sequencing (WES). In one family with autosomal dominant ANCL we identified a **30 bp in-frame insertion in *DNAJC5* encoding the cysteine-string protein alpha (CSPα)**. The insertion leads to a duplication of a central core motif of the cysteine-string domain (CSD) of CSPα p.Y120YC-CCCLCCCFN. The mutation affects palmitoylation-dependent sorting of CSPα in neuronal cell model as do the previously described mutations p.Leu115Arg and p.Leu116del. The mutation is remarkable diagnostically. It was initially missed by Sanger sequencing of *DNAJC5* and by WES, to be identified later by re-analysis of original WES data that are shared within the Consortium. Modified protocol for Sanger sequencing of *DNAJC5* confirmed segregation of the mutation within the family and excluded its presence in other 17 cases studied within the Consortium. Our work identifies novel *DNAJC5* mutation in ANCL and provides cautionary tale about the challenges in identification of even relatively short insertions by current genetic methods.   .

1015W

Clinical features and the pathomechanism of early childhood-onset neurodegenerative encephalopathy arising from biallelic *TBCD* mutations. *N. Miyake[1], T. Chihara[2], M. Miura[3], H. Shimizu[4], A. Kakita[4], M. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University, Yokohama, Kanagawa, Japan; 2) Department of Biological Science, Graduate School of Science, Hiroshima University, Hiroshima, Japan; 3) Department of Genetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan; 4) Department of Pathology, Brain Research Institute, University of Niigata, Niigata, Japan.

   Tubulin folding cofactor D (TBCD) is a member of tubulin specific chaperone and plays an important role for microtubule formation together with other tubulin specific chaperones, TBCA, TBCB, TBCC, and TBCE. We analyzed four families (two Japanese families, a Chinese family, and an Israeli family) with affected siblings showing early onset (before 1 year of age), progressive diffuse brain atrophy with regression, postnatal microcephaly, postnatal growth retardation, muscle weakness/atrophy, and respiratory failure. One family is consanguineous. Based on autosomal recessive model, we identified biallelic *TBCD* mutations in eight affected individuals from the four families by whole exome sequencing. A total of seven mutations were found: five missense mutations, one nonsense, and one splice site mutation resulting in a frameshift. These variants were unreported or extremely rare in control databases, and all missense variants were thought to be pathogenic *in silico* predictions. *TBCD* was ubiquitously expressed in all tissues, but highly expressed in brain, heart, and skeletal muscle in the fetus and adult. This expression pattern was consistent with the affected organs in the patients, including the central nervous system and skeletal muscles. To see the mutational effects of the identified variants, protein-protein interactions between TBCD (wild type and mutant) and other complex components (α/β-tubulin, TBCE, TBCC, and ARL2) were examined by co-immunoprecipitation, and revealed the impaired binding between most mutant TBCD proteins and ARL2, TBCE, and β-tubulin. The *in vivo* experiments using olfactory projection neurons in *Drosophila melanogaster* indicated that the *TBCD* mutations caused loss-of-function effects. Furthermore, the autopsied brain from one deceased individual showed characteristic neurodegenerative findings: cactus and somatic sprouts formation in the residual Purkinje cells in the cerebellum which are also seen in some diseases associated with mitochondrial impairment. These findings might reflect structural and metabolic abnormalities resulting from disrupted mitochondrial transport in neuronal cells caused by TBCD depletion.

1016T

Mutant human proteins linked to familial neurodegeneration cause secondary protein misfolding in the spinal cord. *M.C. Pace, G. Xu, S. Fromholt, J. Howard, B.I. Giasson, J. Lewis, D.R. Borchelt.* Neuroscience, University of Florida, Gainesville, FL.

The relationship between protein homeostasis (proteostasis) and the pathology in neuromuscular disorders characterized by neurodegeneration, such as amyotrophic lateral sclerosis (ALS) and Parkinson's disease (PD), has become an area of intense interest. Prior studies in *C. elegans* demonstrated that the expression of aggregation-prone proteins compromised the folding of other proteins, causing "bystander" or "secondary" misfolding. In prior studies of mice that develop Alzheimer-amyloidosis, we found evidence of secondary misfolding in brains of mice with high amyloid burden using detergent-extraction coupled to bottom-up proteomics analysis. Here, our goal was to translate this methodology to identify secondarily misfolded proteins in transgenic mouse models of spinal proteinopathies. Additionally, we sought to comparatively analyze the effects that different initiating proteinopathies have upon secondary protein misfolding. Our model system used transgenic mice that express mutant superoxide dismutase 1 (SOD1, associated with familial ALS), tau (associated with tauopathies such as frontotemporal dementia with Parkinsonism linked to chromosome 17), and α-synuclein (αSyn, linked to familial variants of PD); mice expressing these genes (including P301L tau, A53T αSyn, and G93A SOD1) develop spinal motor neuron degeneration. In spinal cords of paralyzed mice from these strains of mice, we identified hundreds of proteins that aberrantly fractionate to detergent-insoluble fractions in tissues with abundant pathologic inclusions. Although many of the identified proteins were common to all the models, there were also proteins unique to each model. These findings indicate that multiple types of proteinopathy can induce "secondary" misfolding, and that the vulnerability of a subset of proteins to secondary misfolding is dependent upon the nature of the initial proteinopathy (whether this be mutant forms of tau, αSyn, or SOD1). Future studies will be necessary to determine the degree to which mutations in other pathological proteins associated with neurodegeneration, such as *TARDBP*, *FUS*, and *UBQLN2*, result in the secondary misfolding of proteins.

1017F

Unbalanced translocation causing unbalanced brain: A case of hemimegalencephaly. *A.R. Barone[1], G.M. Mirzaa[2], D. McDonald-McGinn[1], L. Conlin[1], E.H. Zackai[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) Seattle Children's Hospital, Seattle, WA.

Background: Whole exome sequencing in patients with focal cortical dysplasia (FCD), hemimegalencephaly (HME) and megalencephaly has revealed constitutional and mosaic changes in genes in the PI3K-AKT-mTOR pathway and its regulators, such as *PIK3CA, PTEN, AKT3*, and *DEPDC5*. It has been found that up-regulation of the mechanistic target of rapamycin (mTOR) pathway can result in cortical malformations and epilepsy. Here we report a patient with HME whose mTOR pathway gene studies on brain tissue were normal. However, a constitutional unbalanced translocation with a derivative chromosome 16 with a terminal deletion of 695kb in 16p13.3 and a terminal duplication of 1.91Mb in chromosome 1q44 was found. The terminal deletion in 16p13.3 includes 45 genes, most importantly *NPRL3* (nitrogen permease regulator-like-3). **Case Report:** A non-dysmorphic male infant was born by C-section at 36 6/7 weeks gestation to a 35 yo G2P1->2 mother. Growth parameters included weight and length less than 10%ile with head circumference at 20%ile. Seizures began on DOL#1. MRI had abnormalities in the right hemisphere of the brain consistent with HME. He was transferred to our hospital DOL#6 and found to have subclinical status epilepticus. As the only intervention remaining was hemispherectomy, parents elected to withdraw care and the patient expired. Autopsy results of right-sided brain biopsy included: 1) widened cortical ribbon with hypercellularity, 2) parenchymal calcifications, 3) a "chicken wire"-like vascular pattern, and 4) nodules of gray matter heterotopia within the white matter. Biopsy results from the left side of the brain were normal. Genetic testing included microarray and sequencing of targeted PI3K-AKT-mTOR genes from brain samples, blood, and skin. Gene sequencing was unrevealing, but the infant had a constitutional unbalanced translocation resulting in heterozygous deletion of *NPRL3* as detailed above. **Conclusion:** There have been heterozygous *NPRL3* missense, frameshift, and nonsense mutations in patients with focal epilepsy and FCD. NPRL3 is a subunit of GAP activity toward RAGs complex 1 (GATOR1), a negative regular of mTOR. Functional studies of these *NPRL3* mutations have indicated loss of the protein product of the mutated allele, indicating haploinsufficiency. Interestingly, there have been no prior reports of HME with the 16p13.3 deletion. We propose this case as a first example of *NPRL3* haploinsufficiency via unbalanced translocation as causative of the HME.

1018W

Expanding the natural history of *KIF1A* associated disorders (KAND).
L. Boyle[1], W.K. Chung[1,2]. 1) Columbia University College of Physicians and Surgeons, New York, NY; 2) Department of Pediatrics, Columbia University Medical Center, New York, NY.

The gene *KIF1A* encodes for a kinesin responsible for anterograde axonal transport of synaptic-vesicle precursors and neurotransmitters. *KIF1A* has been associated with distinct disorders including the peripheral nervous system disorder hereditary sensory neuropathy IIC (HSN2C), the central nervous system upper motor neuron disorder hereditary spastic paraplegia-30 (SPG30), as well as a complex syndrome with a constellation of symptoms including axial hypotonia, peripheral spasticity, intellectual disability, and variable cerebellar and cerebral atrophy (MRD9). Though pathogenic variants in *KIF1A* with each of these distinct diagnostic categories based on phenotype and mode of inheritance, as we expand our understanding of the natural history of these disorders, these discrete disorders are really a spectrum of clinical severity across KAND currently associated with 30 genetic variants. To gather phenotypic data on mutations carriers, we collected caregiver reported medical history and Vineland Adaptive Behavior Scales (VABS-II) on 20 individuals, age 5 months to 21 years (mean=8 years, median=5 years), 18 of whom have *de novo* mutations in the motor domain of *KIF1A* and two of whom are compound heterozygotes with one mutation in the motor domain and one outside the motor domain. New features not previously observed in KAND include a high prevalence of gastrointestinal reflux (50%) and multiple males with genital abnormalities including micropenis, microorchidism, and cryptorchidism. The majority of individuals have ophthalmologic findings (90%), including optic nerve hypertrophy and atrophy (40%), strabismus (20%), cataracts (10%), and cortical visual impairment (15%). Half of the individuals have seizures, most commonly absence and generalized tonic clonic seizures (25% each). Developmentally, the majority of individuals have adaptive behavior composite scores in the low range on the VAB II (mean=52.9, SD=15.5). The phenotype associated with the recurrent p.E253K variant is particularly severe and includes severe neonatal hypotonia, with 3 of the 5 known individuals dying before 4 years of age. Our findings expand the phenotypic spectrum of KAND.

1019T

An autopsy case of familial amyloid polyneuropathy (FAP) with novel transthyretin (TTR) mutation (TTR, Lys80Arg). *H. Furuya[1], Y. Miyamoto[1], T. Yamamoto[3], Y. Morita[1], Y. Osaki[1], K. Furuta[1], M. Ueda[3], N. Maeda[4], H. Kitoaka[2], T. Iwaki[4], Y. Ando[3].* 1) Neurology, Kochi Medical School, Kochi University, Kochi, Kochi, Japan; 2) Cardiology, Kochi Medical School, Kochi University, Kochi, Kochi, Japan; 3) Neurology, Kumamoto University, Kumamoto, Japan; 4) Neuropathology, Kyushu University, Fukuoka, Japan.

Hereditary transthyretin (TTR) amyloidosis is caused by inheritance of an abnormal TTR gene in an autosomal dominant fashion. Here, we report a Japanese late onset case with familial amyloid polyneuropathy (FAP), caused by a novel TTR mutation of TTR compound heterozygous amyloidogenic TTR (ATTR) with wild type TTR (TTR 30Val / ATTR 80Arg), who shows orthostatic hypotension, electric shock like pain and allodynia in the trunk and severe pre / post micturitional pain with mild neuropathy and cardiopathy without vitreous opacity. Post-mortem examination showed severe amyloid deposition in peripheral nerve, dorsal root ganglion and vascular wall of the choroid plexus, which is compatible with that of severe case of type 1 FAP with mutation TTR (ATTR 30Val / TTR 80Lys). Though, his younger brother possessed compound heterozygous pathological mutation in each allele (ATTR 30Met / ATTR 80Arg), he only showed mild neuropathy. These results indicate that the novel TTR mutation (ATTR Lys80Arg) is the cause of amyloidosis with symptoms of micturitional abdominal pain and without or mild cardiopathy and vitreous opacity, if it exists compound heterozygous with wild type TTR (TTR 30Val / ATTR 80Arg).

1020F

Identification of novel *SNORD118* mutations in seven patients with leukoencephalopathy with brain calcifications and cysts. *K. Iwama[1,2], T. Mizuguchi[1], J. Takanashi[3], H. Shibayama[4], M. Shichiji[5], S. Ito[5], H. Oguni[5], T. Yamamoto[6], A. Sekine[7], S. Nagamine[7], Y. Ikeda[7], H. Nishida[8], S. Kumada[8], T. Yoshida[9], T. Awaya[9,10], R. Tanaka[11], R. Chikuchi[12], H. Niwa[12], Y. Oka[13], S. Miyatake[1], M. Nakashima[14], A. Takata[1], N. Miyake[1], S. Ito[2], H. Saitsu[14], N. Matsumoto[1].* 1) Human Genetics, Yokohama City University, Yokohama, Kanagawa, Japan; 2) Department of Pediatrics, Graduate School of Medicine, Yokohama City University, Yokohama, Japan; 3) Department of Pediatrics and Pediatric Neurology, Tokyo Women's Medical University, Yachiyo Medical Center, Yachiyo, Japan; 4) Department of Neurology, Kameda Medical Center, Kamogawa, Japan; 5) Department of Pediatrics, Tokyo Women's Medical University, Tokyo, Japan; 6) Institute of Medical Genetics, Tokyo Women's Medical University, Tokyo, Japan; 7) Department of Neurology, Gunma University Graduate School of Medicine, Maebashi, Japan; 8) Department of Neuropediatrics, Tokyo Metropolitan Neurological Hospital, Tokyo, Japan; 9) Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan; 10) Department of Anatomy and Developmental Biology, Kyoto University Graduate School of Medicine, Kyoto, Japan; 11) Department of Child Health, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan; 12) Department of Neurology, Kariya Toyota General Hospital, Kariya, Japan; 13) Department of Neurosurgery, Nagoya City University Hospital, Nagoya, Japan; 14) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan.

Leukoencephalopathy with brain calcifications and cysts (LCC) is neuroradiologically characterized by leukoencephalopathy, intracranial calcification, and cysts. Coats plus syndrome is also characterized by the same neuroradiological findings together with defects in retinal vascular development. LCC and Coats plus were originally considered the same clinical entity that was termed cerebroretinal microangiopathy with calcifications and cysts, but evidence now suggests that they are genetically distinct. Mutations in CTS telomere maintenance complex component 1 (*CTC1*) and small nucleolar RNA, C/D box 118 (*SNORD118*) genes have been found to cause Coats plus and LCC, respectively. We recruited eight unrelated families with LCC. These patients typically showed major neuroradiological findings of LCC with no other signs of extra-neurological manifestations. *SNORD118* was examined by Sanger sequencing in these families. Seven out of eight (87.5%) probands carry compound heterozygous mutations, suggesting that *SNORD118* mutations are the major cause of LCC. We identified a total of eight mutations, including four novel. Although some of the variants identified in this study are heterozygously present in public databases with an extremely rare frequency (<0.1%), biallelic mutations were exclusively found in most unrelated LCC families. Importantly these mutations were overlooked in our initial whole exome sequencing (WES) study focusing on only protein-coding genes. *SNORD118* is located in the 3'-untranslated region of the protein-coding gene transmembrane protein 107 (*TMEM107*), which was targeted by our exome capture. Indeed, we could confirm these variants retrospectively by utilizing the annotation database of Ensemble human genes. In summary, point mutations in the non-protein-coding RNA *SNORD118* were causal for a specific neurological phenotype, LCC. Our results indicate that functionally important region such as non-coding RNAs also warrant investigation by whole exome or genome sequencing.

1021W

New homozygous missense mutation in *NT5C2* underlying hereditary spastic paraplegia SPG45. *A. Onoufriadis[1,7], R. Straussberg[2,3,7], O. Konen[3,4], Y. Zouabi[2,3], L. Cohen[2,5], J.W.C. Lee[1], C. Hsu[1], M.A. Simpson[6], J.A. McGrath[1].* 1) St John's Institute of Dermatology, Division of Genetics and Molecular Medicine, King's College London, London SE1 9RT, United Kingdom; 2) Neurogenetic Clinic, Neurology Institute, Schneider Children's Medical Center, Petah Tikva 49202, Israel; 3) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel; 4) Radiology Institute, Schneider Children's Medical Center, Petah Tikva 49202, Israel; 5) Genetic Institute, Schneider Children's Medical Center, Petah Tikva 49202, Israel; 6) Department of Genetics, Division of Genetics and Molecular Medicine, King's College London, Guy's Hospital, London SE1 9RT, United Kingdom; 7) These authors have contributed equally to the work.

SPG45 is a rare form of autosomal recessive spastic paraplegia associated with mental retardation. This condition results from biallelic loss-of-function mutations in *NT5C2* (5'-Nucleotidase, Cytosolic II) although only a small number of cases with mutations have been described and the spectrum of phenotypic abnormalities is not well defined. Detailed phenotyping and mutation analysis was undertaken in three individuals with SPG45 from a consanguineous family of Arab Muslim origin. All three subjects had moderate intellectual disability, attention deficit disorder and magnetic resonance imaging findings suggestive of significant corpus callosum and white matter pathology. Using whole-exome sequencing, we identified a novel homozygous missense mutation in *NT5C2* (c.1379T>C; p.Leu460Pro). Our data expand the molecular basis of SPG45, adding the first missense mutation to the current database of nonsense, frameshift and splice site mutations. NT5C2 plays a key role in purine metabolism in the brain and contributes to the normal development of central white matter structures. *NT5C2* mutations seem to have a broad clinical spectrum and should be sought in patients manifesting either as uncomplicated or complicated HSP.

1022T

Spastic paraplegia type 4: A novel *SPAST* splice site donor mutation and expansion of the phenotype variability. *A. Orlacchio[1,2], C. Montecchiani[1], R. Miyamoto[3], L. D'Onofrio[1], R. Rumore[1], M. Miele[1], F. Gaudiello[1], Y. Izumi[3], C. Caltagirone[4], R. Kaji[3], T. Kawarai[3].* 1) Laboratorio di Neurogenetica, CERC - IRCCS Santa Lucia, Rome, Italy; 2) Dipartimento di Scienze Chirurgiche e Biomediche, Università di Perugia, Perugia, Italy; 3) Department of Clinical Neuroscience, University of Tokushima, Tokushima, Japan; 4) Dipartimento di Medicina dei Sistemi, Università di Roma "Tor Vergata", Rome, Italy.

   Background: Mutations in SPG4/*SPAST* represent the most frequent molecular etiology in an autosomal dominant form of Hereditary Spastic Paraplegia (HSP). Previous studies demonstrated various molecular pathogenesis in SPG, including loss-of-function, haploinsufficiency, dominant-negative effect, and dysregulation of microtubule-severing activity. The aim of this study is to reveal molecular pathogenicity and genotype-phenotype correlations in SPG4. **Methods:** A cohort of patients with spastic paraplegia recruited in Italy and Japan was investigated clinically and genetically. Initial diagnostic approach included targeted resequencing of currently-known HSP genes or resequencing of whole exomes. Biological effects by nucleotide variation were predicted using bioinformatic tools and confirmed by reverse transcription polymerase chain reaction (RT-PCR) experiment. Measurement of SPG4 transcripts were conducted in cultured T cells treated with nonsense pre-mRNA mediated decay (NMD) inhibitor.   **Results:** A novel *SPAST* intronic variation was found in two families, one from Italy and the other from Japan. The variant is located at the third position at 5'-splice-site of intron 6, c.1004+3A>C. RNA secondary structure seems to remain unchanged by the variant; however, skipping of exon 6 was shown by RT-PCR experiment, presumably leading to generation of frame shifts and premature stop codons (p.Gly290Trp*fs**5). Measurement of SPAST transcripts in lymphocytes demonstrated a reduction through NMD. Intra- and inter-familial phenotypic variations were observed, including age-at-onset, severity of spasticity, as well as scoliosis.   **Conclusions:** Our study demonstrated further evidence of allelic heterogeneity in SPG4, dosage effects through NMD, and broad clinical features of *SPAST* mutation.

1023F

A de novo *HNRNPU* gene mutation identified in a patient with symptomatic infection-associated acute encephalopathy and developmental delay. *S. Shimada[1,2], H. Oguni[2], Y. Otani[2], A. Nishikawa[2], S. Ito[2], K. Eto[2], T. Nakazawa[1,4], S. Nagata[2], T. Yamamoto[3].* 1) Department of Pediatrics, Juntendo University Urayasu Hospital, Chiba, MD; 2) Department of Pediatrics, Tokyo Women's Medical University, Tokyo, MD; 3) Institute of Medical Genetics, Tokyo Women's Medical University, Tokyo, MD; 4) Department of Pediatrics, Toshima Hospital, Tokyo, MD.

   [;Introduction];Heterogenous nuclear ribonucleoprotein U (*HNRNPU*) codes for a highly conserved protein that binds RNAs and mediates different aspects of their metabolism and transport. Recent studies have revealed that *HNRNPU* is a causative gene of epileptic encephalopathy, severe intellectual disability, and hemiconvulsion-hemiplegia-epilepsy (HHE) syndrome. A de novo *HNRNPU* mutation was identified in this study. [;Patient];A 5 year-old boy had had global developmental delay since early infancy. At 13 months of age, he developed febrile status epilepticus triggered by parainfluenzae infection, leaving a residual high intensity diffusion in the right hemisphere. Four months later, he started to have frequent focal seizures, characterized by cyanosis and brief tonic contraction of his body. Since then, he had had various types of seizures refractory to all available to our hospital at 5 years of age, he showed stereotypic hand movements, autistic features, episode apena, and hyperpnea followed by falling. [;Results];Next-generation sequencing detected one nucleotide deletion in the coding region of *HNRNPU*, which was negative in both parents, indicating a de novo origin. [;Conclusion];We identified a rare mutation of *HNRNPU* in a patient with symptomatic infection-associated acute encephalopathy followed by intractable epilepsy. The characteristics of patients with *HNRNPU* abnormalities should be furture evaluated for better understanding.

1024W

Naturally occurring human genetic variation suggests *LRRK2* inhibition is a safe therapeutic strategy for Parkinson's disease. *I.M. Armean[1,2], J.L. Marshall[1,2], E.V. Minikel[1,2], D.A. O'Connell[1,2], S. Ashrafzadeh[1,2], M. Chiumiento[1,2], J. Alfoldi[1,2], D.G. MacArthur[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital Research Institute, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

The recent availability of very large collections of sequenced individuals provides an unprecedented opportunity to identify humans who carry naturally occurring loss-of-function (LoF) variation, representing unique "experiments of nature" that serve as *in vivo* models of the effects of the partial or complete inactivation of human genes. Human LoF variants can aid in the identification and validation of genes that would be effective and well-tolerated targets for therapeutic inhibition in both rare and common diseases. Human LoF variants have already been used to validate candidate drug targets, most visibly in the case of *PCSK9* in LDL-lowering therapy and more recently with *APOC3* and *NPC1L1* reducing the risk of coronary heart disease. Parkinson's disease (PD) is a common neurodegenerative disease affecting about 2% of people aged 60 years and older. Genetic causes account for about 10% of PD cases, of which mutations in *LRRK2* are the most prevalent. All known disease mutations in *LRRK2* are missense and occur in critical functional domains of the protein, suggesting that the mechanism of disease is gain-of-function. This observation is supported by evidence that pathogenic mutations increase *LRRK2*'s kinase activity and therefore lead to the hypothesis of *LRRK2* inhibition as a therapeutic strategy for PD. Here we present the results of the functional and phenotypic review of *LRRK2* LoF variants identified in over 123,136 exomes and 15,496 genomes assembled by the Genome Aggregation Database (gnomAD). Firstly, we performed deep manual review to confirm genotype calls and functional predictions of the 144 *LRRK2* LoF variants, resulting in 114 heterozygous high-confidence LoF variants. For a subset of individuals we have confirmed that heterozygous LoFs result in a clear reduction in protein levels. Through collaborations we have been able to obtain electronic health records for 75 samples carrying heterozygous *LRRK2* LoF variants, and to date find no evidence of consistent or severe disease. Our data validate the use of the gnomAD resource for genotype-guided target validation, and suggest that partial therapeutic reduction of *LRRK2* levels will be well-tolerated in humans.

1025T

Phenotypical features and genetic findings in Lithuanian patients with CMTX1. *B. Burnyte, I. Kavaliauskiene, L. Ambrozaityte, A. Morkuniene, A. Matuleviciene, B. Tumiene, V. Kucinskas, A. Utkus.* Dept. of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania.

BACKGROUND: X-linked Charcot-Marie Tooth disease (CMTX1) is caused by mutations in the *GJB1* gene that encodes a polypeptide which is arranged in hexameric array and form gap junctions. To date over 300 different *GJB1* variants are described in the HGMD® database. OBJECTIVE: To evaluate the phenotypical features and genetic findings in the CMTX1 patients. METHODS: Retrospective analysis of phenotypical and genetic data in CMT1X patients RESULTS: In our cohort of 14 confirmed CMTX1 probands out of 336 CMT patients we identified a total of ten *GJB1* pathogenic mutations, three of which are novel. Three unrelated individuals carried a novel missense mutation c.290A>G, p.His97Arg. One patient had a novel nonsense mutation c.195T>A, p.Tyr65*, while another one carried a novel missense mutation c.596G>A, p.Gly199Asp. In 50% of patients the age at disease presentation was in second decade whereas 28.57% of patients reported an onset before 10 years of age. Most common symptoms at onset were distal weakness of upper or lower limbs, unsteadiness and high stepping gait. Episodic transient fluctuating stroke-like episodes were diagnosed in one patient with described mutation. NCV showed a sensory and motor axonal neuropathy in two patients harboring novel mutations. Ten patients presented a mixed NCV pattern and one – demyelinating. CONCLUSIONS: Three novel out of ten different mutations identified in this study expands the up to date heterogeneity of the CMTX1. Analysis of phenotypes in patients with novel and known mutations of *GJB1* gene showed no phenotypic heterogeneity. The frequency of *GJB1* mutations is 4.16% in Lithuanian patients with CMT.

1026F

Lack of CHCHD2 mutations in Parkinson's disease in a Southern Italy population. *G. Iannello[1,2], R. Procopio[1], M. Gagliardi[1], A. Quattrone[1,2], G. Annesi[1].* 1) Institute of Molecular Bioimaging and Physiology. National Research Council Magna Graecia University of Catanzaro,; 2) Institute of Neurology, Department of Medical and Surgical Sciences University Magna Graecia, Catanzaro, Italy.

Parkinson's disease (PD) is one of the most frequent neurodegenerative disorders, and although most PD cases appear to be sporadic, specific genetic defects have been linked to familial PD. Previously, mutations in *SNCA, LRRK2, VPS35, EIF4G1* and *DNAJC13* were considered responsible for the typical examples of autosomal dominant parkinsonism. However, potential genetic causes remain unidentified. Coiled-coil-helix-coiled-coil-helix domain containing 2 (*CHCHD2*) is a previously uncharacterized small mitochondrial protein thought to localize in the mitochondrial intermembrane space .CHCHD2 mutations have recently been identified in families with autosomal dominant Parkinson's disease (ADPD); these patients present with dopa- responsive parkinsonism. Three mutations 182C>T (Thr61Ile), 434G>A (Arg145Gln), and 300+5G>A was identified in 4 Japanese. However, there are no reports of PD patients with *CHCHD2* mutations in Italian populations. In this study, we performed comprehensive *CHCHD2* mutation screenings in familial and sporadic PD cases to assess the frequencies of known and novel rare nonsynonymous mutations. The study assessed 165 ADPD pedigrees and 160 controls. The relevant ethical authorities approved the study, and written informed consent to participate in genetic research was obtained from all subjects. All patients submitted to a standardized neurological examination by 2 movement disorder specialists, and the criteria for diagnosing PD adopted by the participating neurologists was based on the United Kingdom Parkinson's Disease Society Brain Bank Eighteen known PD gene mutations were previously excluded in the familial cases by targeted high-throughput sequencing. The 4 exons and intron-exon boundaries of CHCHD2 were amplified by PCR and sequenced with the ABI 3730 analyzer (Applied Biosystems). Primer sequences and PCR conditions were described previously. We did not identify any PD patients carrying *CHCHD2* mutations, but we found c.*125G>A genetic variant in heterozygous and homozygous state in our study population already reported in dbSNP. Our results suggest that *CHCHD2* gene may not play a major role in familial and sporadic PD patients in our population.

1027W

DNAJC13 familial Parkinson's disease from South Italy. *R. Procopio[1], M. Gagliardi[1], G. Annesi[1], G. Iannello[1,2], A. Quattrone[1,2].* 1) Institute of Molecular Bioimaging and Physiology, National Research Council, Germaneto, Catanzaro, Italy; 2) Institute of Neurology, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy.

Parkinson's disease (PD) is the second most common neurodegenerative disorder, and the most common neurodegenerative form of parkinsonism. A pathogenic mutation (p.N855S) in DNAJC13(DnaJ [Hsp40] Homolog, Subfamily C, Member 13) was linked to autosomal dominant Lewy body PD in a Dutch–German–Russian Mennonite multi-incident kindred, and was found in four additional patients. 5 DNAJC13 (RME-8) is a DnaJ-domain-containing protein that regulates endosomal trafficking and protein recycling The p.N855S mutation causes an accumulation of receptors in endosomal compartments. In this study, we performed a comprehensive screening of DNAJC13 in familial parkinsonism and in familial parkinsonism to assess the frequency of known and novel rare nonsynonymous mutations. All patients were examined and observed longitudinally by movement disorder neurologists and diagnosed with PD according to published criteria. To estimate the frequency of the Asn855Ser (in familial PD), we screened this variant in a southern Italy PD cohort and control subjects. Our population included 165 patients with familial PD, having at least 1 relative among their first degree, second degree, in third degree family members with a formal diagnosis of PD, major PD genes had been analyzed and they were excluded. Genomic DNA was extracted from peripheral blood by standard method. The purified polymerase chain reaction products were sequenced on an ABI 3500 sequencer. Our sequencing analysis of 165 patients with PD identified one carrier of p.Arg903Lys novel mutation in exon 24 of the *DNAJC13* gene, The carrier of p.Arg903Lys has late-onset PD (AAO 65), and his mother have parkinsonism.

1028T

A rare male patient with classic Rett Syndrome caused by MeCP2_e1 mutation. *A. Goji[1], H. Ito[1], N. Tokaji[1], T. Kohmoto[2], T. Naruto[2], R. Takahashi[2], T. Mori[1], Y. Toda[1], M. Saito[2], S. Tange[2], K. Masuda[2], S. Kagami[1], I. Imoto[2].* 1) Department of Paediatrics, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; 2) Department of Human Genetics, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan.

Rett syndrome (RTT) is a severe neurodevelopmental disorder typically affecting females. It is mainly caused by loss-of-function mutations affecting the coding sequence of exon 3 or 4 of the *methyl-CpG-binding protein 2* (*MECP2*) gene. Phenotype characterized by idiopathic intellectual disability, psychomotor regression involving loss of speech and hand use, development of gait problems, and characteristic repetitive hand stereotypies. A pathogenic *MECP2* variant in a male is most often presumed to be lethal. However, several groups of male patients with an *MECP2* mutation have been distinguished, which includes boys with RTT with a *MECP2* mutation diluted by an additional X chromosome (XXY) or somatic mosaicism. We report on 3-year-old male carrying a small duplication mutation in the open reading frame of *MECP2* exon 1. All symptoms of the main criteria of RTT started to appear after 6 months of age, although fewer milestones in psychomotor development, especially language skills, could be evaluated due to an earlier start of regression. The patient also met exclusion criteria. In addition, his phenotype met 7 of 11 supportive criteria, including breathing disturbances when awake, bruxism when awake, impaired sleep patterns, abnormal muscle tone, growth retardation, small cold hands and feet, and inappropriate laughing/screaming spells, although supportive criteria were entirely eliminated from the revised diagnostic criteria for classic RTT. Our molecular diagnostic analyses using targeted exome sequencing followed by Sanger sequencing detected a *de novo* 5-bp duplication within the open-reading frame of *MECP2* exon 1. This mutation, NM_001110792.1:c.23_27dup, caused a frameshift and introduced a stop codon downstream after 35 missense amino acids [p.(Ser10Argfs*36)]. This mutation has been reported to cause classic RTT in the sporadic case of one female patient, but has not been reported to cause MECP2-related diseases in males. His karyotype was consistent with a normal male. No signs of somatic mosaicism were detected. Mutations in *MECP2* exon 1 affecting the MeCP2_e1 isoform are relatively rare causes of RTT in females, and only one case of a male patient with MECP2-related severe neonatal encephalopathy has been reported. This is the first case of a male with classic RTT caused by a rare mutation in the MeCP2_e1 isoform. Therefore, both males and females displaying at least some type of MeCP2_e1 mutation may exhibit the classic RTT phenotype.

1029F

Epidemiology and genetics of Chiari 1 malformation. *B. Sadler[1], T. Kuensting[1], G. Haller[2], M. Smyth[3], D. Limbrick[3], C. Gurnett[1].* 1) Neurology, Washington University, St Louis, MO; 2) Orthopedic Surgery, Washington University, St Louis, MO; 3) Neurosurgery, Washington University, St Louis, MO.

Background: Chiari 1 Malformation (CM1) is a common neurological disorder in which the cerebellar tonsils herniate through the foramen magnum, symptomatically affecting up to 1/1,000 people. Patients present with a wide array of symptoms stemming from compression of neural tissue that can lead to syringomyelia and/or hydrocephalus. In many cases CM1 is idiopathic but it can occur in the context of known genetic syndromes or genomic abnormalities (syndromic), including Ehlers-Danlos Syndrome, Klippel-Feil Syndrome and Neurofibromatosis Type I. We hypothesize that idiopathic CM1 represents a mild form of syndromic CM1, and that many of the same genes causing syndromic CM1 will be involved in idiopathic CM1. Sample/Methods: We performed a chart review of 500 individuals affected with CM1 who were seen at St Louis Children's Hospital by two neurosurgeons between 2003-2015. DNA is being prospectively collected and has been collected from ~150 of these CM1 cases and collection is ongoing. When possible, affected and unaffected family members, were also recruited. Exome sequencing was performed on 98 CM1 cases, including 8 trios consisting of early-onset CM1 cases with unaffected parents. Alignment and calling pipeline includes Burroughs-Wheeler Aligner (BWA) and Genome Analysis Tool Kit (GATK). Results: In the cohort of 500 patients with CM1, 69% had idiopathic CM1. These included individuals with scoliosis and/or syringomyelia. Genetic or genomic causes were identified in 8% of cases, including Marfan Syndrome, Crouzon Syndrome and Alagille Syndrome. 32 (6%) of patients had a known family history of CM1. Idiopathic CM1 cases had the oldest average age at diagnosis, although had a mean tonsilar ectopia reflective of the mean of the entire sample. Examination of sequence data resulted in the identification of rare nonsense and de novo mutations in several genes known to cause skeletal abnormalities. Discussion: The vast majority of CM1 cases are idiopathic. No genes have yet been conclusively identified for CM1. Mutations in skeletal disease genes indicate that pathways involving these genes may be critical for CM1 development.

1030W

Report of phenotypic variability of periventricular nodular heterotopia in a four-generation Caucasian family with a novel *FLNA* mutation. *D. Khattar, RJ. Hopkin.* Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Background: X-linked periventricular nodular heterotopia (PVNH1 [MIM30049]) is characterized by delay in migration of neurons from the periventricular region resulting in nodular clumps lining the lateral ventricles. This is frequently associated with seizures, borderline to normal intelligence, and cardiac or vascular anomalies including coagulopathy. Most cases are the result of a new mutation, but familial cases have occurred. We report a four generation family with PVNH1. **Patients:** The proband was a 9 month female with gross motor delays, hip dysplasia, and failure to thrive. Significant cardiac findings included large PDA necessitating closure, mildly hypoplastic aortic annulus with dilated ascending aorta, and small patent foramen ovale. Brain MRI showed diffuse bilateral subependymal nodular heterotopia with enlarged cisterna magna. The mother of the proband was a 20 year old healthy female with normal cognition without cardiac or epileptic history. She had significant joint laxity. A novel heterozygous mutation, c.6923C>A, in exon 43 of *FLNA* gene [MIM 300019] resulting in p.S2308X was identified. The grandmother, great grandmother, and a maternal aunt of the proband have seizures and periventricular heterotopias, but have not yet had molecular testing. Other family members are also being evaluated. **Summary:** The inter-familial variability is remarkable ranging from near normal to perinatal disability. The mother adds particular emphasis of association with connective tissue findings as part of an extended spectrum of PVNH rather than a separate entity of Ehlers-Danlos Syndrome. Furthermore, the importance of multigenerational family history and screening for cardiac and aortic complications is evident. The cause of the variability is incompletely understood but not directly predicted by the mutation alone.

1031T

Sensory, behavioral, and social phenotypes observed in individuals with Williams syndrome in Japan. *T. Awaya[1,2], R. Kimura[1], M. Nakata[1,2], T. Kato[2], K. Matsushima[3], T. Kato[3], S. Okazaki[4], K. Tomiwa[4,5], T. Heike[2], M. Hagiwara[1].* 1) Department of Anatomy and Developmental Biology, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan; 3) Department of Human Health Science, , Kyoto University Graduate School of Medicine, Kyoto, Japan; 4) Department of Child Neurology, Osaka City General Hospital, Osaka, Japan; 5) Todaiji Ryoiku Hospital for Children, Nara, Japan.

[Introduction] Williams syndrome (WS) is a genetic disorder caused by a microdeleton in 7q11.23. Although WS was originally reported as a neurodevelopmental disorder with distinctive facial features, cardiovascular problems, such as supravalvular aortic stenosis (SVAS) and transient hypercalciuria, it is now gathering significant attention because of its characteristic cognitive features. Several groups reported sensory modulation among people with WS, and its significance in their behavioral problems. Here, we investigated such properties along with social problems in Japanese cohort.   [Methods] We included 54 participants with WS, divided into three groups (Ages 1.5-5, 6-18, 18-59, according to the questionnaire used for the Achenbach System of Empirically Based Assessment (ASEBA), n=13, 22, 19), who answered all the questionnaires below; the Hyperacusis questionnaire (HQ), the Fear Survey Schedule for Children-Revised (FSSC-R), the Short Sensory Profile (SSP), the Childhood or Adult Behavior Checklist (CBCL or ABCL) of ASEBA, and the Social Responsiveness Scale-2 (SRS-2).   [Results] Individuals with WS showed high scores of HQ and FSSCR, as previously reported. We observed >90% individuals with high SSP scores, and 54.7% showed clinical to borderline T-score of total problems (37.7%, 39.6% for the internalizing/externalizing problems, respectively) in CBCL/ABCL. We observed significant correlation between the scores of SSP, CBCL/ABCL, SRS-2. Correlation analysis of SSP, CBCL/ABCL, and SRS-2 subscales indicated that there are several strong correlations among them. Of note, Under-responsive/Seeks sensation features of SSP was strongly correlated to several subscales of SRS-2, and Avoidance demension of SSP was correlated to Externalizing problems assessed with CBCL/ABCL.   [Discussion] The sensory modulation observed in WS significantly correlated with several aspects of behavioral and social problems. Visiospacial deficit among individuals with WS is well-known; however, other sensory problems also associated with their behavioral and social difficulties. We also noticed differences from previous reports for several features, probably due to sociocultural environment. Detailed data collection is needed for further clarification.

1032F

Combining Bionano and exome sequencing identifies a homozygous structural variation in the novel *AGBL3* gene underlying microcephaly.
D. Belandres[1], M. Breuss[1], D. Musaev[1], C. Bautista[1], V. Stanley[1], H. Kayserili[2], J.G. Gleeson[1]. 1) Howard Hughes Medical Institute, Rady Genomics Institute, UCSD, La Jolla, CA 92093 USA; 2) Department of Medical Genetics, Koç Üniversitesi Hospital, Istanbul 34010 Turkey.

Chromosomal structural variations (SVs) remain a challenge to molecularly characterize, especially when under 10 kbp in size. Here we present a combined approach in a family in which the proband presented with microcephaly with simplified gyral pattern (MSGP) in the setting of parental consanguinity. Whole exome sequence analysis from the proband failed to identify a likely cause, so we combined linkage analysis, XHMM, and Bionano's Saphyr™ Optical Mapping based System. The pedigree consisted of a single affected and two healthy siblings, which was studied first with linkage analysis, allowing exclusion of 90% of the genome, and identifying the major peak on chr. 7q33. Bionano Next-Generation Mapping (NGM) was applied to cell lines derived from affected and parent, and identified a deletion of 5-12 kbp encompassing part of the coding region of the *AGBL3* gene. This was found from both Nt.BspQI and Nb.BssSI restriction enzymes used for optical mapping, and was homozygous in the affected and heterozygous in the parent. The deletion was confirmed by review of the exome sequence pile-up, which showed absence of coverage from exons 9-11 in the affected, statistically flagged using the SV detection software XHMM. *AGBL3* encodes a cytosolic carboxypeptidase that removes acidic amino acid side chains from tubulins and other proteins, specifically cleaving polyglutamylation post-translational modifications. Integrating Bionano's NGM technique revealed that the deletion extended into flanking introns, which contained a high-density of repeated dinucleotide elements, suggesting ancestral homologous recombination. Synthesizing data from linkage, exome, and Bionano allowed for the discovery, confirmation, and refinement of a homozygous deletion in a novel gene. We expect that this integrative approach will be a powerful strategy in human genetics, especially for 'exome-negative' cases and we are currently applying it to additional cases.

1033W

Genetics of childhood-onset psychosis. *M. Ameri, A.C Need.* Centre for Psychiatry, Division of Brain Sciences, Department of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London, W12 0NN, UK.

Childhood-Onset Psychosis (COP) is a very rare and debilitating disorder characterised by the onset of psychotic symptoms before age 14. Evidence suggests that it may be more frequently caused by rare (MAF<1%), highly penetrant genetic variants, in comparison to adult-onset schizophrenia. Here we report early findings from a study exploring the potential of exome sequencing as a diagnostic tool in child psychiatry clinics. We hypothesise that some occurrences of COP are the result of known Mendelian disorders manifesting primarily with psychiatric symptoms. In this study, we will extract saliva DNA from 100 trios, perform pedigree analysis, examine rare schizophrenia-associated copy number variants (CNVs) and perform exome sequencing to identify rare inherited or *de novo* SNVs or INDELs. To date we have recruited 19 families, with 9 male and 10 female probands, and generated 3-generation pedigrees for each family. Data on race, ethnicity, age of diagnosis and cognitive function will be presented. All probands have a primary diagnosis of a psychotic illness with 10/19 having a comorbid neurodevelopmental disorder. According to family report, 9/19 probands have a family history of a psychiatric diagnosis. Cognitive assessment (WASI-II) on the first 12 probands shows a variation in their full-scale intelligence quotient (FSIQ) whereby 3/12 scored extremely low (FSIQ<70), 2/12 scored borderline (FSIQ 70-80), 3/12 scored low average (FSIQ 80-90) and 4/12 scored average (FSIQ 90-110). In addition, 5/19 probands are on Clozapine, highlighting their disease severity. Genetic analysis of the first 8 probands indicates that none have common schizophrenia-associated chromosomal deletions or duplications. Exome sequence analysis based on ACMG guidelines suggests that 4/7 sequenced probands may have a mutation causing a Mendelian disease that is known to manifest with psychiatric symptoms. These will be investigated further with clinical assessments. The candidate disease genes are *SCN1A, L1CAM, RET* and *SIX3*. In addition, 2 probands have missense variants in *CLCN2* and *GRIN2A*, which may be contributing to their behavioural difficulties (including aggression), as well as speech delay. The identification of highly penetrant variants that predispose these probands to psychotic and severe neurocognitive phenotypes can have positive implications in the clinic as a diagnostic tool, as well as providing a platform for understanding of the underlying neurobiology of COP.

1034T

Expansion of the molecular and phenotypic spectrum of *CAMTA1*-related neurological disorders. *L.B. Henderson[1], H. Sroka[1], B. O'Connor[2], J.W. Innis[2], M. Byler[3], R.R. Lebel[3], J. Humberson[4], R. McClellan[5], H. Vernon[5], G. Mirzaa[6], D. Castro[7], B. Grysko[8], L. Seaver[8], F. Almesaifri[9], T. Ben-Omran[9], R. Willaert[1], M.T. Cho[1], J. Juusola[1].* 1) GeneDx, Gaithersburg, MD; 2) Division of Pediatric Genetics, Metabolism and Genomic Medicine, University of Michigan, Ann Arbor, MI; 3) SUNY Upstate University Hospital, Syracuse, NY; 4) Pediatric Genetics, University of Virginia, Charlottesville, VA; 5) Kennedy Krieger Institute, Baltimore, MD; 6) Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA; 7) Children's Medical Center of Dallas, Dallas, TX; 8) Spectrum Health Medical Genetics, Grand Rapids, MI; 9) Hamad Medical Corporation, Doha, Qatar.

    The CAMTA1 gene encodes calmodulin-binding transcription activator 1, a conserved calcium-sensitive transcription factor that is highly expressed in brain with enrichment in the cerebellum. Mice lacking *Camta1* display severe ataxia and cerebellar atrophy, indicating a role for this gene in neuronal maturation and survival. In humans, studies using chromosomal microarray have demonstrated that heterozygous intragenic rearrangements in *CAMTA1* are associated with intellectual disability (ID), congenital cerebellar ataxia, and behavioral problems. Point mutations in *CAMTA1* have not been reported to date, though some of the previously reported intragenic rearrangements were predicted to result in a frameshift leading to premature translational termination. Using clinical exome sequencing, we have identified pathogenic loss-of-function sequence variants in *CAMTA1* in nine unrelated patients. All individuals, ranging in age from 2 to 22 years, presented with developmental delays and/or ID. Other common features included abnormal behaviors (8/9), hypotonia (6/9), muscle weakness/myopathy (5/9), and ataxia (4/9). Three individuals were diagnosed with autism. Inheritance could be determined for eight variants: four occurred *de novo* in the affected proband and four were inherited from a parent, three of whom were noted to have some features consistent with a *CAMTA1* disorder (history of hypotonia, speech difficulties, cerebral palsy, poor coordination, autistic features and/or epilepsy). One carrier parent was apparently asymptomatic, thereby suggesting not only variable expressivity but also reduced penetrance of pathogenic variants in *CAMTA1*. In summary, our results expand the clinical knowledge of *CAMTA1*-related disorders and broaden the mutational spectrum to include loss-of-function sequence variants.

1035F

Delineation of a new neurobehavioral syndrome associated with mutations in RFX3. *H. Hodges[1], R. Pfundt[2], T. Gardeitchik[2], B.B.A. de Vries[2], M. Swinkels[3], M. Simon[3], K. vanGassen[3], A. Hamosh[4], C. Applegate[4], L. Immken[5], M. Willing[6], E. Fassi[6], T. Toler[6], K. Gripp[7], L. Baker[7], T. Yu[8].* 1) Division of Developmental Medicine, Department of Medicine, Boston Children's Hospital, Boston, MA; 2) Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands; 3) Division of Biomedical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 5) Department of Medical Genetics, Dell Children's Hospital, Austin, TX; 6) Division of Genetics and Genomic Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO; 7) Division of Medical Genetics, Nemours/A.I. DuPont Hospital for Children, Wilmington, DE; 8) Division of Genetics and Genomics Research, Harvard Medical School, Boston Children's Hospital, Boston, MA.

    Whole exome sequencing in an index case of a 16-year-old male with autism, mild intellectual disability and behavioral concerns revealed a *de novo* mutation in *RFX3*, a gene without established medical significance. Through GeneMatcher we have now identified a total of 8 cases of *RFX3* mutations in association with autism spectrum disorder, behavioral difficulties, and variable cognitive outcomes including mild or moderate intellectual disability. *RFX3* is a member of the regulatory factor X gene family which encodes transcription factors that contain a highly-conserved DNA binding domain. Prior studies in mice prove that *RFX3* regulates the transcription of genes involved in cilia formation and is expressed in a wide array of tissues, including adult and developing brain. Biallelic knockout of *RFX3* in mice causes situs inversus, hydrocephalus, and agenesis of the corpus callosum, although the phenotypic consequences of heterozygous mutation have not been described. Our series includes 7 males and 1 female patient, all with autism, 5 patients with intellectual disability, 4 with diagnosed attention deficit hyperactivity disorder, and 4 with nonspecific behavioral functioning difficulties. Other common features include hypotonia and joint laxity. This case series suggests that RFX3 mutation is responsible for a novel human neurobehavioral syndrome overlapping with autism spectrum disorder.

1036W

Delineation of the phenotype associated with *de novo TBR1* variants in 15 unrelated patients and review of the literature. *S. Nambot[1,2,3], A. Masurel[1,3], J. Thevenon[1,2,3], AL. Bruel[3,4], AL. Mosca-Boidron[2,3,4], C. Mignot[5], S. Chantot-Bastaraud[5,6], J. Metreau[7], T. Bienvenu[8], M. Rossi[9], A. Piton[10], F. Petit[11], N. Le Meur[12], V. Layet[13], D. Amram[14], E. Bhoj[15], D. Li[15], M. Cho[16], E. Fiala[17], W. Meschino[18], S. Hiatt[19], H. Olivié[20], T. Jouan[2], F. Tran Mau-Them[2,3,4], C. Philippe[2,3,4], Y. Duffourd[3,4], E. Tisserant[3,4], L. Faivre[1,2,3], C. Thauvin-Robinet[1,2,3].* 1) Genetic center, CHU, Dijon, France; 2) UF Diagnostic innovation des maladies rares, Laboratoire de Cytogénétique et Génétique Moléculaire, Centre Hospitalier Universitaire de Dijon, Dijon, France; 3) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (FHU TRANSLAD), Centre Hospitalier Universitaire de Dijon et Université de Bourgogne-Franche Comté, Dijon, France; 4) Inserm UMR 1231 GAD, Génétique des Anomalies du Développement, Dijon, France; 5) Service de Génétique et d'Embryologie Médicales, AP-HP, Hopital Trousseau, Paris, France; 6) Service de Génétique et d'Embryologie Médicales, INSERM U933, Paris, France; 7) Service de Neurologie Pédiatrique, Hôpital du Kremlin Bicêtre, Paris, France; 8) Service de génétique et biologie moléculaire, CHU Paris Centre – Hôpital Cochin, Paris, France; 9) Service de génétique, Centre de référence des anomalies du développement, Hôpital Femme-Mère-Enfant, Groupement Hospitalier Est, Bron, France; 10) Laboratoire de diagnostic génétique, Hôpital Civil, CHU de Strasbourg, Strasbourg, France; 11) Clinique de Génétique Guy Fontaine, Pôle de Biologie Pathologie Génétique, CHRU de Lille, Hôpital Jeanne de Flandre, Lille, France; 12) Service de Génétique, CHU de Rouen, France; 13) Service de Génétique, CH Le Havre, France; 14) Unité de Génétique Médicale, CH Créteil; 15) Department of Pediatrics, Division of Human Genetics and Molecular Biology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 16) GeneDx Laboratory, San Francisco, USA; 17) Washington University, Saint Louis, USA; 18) North York General Hospital, Toronto, Ontario, Canada M2K 1E1; 19) Richard Myers Lab/Greg Cooper Lab, HudsonAlpha Institute for Biotechnology, Huntsville, USA; 20) Centre for Developmental Disorders, UZ Leuven, Belgium.

TBR1 encodes a brain-specific T-box transcription factor highly expressed in the cerebral cortex. It plays an essential role in the differentiation of neocortical neurons, neuronal migration and axonal projection during embryonic development. TBR1 regulates several candidate genes for autism spectrum disorders (ASD), such as *GRIN2B*, through interactions with *CASK*, *AUTS2* and *RELN*. Whole-exome sequencing and the molecular inversion probe method in a large series of 3700 probands with ASD identified 7 pathogenic *TBR1* variants in 7 unrelated patients. Besides, one patient with a 2q24.2 deletion encompassing only *TBR1* and presenting intellectual disability (ID), growth retardation, mild facial dysmorphism, pectus excavatum and long hands was fully described. *TBR1* has thus been reported as a high-confidence risk gene for ASD. Nevertheless, fine phenotype descriptions are lacking and *TBR1* is still not reported as a human disease-causing gene in the OMIM database. Here, we report on a full clinical description of 15 unrelated patients, 9 males and 6 females, all carrying *de novo TBR1* variants. The patients were recruited thanks to national and international datasharing, through databases such as PhenomeCentral and GeneMatcher. Clinical and molecular results were confronted to a review of the literature. In total, 28 patients presented with *TBR1* causal variants including 13 truncating and 7 missense variants located in functional domains, as well as 8 CNV (deletion) ranging from 122kb to 12.3Mb. All patients, aged from 1 to 29 years, presented with moderate to severe ID and ASD. Six patients had a history of seizures. Non-specific dysmorphic features were observed in all patients and growth retardation was frequently reported. Heterogeneous limb anomalies were present in half of the patients with almost constant joint laxity. About one third of the patients exhibited hypotonia, gastroesophageal reflux disease, constipation, and small cerebral imaging anomalies, but further analysis of the MRI pictures may highlight characteristic cerebral anomalies. Associated non-recurrent malformations included deafness, coloboma, hypothyroidism and acrocyanosis. This report, based on a review of the literature and reverse phenotyping, provides an accurate description of the phenotype associated with *TBR1* variants and confirms their implication in syndromic ASD. *TBR1* should thus be considered a human disease-causing gene and described as such in the OMIM database.

1037T

***De novo TCF20* pathogenic variants are associated with intellectual disability, dysmorphic features, hypotonia, and neurological deficits with similarities to Smith-Magenis syndrome: Seven new cases further delineate the phenotypic presentation of this new syndrome.** *F. Vetrini[1], J.T Alaimo[2], J.A. Rosenfeld[2], W. Zhu[1], S. Lalani[1,2,4], J. Zhang[2], J. Posey[2], A. Lewis[2,4], L. Burrage[2,4], J. Holder[2,4], B. Graham[2,4], J. Harris[5], J.B. Gibson[5], M. Pastore[5], K.L. McBride[6], M. Komara[7], L. Al-Gazali[7], A. Al-Shamsi[8], E. Fanning[9], K. Wierenga[9], L. Seaver[10], D.M. Muzny[2,3], R.A. Gibbs[2,3], S.H. Elsea[1,2], C.M. Eng[1,2], J.R. Lupski[1,2,3,4], Y. Yang[1,2], P. Liu[1,2].* 1) Baylor Genetics Laboratories, Baylor College of Medicine, Houston, TX 77030, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 4) Department of Pediatrics, Texas Children's Hospital, Houston, TX 77030, USA; 5) 'Specially For Children, Austin, Texas 78731; 6) Nationwide Children's Hospital, Columbus, OH, USA, 43205; 7) Department of Pediatrics, College of Medicine & Health Sciences, United Arab University, UAE; 8) Department of Pediatrics, Tawam Hospital, Al-Alin, UAE; 9) Department of Pediatrics, Section of Genetics, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104; 10) Department of Pediatrics, University of Hawai'i, Honolulu, HI 96826.

Neurodevelopmental disorders are genetically and phenotypically heterogeneous encompassing intellectual disability (ID), autism spectrum disorders (ASDs), structural brain abnormalities and neurological manifestations with a significant number of genes associated. Next generation sequencing can resolve locus heterogeneity underlying clinical phenotypes, that may otherwise be co-assigned as a specific syndrome based on shared clinical features and can link phenotypically diverse diseases to a single locus. By whole-exome sequencing (WES) and chromosomal microarray analysis (CMA), we identified five inactivating single nucleotide variants (one missense, three frameshifts and one non-sense) and two whole-gene deletions of *TCF20* gene in seven patients from unrelated families presenting with a phenotype characterized by intellectual disability with specific language deficits, hypotonia, variable dysmorphic features and other neurological deficits including ataxia and movement disorder. The missense change c.5129A>G (p.Lys1710Arg) occurs in a highly evolutionary conserved position of the functional F-Box/GATA-1 like finger motif of the protein. The variants were all found to be *de novo* in the six cases with available parental samples. To date, a few de novo mutations potentially disrupting TCF20 function in patients with ID, ASD and hypotonia have been reported. TCF20 encodes a transcriptional co-regulator related to RAI1, the dosage-sensitive gene responsible for Smith-Magenis and Potocki-Lupski syndromes. TCF20 shares similar domains with RAI1 that are involved in histone modification and co-transcriptional regulation of a plethora of genes involved in biological pathways likely disrupted in these two syndromes. Therefore, we propose that deleterious mutations in TCF20 are associated with a novel syndrome that shares characteristics with Smith-Magenis syndrome. Together with previously described cases, our findings produce a more detailed picture of the phenotype of TCF20-associated neurodevelopmental disorder.

1038F

Intellectual disability with severe self-injury behavior caused by *THOC2* splice site variant. *N. Ishihara[1], H. Inagaki[2], M. Kawai[2], C. Inuo[1], H. Kurahashi[2].* 1) Dept Pediatrics, Fujita Health University, Toyoake, Aichi, Japan; 2) Div Mol Genet. Inst Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, Japan.

[Introduction] *THOC2* encodes a subunit of the highly conserved TREX mRNA-export complex, and known to cause syndromic intellectual disability (ID). Here we present a boy with *THOC2* splice site mutation presented as ID with severe self-injury behavior. [Subject] Subject was a 3-year-old boy born to healthy non-consanguineous parents. His 7-year-old sister was healthy with normal development. He was pointed out nuchal translucency by fetal ultrasound scan at gestational age of 19 weeks, and received amniocentesis to reveal normal male karyotype. After birth, he showed developmental delay with axial hypotonia, speech delay, and failure to thrive caused by disturbance of oral ingestion and frequent aspiration pneumonia. After performing gastrostomy, he began to gain weight, and behavior problem of punching his own head and/or chin became obvious. [Results] Whole exome sequences were performed with trio DNA samples extracted from white blood cells of subject and his parents. Variants with less than one percent allele frequency were classified into de novo, autosomal recessive, and X-linked forms. Among these, we identified an X-linked, splice site mutation of *THOC2* exon 28 (c.3503+4A>C) of maternal origin. This mutation was neither reported in the public databases such as dbSNP build 147 nor ExAC. RT-PCR analysis of mRNA from subject's white blood cells revealed that this mutation led to intron retention resulted in introducing premature termination codon. Since we could not detect any intron retention product from his mother, we carried out HUMARA assay using DNAs from her white blood cells. The result revealed skewed inactivation (98:2), and that the inactivated X chromosome was inherited to subject. [Discussion] X-linked ID caused by *THOC2* mutations was first reported by Kumar et al. in four different multigenerational families. All four variants led to missense substitutions, a class of mutations for which it is difficult to assign functional effect. Our case is a first report of *THOC2* splice site mutation in which lead to intron retention resulted in introducing premature termination codon. The fact that his mother with whom carrying heterozygous variant had skewed X-inactivation indicate strongly the pathogenesis of the variant. Since our case appeared unique phenotype, decrease of normal mRNA expression may concern an important role of neuronal development, and it would be a key of unraveling the function of *THOC2* in the central nervous system.

1039W

Guidelines for phenylbutyrate drug levels in the management of urea cycle disorders. *Y. Jiang[1], M. Almannai[2], R. Sutton[1,2], Q. Sun[1,2], S. Elsea[1,2].* 1) Biochemistry Department, Baylor Genetics, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.

Urea cycle disorders (UCDs) are heritable conditions that share a common biochemical perturbation of the accumulation of nitrogenous waste in the form of ammonia. UCDs result from a deficiency of one of six enzymes and include a spectrum of phenotypic presentations. Currently, sodium phenylbutyrate and glycerol phenylbutyrate are the primary medications used to treat patients with UCDs, and while a specific therapeutic range has not been determined, it has been demonstrated that phenylbutyrate overdose may result in toxicity. To monitor these drug metabolite levels, a fast and simple ultra-performance liquid chromatography (UPLC-MS/MS) method was developed and validated for quantification of phenylbutyrate (PB), phenylacetate (PA), and phenylacetylglutamine (PAG) in plasma and urine. The separation of all three analytes was achieved in 2 min, and the limits of detection were less than 0.04 µg/ml. Intra-precision and inter-precision were less than 8.5% and 4% at two quality control concentrations, respectively. Average recoveries for all compounds ranged from 100% to 106%. With the developed assay, we evaluated data from 35 patients with confirmed UCDs, including ornithine transcarbamylase deficiency, citrullinemia, argininemia deficiency, argininosuccinate lyase deficiency, carbamoyl phosphate synthetase deficiency, and N-acetylglutamate synthase deficiency, undergoing treatment with phenylbutyrate nitrogen scavengers. A strong correlation between the absolute PA level and the PA/PAG ratio and an inverse correlation between the PA/PAG ratio and glutamine level in plasma were observed. These findings confirm prior studies that the PA/PAG ratio is a good predictor of toxicity. Moreover, all individuals with a PA/PAG ratio ≥0.6 had plasma glutamine levels <1000 µmol/L in our study, suggesting that PA/PAG ratios in the range of 0.6 - 1.5 resulted in good control of glutamine levels in all subjects and PA and PA/PAG ratio were well below toxic levels of 500 µg/mL and 2.5. Thus, these data demonstrate that the PA/PAG ratio not only guide therapy to ensure toxic levels are not reached, but may also be considered in optimal dosing for therapeutic effect.

1040T

Novel *de novo* TAOK1 variants associated with a neurodevelopmental phenotype, macrocephaly, and joint hypermobility. *H.M. McLaughlin[1], K. Langley[1], A. Noyes[1], K. Miller[2], N. Shur[2], S. Desai[3], K. Baranano[3], A. Schreiber[4], A. Erwin[4], A. Blesson[5], K.W. Gripp[5], M. Pearson[6], T. Kleefstra[7], A. Telegrafi[1], L. Henderson[1], K.G. Monaghan[1], L. Kalsner[8], J. Juusola[1].* 1) GeneDx, Inc., Gaithersburg, MD; 2) Albany Medical Center, Albany, NY; 3) Kennedy Krieger Institute, Baltimore, MD; 4) Cleveland Clinic Genomic Medicine Institute, Cleveland, OH; 5) Nemours/A.I. DuPont Hospital for Children, Wilmington, DE; 6) District Medical Group,. Phoenix, AZ; 7) Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 8) Connecticut Children's Hospital, Hartford, CT.

The *TAOK1* gene encodes the ubiquitously expressed thousand and one amino acid kinase 1, a Ste20-related mitogen-activated protein kinase kinase kinase (MAP3Ks) (MIM 610266), which maintains mitotic microtubule dynamics, establishes proper chromosome-microtubule attachment, and ensures the accurate segregation of the chromosomes in both mitosis and interphase human cells. Expression of *TAOK1* has been shown to be concentrated in human brain tissue. Reduction of *TAOK1* expression in cells results in the inability to engage the DNA damage-induced G2/M checkpoint in the final attempt to catch cells that have escaped earlier checkpoints. In an attempt to find *de novo* causes for neurodevelopmental disorders, we looked at a total of 9,302 exome cases with a neurodevelopmental phenotype, and identified *TAOK1* as a strong candidate in six individuals with novel *de novo* heterozygous *TAOK1* variants; an additional affected individual was then identified through Gene-Matcher. Multiple variant types, including missense, nonsense, frameshift, and splice site variants were identified among the seven unrelated patients. The *de novo* heterozygous presence of these variants suggests an autosomal dominant inheritance pattern. The range of neurodevelopmental concerns reported include severe intellectual disability, autism spectrum disorder, learning disabilities with attention issues, behavior problems, language and motor delays, and slow processing. Additional features reported included neonatal feeding difficulties, hypotonia, macrocephaly, and joint hypermobility. No patient was reported to have a seizure disorder. Two of the seven individuals were heterozygous for an additional *de novo* variant. One individual had a de novo variant in *KANK1*, and was the only patient to have multiple facial dysmorphisms, strabismus, and severe intellectual disability with regression. The second *de novo* variant reported was in *NLGN2*, a candidate gene which may also play a role in neurodevelopment. It is unclear if or how these additional variants are impacting the clinical picture. To our knowledge, this is the first report of novel *de novo* variants in *TAOK1* resulting in human disease. While the mechanism by which these variants affect *TAOK1* function has not been fully elucidated, we believe they may be related to the neurodevelopmental concerns, neonatal feeding difficulties, hypotonia, macrocephaly, and joint hypermobility observed in these seven individuals.

1041F

Neurodevelopmental profile for boy with unique 6p deletion. *A. Saba[1], C. Cappello[1], A. Ganges[1], C. Samango-Sprouse[1,2,3].* 1) The Focus Foundation, 820 W Central Ave. #190 Davidsonville, MD 21035; 2) Department of Pediatrics, George Washington University, 2121 I St. NW Washington, DC 20052; 3) Department of Human and Molecular Genetics, Florida International University 11200 SW 8th St. Miami, FL 33199.

Introduction: Interstitial 6p deletion syndrome is a very rare neurogenetic disorder that is typically caused by a de novo mutation, although in some cases it may be inherited. Phenotypic abnormalities commonly present in multiple areas, including neurodevelopment, ocular, auditory, cardiac, and musculoskeletal. We present a 9-year-old male with 6p21.33-6p22.3 deletion who has Childhood Apraxia of Speech (CAS), language-based learning disorder (LLD), ADHD, and global dyspraxia (GD), with intact nonverbal abilities. Statement of Purpose: We describe the neurodevelopmental profile (NDP) of a child with 6p21.33-6p22.3 deletion through serial evaluations to expand the genotypic and phenotypic profile of interstitial 6p deletion syndrome. Methods: NDP was assessed annually using Wechsler Scales (WISC), Leiter International Performance Scale-III (LIPS), Expressive (EOW) and Receptive One Word Vocabulary Test-4 (ROW), Bruininks-Oseretsky Test of Motor Proficiency (BOT-2), The Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery), Child Behavioral Checklist (CBCL), and Sensory Profile for Children (SP). Results: WISC revealed composite scores of 64 and 82 in verbal comprehension and perceptual reasoning, respectively. On LIPS, mean nonverbal IQ was 83. Discrepancies between verbal and nonverbal abilities, along with his severe speech delays and distortions, support the presence of CAS. Standard scores on EOW and ROW were 79 and 80, respectively, which further support the presence of a receptive and expressive language disorder. Neuromotor assessments revealed GD with deficits in motor planning and coordination. On BOT-2, fine motor control was below the 1st percentile (SS= 22). Visual motor integration (SS = 46) and motor coordination (SS = 52) on the Beery were also below the 1st percentile. Total competencies on CBCL were below the 10th percentile and in the clinical range (t-score = 26). Sensory disturbances in the visual and auditory systems were indicated on the caregiver SP. Discussion: To our knowledge, this is the first reported case of 6p deletion in this chromosomal range associated with NDP findings of CAS, LLD, ADHD, and GD. Further studies on 6p21.33-6p22.3 deletion must be conducted to understand the complexity of the genotypic-phenotypic interaction. Our research further supports the importance of chromosomal microarray when complex learning disorders are present in young children. .

1042W

***ARID4A* de novo variants identified by exome sequencing among individuals with neurodevelopmental disorders.** *K.G. Monaghan[1], N. Alexander[1], A. Wilson[2], A. Iglesias[3], M. Azage[4], D.M. Niyazov[4], E. Moran[5], J. Pappas[5], J. Litwin[6], J.M. Graham, Jr.[7], K. Klinard[8], M. Gucsavas-Calikoglu[8], J. Hoffman[1], M. Guillen[1], R. Person[1], R.E. Schnur[1], G. Douglas[1], S. Bale[1], J. Juusola[1].* 1) GeneDx, Gaithersburg, MD, USA; 2) Children's Hospital of New York-Presbyterian, New York, NY, USA; 3) Division of Clinical Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY, USA; 4) Ochsner Hospital for Children, New Orleans, LA, USA; 5) NYU Hospital for Joint Diseases, New York, NY, USA; 6) University of California, San Francisco, San Francisco, CA, USA; 7) Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 8) University of North Carolina, Division of Genetics and Metabolism, Department of Pediatrics, Chapel Hill, NC USA.

Clinical exome sequencing (ES) is an effective tool to identify the genetic basis of neurodevelopmental disorders (NDD). Clinical ES was performed on 11,649 probands referred to our laboratory for NDD. Using trio-based ES, we identified de novo *ARID4A* variants in five unrelated individuals. In addition, we identified one individual with a missense variant, but were unable to confirm that it was de novo because the variant nucleotide was present in a small number of next generation sequencing reads in the unaffected father, indicating possible paternal mosaicism. One frame-shift and five missense variants were identified. The missense variants had damaging in-silico predictors and occurred in evolutionarily conserved amino acids. The 6 individuals in this cohort ranged in age from 6-28 years. They all had developmental delays and behavior abnormalities. Other common features included developmental regression (3/6), autism (4/6), and tone abnormalities (4/6). Intellectual disability was assessed and reported in 3 individuals. Two out of 4 individuals tested by brain MRI had brain anomalies (thin corpus callosum, absent septum pellucidum, and holoprocencephaly in one, pontocerebellar atrophy in another). Two patients had facial dysmorphism. *ARID4A*, located at 14q23.1, is ubiquitously expressed and participates with other proteins in epigenetic complexes to regulate genomic imprinting within the Prader-Willi/Angelman syndrome domain. In addition, through binding to the retinoblastoma protein (RB1), ARID4A represses transcription activity by serving as an adaptor to recruit histone deacetylases (HDAC) to its N-terminus repression domain-1 and provides a second HDAC-independent repression function at its C-terminus. To date, no sequence variants have been published in the *ARID4A* gene. However, a de novo 14q23.1q23.3 deletion including *ARID4A* was reported in an individual with developmental delay, hypoplasia of the corpus callosum, epilepsy, and neuroblastoma, suggesting a haploinsufficiency mechanism of disease. Our data suggest that de novo variants in *ARID4A* may be associated with a spectrum of clinical features including developmental delay, intellectual disability, autism, behavioral abnormalities, tone abnormalities, and brain anomalies. The findings presented in this case series may facilitate the identification of additional affected patients to further the knowledge regarding the phenotype associated with de novo variants in *ARID4A*.

1043T

De novo missense variant in *CSNK2A1* can cause intellectual disability, behavioral problem, and dysmorphic features: The first male child with Okur-Chung neurodevelopmental syndrome. *M. Akahira-Azuma, Y. Tsurusaki, Y. Enomoto, M. Minatogawa, T. Yokoi, K. Kurosawa.* Division of Medical Genetic, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan.

Whole exome sequencing lead to elucidate the various kinds of novel genes with intellectual disability (ID). In 2016, Okur and Chung identified de novo heterozygous mutations in the *CSNK2A1* gene in 5 unrelated females with delayed psychomotor development, intellectual disability with poor speech, behavioral abnormalities, cortical malformations in some patients, and variable dysmorphic facial features. This was named Okur-Chung neurodevelopmental syndrome (OCNDS). To our knowledge, additional case was not reported, especially in males. We report on a 7-year and 4-month old Japanese male in severe developmental delay, speech impairment, past hypotonia, short stature, and dysmorphic features including synophrys, hypertrichosis, down-slanting palpebral fissures, and bulbous nose. Chromosome analysis was 46,XY and microarray did not show any abnormal CNVs. Brain MRI revealed delayed myelinization in cortex and the decreased volume of pituitary gland. He walked alone at 3 years and started to speak meaningful words and two-phrase words at 6-year and 1-month. His clinical features resembled Coffin-Siris syndrome (histone remodeling complex) and Klefstra syndrome *(EHMT1* variant). At 7-year and 4-month of age, he was socialized at school, behaved hyperactive, spoke two phrases and counted one to 20, walked alone with wide gait, but toilet training was not completed. He had a recurrent, de novo c.A593G, p.K198R variant, which has been described previously by Okur et al. Functional studies of the variant and studies of patient cells were not performed. The mutation was found by whole exome sequencing (WES) and confirmed by trio Sanger sequencing. This is the first male case in OCNDS with a recurrent missense variant. Since Okur described 4 missense mutations and 1 splice site mutation, the mechanism of *CSNK2A1* variant could be the gain of function. In addition, somatic mutations in *CSNK2A1* have been implicated in various cancers. However, it may play a different role in germ line. Since phenotype of OCNDS overlaps with Coffin-Siris syndrome and Klefstra syndrome, we assume that *CSNK2A1* may have similar function as those two syndromes. The functional studies of the variant should be performed to confirm. .

1044F

Exome sequencing links cerebellar malformations to known neurodevelopmental disorders. *K.A. Aldinger[1], A.E. Timms[2], V. Chung[1], D. Doherty[3], W.B. Dobyns[1,3].* 1) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 2) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 3) Department of Pediatrics, University of Washington, Seattle, WA.

Cerebellar malformations are common, heterogeneous birth defects that are associated with significant developmental disabilities, including coordination, intellectual and behavioral deficits, and epilepsy. Both non-genetic prenatal and genetic factors have been implicated in the cause of cerebellar malformations. However, few patients receive a genetic diagnosis and, for many patients, no underlying cause is sought or identified. To address the genetic basis of cerebellar malformations, we performed exome sequencing in 101 families with one or more children diagnosed with Dandy-Walker malformation (DWM) or cerebellar hypoplasia (CBLH). We identified an underlying genetic cause in 41 of 101 families. Among the 49 variants identified, 57% (28/49) were *de novo* and 80% (39/49) were missense; 24% (12/49) of these were already known to be pathogenic mutations. Our results include 28 genes previously associated with neurodevelopmental disorders, including 8 genes with mutations found in more than one patient, and 5 novel genes. Intellectual disability was significantly more frequent in CBLH (80%) than in DWM (59%; p=0.035) and significantly more children with CBLH (55%) received a genetic diagnosis than individuals with DWM (26%; p=0.0024). Our data show that exome sequencing is a powerful tool for establishing molecular diagnoses for children diagnosed with common cerebellar malformations. We also implicate several new genes in cerebellar development, and show that cerebellar malformations often occur as one component of a broader developmental brain disorder.

1045W

***De novo* variants at residue 480 in *FAR1* are associated with an autosomal dominant early-onset neurological disorder.** *J. Juusola[1], M.T. Cho[1], N. Alexander[1], L. Brady[2], M. Tarnopolsky[2], S. Sell[3], R. Ladda[3], J. Douglas[4], C. Nowak[4], E. Ulm[5], C. Tian[5], S. Perlman[6], S. Matsumoto[6], A. Calhoun[6], A.V. Drack[6], R. Person[1], B. Baskin[1], M.J. Guillen Sacoto[1], A. Begtrup[1].* 1) GeneDx, Gaithersburg, MD; 2) Department of Pediatrics, McMaster Children's Hospital, Hamilton, ON; 3) Hershey Medical Center, Hershey, PA; 4) Division of Genetics, Boston Children's Hospital, Boston, MA; 5) Neuromuscular Clinic, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 6) University of Iowa, Iowa City, IA.

Exome sequencing (ES) is an established testing methodology for discovery of new associations between genes and human disorders. Novel inheritance patterns and/or phenotypes can emerge after initial characterization of disease-associated genes. Biallelic loss-of-function variants in *FAR1* have recently been implicated in peroxisomal fatty acyl-CoA reductase-1 disorder, which is an autosomal recessive condition characterized by severe intellectual disability (ID), microcephaly, cataracts, epilepsy, and growth retardation. The *FAR1* gene encodes fatty acyl-CoA reductase-1, which plays a role in plasmalogen biosynthesis by catalyzing the conversion of fatty acyl-CoA into a fatty alcohol. Analysis of 19,679 ES trios identified five probands with heterozygous *de novo* variants in the well-conserved amino acid residue 480 in *FAR1*. The *de novo* variants observed in our cases are p.Arg480His (c.1439G>A) (2 cases), p.Arg480Cys (c.1438C>T) (2 cases), and p.Arg480Leu (c.1439G>T) (1 case), none of which were observed in ExAC. The probands range in ages from 3.5 years to 6 years and have many features in common with those of the newly described recessive *FAR1* syndrome, such as spasticity (5/5), cataracts (4/5), and seizures (3/5). However, they have normal brain MRIs, no growth restriction, ID in only 2/5, and 2/4 actually have macrocephaly. All five probands display significant neurological symptoms including spastic di/paraplegia, ankle clonus, and hypo/hypertonia. As a result, they have gait abnormalities and are unable to walk independently (4/5) or are unable to walk at all (1/5). It is possible that heterozygous missense variants at the 480 residue variants confer a gain of function mechanism, rather than the loss of function resulting from biallelic variants. The features of this newly proposed autosomal dominant condition overlap with the recessive condition but appear to potentially differ in some aspects as well. Additional research on the effects of variants in *FAR1* is needed to further elucidate the clinical features and mechanism(s) of pathogenic variants.

1046T

Alpha-thalassemia X-linked intellectual disability (ATRX) syndrome in a Colombian patient. *A. Paredes, J. Rojas.* Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogota D.C, Colombia.

Introduction: Alpha-thalassemia X-linked intellectual disability syndrome is rare genetic disorder caused by mutations on ATRX gene, mapped on Xq21 chromosome, encodes a protein that participates in the pattern of DNA methylation, but is still unclear its association with pathological phenotypes. Characteristically, patients have small head circumference, telecanthus, short nose, tented upper lip, genital anomalies, severe developmental delays, hypotonia, intellectual disability, and alpha-thalassemia. Detection of alpha thalassemia, molecular and functional studies confirm diagnosis, allows monitoring of thalassemia and gastroesophageal reflux, major comorbidities that can cause death. **Case Report:** We present a male patient of 3 years old, referred for study of perineoscrotal hypospadia, microtia, marked global neurodevelopmental delay, behavioral disorder within autistic spectrum and pulmonary hypertension. He has CGHarray, G-banded karyotype 46,XY, brain MRI and echocardiogram normal, abdominal ultrasound reports left renal pyelectasis. Familiar history, two maternal cousins in second grade, one with autism and the other one with hypospadia. We performed the whole exome sequence analysis which reports a variant hemizygous c.2519G>T (pR840I) in ATRX gen. **Discussion:** Prevalence of Alpha-thalassemia X-linked intellectual disability syndrome is not known, results from mutation in ATRX gene, which encodes a protein that belong to the SWI/SNF family of chromatin remodeling proteins, causing abnormalities in the pattern of ADN methylation. The variant c.2519G>T (pR840I) in ATRX gen was identified in this patient; is already recognized in different populations. It is located in the helicase domain, theoretically can lead to delayed psychomotor development together with mild urogenital abnormalities. Mutations in this domain are associated with milder phenotypes than mutations in the PHD-like domain. But due to its unknown prevalence and low knowledge we bring this case report. **Conclusion:** This patient has a variant classified as Variant of Uncertain Significance (VUS) by in silico analysis, however the clinical record of the patient agrees with the phenotype described according to literature reviews, whereby we consider the patient have Mental retardation-syndrome, X-linked caused by mutation in the ATRX gene. The knowledge and early diagnosis of this syndrome could prevent deaths caused by symptoms that can be monitored and management.

1047F

Mutations in *DDX3X* are a common cause of syndromic intellectual disability. *X. Wang[1,2], J. Rosenfeld[1], C. Bacino[1,3], B. Graham[1,3], P. Moretti[1,4], F. Scaglia[1,3], L. Immken[5], J. Harris[5], S. Hickey[6,7], T. Mosher[7], A. Slavotinek[8], J. Zhang[2], J. Beuten[2], M. Leduc[1,2], W. He[2], V. Vetrini[2], M. Walkiewicz[1,2], W. Bi[1,2], R. Xiao[1,2], Y. Shao[1,3], D. Muzny[1,9], R. Gibbs[1,9], M. Eldomery[1], Z. Akdemir[1], T. Harel[1], J. Posey[1], J. Lupski[1,9], C. Eng[1,2], F. Xia[1,2], Y. Yang[1,2], members of the UDN.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Genetics, Houston, TX; 3) Texas Children's Hospital, Houston, TX; 4) Neurology, Baylor College of Medicine, Houston, TX; 5) Specially for Children, Austin, TX; 6) Clinical Pediatrics, the Ohio State University, Columbus, OH; 7) Division of Molecular & Human Genetics, Nationwide Children's Hospital, Columbus, OH; 8) Clinical Pediatrics, University of California, San Francisco, CA; 9) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Pathogenic variants in the *DDX3X* (Xp11.4) gene have been recently reported to cause syndromic intellectual disability (ID) patients predominantly affecting females (38 patients, PMID: 26235985). Here we report 26 individuals with *DDX3X*-related disorders in a cohort of 4839 patients with developmental disorders. Among the 4839 (2152 females, 2687 males) patients referred to our laboratory for clinical whole exome sequencing and had developmental delay (DD) and/or ID, 26 (24 female and 2 male) have pathogenic or likely pathogenic variants in *DDX3X*. A total of 25 unique variants were identified (23 novel and 2 reported previously in patients with ID). The 25 variants include 10 missense, 6 frameshift, 3 splice site, 4 nonsense, and 2 in-frame deletion/duplication changes. In 24 patients, the *DDX3X* variants were confirmed as *de novo*, supporting the variant pathogenicity. Two of the *de novo* variants are mosaic with allele frequencies of 14% and 21% (mutant reads/total reads, 19/137 and 69/327). Altogether, *DDX3X* ranks the third among approximately 450 genes with *de novo* variants in our laboratory (*ARID1B* the first, 43 patients; *ANKRD11* the second, 29 patients). The most frequent clinical presentations for the 24 female patients include DD and/or ID (24/24), hypotonia (17/24), dysmorphic features (17/24), structural brain abnormalities (15/24), movement disorder (15/24), visual impairments (8/24), microcephaly (6/24), and autism spectrum disorder (5/24). Some of these patients had respiratory problems (5/24) and congenital heart disease (4/24), which were not previously reported. Mitochondrial DNA depletion was found in two patients undergoing muscle biopsies, which is another feature that has not been described. On the other hand, skin abnormalities (3/24) are underrepresented in our cohort, in comparison to previous publications (14/38). In addition, we report for the first time a male patient carrying a *de novo* variant. He had phenotypes more severe than that of the reported male patients with inherited variants in *DDX3X*. In summary we identified 26 unrelated patients with causal variants in *DDX3X* and expanded the genotypic and phenotypic spectrum of *DDX3X* related disorders. The collective data suggest that *DDX3X* defects are a frequent cause of syndromic ID in female patients and the causal variants are likely to be loss-of-function.

1048W

It does not have to be the whole exome: Mendeliome sequencing increases the diagnostic yield in patients with unexplained intellectual disability by 30%. *A. Rump[1], L. Gieldon[1], L. Mackenroth[1], A. Kahlert[1], J. Lemke[2], K. Hackmann[1], A. Krueger[1], F. Kuhlee[1], F. Stuebner[1], A. Tzschach[1], A. Schrock[1], N. Di Donato[1].* 1) University of Technology Dresden, Institute for Clinical Genetics, Germany; 2) Institute of Human Genetics, University Clinic Leipzig, Germany.

   Introduction Advances in deciphering the genetic causes for impaired development have always been accompanied by advances in technology. In the past, array-CGH technology has boosted the detection rate significantly; but up to 50% of children with developmental delay still remain undiagnosed. Since the latest technological advance, next generation sequencing (NGS), has the power to improve the diagnostic yield tremendously, we applied this method to 108 index patients with developmental delay or intellectual disability (ID) and pre-excluded genomic imbalances. **Method** Genomic DNA samples of 80 parent-patient trios plus 28 individual patients were analyzed for mutations in 4813 genes, using the TruSight-One gene panel on the MiSeq platform (Illumina, San Diego, CA). After sequencing with median target coverage of 80-fold and mapping to hg19, sequence variants were called with the CLC Biomedical Genomics Workbench (Qiagen, Hilden, Germany). All *de novo* variants, homozygous variants with heterozygosity in both parents and compound heterozygous variants were analyzed as well as hemizygous variants in male patients. These variants were screened for clinical und molecular concordance (i.e. disease-association of the gene, published mutation). In addition, all splice-relevant variants, all frameshift and nonsense-mutations were screened for clinical concordance. The results were discussed in a team of clinicians and molecular geneticists, relevant variants were validated by Sanger-sequencing. **Results** Using the mendeliome in a diagnostic setting, we established a diagnosis in 33 of the 108 index patients (30%). For 12 further patients, we found one or two possibly causative candidates (11%). Five patients (4.6%) showed incidental findings which either made treatment or surveillance necessary. (e.g. homozygous *MUTYH*-mutations, *SDHA*-mutation). Nine patients (8%) have been identified to be carriers for a recessive disease (*PAH*- or *CFTR*-mutations). **Conclusion** Mendeliome next generation sequencing significantly increases the diagnostic yield in patients with ID unsolved by previous routine testing (array-CGH, conventional karyotyping) by at least 30%. However, variant interpretation remains challenging and calls for national and international data exchange and establishment of in vitro pipelines for variant evaluation.

1049T

Exploring the therapeutic potential of CRISPR/Cas9 technology for the treatment of MeCP2 duplication syndrome. *E. Maino[1,2], E.A. Ivakine[1], R.D. Cohn[1,2,3].* 1) Program in Genetics and Genome Biology, The Hospital for Sick Children Research Institute, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Canada; 3) Department of Paediatrics, University of Toronto and The Hospital for Sick Children, Toronto, Canada.

   Methyl-CpG-binding protein 2 (MeCP2) is a multifunctional chromatin protein required for the maintenance of neuronal functions. Duplications of the chromosomal region containing *MeCP2* gene drive the onset of MeCP2 Duplication Syndrome, an ultra rare X-linked condition, characterized by severe intellectual disability, motor dysfunction, recurrent respiratory tract infections and limited life expectancy. As of today, no suitable animal models of the disease are available, hampering the understanding of the disease pathogenesis and limiting the development of novel therapeutical strategies. To study MeCP2 Duplication Syndrome, I am generating the very first mouse model that recapitulates a patient duplication mutation, employing the new CRISVar technology. Based on this mouse model, I will then design a CRISPR/Cas9-one guide approach, recently proven effective in patients' cells by our group, aiming to remove the duplication and to improve disease outcomes. The ultimate long-term goal of the project is to develop gene editing strategies for treatment of MeCP2 Duplication patients, which can be translated to other genetic disorders caused by gene duplications. .

1050F

A 10q23.31 microduplication is associated to autosomal dominant primary microcephaly. *D. Oliveira[1], G.F. Leal[2], L.C. Caires[1], E. Goulart[1], C.M. Musso[1], K. Silva[1], A.L. Sertié[3], J.R. Oliveira[4], A.V. Krepischi[1], A.M Vianna-Morgante[1], M. Zatz[1].* 1) Instituto de Biociências, Universidade de Sao Paulo, São Paulo, Brasil; 2) Instituto de Medicina Integral Professor Fernando Figueira, Recife, Pernambuco, Brasil; 3) Hospital Israelita Albert Einstein, São Paulo, Brasil; 4) Departamento de Neuropsiquiatria, Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil.

　　A family with an autosomal dominant pattern of inheritance of primary microcephaly has been studied clinically and molecularly. The propositus was born to a nonconsanguineous couple, from an uneventful pregnancy. He attended to school, but dropped out of it by the middle of primary couse. Nevertheless, he was able to read and presented normal social skills. His OFC circumference was below the second percentile (SD < 2). Apart from him, his family consists of two other individuals with similar clinical profiles (his mother and sister) and a brother and father with normal head circumferences. Aiming to better understand the biological basis of microcephaly in these individuals, G banding karyotype was performed, but showed no abnormalities. DNA samples of the three affected and two normal patiens were evaluated through array comparative genome hybridization (*SurePrint G3 Human CGH Microarray, Agilent Technologies).* This analysis identified a 382 kb duplication in 10q23.31, segregating with microcephaly and absent in the two normal individuals. This chromosomal alteration, which has not been associated to autosomal dominant primary microcephaly yet, was observed as very rare, due to its absence in public databases of general population (Database of Genomic Variants- DGV; http://projects.tcag.ca/variation/). Additionally, according to DECIPHER ((http://decipher.sanger.ac.uk/), there have been described only patients with microdeletions in this region, with a variable overlapping with the copy number variation here presented. The 10q23.31 duplication involves intirely two genes, *PTEN* and *KLLN1,* and partially two others, *PAPSS2* and *ADA1A*. After informed consent, skin biopsies of two affected individuals were taken and employed for establishing a primary cell culture of fibroblasts. Functional studies are underway in order to better understand the molecular implications of this novel chromosomal microduplication.

1051W

A missense mutation in the CRBN gene that segregates with intellectual disability and self-mutilating behaviour in a consanguineous Saudi family. *A. Sheereen[1], M. Alaamery[1], S. Bawazeer[1], Y. AlYafee[1], S. Massadeh[1], W. Eyaid[2].* 1) Developmental Medicine Department, King Abdullah International Medical Research Centre, Ministry of National Guard Health Affairs. Riyadh. Saudi Arabia; 2) Department of Pediatrics, King Abdulaziz Medical City (KAMC); Ministry of National Guard-Health Affairs (MNGHA), Riyadh. Saudi Arabia.

　　Background Autosomal recessive non-syndromic intellectual disability (ARNS-ID) is an etiologically heterogeneous disorder. Although little is known about the function of human cereblon (CRBN), its relationship to mild cognitive deficits suggests that it is involved in the fundamental processes of human memory and learning. **Objectives** We aim to identify the genetic cause of intellectual disability and self-mutilation in a consanguineous Saudi family with five affected members. **Methods** Clinical whole-exome sequencing was performed on the proband patient, and Sanger sequencing was done to validate and confirm segregation in other family members. **Results** A missense variant (c.1171T>C) in the CRBN gene was identified in five individuals with a severe intellectual disability (ID) in a consanguineous Saudi family. The homozygous variant was co-segregating in the family with the phenotype of severe ID, seizures and self-mutilating behaviour. The missense mutation (p.C391R) reported here results in the replacement of a conserved cysteine residue by an arginine in the CULT (cereblon domain of unknown activity, binding cellular ligands and thalidomide) domain of CRBN, which contains a zinc-binding site. **Conclusions** These findings thus contribute to a growing list of ID disorders caused by CRBN mutations, broaden the spectrum of phenotypes attributable to ARNS-ID and provide new insight into genotype–phenotype correlations between CRBN mutations and the aetiology of ARNS-ID.

1052T

Novel *AHDC1* mutations cause intellectual disability and developmental delay. *Y. Tsurusaki[1], Y. Enomoto[1], K. Ida[2], K. Kurosawa[2].* 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan.

Intellectual disability (ID) is a neurodevelopmental disorder that is characterized by significant limitations in intellectual functioning and adaptive behavior that are apparent before 18 years of age. An IQ score of less than 70 is characteristic. The prevalence of ID has been estimated at 1% of the worldwide population. To date, more than 700 causative genes have been reported for all forms of ID. To elucidate genetic causes in ID and developmental delay, we performed targeted resequencing using a TruSight One Sequencing Panel or whole exome sequencing in two unrelated Japanease pedigrees including a total of three patients. Genomic DNA libraries were sequenced using MiSeq with 151-bp paired-end reads or HiSeq2500 with 101 bp paired-end reads. Mapping to human genome hg19 was performed using Burrows-Wheeler Alignment. Aligned reads were processed by Picard to remove PCR duplicates. Variants were called using the Genome Analysis Toolkit, and annotated by ANNOVAR. Copy number variation analysis was performed using the log2 ratio of read depth derived from next generation sequencing data. We found that *AHDC1* were mutated in a patient from a family. The novel heterozygous mutation c.1167_1168insATAG (p.Pro390Ilefs*128) confirmed in both the patient and his affected brother by Sanger sequencing, but this mutation was not identified in either of the parents. To investigate the mosaicism, we further performed deep resequencing for this mutation using Nextera DNA Library Preparation Kit. Somatic mosaicism was confirmed in the mother in her saliva and blood, respectively. Furthermore, we found 1.86Mb deletions at 1p36.11–p35.3 involving *AHDC1* in a patient from another family. Our study demonstrated that *AHDC1* mutations can cause intellectual disability and developmental delay and the diagnostic utility of next generation sequencing technologies is obvious.

1053F

Mutation in *OASL* gene causing speech delay and intellectual disability. *M. Alfadhel[1], A. Alasseri[2], W. Alharbi[2], S. Al Ghamdi[2].* 1) Genetic Division, Department of Pediatrics, King Abdullah Specialized Children's Hospital, King Abdul Aziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia; 2) Medical Genomics Research Department, King Abdullah International Medical Research Centre, King Abdul Aziz Medical City, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia.

Background: *OASL* (oligoadenylate synthase-like) encodes an interferon-induced protein that is a member of the 2-5A synthetase family. 2-prime,5-prime oligoadenylates (2-5As) bind to and activate RNase L (180435), resulting in general RNA degradation and consequent inhibition of protein synthesis. 2-5As are produced by a well-conserved family of interferon-induced enzymes, the 2-5A synthetases or OASs. **Methods**: Whole exome sequencing followed by RNA sequencing. **Results:** The phenotype is a neuropsychiatric disorder characterized by speech delay, intellectual disability and autistic behavior. MRI scan of the brain showed no abnormalities. Whole exome sequencing revealed a homozygous non-sense mutation in *OASL* gene as follows: c.1054C>T/p.Arg352X **Conclusion:** We report the first human case of a biallelic truncation of the *OASL* gene, and alert clinicians to consider this novel gene variant in any patient with speech delay.

1054W

Only genotype-first approach permits *BRWD3* mutations' diagnosis. *J. Delanne[1,2,3,4], M. Lecat[1], P. Blackburn[5,6], E. Klee[7], C. Stumpel[8], S. Stegmann[8], S. Steven[8], M.S. Mulhern[9], C. Nava[10], D. Heron[10], B. Keren[10], S. Desai[11], S. Naidu[11], C. Poe[2,3,4], M. Chevarin[2,3,4], T. Jouan[2,3,4], J. Thevenon[1,3,4], P. Kuentz[2,3,4], E. Tisserand[3,4], Y. Duffourd[3,4], C. Philippe[2,3,4], L. Faivre[1,2,3,4], C. Thauvin-Robinet[1,2,3,4].* 1) Centre de Génétique et Centre de référence « Anomalies du Développement et Syndromes Malformatifs », Hôpital d'Enfants, Centre hospitalier universitaire de Dijon, Dijon, Dijon, France; 2) Laboratoire de Génétique chromosomique et moléculaire, UF Innovation en diagnostic génomique des maladies rares,, Centre Hospitalier Universitaire de Dijon, Dijon, France; 3) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (FHU TRANSLAD), Centre Hospitalier Universitaire de Dijon et Université de Bourgogne-Franche Comté, Dijon, France; 4) UMR-Inserm 1231 GAD team, Génétique des Anomalies du développement, Université de Bourgogne Franche-Comté, F-21000 Dijon, France; 5) Center for Individualized Medicine, Mayo Clinic, Jacksonville, Florida 32224, USA; 6) Department of Health Sciences Research, Mayo Clinic, Jacksonville, Florida 32224, USA; 7) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 8) Department of Clinical Genetics and School for Oncology and Developmental Biology (GROW), Maastricht University Medical Center (MUMC), Maastricht, The Netherlands; 9) Institute for Genomic Medicine, Columbia University, New York, NY 10032, USA; 10) Assistance Publique-Hôpitaux de Paris, Département de Génétique, Hôpital Pitié-Salpêtrière, Paris, France; 11) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, Maryland, 21205.

Since first description of BRWD3 related phenotype in 2007, only two additional families have been reported with intellectual Disability (ID). Because phenotype appeared nonspecific, only genotype first approach could allow identifying new *BRWD3* variants. Indeed**,** using whole exome sequencing (WES) and international datasharing, we identified six new unrelated cases with *BRWD3* variant (4 males and 2 females). X-inactivation study performed in 1/2 female cases showed skewed X inactivation (100%/0%). Reviewing the 9 previously cases published in the literature, we describe the phenotype of 13 males and 2 females with *BRWD3* variant, representing 10 different variants (6 truncating variants and 4 missenses). Most common features in males (excluding the one with the mosaic variant) included ID (11/12 cases), speech delay (11/12 cases), macrocephaly with high forehead (9/12 cases), long face due to pointed chin (8/12 cases), shyness (6/12 cases), large fingers or toes (6/12 cases), and large ears (5/12 cases). Both females presented with macrocephaly and speech delay, as well as epilepsy that was only reported in one other male. Among these 15 cases, 10 variants (all in males) were maternally inherited from unaffected mothers and 5 cases occurred *de novo* (including the 2 females and the male with a mosaic variant). In males, the 8 BRWD3 variants included 3 frameshifts (including 1 in mosaic), 1 stop-gained, 1 partial deletion and 3 missenses. When one female had a frameshift, the other presented a missense in the WD40 containing repeat domain with complete X-skewed inactivation. Macrocephaly was absent in both patients with missense variants but was present in all patients with truncating variants. This study evidences that the *BRWD3* related phenotype with ID, behavioural disturbances and macrocephaly could affect males and females, remaining mainly unspecific and hardly recognizable for clinicians. Only genotype-first approach permits BRWD3 mutations' diagnosis. This fine clinical description will be so helpful for reverse phenotyping after whole exome/genome sequencing.

1055T

Attempts to elucidate role of *ZBTB11* gene as a novel candidate gene in intellectual disability. *Z. Fattahi[1], T. Sheikh[2], K. Kahrizi[1], R. Harripaul[2], F. Larti[1], N. Bazazzadegan[1], M. Haddadi[3], M. Ansar[4], H.H. Ropers[5], J.B. Vincent[2], H. Najmabadi[1].* 1) Genetics Research Centre, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran, Islamic Republic of; 2) Molecular Neuropsychiatry & Development (MiND) Lab, Campbell Family Mental Health Research Institute, Center for Addiction and Mental Health, Toronto, ON, Canada; 3) Department of Biology, Faculty of Science, University of Zabol, Zabol, Iran, Islamic Republic of; 4) Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan; 5) Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany.

Exploring genes and pathways underlying Intellectual Disability (ID) can provide insights into brain development and function clarifying the complex puzzle of how cognition develops. Nowadays, gene discovery is not a rate-limiting step in ID but biological interpretation of the identified mutations seems indispensable. As part of ongoing systematic studies to identify candidate ID genes, linkage analysis and next generation sequencing revealed *ZBTB11,* as a novel candidate ID gene, in two consanguineous families. *ZBTB11* encodes an understudied transcription repressor and as predicted by three-dimensional modeling, the two identified variants (p.H729Y, p.H880Q) disrupt canonical Zn2+-binding residues of C2H2 zinc finger domains, leading to possible altered DNA binding. This prediction was proved by localization study performed in HEK293T cells transfected by wild-type/mutant ZBTB11-GFP constructs. We found abolished localization for mutants being excluded from the nucleolus where the wild-type recombinant protein is localized. Next, exploring target genes and pathways, ChIP-sequencing was performed. Although ChIP-sequencing results did not fulfill quality criteria, the enriched pathways and biological functions detected with the help of ENCODE data of HEK293 stably expressing eGFP-ZBTB11 is in concordance with its nucleolus localization. The genes with "ribonucleoprotein complex binding" GO molecular function are enriched indicating fundamental role of this gene in cellular pathways. This provides another evidence for the role of mutated *ZBTB11* in ID, since abnormal nucleoli is considered as one of the hallmarks in some developmental disorders with ID. Recently, it is shown that this gene is widely expressed in Zebrafish early development that remains in nervous system afterwards. The zebrafish mutant has shown brain and spinal cord degeneration with prominent apoptosis of central nervous system. In order to provide another model, we started to investigate ZBTB11-ortholog in Drosophila brain by targeting RNAi using UAS/Gal4 system. The drosophila F1 offspring with lower expression of ZBTB11-ortholog in brain were generated. Immunofluorescence study of mushroom bodies in drosophila's brain revealed no prominent structural changes, not excluding the presence of subtle changes but proposing the involvement of ZBTB11 in more basic pathways that should be assessed by the upcoming behavior assays. In conclusion, this study elucidates involvement of *ZBTB11* in ID.

1056F

Inositol monophosphatase 1 (IMPA1) deficiency and cognitive impairment. *T. Figueiredo[1], C.P. Walker[2], G.S. Kobayashi[1], A. Pessoa[3], P. Nóbrega[4], U.S. Melo[1], D.F. Oliveira[1], N. Murphy[2], M. Rafferty[2], F. Kok[1], R. Cho[2], S. Santos[5], M. Zatz[1].* 1) Human Genoma and Stem Cell Research Center, Biosciences Institute, University of São Paulo, São Paulo, Brazil; 2) Psychiatry, University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Neurology Department, State University of Ceará, Fortaleza, Brazil; 4) Neurology Department, Federal University of Ceará, Fortaleza, Brazil; 5) Centre for Public Health Research, Paraíba State University, Campina Grande, Brazil.

Dysfunctions of the inositol cycle has been implicated in a wide variety of human diseases including developmental defects and neurological diseases. IMPA1 has attracted much interest in genetic studies of neuropsychiatric diseases because at therapeutic concentrations, lithium, the main pharmacological treatment for bipolar disorder it is an uncompetitive inhibitor of IMPA1. Although animal models and *in vitro* analysis have contributed to the understanding of the pathophysiology of IMPA1 deficiency, no human disease had been attributed to a malfunction of this protein. We identified a homozygous 5 bp duplication (c.489_483dupGGGCT) in *IMPA1*, leading to a frameshift (p.Ser165Trpfs*10), in a consanguineous family with severe intellectual disability and disruptive behaviors from geographically isolated and very poor region of Northeastern Brazil (Figueiredo *et al.*, 2015). The *IMPA1* mutation generates a restriction site for the *BseYI* enzyme (New England BioLabs), allowing to identify three different bands patterns according to the three different genotypes. We investigated the mutation in 69 family members and identified: 9 affected homozygotes (AH), 21 heterozygotes (HET) and 39 normal homozygotes (NH). Next, we investigated resting Electroencephalogram (EEG) data from 30 participants (17 NH, 9 HET, 4 AH). Brain Magnetic Resonance Imaging revealed no structural abnormalities in affected individual. In resting conditions we observed a parametric reduction in α- and β-band power over right frontal and temporal scalp regions (i.e., WT > HET > HOM). Eyes-closed dominant gamma2 frequency and γ1 and γ2 variability demonstrated a parametric increase across groups in frontal electrodes (i.e., WT < HET < HOM). This pattern is consistent with *IMPA1* mRNA expression patterns which show higher expression in subcortical and prefrontal brain regions. This data suggests that the loss of function associated with *IMPA1* mutation may be related to impaired regulatory mechanisms (i.e., α/β activity) with subsequent dysregulation of high frequency synchrony. We have now generated induced pluripotent stem cells (iPSCs) from affected patients blood samples to further investigate the effects of mutation on disruption of signaling pathway networks and morphofunctional characteristics in neuronal cell derived from iPSCs. Funding: FAPESP, CNPq and FAPESQ. .

1057W

A novel pathogenic variant of the *HECW2* gene in a Japanese boy with global developmental delay, hypotonia, and short stature. *M. Minatogawa[1], M. Maeda[2], K. Yanagi[1], K. Hata[3], Y. Matsubara[4], T. Kaname[1].* 1) Genome Medicine, National Center for Child Health and Development, Tokyo, Japan; 2) Saga Handicapped Children's Hospital, Saga, Japan; 3) Maternal-Fetal Biology, National Center for Child Health and Development, Tokyo, Japan; 4) National Center for Child Health and Development, Tokyo, Japan.

Background: Exome sequencing is an effective strategy to identify genetic causes of developmental delay, intellectual disabilities (ID), and congenital anomalies. Recently, it has been reported that loss-of-function variants of *HECW2* gene are assumed to be responsible for neurodevelopmental disorder with hypotonia, seizures, and absent language [Berko et al., 2017]. We report on a Japanese boy with global developmental delay, autism spectrum disorder, hypotonia, short stature, and mild dysmorphic facial features associated with a novel variant in the *HECW2* gene. **Case Report:** The propositus was a 5-year-old male. He was born at 40 weeks of gestation by normal delivery after uneventful pregnancy. His birth weight was 2998 g, and his length was 48.2 cm. He presented with global developmental delay, hypotonia, and feeding difficulties in his infancy. On closer clinical examination at the age of 1 year old, nothing was to be found but brain MRI revealed arachnoid cysts in both of the middle cranial fossa. G-Banded karyotype showed a normal male karyotype 46,XY. At the age of 5 years, he was referred to our hospital to determine the existence of the underlying disease. At that time, he also showed short stature 99cm (-2.0 SD), mid-facial hypoplasia, autism spectrum disorder, and severe constipation. **Methods:** Written informed consent was obtained from his parents. Whole exome sequencing was performed using Human All Exon V6 kit (Agilent) and Hiseq 2500 (Illumina). The detected variants were confirmed by Sanger sequences. **Results:** A novel heterozygous missense variant was detected. The variant was *de novo* and was not found in the variant database, such as 1000Genome, dbSNP and ExAC. The variant lies within the HECT domain of the protein. **Conclusion:** We found the novel variant in the *HECW2* gene related to neurodevelopmental disorder. A few cases of neurodevelopmental disorder caused by the pathogenic variants of HECW2 have been reported. It has not been known well about the pathogenic variants specific mechanisms. To reveal the phenotypes associated with *HECW2* variants required further analysis. Our report of phenotypes data on the case will be helpful to establish the *HECW2* related disorder.

1058T

De novo LoF mutations in *MED12* cause a syndromic form of X-linked ID in females. *D.L. Polla[1], R.B. Hufnagel[2], A.T. Vulto-van Silfhout[3], R. Pfundt[3], E.M.H.F. Bongers[3], D. Li[5], E.J. Bhoj[5], J.B.G.M. Verheij[4], J.S. Klein Wassink-Ruiter[4], H. van Bokhoven[1], C. Zweier[6], A.P.M. de Brouwer[1].* 1) Department of Human Genetics, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands; 2) Ophthalmic Genetics and Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, Maryland, USA; 3) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 4) Department of Medical Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; 5) Department of Pediatrics, Division of Human Genetics and Molecular Biology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 6) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.

Intellectual disability (ID) is a clinically and genetically heterogeneous disorder affecting 2-3% of the general population. There are many genetic and non-genetic causes of ID; the X-linked forms are most intensively studied. Mutations in the X-linked gene *MED12* are known to cause at least three different, but closely related, syndromic intellectual disability disorders in males: Opitz-Kaveggia syndrome, Lujan-Fryns syndrome and Ohdo syndrome. These hemizygous mutations are invariably missense changes suggesting that loss-of-function (LoF) mutations, i.e. nonsense, frame shift, and canonical splice site variants, are likely embryonically lethal. Carrier females in these families are usually not affected. MED12 is a member of the Mediator complex, a multiprotein complex which has a central role in RNA polymerase II transcription. Mediator regulates signals involved in cell growth, development and differentiation, as well as in the Sonic hedgehog signaling and the extraneuronal gene silencing network. This could explain its role in the different X-linked intellectual disability syndromes. In our study, we found *de novo* LoF mutations in five females, seemingly resulting in an overlap in phenotypic features consisting of ID, speech delay, short stature, tall prominent forehead, low set ears, syndactyly, and severe feeding problems. These features are present in at least four out of the five patients. We hypothesize that even though LoF mutations in males are lethal, LoF mutations in females are not lethal due to X-inactivation and give rise to a clinically distinct intellectual disability phenotype.

1059F

A novel missense variant in the POMK gene causes Walker-Warburg syndrome. *E. Preiksaitiene[1], N. Voisin[2], L. Gueneau[2], E. Benušienė[1], A. Reymond[2], V. Kučinskas[1].* 1) Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; 2) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies) type A, 12, also known as Walker-Warburg syndrome represents the most severe phenotype of dystroglycanopathies (OMIM 615249). We present a new family with three affected individuals manifesting clinical features of Walker-Warburg syndrome due to homozygous nonsense variant in the *POMK* gene. Whereas the first pregnancy of an unrelated Lithuanian couple was uneventful and gave birth to a healthy son, the second pregnancy fetal ultrasound examination at 22-23 weeks of gestation showed internal hydrocephalus of a female fetus. Both infections and molecular genetic tests for aneuploidies were negative. Delivery was induced at 32 weeks of gestation and resulted in a stillbirth with severe hydrocephalus. During the third and fourth pregnancies internal hydrocephalus of male fetuses has been noted at 12-15 weeks of gestation and the pregnancies have been terminated. Exome sequencing of fetal DNA from the third fetus revealed a homozygous nonsense variant c.136C>T (p.(Arg46Ter)) in the *POMK* gene. Sanger sequencing confirmed that this variant was homozygote in the proband and the two other affected siblings and heterozygote in the parents and their healthy son. Only three families with pathogenic variants in the *POMK* gene, causing congenital muscular dystrophy-dystroglycanopathy with brain and eye anomalies have been reported to date. Clinical and molecular findings of this fourth family provide additional details of the phenotypic manifestation of *POMK*-related congenital disorders.

1060W

Clinical heterogeneity in Renpenning Syndrome patients due to c.459_462del mutation in PQBP1 gene: A case report. *J. Rojas, A.I Sánchez.* Cundinamarca, Pontificia Universidad Javeriana, Bogota, Cundinamarca, Colombia.

   Introduction: Mutations in *PQBP1* have been identified in families with non-syndromic X-linked intellectual disability and syndromic cases like Sutherland-Haan, Cerebropalatocardiac, Golabi–Ito–Hall, and Renpenning syndromes (RS) among others. This gene encodes for the polyglutamine tract binding protein1, located in Xp11.2. The most frequently clinical features associated with mutation in *PQBP1* are microcephaly and short stature, however some other characteristics may be seen such as spasticity, small testes, anal atresia or stenosis, cardiac defects, ocular coloboma and cleft palate. RS is a very rare and well-recognized condition of unknown prevalence, characterized by severe cognitive impairment, microcephaly, small stature and hypogonadism. **Case report**: 22 y/o male patient who was the 1st pregnancy of a healthy non-consanguineous couple. Family history was unremarkable. He was born by uncomplicated spontaneous vaginal delivery after a full non-controlled preterm gestation. Birth weight at 36 weeks was 2350g and height was 47cm. Tetralogy of fallot was diagnosed at birth for which he required surgery. Patient was classified as having moderate cognitive impairment. Physical examination revealed microcephaly, prominent ears, long-shaped face, mild facial coarsening and thick lips. Short stature and pectus carinatum were seen with no hypogonadism/small testes. A friendly personality was evident. Karyotype, aCGH and molecular testing for Williams syndrome were negative. Whole exome sequence analysis was performed. A pathogenic hemizygous mutation in *PQBP1* gene was detected: c.459_462del (p.(Arg153Serfs*41). Mother was a carrier. **Discussion**: Approximately 80 X-linked genes have been reported associated with intellectual disability. Mutations in *PQBP1* gene have considerable inter and intrafamilial phenotypic variation. This variant has been previously reported as a cause of X-linked intellectual disability and has been reported as a gain-of-function pathogenic mutation that interacts with FMRP protein. This interaction may suggest an overlapping phenotype of our patient and X fragile syndrome patients. It is important to highlight that the typical features of small testes in RS were not seen in this patient, however he present with some dysmorphic features that to our knowledge, have never been reported with this mutation. **Conclusion**: more genotype-phenotype correlation studies are necessary to characterize the whole phenotype of this very rare syndrome. .

1061T

Intellectual disability and macrocephaly in three sisters from a consanguineous marriage with a novel *SZT2* homozygous mutation. *C. Trujillo[1], M.K. Alwasiyah[2,3], M.I. Naseer[2], M. Abu-Elmagd[2], M.A. Jafri[2], R.A. Bajammal[3], A. Chaudhary[2], H. Bouazzi[4,5], M. Al-Qahtani[2].* 1) Erfan and Bagedo Hospital, Jeddah, Saudi Arabia; 2) Center of Excellence in Genomic Medicine Research;Jeddah, Saudi Arabia; 3) Aziziah Maternity & Children Hospital, Jeddah, Saudi Arabia; 4) Research Unit: Epidemiology, Etiology of Congenital Malformations and Therapeutics, Monastir Faculty of Medicine, Tunisian Republic; 5) Hôpital Necker - Enfants Malades, Medical Genetics Laboratory, Paris Descartes University, France.

   We have identified a large consanguineous Saudi family with 3 sister presenting a similar clinical picture of developmental delay, intellectual disability, high forehead and macrocephaly.
At 3 years of age we saw the third affected sister. She was the born to first cousins parents, by cesarean section due to macrosomia. Her birth weight was 3600 gr, height 52 cm and Head Circumference 36 cm. Apgar score 8 at one minute and 10 at five minutes. She was discharge with no complaints.
At 3 years of age she was not able to walk or stand or sit without support and she cannot crawl. Her language was practically absent and limited to guttural sounds without words. There was no history of seizures or abnormal movements, although one of the sisters had history of seizure.
Her HC was 51 cm in the 95th percentile, Height was 90 cm in the 10th percentile and her weight 11.5 kg below the 5th percentile.
She did not have dysmorphic features or skin stigmata. Cranial nerves were normal, cardiopulmonary normal, and no abdominal masses or organomelia. Routine laboratory CBC, thyroid hormone, liver enzymes, creatinine, ammonia, lactic acid etc were normal. Tandem Mass Spectrophotometry was normal Brain MRI showed prominent extra axial CSF space with wide Sylvian fissure, no abnormal signal intensity, no ventriculomegaly, normal corpus callosum, normal posterior fossa and cerebellum.
Exome sequencing was performed in affected siblings of this family and their parents. Whole exome sequencing data analysis, confirmed by subsequent Sanger sequencing validation, identified a novel homozygous mutation c.9368G>A resulting in a substitution of a conserved glycine residue into a glutamic acid in the exon 67 of *SZT2* gene. The mutation was ruled out in 100 unrelated healthy controls. The missense variant detected in this study has not yet been reported as pathogenic in literature or variant databases. Seizure threshold 2 (*SZT*2) gene located on chromosome 1p34.2 encode protein mainly expressed in the parietal, frontal cortex and dorsal root ganglia in the brain. Previous studies in mice showed that mutation in this gene can confers low seizure threshold, enhance epileptogenesis and in human may leads to facial dysmorphism, intellectual disability, seizure and macrocephaly.
This homozygous *SZT2* variant might be the causative gene that further explain the ID and developmental delay in this Saudi family.

1062F

The phenotypic spectrum of Xia-Gibbs Syndrome. *Y. Jiang[1,2], M.F. Wangler[2,3], A.L. McGuire[4], J.R. Lupski[1,2,5,6], J.E. Posey[2], F. Xia[2], Q. Meng[1,2], M. Murugan[1,2], R.A. Gibbs[1,2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030; 3) Texas Children's Neurological Research Institute, Houston, TX, 77030; 4) Center for Medical Ethics and Health Policy, Houston, TX, 77030; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX,77030; 6) Texas Children's Hospital, Houston, TX,77030.

   Background: *De novo* truncating mutations of AT-Hook DNA Binding Motif Containing 1 gene (*AHDC1*) lead to Xia-Gibbs Syndrome (XGS: OMIM # 615829). Phenotypes of this disorder were described in an initial report of four probands (Xia et al., 2014: PMID 24791903) and a subsequent report of seven patients (Yang et al., 2015: PMID 27148574), and include intellectual disability, developmental delay, hypotonia, speech delay and dysmorphic features. We have established a Xia-Gibbs Syndrome registry in order to more systematically evaluate the clinical spectrum of XGS through collecting detailed clinical information from patient families. Here we report the aggregate evaluation of the clinical spectrum of a cohort of 19 patients including five individuals previously reported.   **Methods:** We obtained institutional review board (IRB) approval at BCM to recruit families to the Xia-Gibbs Syndrome registry. A total of 19 families in the registry submitted a comprehensive clinical survey to allow detailed assessment of phenotypic information.   **Results:** This cohort includes 10 male and 9 female patients at a median age of 8 (3~21) years old. Common phenotypes include: intellectual disability (17/19), hypotonia (17/19), speech delay (19/19), ataxia (13/19), sleep apnea (8/19), abnormal MRI (8/19) and strabismus (8/19). We also assessed the age range of patients who achieve the following developmental milestones: the median age for independent walking (16 patients) is 2.5 (1.5~6) years. The median age for using first words (12 patients) is 2.75 (1~5) years and the median age for using two words together (8 patients) is 3.5 (2~7) years. In addition, seven of the 19 patients were reported to have no speech, and all these individuals were male (p < 0.01). Seizures were reported in 6 out of 19 patients, with median age-of-onset = 4 years (range: 9months~12 years). Scoliosis was reported in 4 out of 19 patients, an increase rate relative to the general population (p<0.001). **Conclusion:** Data from this cohort have expanded our knowledge of the common phenotypes of XGS and provided further information for developmental milestones. In addition, we identified that XGS patients are at an elevated risk of developing seizures and scoliosis, and male patients are more likely to be non-verbal than females. The ability to aggregate data from the observation and care of individuals with XGS is a key step for clinical management, family counseling and molecular research into the condition.

1063W

Updating penetrance estimates for deletion and duplication syndromes with variable phenotypic manifestation. *J.W. Ahn[1], A. Corrigan[2], S. Bint[2], C. Mackie Ogilvie[2].* 1) Genetics Laboratories, Guy's & St Thomas' NHS Foundation Trust, London, United Kingdom; 2) Genetics Laboratories,Viapath, London, United Kingdom.

   Deletion and duplication syndromes that exhibit incomplete penetrance are frequently detected by chromosomal microarray-based testing for copy number variation in the genomes of patients referred for neurodevelopmental disorders. Examples include 1q21.1 (OMIM 612474) and 16p11.2 (OMIM 611913) deletion syndromes, and collectively, these syndromes are often referred to as susceptibility loci, as the expressivity of associated phenotypic traits is highly variable. These syndromes also make up a large proportion of the findings in these patients, e.g. 16p11.2 deletion syndrome represents ~10% of all syndromic findings in this cohort. Due to the clinical importance of this group of syndromes, attempts have been made to estimate penetrance for each of these susceptibility loci, most notably by Rosenfeld et al. (2013; doi:10.1038/gim.2012.164) and Kirov et al. (2014; dx.doi.org/10.1016/j.biopsych.2013.07.022). Respectively, these studies described a diagnostic cohort of ~30,000 cases compared to ~22,000 controls, and ~6800 cases against ~6300 controls. However, both these studies derive penetrance estimates from the incidence of each syndrome, and incidence (number of new cases detected in a given period of time) may vary between populations due to a variety of factors including testing methodology, access to genetic healthcare and population stratification-type effects. The Guy's & St Thomas' Regional Genetics Centre serves a population of ~7 million in South East England and patients referred for neurodevelopmental disorders are routinely tested by chromosomal microarray. We present data for 12 susceptibility loci from a diagnostic cohort of ~30,000 patients. Of these, 11 were found to have a lower incidence than in the diagnostic cohort presented by Rosenfeld et al, and penetrance estimates are accordingly affected. The largest difference was for 1q21.1 duplication syndrome (OMIM 612475) which had a >3-fold reduced incidence in our cohort. These findings have implications for variant interpretation and patient counselling for our diagnostic cohort. Furthermore, they indicate that incidence of these syndromes may also be different in other cohorts, which questions the applicability of previously published estimates across different populations.

1064T

Exome Pool-Seq in neurodevelopmental disorders. *C. Zweier, A. Ekici, C. Thiel, J. Hoyer, A. Wiesener, C. Kraus, A. Reis, B. Popp.* Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.

High throughput sequencing has greatly advanced disease gene identification, especially in heterogeneous entities such as neurodevelopmental disorders (NDDs), comprising intellectual disability (ID) and autism spectrum disorders. Despite falling costs this is still an expensive and laborious technique, particularly when studying large cohorts. To address this problem we developed Exome Pool-Seq as an economic and fast screening technology. Sequencing of 96 individuals can be performed in eight pools of 12 equimolar concentrated samples on less than one Illumina sequencer lane. Calling variants using a ploidy of 24 and performing subsequent validation and segregation analysis by Sanger sequencing, we identified 27 (likely) pathogenic mutations in a pilot study of 96 cases. Twenty five loss-of function or likely pathogenic (previously reported, deleterious prediction) missense variants were identified in 923 established NDD genes (based on SysID database, status November 2016) (*ACTB, AHDC1, ANKRD11, ATP6V1B2, ATRX, CASK, CHD8, GNAS, IFIH1, KCNQ2, KMT2A, KRAS, MAOA, MED12, MED13L, RIT1, SETD5, SIN3A, TCF4, TRAPPC11, TUBA1A, WAC, ZBTB18, ZMYND11*), two in 543 (SysID) candidate genes (*ZNF292*, *BPTF*), and additionally a *de novo* loss-of-function variant in *LRRC7,* not previously implicated in NDDs. Identified mutations included mainly *de novo*, but also X-linked and autosomal-dominantly or autosomal-recessively inherited variants. With a detection rate of 28 %, Exome Pool-Seq achieves comparable results to individual exome analyses but reduces costs by more than 85 %. Compared to other large scale approaches using Molecular Inversion Probes (MIP) or panels of a predefined set of disease genes, it allows flexible re-analysis of data at any time. Exome Pool-Seq is thus well suited for large-scale, cost-efficient and flexible screening in characterized but heterogeneous entities like NDDs.

1065F

Novel mutation in *ARHGEF9* associated with developmental delay and seizures. *E. Fattakhov[1], M. Byler[1], K. Werner[2], R.R. Lebel[1].* 1) Genetics Section, Upstate Medical University, Syracuse, NY; 2) Neurology Department, Upstate Medical University, Syracuse, NY.

This boy was the product of term delivery after an uncomplicated pregnancy in a 22 year old primigravid woman and her 27 year old non-consanguineous partner, in their native Macedonia. The child presented to genetics clinic at age 3.5 years with global developmental delays. Brain MRI, metabolic panels, and SNP array all had normal/negative results before this evaluation. Our whole exome sequencing revealed an alteration, p.E115K (c.3343G>A), in the *ARHGEF9* gene on the X chromosome. Mother left when child was 2 months of age and is likely still in Macedonia. She was unavailable to determine whether the trait is inherited or *de novo*. Because it has not been reported to date, this alteration is considered to be a variant of uncertain clinical significance. However: the amino acid is highly conserved in vertebrates, the alteration is predicted theoretically to be significant, and it has not been found in a large cohort of persons with normal phenotypes. Further, Table 1 demonstrates a close correspondence of the phenotype to the eight cases reported thus far. The E115K mutation has the potential to modify secondary protein structure, given that glutamine contains a polar neutral side chain, while lysine is a basic electrically charged amino acid. The *ARHGEF9* gene typically encodes functional collybistin, a neuronal GDP-GTP exchange factor. Collybistin has been shown to interact with rapamycin complex 1 (mTORC1), which is vital for neural differentiation in the central nervous system. Furthermore, collybistin binds to the neuronal scaffolding protein gephyrin, which allows for the post-synaptic recruitment of GABA$_A$ receptors, thus modulating the robustness of inhibitory synapses. We propose that alteration of *ARHGEF9* is the cause of the seizures and the developmental delay in this patient.

Table 1. *ARHGEF9* **Mutation Phenotype**.

	1	2	3	4	5	6	7	8	9
Seizures	+	+		+	+	+	+	+	
Intellectual disability	+	+	+	+	+	+		+	+
Speech delay	+				+				+
Brachycephaly	+								
Macrocephaly				+					
Trigonocephaly						+			
Flared eyebrows	+								
Low-set ears	+								
Facial dysmorphisms	+		+						
Hyperekplexia		+							
Sensory Hyperarousal			+						
Hyperactivity		+	+						
Disturbed Sleep-wake cycle			+						
Brain MRI: cerebral atrophy		+				+			
Brain MRI: polymicrogyria						+			

1066W

Is incontinentia pigmenti a genetic male disease? *F. Fusco[1], R. Sabbatella[1], V. Valente[1], D. Fergola[1], M.B. Lioi[2], M.V. Ursini[1].* 1) Institute of Genetics and Biophysics "ABT", IGB-CNR, Naples, Italy; 2) University of Basilicata, Potenza, Italy.

Post-zygotic Somatic mosaicism of *NEMO/IKBKG* mutation causes Incontinentia Pigmenti (IP; OMIM#308300) in IP male patients. Indeed, IP is an X-linked dominant disease and generally is lethal in males, and only the females survive. Therefore the transmission of the disease is always from the IP affected mother to the daughter. We report for the first time that IP males are able to transmit their *NEMO/IKBKG* somatic mutation to their IP children by a germline mosaicism for the same mutation. We found this *unusual transmission* of the disease from IP father-to-daughter in two families belonging to the Incontinentia Pigmenti Genetic Biobank, IPGB (http://www.igb.cnr.it/ipgb). We report the Family1 with the IP father and his IP-daughter with classical IP signs, both with novel p.Gln132X mutation of *NEMO/IKBKG* and the Family2 with the IP father had two IP-daughters all with the same NEMO/IKBKG recurrent exon4-10 deletion. Low level mosaicism for the NEMO/IKBKG mutation in the skin (20%; 1%) in the urine (8.3%; 35%) and in germline tissue (16.7%; 35%) was detected in both IP fathers, respectively. We concluded that the somatic mosaicism the IP males includes germline mosaicism that impacts not only on the transmission but also on the recurrence risk of the disease. We are evaluating 38 additonal IP males, teenagers belonging to our IPGB biobank testing them for mosaicism which represent a helpful tool in the genetic counseling of these families.

1067T

A new X-linked form of syndromic intellectual disability on Xp11.22. *D.A. Scott[1,2], C. Grau[1], M. Starkovich[2], M.S. Azamian[1], F. Xia[1,3], S.W. Cheung[1,3], P. Evans[4], A. Henderson[5], S.R. Lalani[1].* 1) Department of Molecular & Human Genetics, Baylor College Medicine, Houston, TX; 2) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 3) Baylor Genetics, Houston, TX; 4) Departments of Pediatrics and Neurology, University of Texas Southwestern Medical School, Dallas, TX; 5) The Newcastle upon Tyne Hospitals, Newcastle upon Tyne, England.

By searching a clinical database of over 60,000 individuals referred for array-based copy number variant (CNV) analyses and online resources, we identified four males from three families with intellectual disability, developmental delay, hypotonia, severe joint hypermobility and relative macrocephaly who carried small, overlapping deletions of Xp11.22. The maximum region of overlap between their deletions spanned ~430 kb and included two protein coding genes, *GSPT2* and *MAGED1*. Deletions of this ~430 kb region have not been previously implicated in human disease. GSPT2 is abundantly expressed in the brain and plays a critical role in preventing translational readthrough. Duplications of *GSPT2* have been documented in individuals with intellectual disability, but the phenotypic consequences of a loss of *GSPT2* function have not been elucidated in humans or mouse models. Changes in *MAGED1* have not been associated with intellectual disability in humans, but loss of MAGED1 function is associated with neurocognitive and neurobehavioral phenotypes in mice. Although the molecular mechanisms by which MAGED1 affects brain development and function have not been fully elucidated, MAGED1 binds and positively regulates the transcriptional activity of a subset of transcription factors including SIM1, NPAS4 and ARNT2 whose deficiencies have been shown to affect brain development and function. All of the affected males in these families inherited their Xp11.22 deletions from their unaffected mothers. Studies performed on DNA from one of these mothers did not show evidence of skewed X-inactivation. These results suggest that Xp11.22 deletions that encompass *GSPT2* and *MAGED1* cause an X-linked form of syndromic intellectual disability and that loss of GSPT2 and/or MAGED1 function may contribute to the intellectual disability and developmental delay seen in affected males.

1068F

Family-based whole exome sequencing for identifying novel variants in primary myopathies. *J. Hwang[1], H. Kim[1], S. Park[1], J. Chae[2], Y. Ahn[1].* 1) Division of Rare Diseases, Center for Biomedical Sciences, National Research Institute of Health, Chungcheongbuk-do, Republic of Korea; 2) Department of Pediatrics, Pediatric Clinical Neuroscience Center, Seoul National University Children's Hospital, Seoul National University College of Medicine, Seoul, Korea.

Primary myopathy is a clinical and genetic heterogeneous muscle disorder, sub-classified into dystrophies, myotonia, congenital myopathies, familial periodic paralysis, mitochondrial myopathies, and metabolic myopathies. Despite clinical and histological investigations, the genetic information about causes of primary myopathy is not fully established. Whole exome sequencing approaches have been used to uncover novel variants as a discovery and diagnostic tool on rare diseases. Here, we conducted family-based whole exome sequencing (130 trio). Integrative functional enrichment analyses allowed us to understand and interpret novel variants better, including analyzing the functional annotations of structural variation and genetic regulation. To study functional implications, follow-up experiments were conducted using the zebrafish model. Our results could provide additional insight into phenotypic convergence and divergence in the diagnosis of muscle disorders. .

1069W

Engineering tissue specific delivery of enzymes for lysosomal disease treatment. *K. Cygnar, A. Baik, C. Schoenherr, A. Murphy, A. Economides.* Regeneron, Tarrytown, NY.

Many lysosomal diseases (LDs) are caused by deficient enzymatic processing of substrates in the lysosome and their resultant accumulation. Disease manifestation varies depending on the affected tissues and the residual enzyme activity, with consequences from mild impairment to neonatal lethality. The advent of enzyme replacement therapy (ERT), in which recombinant enzyme is delivered to clear the accumulated substrate from affected tissues, revolutionized the treatment of LDs. In ERT, tissue targeting and lysosomal trafficking of recombinant enzymes occurs primarily through the cation-independent mannose-6-phosphate receptor (CI-MPR) pathway. However, many clinically relevant tissues express this receptor (*e.g.* skeletal muscle in Pompe disease) weakly, hence rendering the delivery of enzyme to the affected tissues inefficient. Delivering therapeutic amounts of recombinant enzyme to the lysosomes of affected tissues remains a challenge for many LDs due to this unfavorable biodistribution. Here, we present an antibody-guided enzyme delivery technology to create tunable patterns of tissue biodistribution by using antibody targeting to a chosen internalizing cell-surface protein, thereby bypassing the limitations of CI-MPR-mediated delivery of native enzymes. The antibody acts as a guide for the enzyme to reach the lysosomes of the tissues of interest in ERT. As a proof of concept, an antibody against an internalizing surface protein enriched in muscle was fused to alpha glucosidase (GAA). Enzymatic activity, intracellular half-life, lysosomal processing of the enzyme, and antibody binding were unaffected in this format. We demonstrated antibody-dependent, CI-MPR-independent uptake of the antibody-GAA drug in several cell types, including skeletal myoblasts and HEK 293. In a mouse model of Pompe Disease, we also found that the antibody-GAA reduced accumulated lysosomal glycogen in skeletal and cardiac muscle. Importantly, glycogen was significantly reduced in skeletal muscles compared to GAA alone. This technology is generalizable to other LSDs and also amenable to depot-based cross-correctional gene therapy models.

1070T

Mutational spectrum of Duchene Muscular Dystrophy in Colombian patients. *P. Ayala-Ramírez[1], M. Garcia-Acero[1,3], T. Pineda[2], M. Guerra-Torres[1], R. Garcia-Robles[1], T. Buitrago[1], A. Poveda[1], F. Suarez-Obando[1,3].* 1) Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Biotecnologia y Genetica - Biotecgen S.A; 3) Servicio de Genetica, Hospital Universitario San Ignacio.

Introduction: Duchenne muscular dystrophy (DMD) is the most common muscular dystrophy in children, characterized by rapidly progressive muscle weakness and wasting of skeletal, smooth and cardiac muscle. Is a X-linked recessive muscular dystrophy caused by absence of dystrophin, usually due to mutations in DMD gene, 70% of DMD cases are caused by deletions/duplications of one or more DMD exons and 30% of cases have alterations in sequence. The phenotype depends on the location of the mutation in the gene and does not necessarily depend on the size of the deletion or the duplication. The molecular characterization of the disease is necessary for the genetic counseling, the prognosis and for the possible pharmacological management of certain mutations Methodology: descriptive analysis of mutational spectrum of 62 Colombian DMD individuals. For the identification of the exon deletion or duplication of dystrophin was first done multiple ligation-dependent probe amplification (MLPA) and then mutations were identified with direct sequencing. Results: A total of 62 patients with suspected DMD were confirmed by molecular analysis. The average age of the individuals was 9 years with age at onset of symptoms less than 4 years in 79% of patients. Most patients with DMD presented deletions (54.8%) or duplications (24.1%) in the dystrophin gene. The remaining 24.1% corresponds to the set of point mutations, highlighting nucleotide substitutions nonsense and frameshift type as the most common. The variants were distributed throughout the gene and 5 new mutations in *DMD* were identified Conclusions. Mutational analysis of the patients studied confirms the great genetic heterogeneity of DMD, mutations in the *DMD* gene may lead to a broad spectrum of clinical phenotypes that are closely correlated with molecular alterations and with the severity of the altered protein function caused by the mutation, however the phenotype - genotype correlation is uncertain, although the location of the isoforms and in some cases their function is known, the mutation of certain regions does not behave similarly in all patients.

1071F

A combination of capture-based high-coverage NGS and WES analysis uncovers potential deleterious variants in the *NARS2* gene expanding the phenotypical spectrum of combined oxidative phosphorylation deficiencies. *Y. Wang[1,2], F. Vetrini[1], A.V. Dharmadhikari[1], J.A. Rosenfeld[2], V.M. Knight[3], R. Gavrilova[4], R. Lutz[5], S. Agadi[6], Y. Yang[1,2], F. Li[2].* 1) Baylor Genetics, Houston, TX,; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Pediatric Neurology and Epilepsy, Yale University, New Haven, CT; 4) Department of Neurology, Mayo Clinic, Rochester, MN,; 5) Department of Pediatrics, University of Nebraska Medical Center, Omaha, NE; 6) Department of Pediatrics-Neurology, Baylor College of Medicine, Houston, TX.

Mitochondrial respiratory chain diseases represent a large and heterogeneous group of energy deficiency disorders that are caused by both nuclear and mitochondrial-encoded genetic variants. Nuclear genes responsible for combined oxidative phosphorylation (OXPHOS) include proteins involved in mtDNA replication, transcription, RNA maturation and translation. Mitochondrial asparaginyl-tRNA synthetase *NARS2* belongs to a large group of aminoacyl-tRNA synthetases genes necessary for the mitochondrial translation process. Defects in *NARS2* gene are associated with combined oxidative phosphorylation deficiency 24, MIM: 616239, with wide phenotypic variability ranging from a milder form affecting only skeletal muscle to a more severe infantile-onset neurodegenerative disorder. Very few cases have been reported thus far in the literature. Capture NGS reliably detects SNVs and small indels whereas CNVs require sophisticated analysis via an in house-designed algorithm. In combination with whole exome sequencing (WES), we have identified a deletion encompassing exons 8 and 9, eight SNVs (p.R51C, p.S83G, p.A123A, p.R243Q, p.N186S, p.R348W, p.P214L and c.689+1G>A) in a compound heterozygous state in five patients presenting with neurodevelopmental abnormalities of unknown etiology. These five patients presented with variable clinical phenotypes including early onset epileptic encephalopathy (n=3), structural brain abnormalities (n=4), intellectual disability (n=3), myopathy (n=2), hearing loss (n=1), hypoglycemia (n=1), lactic acidosis (n=1), and hepatomegaly (n=1). Among these changes, p.R51C, p.S83G, p.A123A, p.R243Q, p.N186S, p.R348W and c.689+1G>A have never been reported to be associated with disease. The synonymous change p.A123A was predicted to reduce the affinity for the SF2/ASF splicing enhancer factor; others were predicted to be damaging/deleterious to the structure of the NARS protein via *in silico* predictions. All the missense changes potentially affect conserved domains of the protein: two in the anti-codon binding domain (p.R51C, p.S83G) and four potentially affect the catalytic domain (p.R243Q, p.N186S, p.P214L and p.R348W). In conclusion, we report nine *NARS2* potentially deleterious changes in five new patients thus expanding the phenotype and the mutation spectrum of combined oxidative phosphorylation deficiency 24. .

1072W

Adenosylcobalamin synthesis in cultured fibroblasts from patients with isolated methylmalonic aciduria. *D. Watkins, D. Rosenblatt.* Child Health & Human Development, Research Institute, McGill University Health Centre, Montreal, Quebec, Canada.

Methylmalonic aciduria (MMA) occurs in patients with decreased activity of the mitochondrial enzyme methylmalonylCoA mutase, which catalyzes conversion of methylmalonylCoA to succinylCoA, either due to mutations affecting the mutase enzyme itself (the *mut* complementation class) or by mutations affecting synthesis of its adenosylcobalamin (AdoCbl) cofactor (*cblA*, *cblB* and *cblD* variant 2 classes). The mut class is subdivided into *mut°* and *mut⁻* groups, depending on whether or not there is residual enzyme activity. We have investigated the synthesis of cobalamin cofactors from exogenous [⁵⁷Co] cyanocobalamin (CNCbl) in cultured fibroblasts from patients with isolated methylmalonic aciduria: 72 *cblA*, 45 *cblB*, 3 *cblD* variant 2, 194 *mut°* and 53 *mut⁻* patients. Fibroblasts were incubated for 96 hours in medium supplemented with labeled CNCbl bound to human transcobalamin, and cellular cobalamins were extracted in hot ethanol and separated by high performance liquid chromatography. No significant differences in incorporation of labeled CNCbl were detected between the different classes of isolated MMA patients. Synthesis of AdoCbl from exogenous CNCbl was decreased compared to control fibroblasts in all classes. AdoCbl as a percentage of total labeled cellular cobalamin was less in *cblA* (3.3% ± 1.6, mean ± SD), *cblB* (3.0% ± 2.1), and *cblD* variant 2 (3.5% ± 1.8) cells than in either *mut°* (8.1% ± 4.4) or *mut⁻* (8.3% ± 6.9) cells. AdoCbl was 15.3% ± 4.2 of labeled cellular cobalamin in control fibroblasts. There was no significant difference in AdoCbl levels between *mut°* and *mut⁻* cells. Thus, the level of AdoCbl synthesis is cultured fibroblasts can distinguish between disorders involving the synthesis of the cofactor from primary mutase disorders, but cannot distinguish between the two subtypes of mutase deficiency.

1073T

A zebrafish *mut°* model recapitulates key aspects of severe methylmalonic acidemia. *K.T. Ellis[1], M. Arnold[1], R. Sood[2], B. Carrington[2], K. Bishop[2], V. Hoffmann[3], P. Zerfas[4], N. Achilly[1], J. Sloan[1], O.A. Shchelochkov[1], C. Venditti[1].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD; 2) Zebrafish Core, Translational and Functional Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Diagnostic and Research Services Branch, Office of the Director, NIH, Bethesda, MD; 4) Office of the Director, National Institutes of Health, Department of Health and Human Services, Bethesda, MD.

Methylmalonic acidemia (MMA) is an autosomal recessive inborn error of metabolism caused by mutations in methylmalonyl-CoA mutase (MUT), the enzyme responsible for metabolizing methylmalonyl-CoA into succinyl-CoA. Patients with *mut°* type MMA, characterized by a severe deficiency of the MUT enzyme, present neonatally with lethargy, acidosis, and encephalopathy resulting in high mortality. Despite dietary therapy, patients can experience failure to thrive, life-threatening episodes of acidosis, and poorly understood multisystemic complications including renal failure and basal ganglia injury. To create a tractable disease model and investigate disease mechanisms, we generated a zebrafish model of *mut°* MMA using zinc finger nucleases to target exon 2 of the zebrafish *mut* gene. This resulted in a 10 bp deletion (c.326_335del, p.Tyr109*). *mut°* fish presented with growth retardation at 7 days post fertilization (dpf) with growth ceasing at 4.5 mm standard length, the size of unaffected clutchmates at this time. The onset of growth retardation coincided with the depletion of maternally transferred *mut* mRNA present in the egg as identified by RT-PCR. By 19 dpf, *mut°* fish continued to display a significant reduction in size, at 59% of the wildtype standard length (p<0.0001). *mut°* fish showed 100% mortality by 26 dpf (p=0.0016). RT-PCR revealed a decrease in *mut* mRNA levels at 15 dpf, suggesting that the mutant mRNA is degraded by nonsense-mediated decay. Tissue metabolite analysis at 15 dpf displayed a 28-fold increase in methylmalonate and 2-methylcitrate levels compared to age-matched clutchmates (p<.0001), a metabolic hallmark seen in the human disease. Further, at this age, *mut°* hepatocytes exhibited eosinophilic inclusions ranging from 1-3 microns in size that resemble megamitochondria seen in liver and kidney tissues from MMA patients and mouse models. To assess mitochondria in a zebrafish organ sharing function with mammalian kidneys, we performed transmission electron microscopy on the gills of *mut°* fish, an organ playing a key role in acid-base equilibrium. Here, clearing of cristae from the mitochondrial matrix in the gill epithelium of *mut°* fish is observed. These data demonstrate that this novel zebrafish model recapitulates key features of *mut°* type MMA. To characterize the contribution of individual organs to the pathology of MMA, we are generating transgenic lines expressing human *MUT* to restore enzyme function in the liver and muscle.

1074F

Novel biallelic mutations in the *PNPT1* gene encoding a mitochondrial-RNA-import proteinPNPase cause delayed myelination. *A. Kikuchi[1], R. Sato[1,2], N. Arai-Ichinoi[1], T. Matsuhashi[1], Y. Numata-Uematsu[1], M. Uematsu[1], Y. Fujii[3], K. Murayama[4], A. Ohtake[5], T. Abe[6], S. Kure[1].* 1) Department of Pediatrics, Tohoku University Hospital, Sendai, Japan; 2) Department of Pediatric Neurology, Miyagi Children's Hospital, Sendai, Japan; 3) Department of Pediatrics, Hiroshima University Hospital, Hiroshima, Japan; 4) Department of Metabolism, Chiba Children's Hospital, Chiba, Japan; 5) Department of Pediatrics, Saitama Medical University, Saitama, Japan; 6) Department of Nephrology, Endocrinology, and Vascular Medicine, Tohoku University School of Medicine, Sendai, Japan.

Recent studies suggest that impaired transcription or mitochondrial translation of small RNAs can cause abnormal myelination. A polynucleotide phosphorylase (PNPase) encoded by *PNPT1* facilitates the import of small RNAs into mitochondria. *PNPT1* mutations have been reported in patients with neurodevelopmental diseases with mitochondrial dysfunction. We report here two siblings with *PNPT1* mutations who presented delayed myelination as well as mitochondrial dysfunction. We identified compound heterozygous mutations (c.227G>A; p.Gly76Asp and c.574C>T; p.Arg192*) in *PNPT1* by quartet whole-exome sequencing. Analyses of skin fibroblasts from the patient showed that PNPase expression was markedly decreased and that import of the small RNA *RNase P* into mitochondria was impaired. Exogenous expression of wild-type *PNPT1*, but not mutants, rescued ATP production in patient skin fibroblasts, suggesting the pathogenicity of the identified mutations. Our cases expand the phenotypic spectrum of *PNPT1* mutations that can cause delayed myelination.

1075W

***RMND1*-related mitochondrial disease: Phenotypic delineation of four patients including renal manifestations.** *N.T. Le[1], N. Bekheirnia[1], J.A. Rosenfeld[1], A.C. Goldstein[2], D.W. Bartholomew[3], M.T. Pastore[3], C.C. Muraresku[4], E.M. McCormick[4], F. Xia[1], C. Eng[1], Y. Yang[1], F. Scaglia[1], M.R. Bekheirnia[1].* 1) Baylor College of Medicine , Houston, TX; 2) Children's Hospital of Pittsburgh of UMPC, Pittsburgh, PA; 3) Nationwide Children's Hospital, Columbus, OH; 4) Children's Hospital of Philadelphia, Philadelphia, PA.

Background: *RMND1* (Required for Meiotic Nuclear Division 1 homolog) is a nuclear gene which codes for a protein believed to be involved in mitochondrial translation. Pathogenic variants in the *RMND1* gene have been shown to cause recessively inherited infantile-onset mitochondrial disease with neurological, developmental, intellectual, cardiac, gastrointestinal, metabolic, and renal manifestations. **Methods:** We report the clinical, phenotypic, and molecular genetic findings of four patients diagnosed with *RMND1*-related mitochondrial disease. Whole exome sequencing (WES) was used to identify the single nucleotide variants (SNVs) of each respective patient. All SNVs were validated by Sanger sequencing. Patient information was collected as part of the clinical laboratory intake form, and families were consented accordingly. **Results:** We present four patients with recessive mutations in *RMND1* involving the following SNVs: c.485delC, p.P162fs; c.713A>G, p.N238S; and c.533C>T, p.T178M. Developmental delay, sensorineural hearing loss, and renal deficiencies—which include chronic renal failure due to interstitial nephritis, renal tubular acidosis, and end-stage renal disease—were observed in all four patients. Motor delay and structural neurologic findings—such as cortical volume loss, basal ganglia calcification, and white matter abnormalities—were present in three patients. Short stature, intellectual disability, speech delay, absence of speech, seizures, hyperreflexia, and lactic acidosis were observed in half of the patients. **Discussion:** The clinical findings of the four patients demonstrate that mutations in *RMND1* result in a heterogenous array of phenotypic manifestations. In addition to the recent work described by other research groups, the findings of our study depict the clinical spectrum of *RMND1* mutations by delineating the diverse phenotypic manifestations of *RMND1* variants. **Acknowledgment:** This study was supported in part by K12 DK083014 Multidisciplinary K12 Urologic Research Career Development Program.

1076T

Identification of the genetic causes of mitochondrial oxidative phosphorylation (OXPHOS) disease. *S.C. Lim[1], Y. Kishita[1,2], M. Kohda[2,3], Y. Mizuno[1], T. Hirata[3], Y. Yatsuka[1], N.N. Borna[1], H. Harashima[4], K. Murayama[5], A. Ohtake[4], Y. Okazaki[1,2,3].* 1) Division of Functional Genomics & Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University, Saitama, Japan; 2) Intractable Disease Research Center, Graduate School of Medicine, Juntendo University, Tokyo, Japan; 3) Division of Translational Research, Research Center for Genomic Medicine, Saitama Medical University, Saitama, Japan; 4) Department of Pediatrics, Saitama Medical University, Saitama, Japan; 5) Department of Metabolism, Chiba Children's Hospital, Chiba, Japan.

Oxidative phosphorylation (OXPHOS) is a biochemical pathway within mitochondria that generates energy for cellular activities. The proper function of OXPHOS requires many nuclear DNA and mitochondrial DNA (mtDNA)-encoded products, such as the structural components and biogenesis factors of the OXPHOS complexes, and the proteins required for mtDNA replication and expression. Disruption of OXPHOS can cause human metabolic disease, affecting ~1 in 5,000 live births (Skladal et al., 2003). OXPHOS disease can have any age of onset, a wide spectrum of clinical presentations, and generally have poor phenotype-genotype correlations. To-date, more than 250 OXPHOS disease genes have been identified. Nevertheless, these genes account for ~60% of OXPHOS patients, implying that additional causal genes remain undiscovered. DNA samples of 58 early-onset OXPHOS disease patients recruited from Saitama Medical University Hospital and Chiba Children's Hospital in Japan were analyzed with whole exome sequencing (WES) using the Illumina HiSeq platform. Variants that are common in 2,049 healthy Japanese individuals (2KJPN database) were excluded from subsequent bioinformatic and functional analyses. We identified candidate mutations in 6 known OXPHOS disease genes in 6 patients. Additionally, 3 candidate mutations were identified in 2 novel OXPHOS disease genes in 2 patients – one encodes a translocase in the mitochondrial inner membrane and the other encodes a WD repeat domain-containing protein with unknown function in the mitochondria. To confirm the pathogenicity of candidate mutations, functional studies are being performed and include SDS-PAGE and BN-PAGE western blotting, lentiviral-mediated expression of the wildtype gene in patient cultured fibroblasts and analysis of gene knock-out/knock-in cell models generated using the CRISPR/Cas9 system. This research will identify novel OXPHOS disease genes and contribute to the expansion of our knowledge for the genetic basis of OXPHOS disease and the biology of OXPHOS system.

1077F

The clinical spectrum of BCS1L mutations: Case report of a novel mutation and review of the literature. *J. Priestley[1], E. Fernandez-Vizarra[2], P. Mark[3].* 1) College of Human Medicine, Michigan State University, Grand Rapids, MI; 2) MRC-Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK; 3) Division of Medical Genetics, Spectrum Health Medical Group, Grand Rapids, MI.

BCS1L is a ubiquitously-expressed mitochondrial member of the AAA+ (ATPases associated with various cellular activities). It functions as a translocase in the integration of Rieske Fe/S protein during respiratory chain complex III assembly. Since 2001, 40 different *BSC1L* mutations have been identified in clinically affected patients, making them the most frequent cause of isolated mitochondrial complex III deficit. Phenotypes are varied and as mild as Björnstad syndrome (characterized by pili torti and sensorineural hearing loss) or as severe as GRACILE (growth restriction, aminoaciduria, cholestasis, iron overload, lactic acidosis, and early death) syndrome. *BCS1L* mutations are also linked to a third entity: an undefined complex III deficiency, which is a heterogenous-condition generally involving low birth weight, renal and hepatic pathologies, hypotonia, and developmental delay. However, some patients with phenotypes more severe than Björnstad and less severe than GRACILE have no demonstrable reduction in complex III activity or quantity, adding confusion to this intermediate phenotype. Here, we present a female patient initially seen in our clinic at six months of age with failure to thrive, growth restriction, developmental delay, microcephaly, hypotonia, hearing and vision loss, transient lactic acidosis, transient elevation in liver enzymes, and brittle hair – a phenotype both similar to and disparate from those previously described. Subsequent exome sequencing revealed compound heterozygosity in the BCS1L gene: p.Arg184Cys, previously identified in patients with *BCS1L*-associated complex III deficiency and in patients with Björnstad syndrome; and p.Leu280Phe, a previously unreported mutation at a highly conserved amino acid residue. Together with a review of the current literature, these data suggest a clinical spectrum of disease within *BCS1L* mutations that does not fit cleanly into the three aforementioned disease entities. We propose a nomenclature scheme in which Björnstad syndrome is defined as isolated pili torti with sensorineural hearing loss, GRACILE syndrome is defined by its eponymous components and by the homozygous p.Ser78Gly mutation, and all other *BCS1L* mutations and clinical phenotypes are termed "*BCS1L* mitopathies." In doing so, we provide a potential work-up for patients with suspected or known *BCS1L* mutations, so that no key aspects of the *BCS1L* mitopathy are missed.

1078W

Severe lactic acidosis, myopathy, and normal mental status in an infant with biallelic *GTPBP3* pathogenic variants. *H. Vernon[1], C. Peroutka[1], J. Albert[2], K. Doheny[3], K. Hetrick[3], L. Kratz[2], D. Murdock[1], A. Murphy[4], M. Parsons[4], A. Robinson[3], N. Sobreira[1], T. Wang[1], D. Witmer[3].* 1) Inst Gen Med, Johns Hopkins Hosp, Baltimore, MD; 2) Kennedy Krieger Inst, Johns Hopkins Med Inst, Baltimore, MD; 3) Genetics Ctr for Inhrtd Dis Res, Johns Hopkins Univ, Baltimore, MD; 4) Johns Hopkins Univ Dept of Cardiology, Baltimore, MD.

Background: More than 90 nuclear and 37 mitochondrial (mt) genes are involved in post-transcriptional modification of mitochondrial tRNAs (mt-tR-NAs) or in OXPHOS polypeptide synthesis. Pathogenic variants in genes involved in translation of mt-transcripts, including *GTPBP3, MTO1* and *TRMU,* are a recently described cause of mt disease. **Case Study:** Our patient presented with perinatal myopathy and severe lactic acidosis (15 mmol/L) with lactate/pyruvate ratio of 36 (N 10-20) but normal mental status. Also present were persistently increased plasma alanine and 3-hydroxybutyrylcarnitine levels. Free carnitine was normal. Multiple urine organic acid analyses showed markedly increased lactate excretion with moderate increases of 2-hydroxyisovalerate, 2-hydroxybutyrate, 3-hydroxyisobutyrate and 3-hydroxybutyrate, but only trace amounts of acetoacetate, suggesting a mt redox abnormality. There were mild increases of 2-methyl-2,3-dihydroxybutyrate and other hydroxy- and keto-hexanoic acids, absence of significant dicarboxylic aciduria, and trace 3-hydroxyglutarate. By 3 weeks, the urine 2-ketoglutarate and fumarate increased to 3-4 times normal for age. Liver transaminases were unremarkable aside from the immediate perinatal period. There was no hypoglycemia. Cardiomyopathy was first detected at 5 weeks. Rapid whole exome sequencing (WES) from the NICU at 5 weeks revealed 2 heterozygous presumed pathogenic missense variants in *GTPBP3*: c.215C>T(p.Pro721Leu) and c.1175C>G(p.Pro392Arg). c.215C>T(p.Pro721Leu) has a frequency of ~0.8% in African populations (ExAC), though the impacted amino acid, Pro, is highly conserved across species. c.1175C>G(p.Pro392Arg) has not been previously described, and the impacted amino acid (also Pro) is highly conserved across species. At 2.5 years, the patient has non-compaction cardiomyopathy, is ventilator dependent, and has motor developmental delays, but normal cognitive function. **Discussion:** In the setting of significant perinatal instability, rapid WES resulted in a diagnosis for a rare Mendelian disorder in our patient. A cohort of 11 patients with variants in *GTPBP3* was reported in the month our patient was born. Our patient has a novel exon 2 variant and previously unreported biochemical findings. Given variable expressivity in other patients, this case adds to our understanding of *GTPBP3*-associated mt disease and of mt translation defects as an important cause of mt disease to be considered in infants with lactic acidemia.

1079T

Severe leukodystrophy with complete clinical recovery caused by recessive *BOLA3* mutations. *C.A. Stutterd[1,2,3,4], N.J. Lake[4,5], H. Peters[4,6,7], G. Gillies[1], P. Lockhart[1,4], R.J. Taft[8,9,10], M. van der Knaap[11], A. Vanderver[12,13,14], D.R. Thorburn[3,4,5], C. Simons[8], R.J. Leventer[2,4,15].* 1) Bruce Lefroy Centre for Genetic Health Research, Murdoch Children's Research Institute, Victoria, Australia; 2) Department of Neurology, Royal Children's Hospital, Victoria, Australia; 3) Victorian Clinical Genetics Service, Murdoch Children's Research Institute, Victoria, Australia; 4) Department of Paediatrics, University of Melbourne, Victoria, Australia; 5) Mitochondrial Research Group, Murdoch Children's Research Institute, Victoria, Australia; 6) Department of Metabolic Medicine, Royal Children's Hospital, Victoria, Australia; 7) Metabolic Research Group, Murdoch Children's Research Institute, Melbourne, Victoria, Australia; 8) Institute for Molecular Bioscience, University of Queensland, St. Lucia, Queensland, Australia; 9) Illumina Inc, San Diego, California; 10) School of Medicine and Health Sciences, The George Washington University, Washington, DC; 11) Department of Child Neurology, VU University Medical Center, Amsterdam, The Netherlands; 12) Department of Neurology, Children's National Medical Center, Washington, DC; 13) Department of Integrated Systems Biology, George Washington University, Washington, DC; 14) Department of Pediatrics, George Washington University, Washington, DC; 15) Neuroscience Research Group, Murdoch Children's Research Institute, Melbourne, Victoria, Australia.

Aim: To identify the genetic aetiology of a distinct leukodystrophy causing acute neurological regression in infancy with complete clinical recovery. **Methods**: We performed trio genome sequencing to identify the causal variants. Analysis of mitochondrial function in cultured patient fibroblasts was undertaken to confirm the pathogenicity of candidate variants. **Results:** Our patient presented at 18 months with acute right hemiplegia and cognitive regression without obvious trigger. Over five years he had complete recovery of neurological function and continues to have age-appropriate motor and cognitive function at age eight. Cerebral MRI at disease onset revealed extensive, bilateral, heterogeneous T2 hyperintensity involving the periventricular and deep white matter with sparing of the subcortical white matter, basal ganglia, brainstem and cerebellum. There was restricted diffusion at the periphery of the white matter change and no contrast enhancement. MRS of the frontal white matter demonstrated a lactate doublet with decreased NAA. The most recent MRI at age four showed evolution in the diffuse leukodystrophy with decrease in the degree of restricted diffusion. The pattern of white matter abnormality was most consistent with a mitochondrial disorder. Serum and CSF lactate levels were normal, as were mitochondrial respiratory chain enzyme function in liver and muscle. Urine and plasma glycine were elevated. Trio genome sequencing identified compound heterozygous variants in *BOLA3*: one previously reported variant for multiple mitochondrial dysfunctions syndrome 2 (MMDS2), and a novel variant that disrupts a critical residue for Fe-sulfur cluster biogenesis. Analysis of cultured patient fibroblasts demonstrated deficient pyruvate dehydrogenase activity and reduced quantity of protein subunits of mitochondrial complexes I, II, and III, consistent with BOLA3 dysfunction. Previously reported cases of MMDS2 caused by *BOLA3* mutation have leukodystrophy with severe, progressive, life-limiting neurological and multisystem disease. **Conclusions:** We report a novel phenotype for multiple mitochondrial dysfunctions syndrome (MMDS2), associated with complete clinical recovery. We have identified a novel disease-causing variant in *BOLA3* validated by functional cellular studies. Our patient's unique clinical course provides important information on the phenotypic spectrum of MMDS2.

1080F

A viable knockout murine model of *Mmaa* (*cblA*) deficiency provides a platform for microbiome manipulations. *A.F. Lesser[1], O.A. Shchelochkov[1], S. Smith[1], M. Epping[1], P. Zerfas[2], I. Manoli[1], J. Segre[3], C.P. Venditti[1].* 1) Medical Genomics and Metabolic Genetics Branch, NHGRI, National Institutes of Health, Bethesda, MD; 2) Diagnostic & Research Services Branch, OD, National Institutes of Health, Bethesda, MD; 3) Microbial Genomics Section, NHGRI, National Institutes of Health, Bethesda, MD.

Isolated methylmalonic acidemia (MMA) is a group of inherited metabolic disorders characterized by the inability to generate succinyl-CoA from (*R*)-methylmalonyl-CoA as a result of mutations in the methylmalonyl-CoA mutase (MUT) gene or deficiencies in enzymes involved in the synthesis (MMAB or *cblB*) or transfer (MMAA or *cblA*) of its cofactor, 5'-deoxyadenosylcobalamin (AdoCbl). Studies on the bacterial homolog protein MeaB show that it functions as a G-protein chaperone, facilitating the loading of AdoCbl to MUT, as well as the catalytic turnover and rescue of the MUT enzyme. The function of the mammalian protein is not well understood. MMAA patients have an attenuated biochemical and clinical phenotype compared to the classic MUT deficiency, but remain at risk for metabolic decompensations and disease complications including basal ganglia strokes, chronic renal failure and optic nerve atrophy. The treatment for *cblA* MMA includes strict adherence to a protein restricted diet, hydroxocobalamin injections, and in some cases, periodic use of antibiotics to reduce propionate production from the gut microbiome. We previously generated an *Mmaa* deletion mutant allele using homologous recombination and noted neonatal lethality on the C57Bl6;129 background that could be modified by the introduction of FVBN genes. [C57Bl6xFVBN] *Mmaa*[-/-] mice recapitulate the *cblA* patient phenotypes, but uniformly survive until weaning, and experience severe growth failure and reduced survival when transitioned to regular chow. *Mmaa*[-/-] mice also display elevated plasma methylmalonic acid (p<0.0001), decreased 1-^{13}C-propionate oxidation, and the hepatorenal mitochondropathy seen in other forms of MUT deficiency. To explore vitamin responsiveness on the clinical and metabolic phenotype of the *Mmaa*[-/-] mice, the animals were treated with 2.5 mg/kg hydroxycobalamin twice a week for two weeks. An increase in weight (p<0.05) was seen in the treated *Mmaa*[-/-] males. A separate cohort of *Mmaa*[-/-] mice underwent antibiotic treatment for 7 days, which was associated with in an increase in weight (p<0.01) and a decrease in plasma methylmalonic acid (p<0.01). The *Mmaa*[-/-] mice expand the ability to model MMA without the obligate requirement for a rescue transgene. This *cblA* mouse model informs our understanding of the effects of genetic background, dietary interventions, cobalamin response, and microbiome manipulations and shapes our efforts to develop new therapies for propionate oxidation disorders.

1081W

Identification and validation of new hepatic biomarkers in methylmalonic acidemia (MMA). *I. Manoli[1], J.R. Sysol[1], M.W. Epping[1], C. Wang[1], L. Li[1], J.L. Sloan[1], J. Gagne[1], P.M. Zerfas[2], V. Hoffmann[2], J. Cheng[3], R.J. Chandler[1], C.P. Venditti[1].* 1) Organic Acid Research Section, MGMGB, NHGRI, NIH, Bethesda, MD; 2) Division of Veterinary Resources, NIH, Bethesda, MD; 3) Embryonic Stem Cell and Transgenic Mouse Core Facility, NHGRI, NIH, Bethesda, MD.

Methylmalonic acidemia (MMA) caused by methylmalonyl-CoA mutase (*MUT*) deficiency, is characterized by recurrent episodes of life-threatening metabolic ketoacidosis and multiorgan complications. Studies in knockout mice (*Mut*[-/-]) and transplanted MMA patients suggested that extrahepatic organs, mainly the skeletal muscle, are a major source of circulating methylmalonic acid. To study the effects of restoring Mut activity in skeletal muscle on the hepatorenal phenotype of MMA, we generated mice expressing the Mut gene under the control of the muscle creatine kinase promoter (*Mut*[-/-];Tg[INS-MCK-Mut]). Although *Mut* expression in the muscle rescued *Mut*[-/-] mice from neonatal lethality, the *Mut*[-/-];Tg[INS-MCK]*Mut* mice exhibited severe growth failure, hepatic mitochondriopathy, renal dysfunction, and were resistant to obesity when fed a carbohydrate and lipid-enriched diet. We used a fasting challenge to model the metabolic decompensations seen in patients and probed the hepatic adaptations using transcriptomics to identify candidate genes and pathways mediating the stress response. Among the most upregulated genes in the *Mut*[-/-];Tg[INS-MCK]*Mut* mice was fibroblast growth factor - *Fgf21*, a key circulating metabolic regulator. After validation in mice, plasma FGF21 concentrations were measured in a large cohort of patients with different subtypes of isolated MMA (54 *mut*, 8 *cblA* and 5 *cblB*). FGF21 levels correlated with disease subtype severity (*p*=0.0001 between *mut* and *cblA* patients or controls) and showed an inverse relation to height *Z*-scores (*r*=-0.455, *p*=0.03). In contrast to serum methylmalonic acid concentrations, which are influenced by renal dysfunction, plasma FGF21 levels were not not affected by renal function indices (creatinine, cystatin-C , eGFR), but did correlate with markers of secondary mitochondrial dysfunction, including plasma glycine, allanine and urinary isoprostanes (all *p*<0.05). In further support of the use of FGF21 as a biomarker of hepatic mitochondrial dysfunction in MMA, we studied 9 patients before and after liver transplantation, as well as *Mut*[-/-];Tg[INS-MCK]*Mut* mice treated with a liver-directed AAV8 *MUT* gene therapy vector. In both the patients and mice, a decrease in plasma FGF21 was accompanied by a restoration of 1-^{13}C-propionate oxidation, reduced circulating metabolites and, in the mice, weight gain. Our studies demostrate the utility of 1-^{13}C-propionate oxidation and plasma FGF21 as treatment biomarkers for MMA. .

1082T

Identification of large effect variants underlying non-syndromic MA in families segregating the disease. *G.A. Arenas-Perez, A. Reyes-de la Rosa, P. Baeza-Capetillo, C. Garcia-Delgado, V.F. Moran-Barroso, J. Aguirre-Hernandez.* Hospital Infantil de Mexico Federico Gomez, Mexico City, Mexico DF, Mexico.

Introduction: Microtia-atresia (MA) refers to the developmental anomalies of the ear (e.g., shape and size). These may be accompanied by hearing difficulties. It has a variable phenotypic expression and its genetic basis is mostly unknown. Most cases are isolated, with a higher prevalence in populations with Native American ancestry. This has led to suggest that it is a complex disease arising from genetic predisposing variants interacting with environmental factors. A small proportion of cases (~6%) show mendelian inheritance, mostly with incomplete penetrance. The identification of the genes involved in MA in these families may lead to the identification of the genes and variants implicated in the more common sporadic form. **Objective:** To find large-effect variants associated with MA showing mendelian inheritance. **Material and methods:** Records of patients with MA were examined (2007-2015), and cases showing familial segregation were identified. Subsequently, four additional familial cases were identified and in the largest of these families showing MA segregation, whole exome sequencing was performed. The study was approved by the institution's ethics committee, and informed consent was obtained from all participants. **Results:** 242 MA patients were assessed; 22 (9%) had isolated non syndromic MA; 16/22 were male (66%), and 20/22 (83%) were unilateral mostly on the right side. Of these 22 isolated non-syndromic cases, 6 (27%) belonged to families segregating MA, with either an autosomal dominant or an autosomal recessive pattern. In 2016 four additional familial cases were identified. The largest of these latter families segregating MA was chosen for whole exome sequencing. Segregation was consistent with an autosomal dominant model with incomplete penetrance and affected members showed variable expressivity, form prominent helix to grade II MA. After filtering by quality, variant consequence and variant frequency in public databases and in a local database, a single missense variant in *ASIC1* was left (c.475G>A_p. Asp159Asn). **Discussion**: The gender, type and side of MA in this cohort were similar to the data reported in the literature, except for the higher proportion of isolated MA cases. A single candidate variant was observed associated with the disease. This variant was not present in public databases.

1083F

DFNA5 and the infamous skipping of exon 8. *K.T. Booth[1,2,4], H. Azaiez[1,4], K. Kahrizi[3], H. Najmabadi[3], R.J. Smith[1,2].* 1) Otolaryngology-Head and Neck Surgery, University of Iowa, 200 Hawkins Drive, IA; 2) Department of Molecular and Cellular Biology, University of Iowa, Iowa City, Iowa; 3) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 4) Authors Contributed Equally.

Like most cellular processes, RNA-splicing; the formation of mature messenger RNA (mRNA) from transcribed RNA, is a highly conserved and tightly regulated process. Dysregulation of splicing is a common factor underlying many inherited diseases, including deafness. For one deafness-associated gene, *DFNA5*, perturbation of exon 8 splicing results in a constitutively active truncated protein. To date only intronic mutations have been reported to cause exon 8 skipping in patients with DFNA5-related deafness. Here we present three families with post-lingual progressive autosomal dominant non-syndromic hearing loss, segregating novel *exonic* missense mutations in exon 8 of *DFNA5*. Computational analyses predicted these mutations decrease splicing efficiency of exon 8 or oblate it. Mini-gene assays using wild-type and mutant exon 8 of *DFNA5* revealed all three mutations alter normal splicing in-vivo. This is the first report of exonic mutations in the DFNA5 gene to cause deafness. Although at the DNA level the reported mutations here and the six previously reported mutations are unique, at the mRNA level they are indistinguishable; resulting in the skipping of exon 8, yielding the same truncated DFNA5 protein. This study expands the mutational spectrum of DFNA5 related deafness to include missense mutations and highlights the importance of assessing the effects of coding variants influence on splicing.

1084W

Genetics mutation: A novel frameshift mutation in the *USH1G* gene in an Iranian patient with USHER syndrome. *F. Tabei[1], H. Dastsooz[2], S. Mohammadi[2], A. Taghipour[3], F. Kamgarpour[2], M. Alipour[2], S. Tabei[3], M. Fardaei[3].* 1) International Branch of Medical School, Shiraz University of Medical Sciences, Motahari Hospital, Marvdasht, Iran; 2) Comprehensive Medical Genetic Center, Shiraz University of Medical Sciences, Shiraz, Iran; 3) Medical Genetic Department, Medical School, Shiraz University of Medical Sciences, Shiraz, Iran.

Introduction: Usher syndrome, type II is characterized by congenital, bilateral sensorineural deafness, intact vestibular responses, and retinitis pigmentosa. Diagnosis is usually established with the use of electrophysiologic and subjective tests of hearing and retinal function. However, genetic testing covering its related genes can be very helpful if clinical features are inconclusive. The goal of this study was to identify mutation in a patient affected by Usher Syndrome type II. Materials and Methods: The Genetic test was performed using a custom designed Nimblegen chip capturing different genes, including *CDH23, USH1C, PCDH15, USH1G, USH2A, PDZD7, MYO7A, GPR98, DFNB31, CLRN1* followed by Next Generation Sequencing. Results: One novel homozygous frameshift mutation c.993_994insACGGACTGGGCCGC (p.Arg331ArgfsX54) in *USH1G* gene was found and this frameshift mutation can lead to early termination of the amino acid coding, affecting the protein's function. Sanger sequencing was also performed to confirm this mutation in proband and his parents. It is worth noting that this novel mutation is inherited from two unrelated parents, predicting the same origin for this mutation. So, genetic screening for this mutation in their extended family members is necessary to reduce further affected individuals. Conclusions: This mutation is the first large insertion mutation in *USH1G* and since it is occurred exactly adjacent to the same 14 nucleotides in this gene, it can propose a new mechanism for occurrence of this particular mutation in *USH1G*. This work was supported by a grant from the International Branch, Shiraz University of Medical Sciences.

1085T

Leveraging consanguinity in inherited retinal diseases uncovers missing genetic variation: Rare novel disease genes and a multitude of novel variants in known disease genes. *K. Van Schil[1], S. Naessens[1], S. Van de Sompele[1], N. Gruartmoner Roura[1], M. Carron[1], K. Dannhausen[2], S. De Jaegere[1], M. Vanpanteghem[1], F. Coppieters[1], M. Karlstetter[2], B. Tsou[3,4], T. Langmann[2], F. Meire[5], R. Maroofian[6], A.H. Crosby[6], I. Balikova[7], M. Van Lint[8], A. Webster[9,10], M. Michaelides[9,10], B.P. Leroy[1,7,11], E. De Baere[1], IRD study group.* 1) Center for Medical Genetics, Ghent University and Ghent University Hospital, Ghent, Belgium; 2) Laboratory for Experimental Immunology of the Eye, Department of Ophthalmology, University of Cologne, Cologne, Germany; 3) Cell Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; 4) Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY, USA; 5) Department of Pediatric Ophthalmology, Queen Fabiola Children's University Hospital, Brussels, Belgium; 6) Monogenic Molecular Genetics, University of Exeter Medical School, Exeter, United Kingdom; 7) Department of Ophthalmology, Ghent University and Ghent University Hospital, Ghent, Belgium; 8) Department of Ophthalmology, University Hospital of the Free University of Brussels, Brussels, Belgium; 9) UCL Institute of Ophthalmology, University College London, London, United Kingdom; 10) Moorfields Eye Hospital, London, United Kingdom; 11) Division of Ophthalmology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.

Genetic investigations in patient cohorts of consanguineous origin can help to unravel the underlying genetic defect in genetically heterogeneous diseases such as inherited retinal diseases (IRD). A combination of homozygosity mapping and whole exome (WES) or genome sequencing (WGS) already proved to be effective to identify novel IRD genes and causative variants underlying IRD. Here, it was our aim to leverage consanguinity in a cohort of 100 unrelated families of different ethnicities, segregating non-syndromic and syndromic IRD phenotypes. Using an integrative approach combining homozygosity mapping, targeted next-generation sequencing and WES, we identified 76 unique mutations in 44 IRD genes underlying IRD in 74% of the patients. Moreover, we uncovered homozygous mutations in six novel candidate IRD genes in six unrelated families: four of them, *CEP162, EML4, ERICH6*, and *FBN2*, as the primary genetic defect, and two of them, *EML2* and *PATJ*, in combination with other mutations in known IRD genes, in total contributing to 6% of the IRD cohort. Finally, we also demonstrated the importance of copy number variations (CNVs) in IRD, identifying CNVs in 7% of patients, including the first CNV ever described in the *PDE6G* gene. Overall, our study shows the utility of genetic studies in consanguineous pedigrees, further elucidating the hidden genetic variation of genetically heterogeneous diseases such as IRD. It reveals a high mutational load (78%) in 100 unrelated families, consisting of a myriad of novel mutations in previously well-described IRD genes (74%), including CNVs (7%), and a small number of novel IRD genes (4%). Using an exome-oriented approach focusing on different types of hidden genetic variation such as novel mutations in previously described disease genes, novel disease genes, and CNVs, a causal genetic defect remained unidentified in 22% of our IRD cohort, assuming that non-coding mutations will significantly contribute to the overall genetic architecture of IRD and demonstrating that they require genome-oriented approaches.

1086F

In silico analysis and identification of TYR mutations in a Cypriot family. *R. Kalkan[1], P. Tulay[1], M.C. Ergoren[2], B. Turkgenc[3], C. Ogur[4], S.G. Temel[5].* 1) Medical Genetics, Faculty of medicine, Near East University, Nicosia, Cyprus; 2) Faculty of Medicine, Department of Medical Biology, University of Near East, Nicosia, Cyprus; 3) Genetic Diagnostic Center, University of Acibadem, Istanbul, Turkey; 4) Genetic Diagnostic Center, Memorial Health Group, Istanbul, Turkey; 5) Faculty of Medicine, Department of Histology and Embryology, University of Uludag, Bursa, Turkey.

　　Oculocutaneous albinism (OCA) is a rare heterogeneous group of genetic disorders characterized mainly by decreased or absent pigmentation in the skin, hair and eyes. A number of distinct ocular changes are important for the diagnosis of OCA including the reduced melanin amounts in the developing eye leading to abnormalities in the optic system. Mutations in different genes cause different types of OCA in which the prevalence of the different subtypes varies considerably among different ethnic groups. Four-year-old boy was referred our clinic with albinism related characteristics. He had a very light yellow hair and white skin. He was observed to have a pendular nystagmus and mild iris transillumination. Monocular flash responses were observed to be comparable for each eye at the midline, however the responses were asymmetrical. This kind of crossed asymmetry was in consistent with albinism. Sequencing was performed to investigate the possible TYR mutations present in the patient as well as in his relatives . Two variations of homozgous c.575C>A (p.Ser192Tyr) in exon 1 and homozygous c.1205G>A (p.Arg402Gln) in exon 4. were detected for the proband. Both of the parents were shown to be carriers of c.575C>A and c.1205G>A variations. The possible effects of both these variations on the structure and function of tyrosinase were investigated using HGMD, SIFT, MutationTester and PolyPhen. According to the HGMD, c.575C>A (p.Ser192Tyr) and c.1205G>A (p.Arg402Gln) are related with skin pigmentation and reported as a polymorphism. While SIFT and MutationTester has no prediction for these variations, PolyPhen reported them as probable damaging mutations. ClinVar database reported these as benign variants and dbVar reported an association between these variations and presence of freckles. In conclusion, we identified two previously-reported homozygous variations, c.575C>A and c.1205G>A, and family members of our proband are compound heterozygous of these variations. Our proband shows some of the clinical problems of OCA1 which includes hypopigmentation of the skin and hair, nystagmus, iris trans illumination, reduced stereoscopic vision, blonde fundi and crossed asymmetry of monocular flash responses that was in concordant with the variations we detected.

1087W

Genetic therapeutic strategies for Bardet-Biedl syndrome type 1. *M. Cring, A. Drack, V. Sheffield.* University of Iowa, 4171 MERF, Iowa City, IA.

　　Bardet-Biedl syndrome (BBS) is a rare, autosomal recessive ciliopathy that results in retinal degeneration, male infertility, polydactyly, obesity, intellectual disability, renal anomalies, and several other phenotypes. Currently, there are no treatments for this genetically heterogeneous disease. The M390R mutation in BBS1 is the most common genetic lesion implicated in patients. Patients with this mutation typically go blind within the first two decades of life, and males are infertile. Each of these phenotypes are recapitulated in mouse models of *BBS1[M390R]*. The eye and testes are attractive targets for gene therapy and gene editing due to their accessibility for delivery of vectors. Previous work using gene therapy of *BBS1* via subretinal injection in mice has shown that overexpression of BBS1 causes toxicity and further degradation of the retina. One method of overcoming overexpression toxicity of BBS1 is to use CRISPR/Cas9 mediated homologous recombination to edit the native gene. Using ribonucleoprotein (RNP) transfection and an optimized ssODN donor template, we have achieved editing rates of approximately 23% *in vitro*. Continued work is aimed at using CRISPR/Cas9 RNPs delivered *in vivo* and *ex vivo* to rescue male fertility and preserve photoreceptors in a mouse model of BBS1. Additionally, we are preparing an AAV2/5 viral vector expressing N-terminal Flag tagged *BBS1* driven by the endogenous promoter. Previous work has shown that the Flag tag does not interfere with BBsome formation *in vitro*. We have shown that systemic delivery of AAV vectors via tail vein injections results in significant expression of an eGFP reporter gene in the retina. We will test for overexpression toxicity *in vitro* by transfecting or transducing HEK239 cells with our BBS1 constructs and assessing cellular proliferation and cell death. We will also test for overexpression toxicity *in vivo* by delivering the viral vector to *BBS1[M390R/M390R]* and wild-type mice via subretinal injections and comparing retinal thickness between treated and mock injected animals. We will test the effectiveness of rescue of retinal degeneration and visual preservation by immunohistology, electroretinograms, and swim tests. This study will show the feasibility of using genome editing and gene therapy to treat patients with male infertility and retinal degenerative diseases using a gene that requires well-regulated expression.

1088T

Unique mutation spectrums in hearing-impaired Mongolian patients reveal possible migration events and founder effects of common deafness mutations. *Y.H. Lin[1,2,3], J. Erdenechuluun[4], Z. Makhbal[5], D. Bataakhuu[6], Y.H. Lin[1,7], C.J. Hsu[1,3], C.C. Wu[1,8], P.L. Chen[2,7,8,9].* 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 3) Department of Otolaryngology, Taichung Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Taichung, Taiwan; 4) Department of Otolaryngology, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia; 5) Department of Otolaryngology, EMJJ hospital, Ulaanbaatar, Mongolia; 6) Department of Otolaryngology, National Center for Maternal and Child Health, Ulaanbaatar, Mongolia; 7) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 8) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 9) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

More than 60% of patients with sensorineural hearing impairment have a genetic cause. From epidemiological perspectives, mutations of some deafness genes, including *GJB2*, *SLC26A4*, and *MTRNR1*, are much more prevalent than those of other genes worldwide. However, mutations spectrums of these common deafness genes differ significantly across different populations. To trace the origins of certain common deafness mutations in the Asian patients and to explore possible migration events, we performed genetic examination and haplotype analyses in 105 hearing-impaired families of Mongolian ethnicity, and compared their mutation spectrums and haplotypes to those of other East Asian cohorts, including Taiwanese, Japanese, Koreans, and Han Chinese. Dried blood spots samples were collected from Mongolian National University of Medical Sciences (Ulaanbaatar, Mongolia), and were transported to National Taiwan University Hospital (Taipei, Taiwan) for DNA extraction and mutation screening of three common deafness genes, i.e. *GJB2*, *SLC26A4*, and *MTRNR1*. In the Mongolian patients, the more prevalent *GJB2* mutations included c.23+1G>A (allele frequency of 3.33%) and c.235delC (2.86%), the more prevalent *SLC26A4* mutations included c.919-2A>G (3.81%), and the frequency of the m.1555A>G in *MTRNR1* was 1.90%. We also identified several special *GJB2* mutations which had never been reported in other East Asian populations, including c.35delG and c.269T>C (p.L90P). Further haplotype analyses revealed that some *GJB2* variants detected in the Mongolian and other East Asian patients might have the same origin. We also found some unique haplotypes only shared in the Mongolian patients, indicating migration events occurred through Central Asia to East Asia.

1089F

Single base deletion in *ATOH7* gene causes abnormal eye development in a consanguineous family. *M. Ansar, Z. Khan, E. Ullah, S. Sajid.* Department of Biochemistry, Faculty of Biological Sciences, Quaid I Azam University, 45320, Islamabad, Pakistan.

Atonal Homolog 7 (*ATOH7)* is an intronless gene that encodes a 152 amino acid protein. The encoded protein has a basic helix loop helix (bHLH) domain and works as a transcription factor to carry out spatial and temporal regulation of target genes expressed in the retinal ganglion cells (RGCs). ATOH7 is a candidate gene in several ocular defects of optic nerve and retinal vasculature. Till to date nine mutations in *ATOH7* gene have been identified in patients presenting persistent hyperplastic primary vitreous (PHPV), optic nerve aplasia (ONA), non-syndromic congenital retinal nonattachment (NCRNA) and familial exudative vitreoretinopathy (FEVR). For this study a consanguineous family was recruited from a remote village which consisted of six individuals presenting autosomal recessive PHPV. After detailed clinical evaluations blood samples were collected from 12 individuals (including six PHPV patients) and their genomic DNA underwent genome-wide homozygosity mapping with 550K Illumina array. Homozygosity mapping led to the identification of a 9.5Mb homozygous region on chromosome 10:68180999-77687528 (GRCh37/hg19). This homozygous region included the candidate gene *ATOH7*, which was selected for Sanger sequencing in all available individuals of this family. DNA sequencing identified a novel homozygous mutation c.531delG in all affected individuals, while the mutation was in heterozygous form in the unaffected carrier parent. The variant c.531delG segregates in the family and is absent in the healthy controls and public databases. Mutation taster predicts that c.531delG deletion alters the open reading frame (ORF) and produces a premature stop codon p.Ala32Profs*55, which probably result in the formation of a truncated protein. DNA binding and dimerization ability of the *ATOH7* encoded protein resides in the protein region spanning from 42 to 92 amino acids. It is therefore very likely that loss of this region due to c.531delG deletion (p.Ala-32Profs*55) in the affected individuals of this family compromise the role of *ATOH7* encoded protein as transcription factor. However additional functional characterization is required to explore the pathogenic mechanism of ATOH7 mutation identified in this study.

1090W

Molecular genetics of the Usher syndrome in Saudi Arabia: Identification of known and novel mutations by homozygosity mapping and next generation sequencing. *K. Ramzan[1], M. Al Owain[1], S. Al Hazzaa[1], S. Afzal[2], F. Imtiaz[1]. 1) King Faisal Specialist Hospital and Research Centre , Riyadh, Central, Saudi Arabia; 2) King Saud university, Saudi Arabia.*

Usher syndrome (USH) is the most common cause of combined deafness and blindness inherited in an autosomal recessive mode. Molecular diagnosis is of great significance in revealing the molecular pathogenesis and aiding the clinical diagnosis of this disease. However, the molecular diagnosis remains a challenge due to the phenotypic and genetic heterogeneity in USH. Our study aims to comprehensively delineate the genetic basis of this disorder in Saudi Arabia. Consanguineous families are a powerful resource for genetic linkage studies/homozygosity mapping for recessively inherited hearing impairment. Prioritized linkage analysis and homozygosity mapping for was conducted. A next-generation sequencing-based multiplexing assay that encompasses the 120 known hearing loss genes was also used. For genes involved in Usher syndrome, we found mutation in MYO7A (42 families), CDH23 (5 families), PCDH15 (4 families), USH1G (1 family), USH1C (1 family) and USH2A (2 family). The overall results of this study are highly suggestive that the underlying molecular basis of hearing loss in Saudi Arabia is very genetically heterogeneous. The benefit of this study will hopefully provide foundation for knowledge and awareness through screening of carrier status and genetic counseling, thereby having a major impact upon early intervention for and prevention of Usher syndrome in our population.

1091T

***SLC6A6*: Nutritional therapeutic potential of a novel autosomal recessive gene for progressive retinal degeneration and cardiomyopathy.** *E. Ranza[1], M. Ansar[2], M.T. Sarwar[2,3], M. Shetty[4], S.A. Paracha[3], J. Khan[3], I. Kern[5], C.J. Pournaras[6], A. Malcles[7], F.A. Santoni[1,2], P. Makrythanasis[1,2], K. Henry[4], J. Ahmed[3], S.E. Antonarakis[1,2,8]. 1) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 2) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan; 4) Dept. of Biomedical Sciences, School of Medicine and Health Sciences, University of North Dakota, N.Columbia, USA; 5) Pediatric Nephrology and Metabolism Unit, Pediatric Subspecialties Service, Children's Hospital, Geneva University Hospitals, Geneva, Switzerland; 6) Hirslanden Clinique La Colline, Geneva, Switzerland; 7) Department of Ophthalmology, University Hospitals of Geneva, Geneva, Switzerland; 8) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland.*

SLC6A6 (OMIM#186854) encodes a taurine transporter, whose knockout in a reported mouse model causes progressive retinal degeneration, cardiomyopathy and very low taurine levels in blood and other tissues. Taurine deficiency in cats leads to retinal degeneration and cardiomyopathy. To date, *SLC6A6* pathogenic variants have not been linked to any pathology in humans. By a combination of exome sequencing and homozygosity mapping, we have studied a Pakistani consanguineous family (F315), with two children suffering from progressive visual impairment. We have identified a recessively segregating likely pathogenic homozygous missense variant Gly399Val in the 8th transmembrane domain of *SLC6A6*. Gly399 is very well conserved down to zebrafish. Molecular modeling of this variant has indicated that it likely cause displacement of the Tyr138 (TM3) side chain, which is important for recognition and transport of taurine. The two affected children (15-years old boy and 6-years old girl) have very low blood taurine levels (6 and 7 μmol/l), while their heterozygous parents have intermediate levels (24 and 34 μmol/l), and an unaffected non-carrier sibling has normal levels (71 μmol/l) (normal values: 37-127 μmol/l). In-vitro experiments showed that the taurine uptake of HEK-293 cells with mutant (Gly399Val) SLC6A6 was between 11-18% compared to normal. Detailed clinical evaluation showed that the affected boy has complete visual loss, while the affected girl has severe but partial retinal degeneration. In addition, both affected siblings have mild hypokinetic cardiomyopathy with systolic dysfunction. After ethical approval, an oral taurine loading test of 100mg/kg was performed in the family as recommended by the literature and the pharmacological assessment. We observed that the patients receiving a taurine supplementation of 100mg/kg/day (divided in 3 doses) have nearly normal taurine levels in blood. We have initiated a taurine supplementation to both affected individuals with the objective to delay the progression of the visual impairment in the affected girl and to treat the cardiomyopathy in both siblings. We conclude that *SLC6A6* is a novel causative gene for progressive visual loss and cardiomyopathy in humans and provides a potential target for nutritional therapeutic intervention.

1092F

Genetic causes of CHARGE syndrome identified by whole exome sequencing. *D.M. Martin[1,2], A. Moccia[2], A. Srivastava[2], J.M. Skidmore[1], M.M. Wheeler[3], S. Bielas[2], University of Washington Center for Mendelian Genomics (UW CMG).* 1) Pediatrics, Univ Michigan Med Ctr, Ann Arbor, MI; 2) Human Genetics, Univ Michigan Med Ctr, Ann Arbor, MI; 3) University of Washington, Seattle, WA.

CHARGE syndrome is an autosomal dominant, multiple congenital anomaly condition associated with combined vision and hearing loss and malformations in craniofacial structures. CHARGE syndrome affects 1 in 10,000 newborns worldwide and is the second leading cause of deafblindness. Individuals with CHARGE syndrome exhibit variable expressivity of major clinical features, including ocular coloboma, choanal atresia, clefting, and ear abnormalities including hearing loss and inner ear dysplasia. Minor clinical features include global developmental delay, hypogonadotropic hypogonadism, cardiac malformations, and skeletal/renal abnormalities. Pathogenic variants in the ATP-dependent chromatin-remodeling gene *CHD7* are present in the majority of individuals with CHARGE syndrome. However, a fraction (5-40%, depending on the study) of individuals with clinical features of CHARGE syndrome test negative for *CHD7* pathogenic variants, deletions, or duplications, raising the possibility of other genetic and non-genetic etiologies. We hypothesized that non-coding variants in *CHD7* and pathogenic variants in other genes may also cause CHARGE syndrome. To test this, we performed whole exome sequencing (WES) on 28 families with at least one individual who had clinical features of CHARGE syndrome. Pathogenic variants in *CHD7* were present in 15 affected individuals (53.6%), whereas 4 affected individuals (14.3%) exhibited pathogenic variants in genes other than *CHD7* (*EP300, KDM6A, RERE,* and *PUF60*). The remaining 9 affected individuals (32.1%) were not confirmed to have a pathogenic variant, and thus require further analysis. The individuals with pathogenic variants in *EP300, KDM6A,* and *PUF60* exhibit clinical features of Rubinstein-Taybi, Kabuki, and Verheij syndromes, respectively, prompting reassignment of clinical diagnoses. *RERE* encodes Arginine-Glutamic acid dipeptide repeats protein, a broadly expressed nuclear receptor and positive regulator of retinoic acid signaling. Pathogenic variants in *RERE* have been previously reported in two individuals with clinical diagnoses of CHARGE syndrome, one of whom tested negative for *CHD7* pathogenic variants. Taken together, our results show that comprehensive WES can clarify clinical diagnoses and establish new phenotype-genotype relationships. These observations highlight overlapping molecular and epigenetic mechanisms that contribute to the pathogenesis of CHARGE syndrome and related clinical disorders.

1093W

Novel mutations underlying sensorineural hearing loss in Brazil. *K. Lezirovitz[1], G. Vieira-Silva[1], V. Pedroso-Campos[1], E. Pardono[2], J. Sampaio-Silva[1], R. Jesus-Santos[1], L. Reis-Nascimento[1], R.K. Tsuji[3], A.C. Batissoco[1], R.F. Bento[1], J. Oiticica[1].* 1) Otolaryngology Lab/LIM32, School of Medicine Clinics Hospital, University of São Paulo, São Paulo, São Paulo, Brazil; 2) Health Sciences School of São Paulo – FACIS, São Paulo, SP, Brazil; 3) Cochlear Implant Group, Otolaryngology Department, School of Medicine Clinics Hospital, University of São Paulo, São Paulo, SP, Brazil.

Hearing loss is a highly genetic heterogeneous disorder with many causative genes for syndromic and nonsyndromic forms. It has major impacts in language acquisition and communication leading to social, physical and psychological issues. We investigated the prevalence of *GJB2* gene mutations in 339 sensorineural hearing loss Brazilian patients, assisted by the Department of Otolaryngology, São Paulo Clinics Hospital, the largest hospital complex from Latin America. Besides, the prevalence of other known deafness genes were studied in selected cases (*PAX3, MITF, SLC26A4, OTOF*). Novel mutations were considered as probably pathogenic based on at least two bioinformatics tools and absence in databases (ExAC, 1000G). Among the 339 patients, 92% were nonsyndromic and 1,4% has clinical suspicion of Waardenburg syndrome (WS). Sanger sequencing of *GJB2* and screening of *GJB6* deletions revealed biallelic pathogenic mutations in 11% of our samples, the most frequent mutations being c.35delG and p.W24X. We identified a novel pathogenic mutation, c.79_82delGTCCinsAGA (p.Val27fs*8), in compound heterozygosis with c.35delG in one isolated case with prelingual profound deafness. WS has autosomal dominant inheritance and is characterized by pigmentary defects, facial dimorphisms and hearing loss, with a large inter- and intrafamilial expressivity. From the five cases with clinical suspicion of WS, we identified novel probably pathogenic mutations in three of them: one in *PAX3* (c.281C>A, p.G94A) and two in *MITF* (c.1052C>A, p.S351Y; c.644_645insA, p.H215Qfs*11). Segregation analysis confirmed these mutations as probably causative: affected mother and two affected uncles also had p.G94A; p.S351Y was inherited from the father, who presented a blond wick above the right ear; c.644_645insA was *de novo* and the parents has no clinical signs of WS. In eight selected patients with inner ear malformation we found biallelic *SLC26A4* mutations in two (25%): the previously undescribed c.481T>A (p.F161I) and three already described as pathogenic (c.G636T, p.V138F; c.554G>C, p.R185T and c.1615-2A>G). Among four patients with auditory neuropathy, biallelic mutations in *OTOF* were identified in one case and his brother (25%): the novel pathogenic mutation c.3048C>A (p.E1017*) and c.3400C>T (p.R1134*), previously described as pathogenic. Our data highlights the importance of these genes for molecular diagnosis, genetic counselling and better comprehension of genotype-phenotype correlations.

1094T

Unexpected difficulties in discovery of genes involved in hearing loss. *S. Naz[1], N. Sobreira[2], D. Witmer[3], M. Ramzan[1], Baylor-Hopkins Center for Mendelian Genomics.* 1) School of Biological Sciences, University of the Punjab, Lahore, Punjab, Pakistan; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Center for Inherited Disease Research (CIDR), Johns Hopkins University School of Medicine, Baltimore, MD.

The incidence of recessively inherited moderate to severe hearing loss is unknown, though it is believed to be more common as compared to profound deafness. We ascertained three consanguineous families, with three individuals each who manifested moderate to severe hearing loss. We excluded the known deafness genes by linkage analyses. We completed whole exome sequencing in order to identify the pathogenic variants responsible for the phenotype in the participants. Variants with an allele frequency of <0.01 were retained. Apart from all missense, frameshift, in-frame indels, stop-loss and nonsense variants, we examined compound heterozygous variants as well as synonymous and intronic variants located in regions of homozygosity shared by affected individuals in each family. The frequency of the variants was checked in gnomAD. Ambiguous results were obtained in all three cases. In the first family, a homozygous variant (p.P78Q) was identified in the paralogue of a known deafness gene. This co-segregating variant was predicted to be damaging by three of seven *in silico* prediction tools. However, the glutamine residue is fixed in the same corresponding position in one orthologous protein. Comparative genomics indicates the possibility of a cis-suppression effect (Jordan et al. 2015, Nature) since the immediate upstream amino acid is a proline, instead of the serine present in humans. In the second family with three affected cousins, we found a co-segregating, predicted pathogenic variant (p.G>R) in a gene on chromosome 2. The glycine residue is conserved in evolution, but the affected gene is located close to *OTOF*, a known deafness gene. Though, we excluded *OTOF* both by linkage and sequencing, the possibility remains that a disease causing variant is located in an unsequenced regulatory element which affects *OTOF*. In the third family, a predicted pathogenic missense variant (p.Q101P) was identified in the deafness gene *SLC26A4* which was homozygous in the three affected individuals. However, the variant was also homozygous in the unaffected father who manifested no hearing loss even in the fourth decade of his life. These findings highlight challenges for the identification of pathogenic variants in moderate to severe hearing loss. Functional testing is required as well as whole genome sequencing in order to unambiguously elucidate the genetic underpinnings of the disorders in these families. Funded by NHGRI 1U54HG006542, UM1HG006542, USA ☐ HEC 3288, Pakistan.

1095F

Investigation of novel genetic bases for strabismus using linkage analysis and next-generation sequencing. *J. Jurgens[1], M. Brucato[2], D. Avramopoulos[1], E. Pugh[1], D. Witmer[1], H. Ling[1], K. Hetrick[1], P. Zhang[1], K. Doheny[1], C. Ladd-Acosta[2], D. Valle[1], N. Sobreira[1], J. Nathans[3].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD.

Strabismus is a common disorder affecting 1-4% of the population. Despite evidence for high heritability, the genetic basis of strabismus remains poorly understood. Parikh et al. (2003) described a pedigree (19 affected, 26 unaffected) in which strabismus appeared to be inherited as an autosomal dominant trait with affected members in 3 successive generations, but the phenotype was linked to chromosome 7p22.1 under a recessive inheritance model (LOD=4.51). To search for novel genetic causes of strabismus, we first repeated linkage analysis in this family using SNP array data and identified three linked regions with LOD scores>2.0. Two linked intervals on 7p21.1-7p21.3 (LOD=3.3; 11.5 Mb containing 38 genes) and 1p22.3-1p31.1 (LOD=2.5; 3.6 Mb containing 35 genes) were consistent with an autosomal recessive model, while one region on 7q11.21-7q22.1 fit an autosomal dominant model (LOD=2.4; 31.3 Mb containing 214 genes). We performed whole exome sequencing (WES) on 7 affected family members from this pedigree but failed to identify rare (MAF<0.01) functional variants (missense, nonsense, splice site, and indels) co-segregating with strabismus inside or outside of the linked regions. To investigate variants that were missed by WES, we next conducted whole genome sequencing (WGS) on 4 severely affected and genetically distant family members. We began by analyzing WGS in the same manner as WES, but found no variants that had not been identified by WES. Using five algorithms, we searched for CNVs and structural variants in the linkage regions that co-segregated with strabismus, but found none that perturbed coding regions. Currently, we are prioritizing noncoding variants in the linked intervals using functional annotation tools (e.g. ENCODE, Roadmap Epigenomics, GWAVA, CADD, VISTA Enhancer, and miRdSNP) and additional frequency-based databases for filtering (e.g. gnomAD). We are also considering alternative genetic models, including potential oligogenic contributions of variants in the linked interval(s) as well as increased frequencies of disease alleles. Once novel candidate genes and variants are identified, we will proceed with their validation and segregation analysis in additional family members. We anticipate that identification of causal genetic variant(s) in this family will identify relevant biological systems for strabismus. *We thank the Burroughs-Wellcome Fund/MD-GEM, NEI 2T32EY007143-21, and NHGRI 1U54HG006542 for support.*

1096W

Linkage analysis and whole genome sequencing analysis in familial isolated strabismus. *X. Ye[1,2,3], C. Shyr[2,3], N. Roslin[5], D. Giaschi[4], C. Gregory-Evans[4], M. Patel[1,3], W.W. Wasserman[1,2,3].* 1) Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics, Vancouver, BC, Canada; 3) BC Children's Hospital Research Institute, BC, Canada, Vancouver, BC, Canada; 4) Ophthalmology and Visual Sciences, University of British Columbia, Vancouver, BC, Canada; 5) FORGE Canada Consortium.

Strabismus is a misalignment of the eyes, with prevalence as high as 5% in the general population. This complex condition has been associated with many environmental risk factors and genetic components, the latter having been demonstrated through twin studies and family studies. However, we still have limited understanding of strabismus genetics for isolated strabismus, the most common form. We report a seven-generation pedigree with an autosomal dominant pattern of isolated strabismus. We requested relevant clinical records of affected family members and/or arranged detailed phenotyping. Affected members have either esotropic or hypertropic, comitant or non-comitant, strabismus with varying severity. No other ocular or neurological abnormalities segregated in the pedigree. Fourteen family members from three generations were enrolled in total, with nine affected members. The most distantly related individuals are third-degree cousins. Thirteen were enrolled initially for linkage analysis, and an affected second-degree cousin of the index was enrolled later for whole genome sequencing. A high density genome-wide mapping study based on HumanOmni 2.5 SNP array data of thirteen individuals was undertaken and showed a maximum multipoint LOD score of 3.55 on chromosome 14q under parametric linkage models in MERLIN. The identified linkage region on chromosome 14 is a gene poor region, and a parallel whole exome sequencing study on a pair of third-degree cousins did not reveal potential protein-coding variants. To follow up, we conducted a whole genome sequencing study in three second-degree cousins. SNPs extracted from whole genome sequencing of the new participant were combined with the SNP array data and confirm the linkage region. A few non-coding variants were identified by a combination of bioinformatic analyses. The top candidate variants are being subjected to validation focusing on candidate genes associated with the altered regulatory regions. Gene expression levels in lymphoblastoid cell lines and gene knockdown effect in a zebrafish morpholino model are used to examine the candidate genes. Results will be reported at the meeting. Study of this rare family provides an opportunity to reveal genetics of isolated strabismus and suggest diverse strabismic phenotypes can be associated with the same genetic factor.

1097T

A novel homozygous deletion in last exon of CYP1B1 gene cause primary congenital glaucoma in an Iranian female patient. *M. Noruzinia, A. Salehi, M. Ahmadvand, O. Bashti.* Department of Medical Genetics, Tarbiat modares university, Tehran, GA.

Primary Congenital Glaucoma (PCG) is a serious eye disease that is responsible for 15% of blindness in the world. Generally, defects in the trabecular meshwork and anterior chamber angle lead to primary congenital glaucoma. The disease is severe, manifesting at birth or early infancy, associated with intraocular pressure (IOP), and enlarged cornea and globe (buphthalmos). The precise molecular basis of this disabling disease is unclear but some genetic pathogenic mutations are reported in PCG. Mutations in *CYP1B1* and *LTBP2* are two important explanations for PCG. The *CYP1B1* gene harbors more than 70 mutations leading to PCG in different ethnic groups and missense mutations assigned to most cases. Here we report a novel homozygote deletion of a "C" nucleotide at the last exon of CYP1B1 gene in a female patient suffering from PCG. The patient born of a consanguineous marriage and her symptoms are typical for PCG. According to these representations, we employed targeted sequence analysis for genes associated with PCG (ASB10, CYP1B1, FOXC1, LTBP2, MYOC, NTF4, OPA1, OPTN, TBK1, WDR36). Obtained sequences were aligned to human reference genome (GRCh37/h19) using BWA program, analyzed bypicard and GTAK-Lite toolkit to identify variants in the targeted genes relevant to clinical indication. a novel mutation in exon 3 of CYP1B1. linkage analysis further confirmed the absence of the mutation in healthy members of the family. This novel mutation is pathogenic and would be considered in genetic counseling and genetic testing in Iran. Phenotypes in this family and mechanism of mutation pathogenicity are discussed in this article. .

1098F

Identification of novel mutations of *POU3F4* in seven Chinese families with X-linked nonsyndromic hearing loss. *H. Yuan, J. Cheng, Y. Lu.* Medical Genetics Center, Southwest Hospital, Chongqing, China.

DFNX2 deafness, accounting for 50% of X-linked nonsyndromic hearing loss, can be easily distinguished from the other X-linked hearing loss by the remarkable radiological features include partial hypoplasia of the cochlea and dilation of the internal auditory canal, often with a fistulous communication between its lateral aspect. *POU3F4* is the causative gene for DFNX2. We report 7 Chinese family with X-linked non-syndromic hearing impairment in which seven affected males demonstrated congenital profound sensorineural hearing loss. Computer tomography of the temporal bones showed bilateral dilation of the internal auditory canals and with a fistulous communication between its lateral aspect and the basal turn of the cochlea in two affected brothers and normal in their mother. Seven missense mutation in the *POU3F4* gene were identified in these families, and failed to detect in 3700 normal human control. The appropriate use of the genetic testing and prenatal diagnosis has great potential to reduce the recurrence risk in DFNX2 families with genetic defects.

1099W

Next generation sequencing of three families with severe keratoconus identifies putative disease-causing variants. *S.E.M. Lucas[1], B.J. McComish[1], N.B. Blackburn[1,2], E. Souzeau[3], R.A. Mills[3], J.E. Craig[3], S.E. Staffieri[4], D.A. Mackey[5], J.C. Charlesworth[1], K.P. Burdon[1].* 1) Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia; 2) South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, Texas, USA; 3) Department of Ophthalmology, Flinders University, Adelaide, South Australia, Australia; 4) Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, Melbourne, Victoria, Australia; 5) Lion's Eye Institute, Centre for Ophthalmology and Visual Sciences, University of Western Australia, Perth, Western Australia, Australia.

Keratoconus is an eye disease characterised by progressive thinning and protrusion of the cornea causing high myopia and irregular astigmatism. This severely affects vision and quality of life. Keratoconus has a strong genetic basis, but is typically considered a complex disease. Familial examples of Mendelian autosomal dominant and recessive transmission are reported; however, the genetic cause has not yet been identified in most families. Using whole exome (WES) and whole genome (WGS) sequencing, we aim to identify putatively functional variants that segregate with disease in three pedigrees with highly penetrant, severe keratoconus. Family 1 is an extended Australian pedigree with nine affected and seven unaffected family members across four generations. We previously reported two linkage regions, 1p36.23-36.21 and 8q13.1-21.11, and hypothesised a digenic mode of inheritance in this family. WES was obtained for all 16 family members and WGS data for four cases and two unaffected individuals. Family 2 is an extended pedigree from Jordan with eight cases and three unaffected relatives with autosomal dominant keratoconus. WGS was obtained for all 11 family members. Family 3 is an Australian family with autosomal recessive inheritance affecting two siblings from unaffected and unrelated parents. One of the cases has an unaffected child. WGS data were obtained for all five family members. We identified a potentially pathogenic variant in *VPS13D* within the 1p36.23-36.21 linkage region in Family 1. This c.1850C>T variant results in a p.(P617L) substitution, has a frequency of 0.00007 in ExAC, is predicted to be deleterious (CADD score of 34) and occurs at a highly conserved residue (GERP score of 6.07). Very little is known about the function of VPS13D. It may play a role in the regulation of IL-6, which has been reported to be increased in tears from keratoconic eyes. We have identified a putative keratoconus-causing variant in a novel gene, *VPS13D*. Analysis is continuing with the WGS data from all three families to elucidate a potentially functional variant on chromosome 8 in Family 1 and to identify linkage regions and candidate genes in Families 2 and 3. Identifying genes that contribute to keratoconus susceptibility would aid in identifying key biological pathways involved in the disease process. This is essential for developing novel non-surgical treatments with the potential to slow or prevent keratoconus progression.

1100T

Identification of novel candidate genes for recessive visual impairment by analyzing 132 consanguineous families. *M. Ansar[1], M.T. Sarwar[2], H. Shaikh[3], P. Makrythanasis[1], N. Batool[1,2], K. Gul[5], A.A. Khan[4], A. Nazir[2], S.A. Paracha[2], J. Khan[2], S. Qureshi[5], S. Imtiaz[5], Z. Iqbal[6], E. Falconnet[1], P.G. Ribaux[1], Y.M. Waryah[3], M. Guipponi[1,4], C. Borel[1], F.A. Santoni[1,4], A.M. Waryah[3], J. Ahmed[2], S.E. Antonarakis[1,4,7].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan; 3) Molecular Biology and Genetics Department, Medical Research Center, Liaquat University of Medical and Health Sciences, Jamshoro, Pakistan; 4) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 5) Department of Genetics, University of Karachi, Karachi, Pakistan; 6) Department of Neurology, Oslo University Hospital, Oslo, Norway; 7) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland.

Populations that practice consanguinity have a higher prevalence of autosomal recessive disorders due to the extensive genomic regions of homozygosity that include alleles identical by descent. By studying runs of homozygosity (ROH) across the genome, we have observed that the average ROH size of 215 Mb in the offspring of consanguineous families is significantly larger than that of 25 Mb in outbred individuals. Using exome sequencing and genotyping, we have analyzed 132 highly consanguineous families from Pakistan (where consanguinity rate is > 60%), with undiagnosed likely autosomal recessive visual impairment (VI). For each family, there were at least 2 affected individuals; samples from both parents and offspring were collected after informed consent; there were 400 affected individuals in this study. Although the diagnostic yield in visual impairment families is considerably high, the study of consanguineous families continues to provide the opportunity to discover new recessive genes. In 62% of the VI families, we have identified known or novel likely pathogenic variants in genes which are already reported to cause visual impairment (in a similar number of consanguineous families with intellectual disability from the same geographical origin the diagnostic yield is approximately 28%). In 20% of the VI families, we failed to identify any causative or candidate variants in the coding regions covered by exome sequencing. However, in 18% of the VI families we have found likely damaging variants in 24 novel candidate genes. Some of the candidate genes are likely the targets for potential therapeutic interventions. All of the candidate genes/variants have been submitted to matchmaking efforts. We conclude that the discovery of novel autosomal recessive genes for visual impairment can be accelerated by analyzing large cohorts of consanguineous families. This has important implication in correct molecular diagnosis, family planning, and may reduce the incidence of autosomal recessive disorders in consanguineous couples. Gene discovery of novel candidates is the first step in understanding the molecular pathophysiology of a disease and may contribute to innovative treatment.

1101F

Genomic analysis of inherited hearing loss in the Palestinian population. *A.I. Abu-Rayyan[1,2], L.A. Kamal[1], C.J. Canavati[1], Z. Brownstein[2], S. Gulsuner[3], T. Walsh[3], M.C. King[3], K.B Avraham[2], M.N Kanaan[1].* 1) Biological Sciences , Bethlehem University , Bethlehem , Palestinian Territory; 2) Department of Human Molecular Genetics & Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Department of Medical Genetics and Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Childhood onset hearing loss (HL) in Palestine is more frequent than elsewhere, in part due to the high rate of consanguineous marriage in this population. The goal of our project is to estimate the proportion of childhood onset HL in Palestine that is genetic and to identify and characterize all the genes involved. Through schools and clinics for the deaf in the West Bank and Gaza, we enrolled 364 families with HL, confirmed by audiology, either syndromic (21 families) or non-syndromic (343 families). As a first screen, a deaf proband from each family was sequenced for GJB2 and 19 other mutations that are relatively common among Palestinians with HL. Only 11% of families (40) had mutations in GJB2, which is responsible for most inherited HL worldwide. HL in an additional 18% of families (65) was due to other founder alleles. Of the unresolved families, 120 were sequenced using our custom gene panel of 246 human and mouse deafness genes (PMID: 21917145). Novel candidate variants were evaluated by co-segregation with HL in fully sampled families and in Palestinian hearing controls. For genes expressed in accessible tissue, potential splice variants were evaluated by transcript analysis of participant RNA. All variants interpreted as causal were screened in the entire cohort of 364 families. Ultimately, HL in 45% of families (164) was explained by 74 different mutations, most not previously reported, in 29 genes. Results were submitted to LOVD and ClinVar. Genes responsible for non-syndromic HL were GJB2 (40 families), MYO15A (26), TRIOBP (14), MYO7A (11), TMPRSS3 (10), CDH23 (7), TECTA (6), OTOA (3); CACNA1D, CLDN14, ESRRB, ADGRV1, LOXHD1, MYO6, PJVK, PTPRQ, and PTRH2 (2 each); and EPS8L2, MYH9, PCDH15, POU3F4, TMC1, and TMHS (1 each). Genes responsible for syndromic or other HL phenotypes were SLC26A4 (Pendred signs, highly variable across families, 12 families), PAX3 (Waardenburg-1, 3), GPSM2 (CMCS, 2), OTOF (auditory neuropathy, 2), PTRH2 (IMNEPD, 2) and USH1G (RP, 1). The 38% of families (139) that are unsolved singleton cases provide an upper estimate for the proportion of hearing loss in the population that may be due to non-genetic causes. The 16% of families (60) with multiple deaf children and no pathogenic mutation in any known deafness gene represent a highly informative resource for identification and characterization of additional HL genes and of distant non-coding regulatory mutations. Supported by NIH R01DC011835 to KBA and MNK.

1102W

Siblings with Perralut syndrome and *LARS2* mutation who presented with neurologic abnormalities. *R. Kosaki[1], R. Horikawa[2], E. Fujii[3], K. Kosaki[4].* 1) Division of Medical Genetics, National Center for Child Health and Development, Tokyo, Japan; 2) Division of Endoclinology and Metabolism, National Center for Child Health and Development, Tokyo, Japan; 3) Department of Perinatal Medicine and Maternal Care, National Center for Child Health and Development, Tokyo, Japan; 4) Center for Medical Genetics, Keio University School of Medicine.

   Perrault syndrome represents a genetically heterogeneous disorder characterized by sensorineural hearing loss in males and females and ovarian dysfunction in females. Causative genes include *HARS2*, *HSD17B4*, *CLPP*, *C10orf2*, and *LARS2*. Some patients with Perrault syndrome exhibit neurologic features including learning disability, cerebellar ataxia, and peripheral neuropathy and are classified as type 2 and are clinically separate from those without neurological symptoms other than a healing loss (type 1). To date, all reported patients with *LARS2* mutations (15 patients in 8 families) have been classified as type 1. Here, we report female siblings with biallelic mutations in *LARS2*, p.Glu294Lys and p.Thr519Met, who were classified as type 2. The proposita developed progressive sensorineural hearing loss at 18 months and pervasive developmental disorder at 8 years, with repetitive behavior, insistence on sameness, attention deficit, tic, irritability, and an ataxic gait. At age 15 years, she was diagnosed as having primary amenorrhea with elevated FSH and LH and a decreased estradiol; ultrasound and magnetic resonance imaging examinations revealed a small uterus and no detectable ovaries. The proposita's younger sister presented with neonatal sensorineural hearing loss and a mild delay in motor and speech development. She was diagnosed as having primary amenorrhea with endocrinologic and radiographic findings that were comparable to those of her sister. She had difficulty with reading comprehension, and had trouble with open-ended test questions at 12 years of age. We concluded that Perrault syndrome patients with *LARS2* mutations are at risk for neurologic problems, despite previous notions otherwise. .

1103T

A homozygous truncating variant of *KCNE1* (p.Tyr46*) associated with deafness in the absence of a long QT interval reveals a novel genotype-phenotype correlation. *R. Faridi[1,2], S. Boyden[1], R. Tona[1], M. Assir[4,5], M. Shahzad[4], A. Khan[2], R. Morell[3], A. Rehman[1], S. Riazuddin[5], T. Friedman[1].* 1) Laboratory of Molecular Genetics, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD 20892; 2) National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore 54550, Pakistan; 3) Genomics and Computational Biology Core, NIDCD, NIH, Bethesda, MD 20892, USA; 4) Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad, Pakistan; 5) Allama Iqbal Medical Research Centre, Jinnah Hospital Complex, Lahore 54550, Pakistan.

   KCNE1 encodes a regulatory subunit that can associate with the KCNQ1 voltage-gated potassium channel comprised of four subunits encoded by *KCNQ1*. *KCNE1* and *KCNQ1* appear to be necessary for both normal hearing and cardiac ventricular repolarization. Mutations in these two genes are associated with Jervell and Lange-Nielson syndrome (JLNS1, MIM 220400 and JLNS2, MIM 612347, respectively). JLNS is an autosomal recessive disorder characterized by congenital, bilateral profound sensorineural hearing loss and long electrocardiographic QT intervals (>500 milliseconds), and syncopal events or sudden cardiac death in the first or second decade of life if left untreated. Heterozygous normal-hearing carriers can have long QT intervals as part of Romano-Ward syndrome. We ascertained a large consanguineous family (PKDF461) segregating non-syndromic deafness, which showed evidence for linkage to chromosome 21q. Whole genome sequencing identified a rare stop gain variant (p.Tyr46*) of *KCNE1* (allele frequency of 0.000008123 in gnomAD) co-segregating with the deafness phenotype in this family. Family PKDF461 has four deaf individuals who are homozygous for the p.Tyr46* allele, three normal-hearing heterozygous carriers and four normal-hearing individuals that are homozygous for the wild type tyrosine-46 residue. Biallelic null pathogenic variants in *KCNE1* have not been previously reported. In family PKDF461, homozygotes and heterozygotes for p.Tyr46* have a normal QT interval and there have been no sudden unexplained deaths. The absence of a cardiac phenotype in the presence of a null allele of human *KCNE1* may be explained if in the heart (but not the ear), wild type *KCNQ1* can function adequately in the complete absence of a *KCNE1* beta-subunit, but does not function properly in the presence of a dominant negative or gain of function variant of *KCNE1*, which may negatively effect the *KCNQ1* function when biallelic, as in JLNS patients. We hypothesize that another member of the KCNE gene family may compensate for loss of *KCNE1* function in the heart but not the ear of our patients. Although several missense mutations of *KCNE1* are associated with JLNS2, our study indicates a more complex genotype-phenotype relationship, as a biallelic null variant of *KCNE1* is associated with non-syndromic deafness.

1104F

Variation in *MERTK* in patients with retinal dystrophy. *C. Jespersgaard[1], H.G. Gellert-Kristensen[1], M. Fang[2], M. Bertelsen[1], X. Dang[2], H. Jensen[3], Y. Shen[2], N. Bech[3], I. Dai[2], T. Rosenberg[3], J. Zhang[2], L.B. Moeller[1], Z. Tümer[1], K. Broendum-Nielsen[1], K. Groenskov[1].* 1) Clinical Genetics Department, Kennedy Center, Rigshospitalet, Glostrup, Denmark; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) The National Eye Clinic, The Kennedy Center, Department of Ophthalmology, Copenhagen University Hospital, Glostrup, Denmark.

Retinal dystrophy encompasses a range of diagnoses and modes of inheritance, and more than 250 genes have been associated so far. Pathogenic sequence variations in one of these, MER tyrosinase kinase protooncogene; *MERTK*, are known to cause retinitis pigmentosa type 38. MERTK associated retinitis pigmentosa comprise around 1 % of patients displaying a recessive inheritance. Previously 60 different disease causing or probably disease casing variants have been reported in HGMD® Professional 2017.1. Among these are four gross deletions and recently Evans et al reported an additional patient with a *MERTK* deletion encompassing exon 6 to 8. One of the reported gross deletions has been shown to be a founder in the Faroe Islands and is responsible for approximately 30 % of all retinitis pigmentosa cases within that population. In a study with targeted sequencing of 804 individuals clinically diagnosed with retinal dystrophy we found 6 individuals who were homozygous or compound heterozygous for variants in MERTK. The individuals had one of the following diagnoses: autosomal recessive retinitis pigmentosa; autosomal dominant retinitis pigmentosa; X-linked retinitis pigmentosa; Leber's congenital amaurosis; Usher syndrome; Bardet-Biedl syndrome; cone- or cone-rod dystrophy; unspecified macular dystrophy or Stargardt disease; congenital stationary night blindness or age-related macular degeneration. All individuals gave written informed consent for genetic analysis. We analyzed our patient cohort for MERTK variants by analyzing all the SNVs as well as analyzing the data looking for gross deletions or duplications to elucidate the contribution of MERTK within our patient population. We found gross deletions encompassing several exons in heterozygous as well as homozygotic form in our patient cohort. This demonstrates that CNV screening of MERTK should be considered in cases displaying recessive inheritance.

1105W

The NIH Oculocutaneous Albinism Natural History Study. *D.R. Adams[1], W.M. Zein[2], B.P. Power[1], T. Nguyen[3], C. Brewer[4], C. Zalewski[4], J. Bryant[1], W.A. Gahl[1], B.P. Brooks[2].* 1) NHGRI, NIH, Bethesda, MD; 2) NEI, NIH, Bethesda, MD; 3) Clinical Center, NIH Bethesda, MD; 4) NIDCD, NIH Bethesda, MD.

Oculocutaneous Albinism (OCA) is a syndrome of hypopigmentation and aberrant visual development resulting from decreased production of melanin. It is pan-ethnic and constitutes a significant minority of inherited vision impairments. While most isolated OCA is caused by mutations in two genes (*TYR* and *OCA2*), growing lists of additional loci have been described in recent years. Recent clinical trials and advances in the understanding of melanocyte biology have reinvigorated efforts to look for rational treatments of the visual manifestations of OCA. The OCA natural history study at the NIH Clinical Center enrolls individuals with albinism, seeking to expand understanding of the clinical manifestations and underlying biology. The study has four primary goals: 1) To clinically and comprehensively characterize known OCA types; 2) To use participants' cultured melanocytes to study pigment biology, variability in pigment formation related to genotype, and response to proposed treatments; 3) To recruit study participants with hypopigmentation not due to known albinism-causing genes; and 4) to evaluate methods of quantifying eye pigmentation, skin pigmentation and other clinical parameters that may serve as outcome measures in future treatment studies. To date, 125 persons with albinism have been enrolled. Clinical histories, biological specimens and laboratory testing have provided collect data for targeted aspects of OCA including vision, hearing, dietary vitamin D sufficiency, and visual adaptation. The presented work includes highlights from a current effort to collate data from the last 9 years of the study. These data refine the phenotype, contribute to the phenotypic presentation of novel genotypes, and lay the groundwork for future therapeutic studies and clinical guidelines.

1106T

Novel *PXDN* mutations cause microphthalmia and anterior segment dysgenesis. *N. Chassaing[1,2], C. Zazo-Seco[2], P. Bitoun[3], O. Roche[4], J. Plaisancié[1,2], N. Ragge[5,6], P. Calvas[1,2].* 1) Department of Medical Genetics, Purpan Hospital, Toulouse, France; 2) UMR 1056 Inserm -Toulouse University, Toulouse, France; 3) SIDVA 91, Ophthalmic Genetics, 1 rue de la Cour de France, 91460 Juvisy s/orge, France; 4) Department of Ophthalmology, IHU Necker-Enfants Malades, Paris Descartes University, 75015 Paris, France; 5) Dept of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK; 6) West Midlands Regional Clinical Genetics Service and Birmingham Health Partners Birmingham Women's Hospital NHS Foundation Trust, Birmingham UK.

Anterior segment dysgeneses encompass a wide spectrum of developmental disorders, including iris corectopia, Peters and Rieger anomalies. Microphthalmia refers to a reduced axial length of the globe at least 2 standard deviations below the mean for age. Microphthalmia is classified as simple or, complex, when associated with anterior and/or posterior segment dysgenesis. Numerous genes have been shown to underlie these ocular developmental anomalies. Biallelic mutations in *PXDN* gene were first reported in 2011 in two consanguineous families affected by congenital cataract, corneal opacity, and developmental glaucoma and in one family affected by developmental glaucoma and severe corneal opacification. All affected family members displayed isolated ocular involvement. Since then, 5 individuals with *PXDN* mutations have been reported, expanding the phenotype to anterior segment dysgenesis and microphthalmia associated with developmental delay in two patients. We report herein, 3 novel patients with ocular involvement due to *PXDN* mutations. In a female patient suffering from microphthalmia and Peter's anomaly, whole genome sequencing identified a maternally inherited mutation affecting the initiation codon (c.1A>G [p.Met1?]) and a paternally inherited frameshift insertion (c.697_698insCTCAG [p.Val223AlafsTer11]). In a second family, two brothers were affected by ocular developmental anomalies, one with congenital cataract, microcornea, and anterior segment dysgenesis, while the other had severe unilateral microphthalmia, with contralateral anterior segment dysgenesis. Targeted sequencing of a panel of 24 genes involved in ocular developmental anomalies led to the identification of a maternally inherited frameshift mutation (c.4216_4225dup [p.Arg1409ProfsTer2]), and a paternally inherited missense mutation (c.2276C>T [p.Ser759Leu]). The missense variant was rare in the ExAC database (1/120.000 alleles), affected a very conserved amino-acid, and was considered by *in silico* analysis to be deleterious. Interestingly, although heterozygous carriers reported so far have been asymptomatic, the heterozygous mother of the two affected sibs had goniodysgenesis suggesting that *PXDN* mutations may lead to mild ocular developmental defects in the heterozygous state. In conclusion, we report herein 3 novel patients with compound heterozygous mutations of the *PXDN* gene highlighting the importance of screening this gene more widely in patients with anterior segment dysgeneses.

1107F

Comprehensive analysis of CNVs in patients with congenital eye malformations by targeted next generation sequencing. *M. Corton[1,2], D. Aguilera-García[1], L. Rodrigues Jacy da Silva[1,3], M. Tarilonte[1], P. Ramos[1], A. Gomez[1], C. Villaverde[1,2], J. Rosell[4], V. Lopez[5], M.J. Ballesta[5], E. Guillen[5], M.J. Trujillo-Tiebas[1,2], F. Blanco-Kelly[1,2], C. Ayuso[1,2].* 1) Genetics Department. IIS-Fundacion Jimenez Diaz University Hospital, Madrid, Madrid, Spain; 2) Centre for Biomedical Network Research on Rare Diseases (CIBERER), ISCIII, Madrid; 3) University of Mogi das Cruzes, São Paulo, Brazil; 4) Genetics Department. University Hospital Son Espases. Palma de Mallorca, Spain; 5) Medical Genetics Department. University Hospital Virgen de la Arrixaca, Murcia, Spain.

Introduction: Congenital ocular malformations are very highly heterogeneous conditions displaying a wide spectrum of overlapping phenotypes that can occur isolated or taking part of syndromes. They include panophthalmic disorders such as microphthalmia, anophthalmia and coloboma (MAC) spectrum and aniridia, and anomalies restricted to the anterior or posterior segment. Mutational load in the associated genes is not well known due to lack of systematic molecular characterization of large cohorts. Our aim was to perform a comprehensive molecular screen for SNVs and CNVs in genes causing ocular malformations using a custom next-generation sequencing (NGS) gene panel in a large cohort of Spanish patients. **Material and Methods**: A cohort of 130 patients with several ocular phenotypes were analysed, including syndromic and isolated forms of MAC, aniridia, anterior segment dysgenesis, nerve optic hypoplasia and albinism that mostly were previously tested by Sanger sequencing and MLPA for the most prevalent genes. These includes control cases with previously known pathogenic variants and CNVs. A set of up to 260 genes were analyzed by means of custom gene panels (Haloplex and SureSelect), followed by sequencing in Illumina platforms. SNVs were analyzed following an in-house bioinformatic pipeline. CNVs identification was performed prioritizing statistically significant regions in the read depth. Custom Agilent CGH arrays were used to validate CNVs and to refine chromosomal breakpoints. **Results:** Currently we have findings for 88 patients with a variety of phenotypes and additional 48 patients are now being analyzed. A total of 19 pathogenic SNVs spanning 14 different genes and 13 additional VUS were identified, validated and segregated in the families. Our specific bioinformatic method in combination with the high depth of coverage (>400X) allowed the detection of 7 structural variants: a duplication of *GDF3,* complete deletions of *FOXC1, GJA8, HPS5* and *OTX2-SIX6 and* multi-exon deletions of *PAX6*. **Conclusions:** Our study allowed the characterization of 24% of the patients (40% of the patients with MAC). Our results have shown that a NGS panel is a good strategy for the genetic analysis of these patients since it permits the co-detection of SNVs and CNVs in a single design when a homogeneous coverage between samples is ensured. **Grants**: ISCIII (CP12/03256), MINECO (SAF2013-46943-R), Fundación Mutua Madrileña.

1108W

Genetic characteristics of an international large cohort with Stargardt disease: The progression of atrophy secondary to Stargardt disease (ProgStar) study. *K. Fujinami[1,2,3,4], R.W. Strauss[3,4,5,6,7], J. Chiang[8], I. Audo[9,10], P.S. Bernstein[11], D.G. Birch[12], S.M. Bomotti[13], A.V. Cideciyan[14], C.M. Dhaenens[15], A.M. Ervin[5], S. Kohl[16], M.J. Marino[17], J.A. Sahel[9,10,18,19], S. Mohand-Said[9,10], J.S. Sunness[20], E.I. Traboulsi[16], S. West[5], R. Wojciechowski[13], C. Zeitz[9], E. Zrenner[21,22], M. Michaelides[3,4], H.P.N. Scholl[5,7].* 1) Laboratory of Visual Physiology, Division for Vision Research, National Institute of Sensory Organs, National Hospital Organization, Tokyo Medical Center, Tokyo, Japan; 2) Department of Ophthalmology, Keio University, School of Medicine, Tokyo, Japan; 3) UCL Institute of Ophthalmology, London, United Kingdom; 4) Moorfields Eye Hospital, London, United Kingdom; 5) Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland, USA; 6) Departments of Ophthalmology, Johannes Kepler University Linz, Linz, Austria; 7) Department of Ophthalmology, University of Basel, Basel, Switzerland; 8) Casey Molecular Diagnostic Laboratory, Portland, Oregon, USA; 9) Sorbonne Universités, UPMC Univ Paris 06, INSERM U968, CNRS UMR 7210, Institut de la Vision, Paris, France; 10) CHNO des Quinze-Vingts, DHU Sight Restore, INSERM-DGOS CIC 1423, Paris, France; 11) Moran Eye Center, University of Utah, Salt Lake City, Utah, USA; 12) Retina Foundation of the Southwest, Dallas, Texas, USA; 13) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 14) Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 15) Biochemistry and Molecular Biology Department, UF Génopathies, CHU Lille, Lille, France; 16) Molecular Genetics Laboratory, Institute for Ophthalmic Research, University of Tuebingen, Germany; 17) Cole Eye Institute, Cleveland Clinic, Cleveland, Ohio, USA; 18) Fondation Ophtalmologique Rothschild, Paris, France; 19) Department of Ophthalmology, The University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA; 20) Richard E Hoover Low Vision Rehabilitation, Services, Greater Baltimore Medical Center, Baltimore, Maryland, USA; 21) Center for Ophthalmology, Eberhard-Karls University Hospital, Tuebingen, Germany; 22) Werner Reichardt Centre for Integrative Neuroscience, University of Tuebingen, Germany.

Purpose Stargardt disease 1 (STGD1; MIM 248200), the most prevalent inherited macular dystrophy, is an autosomal recessive condition caused by pathogenic sequence variants in the *ABCA4* gene (ATP-binding cassette subfamily A member 4; MIM 601691). We describe the genetic characteristics of the cohort enrolled in the international multi-center progression of STGD1 study (ProgStar) and determine ethnic variations in allele frequency. **Methods** 275 subjects with a clinical diagnosis of STGD1 and harboring multiple pathogenic *ABCA4* variants were enrolled from nine centers in the USA and Europe. All the detected variants were assessed with *in silico* molecular genetic analysis. Subjects were classified into four ethnic subgroups (European/Middle Eastern, African, Asian, and Other/Unknown) based on questionnaire responses. Comparisons of allele frequency were performed with categorical testing of the independence of each prevalent allele. **Results** 195 pathogenic sequence variants were identified in the total cohort, including missense (123), splice site alteration (32), frame shift (19), stop (19), in-frame deletion (1), and large exonic deletion (1). 46 variants were novel. Exclusively missense variants were detected in 139 patients (50.5%). The cohort comprised 196 European/Middle Eastern patients, 18 African (by descent), 14 Asian and 47 with other/unknown ethnicity. The three most prevalent variants in the total cohort were p.G1961E (15.3%), p.G863A (7.3%), and c.5461-10 T>C (4.9%). The most prevalent variant in each ethnicity was p.G1961E (15.3%) in European/Middle Eastern, p.R2017H (13.9%) in African, and p.G1961E (39.3%) in Asian, respectively. The ethnic subgroup analysis revealed a statistically significant ethnic differences in the allele frequencies of p.G1961E, p.R2107H, p.E2096K, p.V1693I, c.4253+4C>T, c.2977_2984delGACATTGA, p.G991R, p.V989A, p.V643M, and p.S95P. **Conclusions** The spectrum of *ABCA4* variants was documented in a well-characterized large cohort with STGD1. Approximately half of the cohort harbors only missense variants, which is in keeping with the relatively mild phenotype of the ProgStar cohort. Subgroup analysis reveals the distinct prevalent alleles in each ethnicity, although the number of subjects for some subgroups is limited.

1109T

First systematic molecular analysis of 45 Greek patients with retinal dystrophies by next generation sequencing reveals 21 novel mutations in 30 genes and establishes a wide spectrum of distinct retinal degenerative diseases. *S. Kamakari[1,5], V. Kokkinou[1], G. Koutsodontis[1], A. Anastasakis[2], S. Koukoula[3], M.K. Tsilimbaris[4], I. Datseris[5].* 1) OPHTHALMIC GENETICS UNIT, OMMA OPHTHALMOLOGICAL INSTITUTE OF ATHENS, ATHENS, Greece; 2) Athens Eye Hospital, Glyfada, Athens; 3) Ophthalmica Institute of Ophthalmology and Microsurgery, Thessaloniki; 4) Department of Ophthalmology, University of Crete School of Medicine, Heraklion; 5) OMMA Ophthalmological Institute of Athens, Athens.

Purpose: Hereditary retinal dystrophies (RD) are characterized by clinical variability and pronounced genetic heterogeneity. The aim of this study was to molecularly diagnose 45 Greek patients with different forms of hereditary retinal dystrophies. **Materials and Methods:** 45 unrelated Greek patients were analyzed by Next Generation Sequencing (NGS), 13 of them using a 105 retinal gene panel as described (Ellingford JM et al. J Med Genet 2016; 53:761-767) and 32 of them using whole exome sequencing comprising a 287 ophthalmic gene panel as described (Haer-Wigman L et al, Eur J Hum Genet 2017; 25:591–599). In addition, PCR and Sanger sequencing, MLPA and array CGH were ised in 4 cases where no mutations or a single mutation was identified.. **Results:** A total of 43 distinct mutations were detected in 30 RD genes in this cohort of 45 Greek patients including 21 novel potentially pathogenic mutations in the *ABCA4, CAPN5, PRPF31, SPATA7, MERTK, FAM161A, ALMS1, CDHR1, MYO7A, USH2A, CNGB1, CRX, LCA5* and *PROM1* genes. The detection mutation rates were 38.5% (5/13 patients) and 71.9% (23/32 patients) for the 105 and 287 gene panels, respectively. Futhermore, in 4 cases with no mutation or a single identified mutation, complete diagnosis was achieved using additional methods thus increasing the mutation detection rates for the two subgroups to 46.1% and 81.2%. Final diagnoses of participating subjects included retinitis pigmentosa (9 cases), Usher syndrome (4 cases), cone-rod dystrophy (4 cases), Leber congenital amaurosis (4 cases), Bardet-Biedl syndrome (2 cases) and 9 single rare cases of isolated or syndromic inherited retinopathy including a case of nephrotic syndrome type 9 due to mutations in the *ADCK4* gene and a case of Knobloch syndrome due to mutations in the *COL18A1* gene. Notably, genetic findings in several families were not consistent with the initial clinical diagnosis. **Conclusions:** This is the first systematic study to investigate the molecular identity of 45 Greek patients with various subforms of RD by the use of NGS technology targeting two panels of 105 and 287 RD genes and additional methods as required. The overall mutation detection rate was 71.1% (32/45 patients). A plethora of novel mutations was documented further expanding the genetic heterogeneity. The molecular identification established the complete diagnosis of the patients thus drastically contributing to family making decision, prognosis and candidacy to current and future treatments.

1110F

Novel *RS1* gene mutations of X-linked retinoschysis Lithuanian patients.
V. Kucinskas[1], L. Ambrozaityte[1,2], L. Cimbalistiene[1,2], R. Strupaite[3], R. Asoklis[3], A. Utkus[1,2]. 1) Dept. of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; 2) Center for Medical Genetics, Vilnius University Hospital Santaros klinikos, Vilnius, Lithuania; 3) Centre of Eye Diseases, Clinic of Ear, Nose, Throat, and Eye Diseases, Faculty of Medicine, Vilnius University.

   X-linked juvenile retinoschisis (XLRS, MIM 312700) is a common early onset macular degeneration characterized by mild to severe loss in visual acuity, splitting of retinal layers, and a reduction in the b-wave of the electroretinogram (ERG). The prevalence of XLRS ranges between 1:5000 to 1:20000 males. The *RS1* gene (MIM 300839) associated with the disease encodes retinoschisin, that is expressed and secreted from photoreceptors and bipolar cells, and it binds strongly and specifically to the surfaces of many cells in the retina. Over 200 disease-causing mutations in the *RS1* gene are known with most mutations occurring as non-synonymous changes in the major protein unit, i.e. discoidin domain. First XLRS Lithuanian patients have been diagnosed due to clinical findings and confirmed molecular diagnosis. Comprehensive ophthalmological examinations, including best-coreccted visual acuity (BCVA), color vision assessment with Ishihara plates, slit-lamp, fundus examination, spectral domain optical coherent tomography (SD-OCT) with fundus autofluorescence images and full-field electroretinography, have been performed. *RS1* gene (NM_000330.3) has been Sanger sequenced for the three probands and pathogenic mutations identified. NM_000330.3:c.599G>T (p.R200L) mutation was detected in one case, *in silico* analysis showing to be pathogenic (desase causing; deleterious; probably damaging). Human Genome Mutation Database (HGMD) involves three other different mutations at the same position (CM095237, CM981767, CM981768) supporting the pathogenicity of the identified variant. NM_000330.3:c.(92_97)insC (p.W33fs) mutation creating a frameshift was identified for another proband, again *in silico* analysis indicating the variant is possibly damaging. The third case was identified with a pathogenic mutation NM_000330.3:c.422C>G (p.R141H), HGMD CM981753. The penetrance of XLRS is almost complete but clinical expression is highly variable, though Lithuanian XLRS patients are presented with similar phenotype. Functional studies would support the characterization of the specific identified variants on the effect of retinoschisin expression, subcellular localization, and protein structure.

1111W

Genetic analysis of children and families with heritable retinal dystrophies in Costa Rica: Identification of an X-linked mutation.
D.J. Wolff[1], I. Znoyko[1], W.B. Glen[1], J. Martinez Arguedas[2], R. Badilla-Porras[3], M.M. Peterseim[4]. 1) Pathology/Lab Med, Med Univ South Carolina, Charleston, SC; 2) Ophthalmology, National Children's Hospital, San José , Costa Rica; 3) Clinical Genetics and Metabolism, National Children's Hospital, San José , Costa Rica; 4) Ophthalmology, Med Univ South Carolina, Charleston, SC.

   The retinal dystrophies (RDs) are a group of diseases with phenotypic and genotypic heterogeneity with over 200 known disease-causing genes. Our study aimed to identify underlying genetic defects resulting in RD in the Costa Rica population that has known founder effects. Whole genome autozygosity mapping (Illumina 850K SNP microarray) was performed on 31 affected children from 23 families. Affected persons had a median 5% homozygosity with many known recessive RD genes mapping within homozygous regions. While no homozygous structural aberrations were detected involving autosomal genes, a 44 kb hemizygous deletion of the X chromosome (153,414,836-153,458,779) was detected in two brothers. Whole exome sequencing (Agilent SureSelect V6) confirmed that the deletion included exons 2 - 6 of OPN1LW and exons 1 - 4 of OPN1MW. The high level of homology between these genes predisposes to unequal recombination and deletion/hybrid gene formation. Loss of function of the OPN1LW (red) and OPN1MW (green) opsin photopigment genes results in blue cone monochromacy which is characterized by reduced vision, color vision impairment and variably, photophobia, myopia and nystagmus. The brothers had symptoms starting age 5 including nyctalopia but no nystagmus; both had subnormal vision (20/40-60; 20/150), some changes in retina appearance and myopia. The maternal grandfather was also reported to have a history of nyctalopia since childhood, retinal degeneration and severely impaired vision (20/200). Identification of this X-linked mutation shows the utility of our approach to identify recessive mutations in the Costa Rican population. Autozygosity mapping in conjunction with whole exome sequencing is ongoing for the 23 families to search for additional mutations with the ultimate goal of identifying pathogenic mutations to assist with better patient care and potentially identify individuals who may be candidates for vision-sparing therapies.

1112T

Unclassifiable brachydactyly (brachydactly E+A2) due to a novel missense mutation in *IHH*. *H. Numabe[1,2], K. Yura[3,4], T. Maeda[2], Y. Morishima[1], N. Inagaki[1], M. Wakai[1], S. Suzuki[2], H. Kawashima[1,2], G. Nishimura[5].* 1) Clinical Genetics Center, Tokyo Medical University, Shinjuku-ku, Tokyo, Japan; 2) Department of Pediatrics, Tokyo Medical University, Shinjuku-ku, Tokyo, Japan; 3) Faculty of Core Research, Ochanomizu University, Bunkyo-ku, Tokyo, Japan; 4) School of Advanced Science and Engineering, Waseda University, Shinjuku-ku, Tokyo, Japan; 5) Department of Radiology, Tokyo Metropolitan Children's Medical Center, Fuchu-shi, Tokyo, Japan.

;The proband was an 11-year-old girl who was referred because of short digits and toes. She showed mild facial dysmorphism, including hypertelorism and a high arched palate as well as overt brachydactyly. The physical findings were otherwise unnoticeable. Her mother was similarly affected. Based on the Japanese standardized MCPP (metacarpo-phalangeal pattern profile), the brachydactyly was characterized by extreme shortness of metacarpals of 4th and 5th fingers, proximal phalanges of 1st fingers, middle phalanges of 2nd and 5th fingers, and distal phalanges of 1st and 3rd fingers. The pattern was classified into the combination of type E and type A2 on Bell's classification. Whole exome analyses of her and her mother's genomes revealed a heterozygous missense variant (c.G1222A; p.G408R) in Indian Hedgehog gene (*IHH*). The variant was likely to be pathogenic.

1113F

A novel missense variant in *IRF6* gene implicated in causing Van der Woude syndrome. *N. AlDhaheri[1,4], N. Beck[1,2,4], E. Davidson[3,4], K. Seifert[3,4], R. Redett[3,4], J. Hoover-Fong[1,2,4].* 1) McKusick-Nathans Institute of Genetic Medicine; 2) Greenberg Center for Skeletal Dysplasias; 3) Department of Plastics and Reconstructive Surgery; 4) Johns Hopkins University School of Medicine, Baltimore, MD, United States.

Background: Orofacial clefts (OFC) are the most common craniofacial birth defect affecting ~ 1 in 700-1000 individuals worldwide. About 30% are syndromic and Van der Woude syndrome (VWS) is the most common monogenic OFC diagnosis accounting for ~ 2%. VWS is a dominant condition presenting with cleft lip and/or palate, bifid uvula, hypodontia and paramedian lower lip pits or a small cyst capping a sinus tract from a mucous gland. VWS has phenotypic variability ranging from an isolated lip pit to bilateral cleft lip and palate with multiple pits or cysts and hypodontia. Penetrance is estimated at 92%. Heterozygous loss of function mutations in the Interferon Regulatory Factor 6 gene (*IRF6*) are identified in~70% of VWS patients; over 300 *IRF6* mutations have been described. Gain of function mutations in *GRHL3* account for another 5% of VWS patients with remaining cases without genetic etiology. **Case Report:** We report a patient prenatally diagnosed with cleft lip and palate. On postnatal exam, she also had a lower lip pit with sinus tract and negative family history for cleft lip/palate and lip pits. Molecular testing of *IRF6* and *GRHL3* revealed a novel missense variant c.232A>G (p.Lys78Glu) in exon 4 of *IRF6,* reported to be of uncertain significance. This variant is not present in ExAc or ClinVar and Polyphen predicted it to be 'probably damaging'. Analysis for the variant in the patient's unaffected parents revealed it was *de* novo, further supporting its pathogenicity. No functional studies exist for this variant. 90% of the mutations in *IRF6* causing VWS or an allelic condition known as popliteal pterygium syndrome are missense or truncation mutations with the remaining 10% creating a new start codons or affecting splicing. The frequency of missense mutations was significantly increased in the DNA-binding and protein-binding domains, encoded by exons 3–4 and 7–9, respectively, and these domains were highly conserved. Furthermore, *IRF6* variants rarely occur in control subjects, implying *IRF6* does not tolerate mutation burden. **Conclusion:** The novel missense mutation in a highly conserved domain of *IRF6* identified in our patient with strong clinical features of VWS strongly supports the pathogenicity of this variant. This report should allow reclassification of this variant from that of unknown significances to a pathogenic variant.

1114W

Origin of *EIF4A3* pathogenic expansion, the causative mechanism of a craniofacial syndrome (RCPS). *G.S.P. Hsia, L. Alvizi, L.A. Brito, C.M. Musso, R.C.M. Pavanello, M. Zatz, M.R. Passos-Bueno.* Centro de Pesquisa sobre o Genoma Humano e células tronco, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil.

Richieri Costa-Pereira syndrome (RCPS) is an autosomal-recessive acrofacial dysostosis, characterized mainly by midline cleft mandible and Robin sequence. RCPS patients can also show intellectual disability and speech impairment. Mutations in *EIF4A3*, mainly expansion at its 5'UTR region (UCSC Genome Browser on Human Feb. 2009 - GRCh37/hg19 - Assembly, chr17:78,109,013-78,120,982), cause RCPS. Sequencing analysis of this region showed multiple allelic patterns, which varies in number and organization of motifs: TCGGCAGCGGCAGCGAGG (CA-18nt), TCGGCAGCGGCACAG-CGAGG (CACA-20nt) and TCGGCAGCGGCGCGCAGCGAGG (CGCA-20nt). RCPS patients present all 3 types of motifs within the same allele and a total of 14 to 16 motifs (CGCA-20nt is the most abundant). Unaffected individuals present lower number of motifs, however it is still unknown if they have CG-CA-20nt motifs and which is the most prevalent allele pattern. To get insights in the arisen of this pathogenic expansion, we characterized *EIF4A3* allele variability in 380 samples of individuals unrelated to RCPS families (control sample, CS) by Sanger Sequencing method. The results revealed that almost 50% of our sample presented 7 or 8 motifs. Most of the individuals (85%) are heterozygous and follow the Hardy-Weinberg equilibrium. We identified 45 different allelic patterns (APs), being the most common compound by 4 motifs CACA-20nt followed by 1 CA-18nt, 1 CACA-20nt and 1 final CA-18nt. Furthermore, 8 heterozygous controls presented CGCA-20nt motifs (9 to 10), which previously was only reported in RCPS patients. Next, we analyzed the haplotype of 16 samples (4 controls; 5 with CGCA-20nt motif; United Kingdom RCPS patient and 6 Brazilian RCPS patients with different expansion structures) using 5 SNPs markers (rs11150824, rs2289534, rs3829612, rs10782008 and rs12943620) flanking *EIF4A3*. The results revealed two haplotypes associated with CGCA-20nt motif. Based on these results, we hypothesized that the pathogenic expansions might have arisen from an unequal crossing-over event between two alleles. The detection of alleles containing the CGCA-20nt motif in the control population suggests that pathogenic alleles can arise independently, and this expanded allele may not be specific of the Brazilian population. It is possible, therefore, that this disease is misdiagnosed in other regions of the world. FAPESP, CNPq.

1115T

Craniosynostosis: Expanding the phenotype of 3 rare syndromes. *E.H. Zackai[1], E. Bhoj[1], J. Taylor[2], D.M. McDonald-McGinn[1], M.H. Harr[1], K. Grand[1], A. Santani[3], D. Li[3], C. McDougall[1].* 1) Division of Human Genetics; 2) Division of Plastic and Reconstructive Surgery; 3) Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA.

Craniosynostosis affects 1/2250 births. An underlying genetic cause is identified in ~24% of cases with mutations in 6 genes (*FGFR2, FGFR3, TWIST1, TCF12, ERF,* and *EFNB1*). At least 52 other genes have been associated with craniosynostosis. Of these, 20 core genes have craniosynostosis as a common feature (>50% of patients). The remaining genes have craniosynostosis in a minority (Miller et al., 2017, J Med Genet). Utilizing whole exome sequencing (WES) in craniofacial clinic, we wish to add 3 diseases (Xia-Gibbs syndrome, Ayme-Gripp syndrome and *KMT5B*-related syndrome) to the latter category and expand their phenotypes to include craniosynostosis. Patient 1 is a 14 year old female with bicoronal craniosynostosis, developmental delay (DD), intellectual disability (ID), atrial septal defect, short stature, lumbar lordosis, joint laxity, and patellar dislocations. WES identified a heterozygous *de novo* pathogenic mutation in *AHDC1* (c.2473C>T;p.Q825X) consistent with Xia-Gibbs syndrome characterized by failure to thrive, hypotonia, DD, moderate ID and dysmorphic features. Two other patients with Xia-Gibbs syndrome had craniosynostosis – one with sagittal and the other with bicoronal and metopic (Yang et al., 2015, Cold Spring Harb Mol Case Stud; Miller et al., 2017, J Med Genet). Patient 2 is a 20 year old male with sagittal and coronal craniosynostosis, bilateral mixed hearing loss, ID and dysmorphic features. WES found a mutation in *MAF* (c.161C>T;p.S54L) causing Ayme-Gripp syndrome characterized by ID, sensorineural hearing loss, brachycephaly, flat facies, short stature, seizures and congenital cataracts. Another patient with Ayme-Gripp syndrome and coronal craniosynostosis has the same *MAF* mutation (p.S54L) as our patient (Ayme & Philip, 1996, Clin Dysmorphol; Niceta et al., 2015, Am J Hum Genet). Patient 3 is a 4 year old female with moderate global DD, autism, and sagittal and bilateral lambdoid craniosynostosis. WES identified a heterozygous *de novo* pathogenic mutation in *KMT5B* (c.598T>C;p.C200R) causing a newly described syndrome characterized by DD, ID, autism and seizures (Stessman et al., 2017, Nature Genet). We know of one additional patient with craniosynostosis who has a mutation in *KMT5B* (Personal Communication). Through the use of WES in the craniofacial clinic we have expanded the phenotypes of 3 rare syndromes to include craniosynostosis. Our cohort illustrates the utility of WES in ascertaining the etiology of craniosynostosis.

1116F

Disorders of sexual development in genetics pediatrics. Three different ambiguous genitalia cases report from Hospital Para El Nino Poblano, México. *J. Aparicio[1,2], F. Cuellar-López[3], ML. Hurtado-Hernández[4], A. Atriano Ponce[5], H. Marodan Garcia[5], S. Chatelain-Mercado[6].* 1) Genetics, Benemerita Universidad Autonoma de Puebla; 2) Genetics; 3) Urology and Pediatric Surgery; 4) CitogenÉtics; 5) Endocrinology, Hospital para el NiÑo Poblano; 6) Biotechnology, Universidad AutÓnoma Metropolitana, MÉxico.

Five pediatric patients with three different disorders of sexual development are reported in this study; the first three male patients (16 years, 4 years and 2 months old, respectively) were diagnosed as having diphallia. These 3 patients had real diphallia, well developed penises, urinarious meatus, and both testicles and one of the case, vessel duplication was reported by urology. All the patients have normal cytogenetic analysis, 46XY. The fourth patient was 2 years old, with hyperplasic clitoris, hyperpigmented tissue similar to labia major (large lips) and internal female organs identified as vagina, uterus and both ovaries. A chimera with two different cells lines [46,XX (48%) and 46,XY (52%)] by cytogenetic studies was reported. And the last child was 2 years 8 months old patient with chromosome translocation, between chromosome Y and 7 chromosomes t(7;Y). Hypospadias peneescrotal, unilateral cryptorchidism, urinary meatus stenosis and malformed scrotum were diagnosed together with vessel duplication. Key words: Ambiguous genitalia, diphallia, hypospadias, cryptorchidism, chromosome translocation and chimera.

1117W

Microdeletion of 17q21.31 causes a novel malformation syndrome. *K. Kurosawa, M. Minatogawa, T. Yokoi, Y. Enomoto, K. Ida, N. Harada, J. Nagai, Y. Tsurusaki.* Kanagawa Children's Med Ctr, Yokohama, Japan.

17q21.31 deletion syndrome, also known as Koolen–De Vries syndrome or KANSL1-related disorder, was a genomic disorder characterized by distinctive facial appearance, intellectual disability, hypotonia, and characteristic behavior. It is usually caused by a recurrent 0.450–0.600 Mb deletion on chromosome 17q21.31, encompassing KANSL1.To date, excluding the KANSL1-related disorder, no syndromic genomic rearrangement was known in the 17q21.31. Here, we report a patient with intellectual disability, dysmorphic facial appearance, hydrocephalus, congenital heart defect, and epilepsy, associated with de novo 17q21.31 deletion, proximal to the critical region of KANSL1-related disorder. The patient was an 8-year-old boy, referred for multiple congenital malformations and severe developmental delay. He was born to non-consanguineous healthy parents at 38 weeks of gestation after uneventful pregnancy. His birth weight was 2450 g (-1.3 SD), length 47 cm (-0.9 SD), and OFC 30.5 cm (-2.0 SD). Developmental milestone was severely delayed, rolling over at 9 months, sitting and crawling at 1 year and 5 months, walking alone at 3 years. At age of 8 years, he spoke no meaning words. His weight was 19.0 kg (-1.5 SD), height 113.6 cm (-2.4 SD), and OFC 51.4 cm (0 SD). He exhibited distinct facial appearance including long face, ptosis, malformed ears, strabismus, and high arched palate. Microarray revealed 0.2 Mb deletion on 17q21.31 proximal to the large cluster of low copy repeats. The deletion interval encompassed 8 RefSeq genes including NAGS, TMEM101, LSM12, G6PC3, HDAC5, C17orf53, ASB16, and TMUB2. In the deleted genes, HDAC5, encoding histone deacetylase 5, is the only gene with a high score of pLI (1.0). HAD5 plays an important role in transcriptional regulation, cell cycle progression and developmental events, which is likely to be consistent with the clinical phenotype of our patient. According to DECIPHER database, two individuals with ptosis and intellectual disability common to our patient were registered, but detailed clinical information was unavailable. These results suggested that haploinsufficiency of HDAC5 might cause distinct phenotype. Additional cases with loss of function mutation in HDAC5 are required for establishing a novel malformation syndrome.

1118T

Phenotypic and genotypic spectrum in Richieri-Costa-Pereira syndrome.
D. Bertola[1,2], G. Hsia[2], L. Alvizi[2], A. Gardham[3], G.L. Yamamoto[1,2], R.H. Honjo[1], L.A.N. Oliveira[1], R.C. Di Francesco[4], B.A. Perez[5], C.A. Kim[1], M.R. Passos-Bueno[2]. 1) Pediatrics, Instituto da Crianca, Sao Paulo, Brazil; 2) Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, Instituto Biociências – Universidade de São Paulo, SP, Brazil; 3) North East Thames Genetics Service, Great Ormond Street Hospital, London, England; 4) Departamento de Otorrinolaringologia – Hospital das Clínicas da FMUSP – Universidade de São Paulo, SP, Brazil; 5) Departamento de Genética – Universidade Federal de São Paulo, SP, Brazil.

Richieri-Costa-Pereira syndrome (RCPS) is a rare autosomal recessive acrofacial dysosotosis described mainly in Brazilian individuals. Cardinal features of this syndrome include Robin sequence, cleft mandible, laryngeal anomalies and limb defects. An expansion in both alleles of a complex repeated motif composed of 18 or 20 nucleotides in the 5' untranslated region of *EIF4A3* have been shown to cause this syndrome, commonly with 15 or 16 repeats. The only patient with mild clinical findings harbored a 14 repeat-expansion in one allele and a point mutation in the other allele. This proband is described here in more details, along with his affected sister, as well as 4 new individuals with RCPS, including a patient from England, of African ancestry. The clinical findings ranged from a severe infant, requiring tracheostomy and gastrostomy due to striking mandible compromise, to individuals showing mild facial dysmorphisms, dental and skeletal anomalies, and, sometimes, asymptomatic structural laryngeal anomalies. Although the most frequent mutation in the present study was the recurrent 16-repeat expansion in *EIF4A3*, there was an overrepresentation of the 14-repeat expansion. Even though the number of patients reported is small, the ones that showed mild clinical findings presented the shorter repeat-expansion in *EIF4A3*, suggesting that the number of these expansions could play a role in phenotypic delineation. It is possible that this syndrome has been underdiagnosed, especially in the cases in which the individuals do not present the full-blown phenotype. Moreover, the analysis of sequencing techniques, including next-generation sequencing, does not routinely detect expansions in the 5' untranslated region of a gene. In these mild cases, a clinical suspicion of RCPS should be taken in individuals presenting microretrognathia or Pierre-Robin sequence and an active search for subtle dental, acral and laryngeal anomalies should be performed. In order to confirm the clinical hypothesis, a targeted genetic analysis is required. FAPESP (2013/08028-1; 2015/21783-9), CNPq (302605/2013-4).

1119F

Combination of *UBR1* and *UBR5* mutations in a severe form of Johanson-Blizzard Syndrome with total agenesis of lateral nasal process and situs inversus.
A. Cetinkaya[1], G. Akay Tayfun[2], B. Erguner[3], A. Tutun[4], MB. Mutlu[1], N. Karadag[4], G. Karatekin[4], A. Karaman[1], NA. Akarsu[5]. 1) Medical Genetics, Zeynep Kamil Women and Children's Hospital, Uskudar, Istanbul, Turkey; 2) Pediatric Genetics, Zeynep Kamil Women and Children's Hospital, Uskudar, Istanbul, Turkey; 3) Advanced Genomics and Bioinformatics Research Center (IGBAM), The Scientific and Technological Research Council of Turkey (TUBITAK-BILGEM), Kocaeli, Turkey; 4) Neonatal Intensive Care Unit, Zeynep Kamil Women and Children's Hospital, İstanbul, Turkey; 5) Department of Medical Genetics, Gene Mapping Laboratory, Faculty of Medicine, Hacettepe University, Ankara, Turkey.

Johanson-Blizzard Syndrome (MIM 243800) is a rare autosomal recessive disorder characterized by hypoplasia/aplasia of alae nasi and impairment of exocrine pancreatic function. Oblique facial clefts have hitherto been defined as the most severe facial phenotype in this syndrome. Here, we present a newborn with total agenesis of facial structures originating from the embryonic lateral nasal process — resulting in bilateral oblique facial clefts which include medial portion of the lower eyelids, lateral nasal walls and nasolacrimal ducts. Additionally left atrial isomerism, laterality defects involving liver and stomach; and bilateral dysplastic kidneys have been observed. The newborn, whose parents were first degree relatives,died hours after birth due to severe congenital defects. A previously undefined homozygous mutation c.2380-1G>C in *UBR1* was identified by exome sequencing. This is predicted to disrupt the consensus splice acceptor site in intron 21, for which another pathogenic point mutation, c.2380-1G>A, has previously been reported [Sukalo M., et al. Hum Mutat. (2014) 35: 521-531]. The exceptionally severe phenotype reported here prompted us to look for another mutation in a possible modifier gene in the UBR1 interactome. We utilized the exome sequencing data and STRING protein assciation networks to identify a heterozygous missense variant, c.2323A>G p.(Thr775Ala), in *UBR5* which is conserved during evolution and is "disease causing" according to Mutation Taster and "possibly damaging" according to Polyphen-2. The variant was not present in large population databases, 1000 Genomes Project, ESP, ExAC and IGBAM in-house database (n=1182 individuals). *UBR5*, like *UBR1*, encodes for a protein that functions in the N-end rule pathway, which directs proteins to ubiquitinylation and subsequent protein degradation by recognizing the amino acid residues and modifications in the N-terminal of polypeptides. The high expression of UBR5 in the lateral wall of the snout during the embryonic development of mice may hint a cooperative action of UBR1 and UBR5 during embryonic development [Kinsella E., et al. PLoS One. (2016) 11: e0157079]. The conserved UBR5 variant could be the first example to highlight the importance of modifier genes of N-end rule pathway in determining the wide phenotypic spectrum of individuals with Johanson-Blizzard Syndrome.

1120W

SHOX duplication in a Kabuki syndrome patient: A possible effect on clinical phenotype. *A.P. Marques-de-Faria[1], J.R.M. Prota[1], C.R. Lincoln-de-Carvalho[1], B.S. Carvalho[2], I. Lopes-Cendes[1].* 1) Department of Medical Genetics – Faculty of Medical Sciences- University of Campinas (UNICAMP); 2) Department of Statistics - Institute of Mathematics, Statistics and Scientific Computing - University of Campinas (UNICAMP).

   Kabuki syndrome (KS) is a multiple malformation syndrome usually caused by pathogenic variants in genes *KMT2D* or *KDM6A*. It is characterized by a typical facial gestalt, developmental delay and growth deficiency. The diagnostic suspicion is usually due to cardinal facial features such as elongated palpebral fissures with eversion of distal third of the lower eyelid, arched eyebrows, short columella with depressed nasal tip and prominent ears. After the first description, the Kabuki phenotype has been expanding as well as the mutational findings. In most cases the pathogenic variants are private and "de novo" consistent with sporadic occurrence. In general, the hypothesis of KS is unequivocal due to its peculiar facial signs despite the wide clinical spectrum. Recently, the expanding phenotype included microphthalmia and overlapped the CHARGE syndrome, which should be considered as a differential diagnosis. We describe a female child of healthy and non-consanguineous parents, birth weight 2935g, length 48.5 cm, with generalized hypotonia, microcephaly, unilateral microphtalmia, up-slanted and elongated palpebral fissures, laterally sparse eyebrows, anteverted nostrils, downturned corners of the lips, prominent ears, deep palmar creases and prominent finger pads. Patient evolved with adequate body weight and stature, neurodevelopmental delay, dysphagia and constipation. At first dysmorphological exam, KS was considered, despite the normal stature. However, in view of differential diagnosis, SNP-array was performed and detected Xq22.33 duplication, including only SHOX gene. As the latter is usually associated with tall stature and does not justify the Kabuki phenotype, an exome sequencing was performed and revealed a missense variant in *KMT2D,* previously registered as disease causing mutation, interpreted as pathogenic and also related with microphtalmia, confirming the initial hypothesis. The normal stature, which is uncommon in KS, although described in other cases, could be justified by SHOX duplication. Structural variants modulating a specific clinical sign of a monogenic syndrome are unusual and could represent an additional factor to consider in face of an atypical clinical phenotype. It is expected that after the increased use of exome sequencing and other diagnostic application of next generation sequencing associated with chromosomal microarray, these complementary genetic findings should be disclosure more frequently. .

1121T

A novel mutation in PDE3A gene in a 7-year-old female patient with dysmorphic features, developmental delay, short stature, and unilateral brachydactyly without high blood pressure. *A. Alali, J. Greally.* Pediatrics-Genetics, Montefiore Medical Center, Bronx, NY.

   Background: The hypertension and brachydactyly syndrome, HTNB; OMIM 112410, is an autosomal dominant syndrome characterized by brachydactyly type E and severe salt-independent but age-dependent hypertension resulting from gain of function mutation in PDE3A gene. **Case report:** A 7 year-old female child first seen at the age of 5 years for dysmorphic features and developmental delay.Prenatal course was complicated with late poor weight gain (IUGR) of unclear etiology. Our patient was born at 36 weeks gestation to a 35 years old G3P2002 mother by caesarean section due to a concerning fetal heart tracing but with no further complications. Birth weight was 2085 grams. The family history was significant for a maternal aunt who had isolated brachydactyly.Our patient was diagnosed with congenital hypothyroidism and started on synthroid. She walked at the age of 12 months and had her first word also at the age 12 months. She spoke in full sentences at age of two but had fair acdemic performancein a special education class.Her length was consistently less than 3rd percentile .Her physical exam was significant for missing distal digital of first and second fingers and third finger brachydactyly, blue sclera bilaterally, low-set anteriorly rotated ears with her left ear more prominent than her right ear, wide mouth , thin upper lip ,short moderately smooth philtrum, high nasal bridge , and bulbous nasal tip.The patient had surgeries for craniosynostosis at age of 2 years and hiatal hernia repair at age of 1 year.Her primary genetic work up including chromosomal analysis and array comparative genomic hybridization (CGH) was normal.On WES, she was heterozygous for the D359N variant of uncertain significance in *PDE3A* gene (c.1075 G>A) inherited from her mother who has no brachydactyly or high blood pressure. This variant is in exon 3 of the *PDE3A* gene, while all the six known mutation associated with hypertension and brachydactyly syndrome (HTNB) reported previously were in exon 4. A family history and blood pressure readings overtime were elucidated from PCP with most readings wre within the range 95/65 mmHg to 92/52 mmHg. Her weight and stature are consistently less than 3rd per centile for age and gender. **Conclusion:** HTNB syndrome although rare, should be considered in any patient presenting with high blood pressure, especially at young age with brachydactyly and short stature .FH is a key point in the light of its different expressivity.

1122F

A unique contiguous gene syndrome encompassing Potocki-Shaeffer syndrome and aniridia. *G. Delplancq[1], M.-A. Boukebir[2], D. Amsallem[3], P. Kuentz[4], L. Thines[2], J.-L Bresson[4], L. Van Maldergem[1].* 1) Department of Human Genetics, Besançon University Hospital, Franche-Comté University, France; 2) Department of Neurosurgery, Besançon University Hospital, Franche-Comté University, France; 3) Department of Pediatric Neurology , Besançon University Hospital Franche-Comté University, France; 4) Department of Cytogenetics, Besançon University Hospital Franche-Comté University, France.

Introduction: Potocki-Schaffer is a contiguous gene syndrome associated with deletions in the 11p11-p12 region. It combines *EXT2*-related multiple exostoses with *ALX4*-parietal foramina. Delayed milestones or intellectual deficiency, craniosynostosis and facial dysmorphia have also been described. A single patient with Potocki-Shaeffer syndrome plus *PAX6*-related aniridia has been described so far. We describe a second patient which confirms this specific contiguous syndrome and characterize the corresponding 30 Mb microdeletion alongside with expansion of the phenotype that encompasses both proximal p11 and *PAX6*.   Materials and methods: A 9 month-old girl born prematured at 32 week to Algerian non consanginous parents was referred to our genetics department by the neurosurgeon because of a enlarged parietal foramina. Her parents aged 49 and 45 at the time of conception were healthy. She had bilateral aniridia and microphtalmia, a thin corpus callosum and delayed milestones. A brain CT-scan indicated a large pulsatile hole of 6 cms on both parietal bones. On the basis of the association of parietalis foramina and aniridia, a diagnosis of probable 11p11-13 microdeletion syndrome was made. Karyotype on lymphocytes confirmed segmental monosomy: 46,XXdel(11) (p11.2). Array-CGH indicated a 11p11.2 – 11p15.1: 18400975-48014598 30Mb deletion [hg19] arising *de novo*.   Conclusion: The current observation confirms combined Potocki-Shaeffer-aniridia as a viable and specific new contiguous gene syndrome amenable at diagnosis on clinical grounds.

1123W

Novel pathogenic variants of *EP300* in two Japanese patients with Rubinstein-Taybi syndrome type 2. *T. Kaname[1], K. Yanagi[1], M. Iso[1], Y. Kuroki[1], M. Minatogawa[1], T. Tohma[2], Y. Matsubara[3].* 1) Dept Gen Med, Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Tokyo, Japan; 2) Wanpaku Clinic, Okinawa, Japan; 3) National Center for Child Health and Development,Tokyo, Japan.

Background: Rubinstein-Taybi syndrome (RSTS) is a rare autosomal dominant congenital disorder characterized by distinctive facial features, broad thumbs, postnatal growth deficiency and mental retardation. Patients bearing a mutation in CREBBP are diagnosed as RSTS 1 (OMIM #180849), who shows classical feature of RSTS. CREBBP is major cause of RSTS. *EP300* is another causative gene for RSTS (RSTS2; OMIM #180849). Mutation in *EP300* is minority in the patient of RSTS. Though clinical feature of RSTS2 seems milder phenotype than that of RSTS1, it has been unclear. We present two cases of RSTS2 with mild phenotype and severe mental retardation. **Patients:** First patient is 2-yr old Japanese girl. Second patient is 8-yr old Japanese boy. Both patients have severe mental retardation, microcephaly, distinctive facial features with grimacing smile, broad thumbs and halluces. Epilepsy and ataxic gait are not observed in both patients. Postnatal growth deficiency is not observed in the first patient but second patient (-3SD). Clinical features are mild compare to typical RSTS. **Results of genetic testing:** Genomic DNA obtained from blood was examined by exome sequencing using a next generation sequencer. First patient has an A to G transition at splicing acceptor site of intron 25 of *EP300*. RT-PCR analysis showed the transition provoked non-canonical splicing in peripheral blood monocyte cells. Second patient has a missense variant in the exon 27 of *EP300*. SIFT and Polyphen predicted as disease causing. Both variants were not found in public database, dbSNP, 1000 Genome, ExAC, HGVD, integrative Japanese Genome Variation Database (https://ijgvd.megabank.tohoku.ac.jp) , and in-house exome data. **Conclusion:** We found pathogenic variants in the *EP300* gene in the patients. They have mild phenotype, especially mild facial features, and have a grimacing smile. According to reports of patients with *EP300* variants, RSTS 2 has milder phenotype than RSTS 1 and grimacing smile might be one of the features of RSTS 2.

1124T

Identification of *AFF4* missense mutation in a girl with Cornelia de Lange syndrome (CdLS) like phenotype and obesity. *E. Nishi[1], N. Miyake[2], N. Matsumoto[2], N. Okamoto[1].* 1) Department of Medical Genetics, Osaka Women's and Children's Hospital, Osaka, Japan; 2) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

CHOPS syndrome (MIM;616368) is a multiple congenital anomaly syndrome, phenotypically overlaps Cornelia de Lange syndrome (CdLS), characeraized with cognitive impairment, coarse facies, heart defects, obesity, pulmonary involvement, short stature, and skeletal dysplasia, caused by heterozygous mutation in *AFF4* (MIM;604417) on chromosome 5q31 [Izumi,et al., 2015]. *AFF4* encodes a scaffold protein that functions as a core component of the super elongation complex (SEC), which is involved in transcriptional regulation during embryogenesis. In 3 unrelated children with CHOPS syndrome, 3 different *de novo* heterozygous missense mutations in *AFF4* (T254A,T254S, and R258W), all of which affected conserved residues in the ALF homology domain were identified [Izumi,et al., 2015]. Here we report a 7 years old girl with delayed psychomotor development with profound intellectual disability, short stature, obesity, spinal lipoma, and somewhat variable dysmorphic facial features, including round face, hypertelorism, thick eyebrows and hair, long eyelashes, short nose, and downturned corners of the mouth ,clinically diagnosed Cornelia de Lange syndrome, who was found to have de novo heterozygous missense mutations in *AFF4*, R258W, by whole exome sequencing. Our case represents the fourth reported case of CHOPS syndrome due to *AFF4* mutations, and provides further evidence to support the notion that *AFF4* mutation cause CHOPS syndrome. Common clinical features among the patient described herein and the 3 patients previously described by Izumi et al. include IUGR, profound developmental delay/intellectual disability, short stature, obesity, congenital heart disease, brachydactyly, and somewhat variable dysmorphic facial features, including round face, hypertelorism, thick eyebrows and hair, long eyelashes, short nose, and downturned corners of the mouth. In our patient, hearing loss, cataract, and lipoma which was reported one by one in patients with CHOPS syndrome was found. So far, she have not showed the chronic lung disease which was found in the all three patients with CHOPS syndrome. This report would be important to further clarify the clinical picture of CHOPS syndrome.

1125F

A novel genetic syndrome with *RAB11B* mutation. *N. Okamoto[1], F. Miya[2,3], T. Tsunoda[2,3], M. Kato[4], S. Saitoh[5], M. Yamasaki[6], Y. Kanemura[7,8], K. Kosaki[9].* 1) Dept Medical Genetics, Osaka Women's and Children's Hospital, Izumi, Osaka, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University; 3) Laboratory for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN; 4) Department of Pediatrics, Showa University School of Medicine; 5) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences; 6) Department of Pediatric Neurosurgery, Takatsuki General Hospital; 7) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization; 8) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization; 9) Center for Medical Genetics, Keio University School of Medicine.

[INTRODUCTION] Rab (Ras-related in brain) proteins are included in the Ras superfamily of small GTPases. The Rab family proteins regulate variable intracellular membrane trafficking processes. They have critical roles in regulating exocytotic and endocytotic pathways. The isoforms Rab11a, Rab11b, and Rab11c/Rab25 constitute the Rab11 subfamily. RAB11B has 4 conserved domains important for GTP binding. Rab11b is highly expressed in brain, heart and testes. We report a boy with a novel pathogenic variant in *RAB11B*. **[CLINICAL REPORT]** The 6-year-old male was the first child of healthy and non-consanguineous Japanese parents. He was noted to have laryngomalacia after birth. He was hypotonic and his developmental milestones were markedly delayed. Brain MRI revealed hypoplasia of corpus callosum and cortical dysplasia of the frontal lobe. Physical examination identified microcephaly (-2.5SD) and dysmorphic features including epicanthal folds, flat nasal bridge, short philtrum and, high arched palate. His testes and penis were small. Endocrinological studies revealed hypogonadotropic hypogonadism. His fingers were short and slender. He showed spasticity of lower extremities. Ophthalmological investigation revealed strabismus and bilateral optic nerve atrophy. He showed severe intellectual disability and spoke a few meaningful words. His physical growth was within normal range. **[METHOD]**With the approval of our institutional ethics committee, the samples were analyzed using WES. Precise methods have been reported in another article (Okamoto et al. J Hum Genet 2014). **[RESULTS]** The patient had a de novo variant in exon2 of the *RAB11B* (NM_004218: c.G64A:p.V22M). This variant is predicted to be pathogenic by in silico analyses. It was not found among public databases. **[DISCUSSION]** There have been no reports on the association of the*RAB11B*variant and genetic disorders. We suppose that abnormality of *RAB11B* causes a novel neurogenetic syndrome with characteristic facial features, cortical dysplasia, optic atrophy and hypogonadism.

1126W

A case report of novel mutation in NSD1 gene, which causes Sotos Syndrome. *J. Prieto[1,2], P. Sanchez[2], A. Paredes[1].* 1) Inst de Gen Humana, Univ Javeriana, Bogota D.C, Colombia; 2) Hospital La Victoria E.S.E, Bogotá D.C, Colombia.

   Introduction: Sotos syndrome is characteristically identify as an autosomal dominant disorder of early overgrowth, caused by mutations in the NSD1 gene, located at chromosome 5q35.3, which encodes nuclear receptor-binding SET domain-containing protein 1, responsible of 90% cases. Patients presents length at or greater than 97th percentile usually during the first four years, associate with macrocephaly, dolichocephaly, bossing forehead, Malar flushing, prognathism, appearance of hypertelorism due to bitemporal narrowing, variable intellectual disability, phenotype changes in adulthood. The diagnosis is made by the identification of heterozygous NSD1 pathogenic variant. **Case presentation:** Patient of 8 year old, who was remitted to our service with diagnosis of high height since 1-year-old, bone age X-rays bone age 6, 5 years and chronological age 5 year 4 months. At physical examination is evident macrocephaly, frontal bossing, hypertelorism, prognathism, thorax pectus excavatum; extension studies reports progesterone: <0,03nmol/l, somatomedin C: 35,0ng/ml, growth hormone: 0,495ng/ml, ACTH: 11,4pg/ml, prolactin: 6.98ng/mmol, TSH: 1,23uUI/ml, T4L: 1,61ng/dl, estradiol: <5pg/ml, cortisol (am): 9,75ug/dl, Fasting glucose: 78,9mg/dl, RNM evaluates Turkish chair as normal, IQ 86. These findings suggest Sotos syndrome. Molecular study confirms the diagnosis, in the sequence analysis of the NSD1 gene, exons 2-23, found a heterozygous for the c.6437dupG, variant of uncertain significance, but resulting from a **Frameshift** mutation. **Discussion:** Sotos syndrome occurs in 1:14000 live births. The diagnosis is establish by the identification of a heterozygous pathogenic variant in NSD1 gene, the sequence analysis identify until 93% patients. NSD1 gene is located at 5q35.3 chromosome, encodes a histone methytransferase, transcriptional intermediary factor. Mutations of this gene have been associated with Sotos and Weaver syndromes. At the moment more than 100 pathogenic variants have been published. The novel variant c.6437dupG identify in this patient, haven't been reported in the databases and was never found in ExAC nor 1000G. Therefore we propose a novel mutation, frameshift that causes a strongly truncated protein, result in a clear pathologic phenotype. **Conclusions:** Is the firts time this varint is repored in a patient with the phenotype of Sotos syndrome. Funtional studies will be necessary in future to confirm pathogenicity of this mutation.

1127T

***SATB2*-associated syndrome: Differential diagnosis and genotype-phenotype correlations by detailed facial dysmorphism analysis using facial recognition technology in 38 individuals.** *Y. Zarate.* Arkansas Children's Hospital, Little Rock, AR.

   SATB2-associated syndrome (SAS, Glass syndrome) is a multi-system disorder characterized by significant neurodevelopmental compromise and craniofacial anomalies. Despite the reported high frequency of subtle dysmorphic features, they are not thought to be distinctive enough to allow for a clinical diagnosis of SAS. To determine how recognizable this syndrome could be based on facial dysmorphism alone and to further delineate commonly shared dysmorphic features among affected individuals, an objective facial analysis using automated facial recognition technology was conducted on a large population of individuals diagnosed with SAS. 2D facial photographs were analyzed using Face2Gene (FDNA Inc; Boston, MA) technology to automatically detect and evaluate subtle craniofacial dysmorphology and compared to a previously stablished SAS mask generated from prior published cases. A total of 58 2D facial photographs of sufficient quality from 38 individuals (28 previously unreported) were analyzed using the facial recognition technology. Based on facial dysmorphism analysis alone, from a list of possible diagnoses, SAS was suggested in at least the at least the top 20 for 21 individuals (55%), at least the top 5 for 7 individuals (18%), and as the top choice in 2 individuals (5%). Some common dysmorphic features automatically detected by the facial technology analysis included anteverted nares, broad nasal tip, low nasal bridge, downslanted palpebral fissures, epicanthus, frontal bossing, hypertelorism, and long philtrum. When multiple photographs of the same individual at different ages were analyzed, the possibility of SAS to be suggested at least within the top 20 of possible diagnoses was higher (11/14=79%) compared to those with a single photograph analyzed (10/24=42%). Similarly, individuals with intragenic mutations were more likely to have SAS suggested at least within the top 20 list (15/25=60%) compared to those with large deletions (3/11=27%). Besides SAS, the next 2 most commonly suggested diagnoses in the differential by the facial photograph analysis were Cerebrocostomandibular syndrome and Mucolipidosis type IV. In summary, this study provides evidence that SAS does have some common dysmorphic features that allow facial recognition technology to suggest this diagnosis within the top 20 of possible diagnoses in over half of individuals. Individuals with point mutations appear to have more recognizable features compared to those with large deletions.

1128F

Noonan-like syndrome with loose anagen hair: Expanding the phenotype. *A. Belonis, D. Weaver.* Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN.

Noonan-like syndrome with loose anagen hair (NS/LAH) is a condition characterized by Noonan-like facies, congenital heart defects, darkly-pigmented skin, brittle nails, short stature, intellectual disability and sparse, thin, slow-growing hair that is easily plucked. The clinical phenotype is caused by a missense mutation in the *SHOC2* gene, c.4A>G, p.Ser2Gly, which is the only mutation reported to cause this condition. The patient is now a 4-year-old female who presented with a number of vascular abnormalities, including skin, hepatic, and extrahepatic hemangiomas, as well as vascular changes of the subdural membrane. These findings appear to expand the phenotype of this condition. At the time of the initial presentation, her genotype was not known, but it was later determined on whole exome sequencing: It identified a c.4A>G mutation in the *SHOC2* gene. She has a phenotype consistent NS/LAH, although she also has several features that have not been previously reported. An abdominal CT scan and an abdominal MRI both demonstrated findings consistent with multiple hepatic hemangiomas, hepatomegaly, and an extrahepatic, extra-adrenal mass suggestive of a hemangioma. At 6 months of age, she was hospitalized for increased intracranial pressure, with findings of papilledema and a bulging anterior fontanel. She required a decompression, followed by placement of an extraventricular drain (EVD), and later had a ventriculoperitoneal (VP) shunt placement. During her EVD placement, her visualized subdural membrane was described as thick, white, and riddled with distended veins. During VP shunt placement, an increased vascular pattern was noted on the surface of her liver. Additionally, the burr hole used for the initial decompression has not yet closed. There have been two patients with *SHOC2* c.4A>G mutation reported with hemangiomas. There is another reported patient with a familial *PTPN11* mutation c.1226G>C, p.Gly409Ala as well as the NS/LAH mutation in *SHOC2* (c.4A>G), who had characteristic features of NS/LAH and a hepatic hemangioma; the authors proposed that the atypical phenotype in this patient was the result of an additive effect, with *PTPN11* c.1226G>C mutation acting as a modifier. Our patient's hepatic hemangioma was initially diffuse and, with treatment, has decreased in size on follow-up imaging. Additionally, our patient's abnormal subdural membrane venous finding appears to be another intracranial vascular anomaly in this condition.

1129W

A Japanese sporadic case of Adams-Oliver syndrome with a novel missense variant in *DLL4*. *M. Nagasaka[1,2], M. Taniguchi-Ikeda[2,3], H. Inagaki[4], U. Ouchi[4], M. Morioka[2], T. Toda[5], H. Kurahashi[4], K. Iijima[2].* 1) Takatsuki General Hospital, Osaka, Japan; 2) Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan; 3) Division of Genetic Counseling, Kobe University Hospital, Kobe, Japan; 4) Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, Japan; 5) Department of Neurology/Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe, Japan.

Adams-Oliver syndrome (AOS, OMIM; 100300) is a rare genetic disease. Various clinical manifestations have been reported, including aplasia cutis congenita of the scalp, terminal transverse limb defects, and cutis marmorata with vascular anomalies such as congenital heart defects. The etiology of AOS has remained largely unknown but defective Notch signaling during vascular formation has been suggested. Here we describe a sporadic Japanese newborn case with clinically diagnosed AOS. The proband was born at 34 weeks, 4 days gestation with a birth weight of 1334 g (-2.7 SD), height of 39.6 cm (-2.2 SD), and occipitofrontal circumference of 26 cm (-2.5 SD). During pregnancy, fetal growth restriction and single umbilical cord artery were present. The patient was transferred to the neonatal intensive care unit at Kobe University Hospital soon after birth, presenting with bleeding around aplasia cutis congenita of the skull. Limb abnormalities, such as bilateral syndactyly of the lower extremities and brachydactyly, and cutis marmorata were also observed. Clinical presentation suggested a diagnosis of typical AOS. Trio whole-exome sequencing identified a *de novo*, novel, heterozygous missense variant in the *Delta-like 4 ligand* gene (*DLL4* c.572G>A, p.Arg191His) in the patient. *DLL4* functions as a requisite ligand for NOTCH1 receptor, which is essential for vascular formation. *DLL4* has recently been identified as a causative gene of an autosomal dominant type of AOS. Amino acid substitution of Arg 191 to His was predicted by molecular models to interfere with direct binding between DLL4 and NOTCH1. The *in silico* variant analysis tools, SIFT and PolyPhen-2 produced scores of 0 (deleterious) and 1.0 (damaging), respectively. By modeling the DLL4 structure and the R191H variant using MODELLER9v8, the side chain of histidine was shown to be shorter compared to that of arginine. Therefore, the Arg191His variant might lead to weakening of the electrostatic interaction between DLL4 and the receptor and it may have reduced access to receptor binding site residues, resulting in the AOS phenotype. Variants in *DLL4* have been recently identified in several families, with these patients showing relatively mild phenotypes compared to autosomal recessive cases. The case described here showed gradual recovery from skull defects after birth, suggesting that mutation in *DLL4* has less effect after birth, leading to the milder phenotype of the case.

1130T

Redefining phenotypic spectrum of constitutional *CDK13* mutations: Three patients without cardiac defects. *T. Uehara[1], H. Suzuki[1], Y. Yamaguchi[1], T. Takenouchi[1,2], K. Kosaki[1], K. Kurosawa[3], S. Mizuno[4].* 1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 2) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan; 3) Division of Medical Genetics, Kanagawa Children's Medical Center, Kanagawa, Japan; 4) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Aichi, Japan.

   Precise regulation of the cell cycle is a prerequisite for human development and growth. Cyclins and cyclin-kinases are key regulators of cell cycle and constitutional mutations in those molecules lead to developmental disorders. Recently, seven patients with de novo constitutional missense mutations in the *CDK13* gene from the cyclin-dependent kinase family were identified through trio exome analysis of a large cohort of 610 patients with congenital cardiac disease accompanied with intellectual disabilities and other extra-cardiac features. Here we report 3 patients with de novo constitutional missense mutations in the *CDK13*. These 3 patients had intellectual disabilities but did not have congenital heart diseases. Exome analysis of the patients and their parents revealed the following: Patient 1 was a 6-year-old male who had de novo heterozygous CDK13 p.G717Q mutation. He showed severe intellectual disability, prenatal and postnatal growth retardation, craniosynostosis, long eyelashes, cupped ears, teeth anomalies, and strabismus. Patient 2 was a 5-year-old male who had de novo heterozygous CDK13 p.G717Q mutation. He showed intellectual disability, cleft palate, dark skin, hypertrichosis, long and wide nose, posteriorly rotated ears, and hypertelorism. Patient 3 was a 1-year-old female who had de novo heterozygous CDK13 p.A842S mutation. She showed intellectual disability, craniosynostosis, hearing impairment, down-slanting palpebral fissures, flat nose, low-set ears, and small chin. We here redefined the phenotypic spectrum of the CDK13 constitutional defect. Cardiac defect is not an essential feature of the syndromic intellectual disability syndrome associated with constitutional *CDK13* mutations. Long or busy eyebrow and earlobe anomaly, and craniosynostosis may represent a diagnostic clue to this newly identified syndrome.

1131F

***IRF6* mutational screening in a sample of Mexican patients with non syndromic cleft lip palate.** *J. Velazquez, B. Estandia-Ortega, M.A. Alcantara-Ortigoza, A. Gonzalez-del Angel.* Genetics, Instituto Nacional de Pediatria, Mexico, Distrito Federal, Mexico.

   Van der Woude Syndrome (VWS, MIM#119300) is caused by pathogenic variants (PV) on *IRF6* gene (~70% VWS cases). Clinical presentation includes cleft lip/palate and lower lip pits, but lower lip pits can be unnoticed or absent at clinical examination, which in some cases difficult its distinction from non-syndromic cleft lip palate (NSCLP). The VWS has been identified in the 0.4-3.7% of patients catalogued as NSCLP. The misdiagnosis of VWS had important implications for genetic counseling or as a confounder for NSCLP association studies. We screened for *IRF6* PV in Mexican mestizo NSCLP patients recruited for genetic association studies. 172 Mexican-Mestizo unrelated NSCLP patients evaluated for a clinical geneticist to exclude those with clinical data suggesting a syndromic form. The nine *IRF6* exons were PCR amplified and sequenced by Sanger method. Identified variants were searched in public databases. To date, complete *IRF6* sequencing has been achieved in 96 individuals, 76 patients sequencing has a progress of ~70%. No PV have been identified. 9 reported non-PV were identified, all in Hardy-Weinberg equilibrium, allelic frequencies 0.14-0.44 which did not show statistical differences with those reported in Mexican ancestry individuals from LA (1000 Genomes project). Screening of VWS in NSCLP patients have been performed in Asian, Caucasian, African and Latin-American populations, but it is the first report in Mexican mestizos and the largest in Latin-America. Although we have not yet completed sequencing, seems unlikely to identify the VWS in our sample, thus our results indicate that VWS could be infrequent in clinically catalogued NSCLP patients. This can be explained by the careful clinical evaluation in order to exclude any minimal expression of VWS. In similar studies, was reported that in *a posteriori* revision of NSCLP-catalogued cases with an *IRF6* PV, they have a minimal expression of lip pits (42%), besides that *IRF6* PV carriers has a pedigree suggesting autosomal dominant inheritance (42%), PV carriers with none of the last characteristics were 43%. In our sample only 5.8% had an affected first degree relative (5 of 10 cases with complete *IRF6* sequence and none with a PV). In our study, minimal expression of VWS was discarded from patients and the proportion of familial cases were low. This diminishes the possibility of assortment bias for NSCLP association studies performed in this sample and remarks the importance of a careful clinical evaluation.

1132W

Urine-derived podocytes-like cells: From a diagnostic to a CRISPR/Cas9 gene therapy perspective in Alport syndrome. *A.M. Pinto[1,2], S. Daga[1], M. Baldassarri[1,2], C. Lo Rizzo[1,2], C. Fallerini[1], V. Imperatore[1], I. Longo[1,2], E. Frullanti[1], L. Massella[3], C. Pecoraro[4], G. Garosi[2], F. Ariani[1,2], M.A. Mencarelli[1,2], F. Mari[1,2], A. Renieri[1,2].* 1) Medical Genetics, Department of Medical Biotechnology, University of Siena, Policlinico 'Santa Maria alle Scotte', Siena, Italy; 2) Medical Genetics, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 3) Division of Nephrology and Dialysis, Bambino Gesù Children's Hospital, IRCCS, Rome; 4) Pediatric Nephrology Unit, Santobono-Pausilipon Hospital, Naples, Italy.

   Alport Syndrome (ATS) is a rare genetic disorder caused by mutations in collagen IV genes, leading to ultrastructural lesions of the Glomerular Basement Membrane (GBM) up to end-stage renal disease. Podocytes, the main cellular component of the glomerular structure, are the only cells able to produce the three collagens IV alpha chains and thus, they are key-players in ATS pathogenesis. However, transcripts-based diagnostic approaches and podocytes-targeted therapeutic strategies, have been hampered by their inaccessibility. Modifying previously established protocols for urine cells isolation, we were able to grow and isolate podocytes-like cells from ATS patients, in a totally non-invasive way. RT-PCR analysis indeed revealed *COL4A3*, *COL4A4* and *COL4A5* expression, for all the isolated cell lines. RNA studies performed on patients-derived podocytes harboring intronic variants of uncertain significance (VoUS) in *COLIV* genes, led to the identification of aberrant splicing patterns leading to premature stop codons and truncated proteins allowing us to define the pathogenic role in a number of VoUS involving +2, +5 positions and exon-intron deletions. We designed a podocytes-like cells co-infection with both a SV40NLS-dCas9-AAV2 and a gRNA-HDR-AAV2 (pAAV-COL4-HDR) construct harboring the homology arms, packed between a gRNA expression cassette and a GFP expression cassette to "snip out" the most frequent single nucleotide *COL4* genetic mutations and stably reverting the phenotype. We are currently running this approach on variants frequently observed in Eastern Europe: p.Gly1045Val, p.Gly1233Arg and p.Gly695Arg in *COL4A3*, p.Gly370Glu in *COL4A4* and p.Gly624Asp, p.Arg1410Cys in *COL4A5*. As proof of principle we demonstrated that it is possible to isolate podocytes-like cells from Alport patients urine samples thus providing a novel easily-available tool for a transcript-based diagnostic approach aiming to provide the family with a proper recurrence risk. In addition, we built on the availability of disease-relevant cell lines to engineer a gene therapy strategy opening up the possibility of a personalized transformative gene-based medicine.

1133T

The expansion of heterotopic bone in fibrodysplasia ossificans progressiva is Activin A-dependent. *A.N. Economides[1], J. Upadhyay[1], L. Xie[1], L. Huang[1], N. Das[1], R. Stewart[2], M. Lyon[2], K. Palmer[2], S. Rajamani[1], C. Graul[2], M. Lobo[2], T. Wellman[2], E. Soares[3], M. Silva[2], J. Hesterman[2], K. Nannuru[1], V. Idone[1], A.J. Murphy[1], S.J. Hatsell[1].* 1) Regeneron Pharmaceuticals, Inc., 777 Old Saw Mill River Road, Tarrytown, NY; 2) InviCRO, LLC, 27 Drydock Avenue, Boston, MA 02210; 3) College of Holy Cross, 1 College Street, Worcester, MA 01610.

   Fibrodysplasia Ossificans Progressiva (FOP) is a rare autosomal dominant disorder that is characterized by episodic, progressive, and cumulative heterotopic ossification (HO) in tendons, ligaments, and a subset of skeletal muscles over a patient's lifetime. FOP is caused by missense mutations in the type I Bone Morphogenetic Protein (BMP) receptor-encoding gene, *ACVR1*. We have shown that HO in FOP requires activation of mutant ACVR1 by Activin A. Activin A – a BMP/TGFß family member which is not osteogenic – induces formation of HO in FOP mice, whereas its inhibition blocks the formation of HO. Here we extend our previous studies by piecing together a 'natural history' of developing HO lesions, and determining where in the continuum of HO Activin A is required, using imaging (T2-MRI, µCT, ^{18}F-NaF PET/CT) coupled with pharmacologic inhibition of Activin A at different times during the progression of HO in FOP mice (MGI:5763014). As disease progresses, expansion of HO lesions comes about through growth and fusion of independent HO events that tend to arise within a neighborhood of existing lesions, indicating that already formed heterotopic bone lesions likely trigger the formation of new lesions. The process appears to be depended on Activin A, as inhibition of this ligand not only suppresses the growth (and can even cause partial regression) of nascent HO lesions but also stops the emergence of new HO events. These results provide evidence for a model where HO is triggered by inflammation, and where existing or developing HO lesions become 'self-propagating' by a process that may involve re-initiation of inflammation (by already formed heterotopic bone lesions) and that also requires Activin A. This new data extends the potential utility of prophylactic treatment with inhibitors of Activin A as a therapy for FOP that might not only stop the emergence of new lesions but also limit the growth and expansion of the heterotopic bone field in subjects with ongoing disease activity.

1134F

Diamond Blackfan "anemia" due to *RPL26* mutation in a family with radial ray and congenital anomalies. *L. Mehta[1], L. Karger[1], B. Webb[1], J. Liao[1,2], L. Edelmann[1,2].* 1) Dept. of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Sema4, a Mount Sinai venture, Stamford, CT.

Heterozygous loss of function mutations in ribosomal protein genes cause Diamond Blackfan anemia (DBA). Typically DBA presents with hematological findings of pure red cell aplasia and macrocytic anemia. Growth retardation and congenital anomalies, commonly of limbs, heart or genitourinary system are present in 30-50% of patients with DBA. We evaluated a 63 yo man, with right triphalangeal thumb, right stenotic ear canal and high myopia, height 5'10", with no other significant medical history. His 52 yo brother had mildly hypoplastic left thumb. This brother's older son had preauricular pits, myopia, and learning disabilities; a younger son had bilateral hypoplastic thumbs and forearms with radioulnar synostosis, right congenital deafness, horseshoe kidney, short stature, and autism. A third brother, age 41 y, had a right triphalangeal thumb. The parents of the patient were not noted to have any anomalies. His 28 yo daughter was unaffected. Townes Brocks and Okihiro syndromes were considered, but *SALL1* and *SALL4* sequencing as well as chromosome microarray were normal. Whole exome sequencing showed a likely pathogenic frameshift variant in *RPL26*, c.341delA, p.D114Afs*15 (NM_000987.3), in the proband and his 41 yo brother, but not in the proband's daughter. Other affected family members declined genetic testing. The variant has not been reported in DBA patients, nor is it present in dbSNP, ESP or ExAC databases. Located in the last exon of the gene, it is likely to disrupt gene function due to changing the reading frame and protein truncation. The patient had normal blood counts, MCV and Hb electrophoresis and elevated erythrocyte deaminase (eADA) of 1083 mu/g Hb (ref range 400-900). He had a normal echocardiogram and renal sonogram. A *de novo* frameshift mutation in *RPL26* has been previously reported in a female infant with shortened forearms, absent thumbs, absent left kidney and absent/narrow external auditory meatus. This patient had severe anemia and neutropenia (Gazda et al, 2011). RPL26 depletion in HeLa cells was observed to affect maturation of ribosomal RNA subunits. While our patient did not manifest hematological abnormalities he was found to have an elevated eADA. Mutations in ribosomal protein genes should be considered in the differential diagnosis of relevant congenital anomalies particularly of the forearms and radial ray, even in the absence of anemia. EADA may be a good screening test in such cases.

1135W

A novel genetic disorder characterized by severe developmental delay and dysmorphism, recurrent pancreatitis, and organomegaly. *M. Morimoto[1,2], X. Chepa-Lotrea[1,2], M. Sincan[1,2], D. Draper[1,2], C.F. Boerkoel[1,2], G. Golas[1,2], D.R. Adams[1,2], W.A. Gahl[1,2,3], M.C.V. Malicdan[1,2,3].* 1) NIH Undiagnosed Diseases Program, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Section of Human Biochemical Genetics, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Introduction: The endoplasmic reticulum (ER), composed of a continuous membranous system, is vital for protein (rough ER) and lipid (smooth ER) synthesis and secretion, lipid bilayer assembly, carbohydrate metabolism, and calcium homeostasis. Disruption of the ER by either gene mutations or misfolding of cellular proteins can lead to neurodegeneration, metabolic or inflammatory disorders. Here we present a child with multisystem abnormalities and identify a gene encoding an endoplasmic reticulum transmembrane protein that may be associated with this novel genetic disease. **Methods:** Extensive clinical evaluation was performed on the patient through the NIH Undiagnosed Diseases Program (UDP). SNP array and whole exome sequencing analyses were performed and candidate variants were validated by Sanger sequencing. Expression analyses were performed in patient dermal fibroblasts using quantitative PCR, western blot, and immunocytochemistry. **Results:** A 5-year-old girl presented to the UDP with a complex multisystem clinical phenotype characterized by severe global developmental delay, hypotonia, recurrent bouts of pancreatitis, chronic liver inflammation, enlarged kidneys, and hepatosplenomegaly. She also exhibited dysmorphic features including brachycephaly, low-set ears, dental crowding, joint contractures, and club foot. Compound heterozygous nonsense mutations were identified in *CCDC47* in the patient. Expression studies in patient dermal fibroblasts showed nearly negligible expression of *CCDC47* mRNA and no protein expression compared to unaffected controls. These results, together with the previously described role of CCDC47 in embryogenesis and endoplasmic reticulum-associated degradation (ERAD), suggest that the lack of CCDC47 protein expression may be contributing to the pathogenesis of this disease through dysregulation of this pathway. **Conclusion:** We report the first patient with a novel genetic disease and hypothesize that biallelic mutations in the endoplasmic reticulum transmembrane protein encoding gene *CCDC47* may underlie the pathogenesis of this disease. We plan to further explore this hypothesis through further studies in neural progenitor cells derived from patient-specific induced pluripotent stem cells.

1136T

A growing need for reverse clinical genomics: Demonstrated by phenotypic characterization of CDK13-related disorders. *B. Bostwick[1], S. McLean[1,11], J. Posey[1], H. Streff[1], K. Gripp[2], A. Blesson[2], N. Powell-Hamilton[2], J. Tusi[2], D. Stevenson[3], E. Farrelly[3], L. Hudgins[3], Y. Yang[1,4], F. Xia[1,4], X. Wang[1,4], P. Liu[1,4], M. Walkiewicz[1,4], D. Grange[5], M. Andrews[5], M. Hummel[6], S. Madan-Khetarpal[7], E. Infante[7], Z. Coban-Akdemir[1], J. Jefferies[8], J. Rosenfeld[1], L. Emrick[1], K. Nugent[1,11], J. Lupski[1,9,10,12], J. Belmont[1,10], B. Lee[1], S. Lalani[1], Members of the Undiagnosed Diseases Network.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX USA; 2) Division of Medical Genetics, A.I. duPont Hospital for Children/Nemours, Wilmington, DE, USA; 3) Division of Medical Genetics, Stanford University School of Medicine, Stanford, CA; 4) Baylor Genetics Laboratories, Houston, TX USA; 5) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, MO USA; 6) Department of Pediatrics, Section of Medical Genetics, West Virginia University Health Sciences Center, Morgantown, WV; 7) Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA, USA; 8) The Heart Institute, Cincinnati Children's Hospital Medical Center; 9) Texas Children's Hospital, Houston, TX, 77030, USA; 10) Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030, USA; 11) Department of Pediatrics, Baylor College of Medicine, San Antonio, TX, 78207 USA; 12) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030, USA.

Historically, most disease gene discovery occurred *after* the gathering of a large cohort of individuals with a remarkably similar phenotype. It was common for a syndrome to be identified, named, and well-delineated prior to gene discovery. With the advent of next generation sequencing, many new gene discoveries are originating from large disease-specific sequencing cohorts often containing thousands of individuals. In these new studies, syndromes are identified first molecularly instead of phenotypically. This experimental design has created a new demand for 'reverse clinical genomics,' where disease-gene directed cohorts are then studied phenotypically to delineate the syndromic associations and appropriate medical management. Here, we use this approach to provide comprehensive phenotypic characterization of a newly described CDK13-related disorder. *De novo* missense variants in *CDK13* were first described in late-2016 as the cause of syndromic congenital heart defects in 7 individuals ascertained from a large congenital cardiovascular malformations cohort. We aimed to further define the phenotypic spectrum of this newly described disorder by recruiting 9 additional individuals with *CDK13* pathogenic variants detected by clinical or research exome sequencing. This approach minimized ascertainment bias as recruitment was non-phenotype driven. Our phenotypic characterization demonstrated greater than expected phenotypic heterogeneity, including 33% (3 of 9) of individuals without structural heart disease on echocardiogram. There was a high penetrance for a unique constellation of facial dysmorphisms, suggesting a potentially recognizable 'facial gestalt' with some resemblance to Kabuki syndrome. In addition, a wide variety of CNS anatomical abnormalities (91%), sacral abnormalities (44%) and renal anomalies (19%) were present. Global developmental delay was universal. Molecularly, two individuals had novel *de novo CDK13* variants (p.Asn842Asp, p.Lys734Glu), while the remaining 7 unrelated individuals had a recurrent, previously published p.Asn842Ser variant. Overall, our study demonstrates that extensive phenotyping using a genotype-first approach to subjects with pathogenic *CDK13* variants enables syndrome delineation and recognition of greater phenotypic heterogeneity than revealed by initial phenotype-driven cohort analysis. The aggregated clinical phenotyping information is used for the initial construction of clinical management guidelines.

1137F

Phenotypic variability in a cohort of patients with VACTER-L association. *O.M. Moreno[1], A.I Sánchez[1], A. Herreño[1], Y. Vargas[2], M. Olaya[2], G. Giraldo[1], J.C Prieto[1], A. Rojas[1].* 1) Instituto de Genética Humana, Pontificia Universidad Javeriana Bogotá, Bogotá, Colombia; 2) Hospital Universitario San Ignacio.

Introduction: VACTER-L association comprises different multisystem congenital malformations including vertebral anomalies, anal atresia, cardiac defects, trachea-oesophageal (TE) fistula with or without oesophageal atresia, renal malformations and limb defects. Prevalence is 1:7000 to 1:40000 individuals. This heterogeneous condition has many overlapping defects with other malformation syndromes and it is usually considered as a diagnosis of exclusion. Fanconi anemia (FA) patients often exhibit birth defects suggesting VACTERL association. Because of this, FA must be considered among one of the first differential diagnosis among VACTER-L patients. **Objective**: The aim of this study was to determine the frequency of VACTER-L characteristics in patients who met criteria for this condition and to identify abnormalities in genetic tests including conventional cytogenetics analysis, molecular cytogenetics (aCGH) and chromosome fragility. **Methods:** This is a prospective study. Data regarding clinical features was obtained by physical examination or from clinical reports, and then it was analyzed to describe all clinical aspects. Complete cytogenetic investigation based on at least *25* metaphases with G-band staining was performed in all patients, as well as chromosome fragility studies for FA using standard procedures. **Results**: We evaluated 18 patients who met criteria for VACTER-L association, 11 were female and 7 were male. Age range between 1 month and 16 years. Ten patients presented with vertebral anomalies. Anal atresia was evident in 6 patients. Congenital heart defects were seen in 13 patients. Oesophageal atresia was diagnosed in 9 patients, six of them having TE fistula. Eight patients had renal anomalies. Limb defects were evident in 7 patients. Normal karyotypes were reported for all patients. Chromosome fragility studies were negative. aCGH results are running. **Discussion and conclusion:** VACTER-L association is characterized as a heterogeneous condition with non-random pattern of defects. All patients in our study exhibit a wide clinical spectrum. Chromosome aberrations were ruled out in all of them and chromosome breakages and fragility for FA were negative. Because abnormalities in aCGH have been identified in patients with VACTER-L association, we are also performing this test, however results are still pending. More association studies are necessary for genotype-phenotype correlations to determine the genetic causes underlying this condition.

1138W

Truncating *MAGEL2* mutations produce fetal lethality in mice and may recapitulate pathogenesis of Schaaf-Yang syndrome. *Y. Negishi[1], D. Ieda[1], T. Miyamoto[2], I. Hori[1], A. Hattori[1], Y. Nozaki[3], H. Komaki[4], J. Tohyama[5], K. Nagasaki[6], H. Tada[7], H. Oishi[2], S. Saitoh[1].* 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Comparative and Experimental Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya Japan; 3) Department of Pediatrics, Jichi Medical University, Tochigi, Japan; 4) Department of Child Neurology , National Center Hospital, National Center of Neurology and Psychiatry (NCNP) , Tokyo , Japan; 5) Department of Child Neurology Nishi-Niigata Chuo National Hospital Niigata, Japan; 6) Division of Pediatrics, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; 7) Department of Pediatrics, Chibaken Saiseikai Narashino Hospital, Narashino, Japan.

   MAGEL2 is a paternally expressed gene located in 15q11-q13, which is responsible for Prader-Willi syndrome (PWS), and truncating mutations on the paternal allele of *MAGEL2* have recently been reported as a cause of Schaaf-Yang syndrome (SHFYNG). The mechanism underlying the disease is considered to be the loss of function of *MAGEL2*; however, the deletions of the entire paternal copy of *MAGEL2* appear to cause very mild phenotype, and there are many uncertainties regarding the role of *MAGEL2* in PWS phenotypes. We sequenced *MAGEL2* in 105 patients suspected PWS clinically, but without a genetic alteration specific for PWS, and detected truncating mutations in *MAGEL2* in 6 patients, including a pair of siblings. Of the 4 sporadic patients, 3 patients for whom samples were obtained from their parents were detected to have *de novo* mutations, and the mutation alleles were paternally derived in all 4 patients. For the 2 siblings, the father was a carrier and the father's mutation was maternally derived. The patients with *MAGEL2* mutations shared several features with PWS, such as neonatal hypotonia, poor suck, and obesity; however, there were also unique findings, including arthrogryposis and a failure to acquire meaningful words. Additionally, a history of cryptogenic encephalopathy was confirmed in 3 of the 6 patients, and this could be an additional feature of SHFYNG. Because *MAGEL2* is a single exon gene, nonsense mutation-mediated mRNA decay does not induced, and we hypothesized that the truncated protein was related to the pathogenesis. Consequently, we next produced transgenic mice that overexpressed the N-terminal of *Magel2* that was conserved in the patients. However, most of Tg-positive mice died in utero or in early neonatal period. While a single Tg-positive mouse was produced, it died several days after birth. These results suggest the possibility that truncated MAGEL2 protein could have toxic effects resulting in mostly fetal lethality in mice, and may explain why SHFYNG shows much more severe phenotype than PWS.

1139T

Case report Kleefstra syndrome in a Colombian patient. *A. Nova[1], G. Giraldo[2], J. Montoya[3].* 1) Instituto Genética Humana, Pontificia Universidad Javeriana, Bogotà, Colombia; 2) Servicio Genética Médica, Clínica Universitaria Bolivariana, Medellìn, Colombia; 3) Servicio Genética Médica,Hospital Universitario San Vicente Fundación, Medellín, Colombia.

   Kleefstra syndrome (KS), is caused by haploinsufficiency of the EHMT1 gene. And is one of the first clinically recognizable Subtelomeric Deletion syndrome . The syndrome is characterized by intellectual disability, hypotonia, and typical dysmorphisms, and may be associated with congenital heart, renal defects and epilepsy. The evolution of the behavioral profile of KS was one of the last described characteristics of the syndrome. In this report, the characterization of a patient with kleefstra syndrome by microdeletion, with molecular diagnosis by means of comparative genomic hybridization, and changes suggestive of MRI leukomalacia, from an early age, in whom no behavioral changes have yet been shown, described in other reports at later ages, in patients with brain alterations similar, confirming the neuroregressive picture of this syndrome and the participation of the EHTM1 gene as the cause of this syndrome. It is also highlighted by the case presented here, the need to report the cases found in our Colombian population, in order to be able to know more broadly the condition of this pathology, its epidemiology, its characteristics, natural history and forecast, in order to be able to carry out Comprehensive management of these patients.

1140F

Acid ceramidase deficiency (Farber disease): A qualitative research study documenting the clinical impact of symptoms in a diverse population of patients and caregivers. *A. Solyom[1], B. Johnson[2], P. Tanpaiboon[3], D. Tetzl[4], K. Coyne[5].* 1) Enzyvant, Basel, Switzerland; 2) Roivant Sciences, Durham, USA; 3) Children's National Medical Center, Washington, DC, USA; 4) Enzyvant, New York, USA; 5) Evidera, Bethesda, USA.

Introduction Acid ceramidase deficiency (Farber disease; Farber lipogranulomatosis) is caused by mutations in the *ASAH1* gene and is considered an ultra-rare disease. Since 2014, Enzyvant has identified a cohort of patients from 30 countries on 6 continents, with a much broader phenotypic spectrum and demonstrating a potentially higher incidence than previously thought. The cardinal symptoms of acid ceramidase deficiency are: arthritis, subcutaneous nodules, and hoarseness. Farber disease is most often misdiagnosed as Juvenile Idiopathic Arthritis.   Objectives Conduct a qualitative research study to understand the symptoms and impact of Farber disease from the perspective of patients and caregivers by conducting one-on-one interviews for concept elicitation, content validity, and patient interpretation of newly developed patient-reported outcome (PRO) tools.   Methods This study consisted of two parts: semi-structured qualitative interviews to understand Farber disease symptoms and to evaluate new PRO measures to capture the impact of symptoms on patients including: subcutaneous nodules, voice, overall pain, arthritis, ability to move joints, ability to perform daily tasks, and level of fatigue. Participants were asked to rate the symptoms' impact on a scale of 0–10, with 0 meaning no impact and 10 meaning greatest impact.   Results Eight interviews were conducted: 2 adult patients (attenuated phenotype); 1 caregiver of deceased child (severe phenotype); 1 transplanted adult patient (moderate phenotype); 1 caregiver of transplanted pediatric patient (moderate phenotype); 1 caregiver of a pediatric patient (moderate phenotype); and 1 caregiver/patient dyad of a non-transplanted pediatric patient (moderate phenotype). Interviews revealed the symptoms of acid ceramidase deficiency have measurable clinical impact across a broad spectrum of phenotypes and symptom severity. The symptoms with the highest-rated impact were the ability to move joints, average impact rating of 6.6 (range 1 to 10); the ability to perform daily activities and subcutaneous nodules, both rated 5.9 (range 1 to 10 for daily activities and 3 to 9 for nodules); voice impact was rated 5 (range of 3 to 9).   Conclusion The information gathered provides a better understanding of methods useful for measuring symptom impact in Farber disease, potentially in the context of future therapeutic trials, and may allow better discussion of symptom impact between physicians, patients and caregivers.

1141W

Natural history study design including retrospective and prospective components to address the lack of clinical data on acid ceramidase deficiency presenting as Farber disease. *D. Tetzl[1], B. Johnson[2], A. Solyom[3].* 1) Enzyvant, New York, USA; 2) Roivant Sciences, Durham, USA; 3) Enzyvant, Basel, Switzerland.

Introduction Mutations in both alleles of the *ASAH1* gene cause acid ceramidase deficiency (Farber disease; Farber lipogranulomatosis) and accumulation of the pro-apoptotic and pro-inflammatory sphingolipid, ceramide. A cohort of patients (some of whom are since deceased) from 30 countries on 6 continents has been identified since 2014, with clinical presentations indicating a much broader phenotypic spectrum and potentially higher incidence than previously thought.   Objectives and Methods We developed a natural history study structure and design to enable systematic collection of information on living patients who have or have not undergone hematopoietic stem cell transplantation (HSCT), and on deceased patients using medical records. The study will also function to evaluate clinical assessment tools used for other diseases with similar symptoms, as well as those created specifically for Farber disease. The pre-existing clinical assessment tools include (among others) the Childhood Health Assessment Questionnaire (CHAQ), and Wong-Baker Faces Pain Rating Scale, 6-minute Walk Test, Pulmonary Function Testing, and Joint Range of Motion measurements. We developed a unique patient reported outcome measure related to Farber disease symptoms including physical impairment, pain, and emotional distress, and a method for the assessment of change in size of subcutaneous nodules.   Conclusion Consideration of the broad spectrum of ages and heterogeneous phenotypes in a small population of patients available to participate in a study of the natural history of acid ceramidase deficiency led to the design of a cross-sectional cohort study with retrospective and prospective components. By including deceased patients and patients having undergone HSCT in the study design, the amount of data that can contribute to the understanding of this very rare disease is substantially increased. Furthermore, the specific prospective assessments will allow us to follow disease progression over time, and to evaluate the degree to which the assessment tools and techniques may be appropriate for registering change in certain symptoms over time. Enzyvant is developing enzyme replacement therapy for Farber disease, and it is our hope that the data collected will contribute to the selection of methods for measuring therapeutic efficacy in any future clinical trials.

1142T

Rare syndrome: First reported Egyptian sibs with 3MC(1) syndrome and detection of novel *MASP1* gene mutation in the family. *A.N. Khalaf[1], S.M Kotb[1], M.M. Fata[1], E.M. Abdalla[1], K.M. Nabil[3], F.S. Alkurraya[2].* 1) Human Genetics, Medical Research Institute, Alexandria, Egypt; 2) KFSHRC, Riyadh, SA; 3) Ophthalmology Dept, Faculty of medicine, Alexandria, Egypt.

Here we are reporting the first Egyptian male sibs with the rare 3MC(1) syndrome. Two male sibs born to consanguineous parents, aged 4 and 2 years, presented with dysmorphic features, cleft lip and congenital heart disease. Pregnancy history and delivery history were irrelevant. Both patients showed motor and speech developmental delay. The elder brother had history of correction of bilateral inguinal hernia and cleft lip at age of 6 months. On examination, this patient showed bilateral microphthalmia with narrow palpebral fissures, severe ptosis and absent lid creases, distorted nose with wide nasal root and bridge, corrected incomplete right cleft lip, very long low-set ears, occipital whorl and abnormal irregular teeth with hypodontia. In addition, bilateral syndactyly of second and third toes and single crease on fifth left finger were also noted. Anthropometric measures were within normal percentiles. Echocardiography revealed PDA and VSD while ABR, CT brain and karyotype were normal. Ophthalmological examination showed mixed astigmatism, prominent schwalb line and alternating esotropia. For the younger similarly affected brother there was a history of congenital clenched hands and deformed feet which gradually improved till completely resolved. On examination, he showed dysmorphic features similar to his brother with corrected right oblique incomplete cleft lip and abnormal helical folding of left ear which is also distorted. Anthropometric measures were also normal. CT brain showed ectatic ventricles, ABR revealed serous otitis media and echocardiography showed PDA, VSD and ASD secondum. Most propable suspected diagnosis was 3MC(1) syndrome ,This syndrome which was first described by Michels et al. with the eye lid triad of blepharophymosis, blepharoptosis and inverse epicanthus in association with limitation of upward gaze, hypertelorism, cleft lip and palate, Various skeletal anomalies, mild intellectual ability and mixed hearing loss. Those findings were greatly matched with our cases with the typical ocular findings, cleft lip and palate. Thereafter, molecular testing was requested for the whole family revealing a novel homozygote mutation in *MASP1* gene in the affected sibs while the parents were heterozygote for the same mutation ,confirming our diagnosis. This aided us in giving proper genetic counseling and offering prenatal genetic testing in future pregnancies for the family. .

1143F

Expanding the phenotype of *DST*-related disorder: A case report suggesting a genotype/phenotype correlation. *M. Pinelli[1,2], G. Cappuccio[1,2], A. Torella[2,3], M. Alagia[2], R. Auricchio[2], A. Staiano[2], V. Nigro[2,3], N. Brunetti-Pierri[1,2].* 1) Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; 2) Department of Translational Medicine, Section of Pediatrics, Federico II University, Naples, Italy; 3) Department of Biochemistry, Biophysics and General Pathology, Università degli Studi della Campania 'Luigi Vanvitelli', Naples, Italy.

The gene *DST* encodes for the large protein BPAG1 involved in hemidesmosomes. Its alternative splicing gives rise to tissue-enriched isoforms in brain, muscle and skin. The few patients described so far with bi-allelic mutations in *DST* gene have either a skin phenotype of epidermolysis bullosa simplex (EBS) or a neurological phenotype. Here, we report a 17-year-old individual presenting with a complex phenotype with both skin and neuronal involvement, in addition to behavioral and gastrointestinal issues, iris heterochromia, cataract, hearing impairment, syringomyelia, osteoporosis, and growth hormone deficiency. Family-trio whole-exome sequencing revealed that she was a compound heterozygous for two variants in the *DST* gene with highly-predicted functional impact: c.3886A>G (p.R1296X) in exon 29 and c.806C>T (p.H269R) in exon 7. Interestingly, exon 7 is included in the neuronal isoform whereas exon 29 is expressed in both skin and neuronal isoforms. This individual herein described is the first case with a mutation affecting an exon expressed in both the neuronal and skin isoforms thus, potentially explaining the more complex and expanded phenotype, that includes several previously unreported findings in patients harboring *DST* mutations. .

1144W

Genomic approaches to investigate children born small for gestational age (SGA) without catch up-growth. *B.L. Freire, T.K. Homma, M. Funari, A. Malaquias, R. Ronjo, G. Vasques, A. Canton, C. Kim, D. Bertola, A. Jorge. University of Sao Paulo, São Paulo, São Paulo, Brazil.*

   Background: Short stature is a complex and heterogeneous disorder influenced by genetic factors. It may be present in non-syndromic or syndromic forms, with a variable clinical spectrum making difficult the clinical diagnosis. **Aim:** We conducted a genomic approach to establish the etiology of growth disorder in children born SGA without catch-up growth. **Methods:** A total of 197 children born SGA were evaluated clinically, laboratory and radiologically by professionals with expertise in dysmorphology. Clinical diagnosis was unable to be obtained in 152 patients. According to the presence/absent of dysmorphic features, developmental delay or intellectual disability they were classified as syndromic (S=65) or non-syndromic (NS=87). We conducted SNP/CGH microarray analyses in 71 patients (S=48/NS=23) and whole exome (n=38)/targeted panel (n=58) sequencing in 96 patients (S=48/NS=48). Identified copy number (CNVs) and single nucleotide (SNVs) variations were classified according to ACMG guidelines. **Results:** Among Non-Syndromic patients, we did not found any pathogenic CNVs, however we identified 9 (19%) pathogenic SNVs using whole exome sequencing/target panel in *IGF1R (2x), IHH (2x), ACAN, NF1, NPR2, PTPN11, LTBP3* genes. Among syndromic group, we found 13 (27%) CNVs and 8 (16.5%) SNVs classified as pathogenic or probably pathogenic, including variants in *COL2A1 (2x), SRCAP (2x), ANKRD11, NPR2, POC1A, BRCA1.* All identified SNVs are extremely rare or absent in public database, were predicted to be deleterious and segregated with the phenotype. The genes involved are associated with rare diseases (i.e. KBG syndrome), or genetic syndromes with high phenotypic variability (i.e. Noonan syndrome), which prevented their clinical recognition. Additionally, we identified novel candidate genes for short stature of prenatal onset: *DGCR8* and *RAB3IP*. The variants identified in these genes are loss-of-function, one of them was identified in homozygous state (*RAB3IP*) and one was a heterozygous de novo mutation (*DGCR8*). All novel candidate genes are involved in critical cellular function pathway and, based on ExAC database, they are intolerant to LoF variation. **Conclusion:** The great heterogeneity of short stature phenotype difficults the clinical diagnosis of children born SGA.The genomic approaches are effective to establish the etiology of growth impairment of a larger number of previously undiagnosed children born SGA and also identified novel genes involved in this phenotype.

1145T

Genetic diagnosis and clinical characteristic reviews in neonate patients of *KMT2D* gene mutation caused Kabuki syndrome. *B. Wu[1,2], Y. Su[3], H. Wang[1], X. Dong[1], P. Zhang[1], W. Zhou[1,2]. 1) Molecular Diagnosis Laboratory of Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defects, The Translational Medicine Center of Children Development and Disease of Fudan University, Shanghai 201102, China; 2) Key Laboratory of Neonatal Diseases, Ministry of Health, Shanghai 201102;; 3) Department of Neonatology, People's Hospital of Xinjinag Uygur Autonomous Region ,Urumqi 830001.*

   Objective To investigate the clinical and genetic features of Kabuki syndrome caused by *KMT2D* mutation. Summarize the clinical features in neonate. **Method** Using Whole-Exome Sequencing (WES) and Clinical panel deep sequencing,;combined with data analysis pipeline established by molecular diagnostic center of Children's Hospital of Fudan University. We summarized the clinical and molecular features of 8 children with *KMT2D* mutations. Databases including PubMed,;CNKI,;and VIP were searched to collect literature of KS, which describe the clinical features of neonatal period from April 2012 to April 2017. **Result** Four males and four females were diagnosed in our laboratory. Mutations were located in exon 11, exon14, exon34, exon 39 (3cases), exon 51 and exon 53 respective. The types of mutations were two stop gained, four missenses and two frameshifts. Mutation of c.C11119T (p.R3707X), c.12697C>T (p.Q4233X), c.16498C>T (p.R5500W) and c.16273G>A(p.E5425K) were reported as pathogenic mutations and had recorded in Human Gene Mutation Database (HGMD). Mutation of c.G4139T (p.C1380F), c.3495delC (p.Pro1165LeufsTer47), c.10881delT (p.Leu3627Argfs Ter31) and c.12560G>A (p.G418E) were novel, which predicted as harmful variants by SIFT, polyphen 2 and MutationTaster software. Seven of them were from new born baby to 4 months infant, all with de novo mutation confirmed with parents data. Only one 8 years female patient with congenital and development delay who have clinical Kabuki syndrome diagnosis. This patient was confirmed carried c.C11119T (p.R3707X) pathogenic mutation by using panel sequencing, and not yet test parent. A total of 18 literatures,;involving 32 neonates were included. The most common clinical features are as follow: feeding problem in 18 cases, cardiac dysplasia in 18 cases, special appearance in 17 cases, skeletal dysplasia in 14 cases, hypoglycemia in 10 cases and hypotonia in 9 cases. **Conclusion** The typical clinical features of KS are not shown in the neonatal period. It's need to consider this disease when the newborn has feeding problem, abnormal cardiac morphology, special appearance and other clinical features. Genetic test can help to diagnose earlier in clinical. Early diagnosis can provide accurate information to clinic, may help patients to acquire appropriate treatment and family genetic counseling.

1146F

De novo truncating mutations in X-linked *MED12* are associated with Hardikar syndrome in females. *D. Li[1], E. Bhoj[1], M. Harr[1], J. Ebrahimzadeh[2], J. Richards-Yutz[2], A. Ganguly[2], E. Rand[3], J. Jaeken[4], D. Cassiman[5], P. Chow[6], A. Hing[7], S. Vergano[8], M. Van Dyck[5], E. Zackai[9,10], H. Hakonarson[1,9,10].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 2) Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA; 3) Division of Gastroenterology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 4) Center for Metabolic Diseases, University Hospital Gasthuisberg, Leuven, Belgium; 5) Department of Paediatric Nephrology, University Hospital Gasthuisberg, Leuven, Belgium; 6) Division of Genetics, Seattle Children's Hospital, Seattle, Washington, USA; 7) Division of Craniofacial Medicine, Seattle Children's Hospital, Seattle, Washington, USA; 8) Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Norfolk, Virginia, USA; 9) Department of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 10) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

Hardikar syndrome (MIM 612726) is a multiple congenital anomaly syndrome characterized by hepatic or biliary anomalies, intestinal malrotation, genitourinary abnormalities, orofacial clefts, and pigmentary retinopathy. Hardikar syndrome is an extremely rare disorder of unknown etiology, with only four individuals being all females reported in the literature, suggesting X-linked dominant and lethality in males. Here we performed exome sequencing in two published patients and two additional previously unpublished patients with a clinical diagnosis of Hardikar syndrome, and identified three *de novo* novel nonsense and frameshift indels in *MED12* in three patients with all having cleft palate, congenital heart disease (i.e. ventricular septal defect, aortic coarctation, or patent ductus arteriosus), biliary abnormalities (i.e. cholestasis, dilated bile ducts, or choledochal cyst), renal abnormalities (i.e. hydronephrosis or absent gall bladder), and retinal pigmentation. Hemizygous missense mutations in *MED12* cause the MED12-related disorders of FG syndrome, Lujan syndrome, and Ohdo syndrome in males, whereas females carrying heterozygous missense mutations are most often unaffected. The pattern of X-inactivation was analyzed based on the methylation pattern at the androgen receptor locus on all four studied individuals with a clinical diagnosis of Hardikar syndrome, 3 of which harboring the *MED12* loss-of-function (LoF) mutation showed extreme skewed X-inactivation (99:1), suggestive of some genetic heterogeneity. These LoF *MED12* mutations identified in Hardikar syndrome raise suspicion for a male embryonic-lethal mechanism, accounting for the female preponderance of patients affected by this syndrome.

1147W

A novel homozygous mutation in MMP21 is associated with heterotaxia and cardiac defects. *L. Aspit[1,5], O. Reish[2,3], L. Benyamin[2], V. Sheffield[4], R. Parvari[1,5].* 1) Microbiology, Immunology and Genetics, Ben Gurion University of the Negev, Beer Sheva, Israel; 2) Genetic Institute, Assaf Harofeh Medical Center, Zerifin, Israel; 3) The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; 4) Department of Pediatrics, Division of Medical Genetics, University of Iowa, Iowa City, IA, USA; 5) National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer Sheva, Israel,.

Introduction: Heterotaxia is a rare congenital autosomal recessive defect in which the major visceral organs are distributed abnormally within the chest and abdomen. The mixed organ orientation found in situs ambiguus often leads to severe complications mainly of the heart and/or lungs. We have characterized a consanguineous Bedouin family from Israel with 2 affected siblings with heterotaxia and complicated heart defects and detected a novel MMP21 mutation. Materials and Methods: Genotyping of both patients and their parents was performed using Affymetrix microarrays and kinSNP analysis. Regions of homozygosity were identified. Exome sequencing was performed on 1 affected DNA. Results: Exome sequencing revealed 50 potential variations within the homozygous regions. Following bioinformatics predictions, segregation in the family and analysis of prevalence in the Bedouin population, one candidate remained - a homozygous variation in the MMP21 gene that is involved in the breakdown of extracellular matrix for normal physiological processes, such as embryonic development, reproduction, and tissue remodeling. The variation causes an amino acid change that is predicted to be highly damaging. Conclusions: The novel MMP21 mutation presented here, reinforce the role of this gene that only recently has been described with association to heterotaxia and is in line with an ENU-induced mutant mice with LR asymmetry and major cardiac defects. Further studies are required to verify whether this gene is a candidate for complicated heart defects without LR asymmetry.

1148T

Exome-first approach identified two novel indels and gene deletions in Mowat-Wilson Syndrome cases previously diagnosed as clinical Angelman syndrome. *F. Gosso[1,3], C. Rohr[2,3], B. Brun[2,3], G. Mejico[1,3], F. Madeira[1,3], F. Fay[1,3], M. Klurfan[4], M. Vazquez[2,3].* 1) Cibic S.A., Rosario, Santa Fe, Argentina; 2) INDEAR, Rosario, Santa Fe, Argentina; 3) Heritas, Rosario, Santa Fe, Argentina; 4) Casa Angelman, Buenos Aires, Argentina.

Introduction Mowat-Wilson Syndrome (MWS) is caused by haploinsufficiency of *ZEB2* (ZFHX1B) on chromosome 2q22.3 [1]. MWS resembles Angelman and other Angelman-like Syndromes (ALS) in that all individuals have moderate-to-severe intellectual disability with minimal speech. Microcephaly, seizures or EEG abnormalities have been observed in up to 90% of affected individuals. According to the Mowat-Wilson Syndrome Foundation, 213 patients worldwide received genetic confirmation so far. Since MWS is misdiagnosed with AS and other ALS during early infancy, it is very important to develop a first-tier single genetic testing that covers all types of genetic mutations including SNVs, INDELs and CNVs to distinguish between them. Whole-exome Sequencing (WES) or clinical-exome Sequencing (CES) was recently proposed as a first-tier clinical diagnostic test in children with intellectual disabilities [2]. We propose WES/CES as differential diagnosis for patients with clinical AS and normal methylation results. **Methods** Whole-exome sequencing (WES) was performed on genomic DNA extracted from venous blood sample from two patients after informed consent was obtained. WES was performed using a mean on-target coverage for both cases of 160X. **Results** Two novel pathogenic mutations were identified in both patients with clinical AS, whom tested normal for 15p11.2-q13 methylation assay. Patient 1: a novel truncating variant was identified (NM_014795.3:c.2177_2180delCTTT, NP_055610.1:p.Ser726TyrfsTer7). Patient 2: a novel pathogenic deletion of 1.79 Mb (chr2:144209542-146007670) this result was also confirmed by comparative genomic hybridization (CGH) array. **Conclusion** WES/CES relevance as first-tier single test has been recently highlighted as differential clinical diagnostic test for a wide range of neurodevelopmental disorders. Using WES in two patients with clinical diagnosis of AS (normal 15p11.2-q13 methylation result) we were able to describe two novel mutations that unequivocally differentiate between MWS, AS caused by a mutations in UBE3A gene and ALS. Here we propose the use of WES/CES to diagnose and differentiate MWS from AS caused by SNVs, INDELs and CNVs and from other ALS in newborn/infants and young children with AS suspicion whom tested normal for 15p11.2-q13 methylation test. **References.** [1]Mowat DR *et al*. (1998). *J Med Genet*, 35:617-623. [2]Monroe GR *et al* (2016). *Genet Med*. 18:949-56.

1149F

Homozygous truncating variants in *TMCO1* contribute to a spectrum of craniofacial, skeletal and neurodevelopmental abnormalities. *J. Ji[1,2], C. Quindipan[1], G. Raca[1,2], X. Gai[1,2], S. Saitta[1,2], D. Schweitzer[3,4].* 1) Pahthology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 2) Department of Pathology, Keck USC School of Medicine, Los Angeles, CA; 3) Craniofacial and Cleft Center, Division of Plastic and Maxillofacial Surgery, Children's Hospital Los Angeles, Los Angeles, CA; 4) Department of Medical Genetics, Children's Hospital Los Angeles, Los Angeles, CA.

TMCO1 (transmembrane and coiled-coil domains protein 1) is expressed in multiple tissues and is critical to early fetal development. Pathogenic variants in *TMCO1* were first described in 11 individuals of Old Order Amish ancestry with craniofacial dysmorphism, skeletal anomalies, and intellectual disability. Since then, a few patients of Turkish ancestry have been described. We report 3 individuals from 2 unrelated consanguineous families with homozygous truncating variants in the *TMCO1* gene. Patient 1 is a 7 year-old Kuwaiti female with bilateral cleft lip and palate, facial features consistent with *TMCO1* dysmorphisms, cardiac defect, duplicated thumb and great toe, intellectual disability, brain structural abnormalities, anxiety and phobias. Trio-exome sequencing (Patient 1 and both parents) revealed a homozygous variant in *TMCO1*: NM_019026.4:c.646_647del (p.Ala216Profs*53). Patients 2 and 3 are siblings of Mexican ancestry. Patient 2 is an 8 year old male with bilateral cleft lip and palate, dysmorphic features, profound bilateral sensory neural hearing loss (SNHL), severe pectus carinatum, scoliosis, segmental vertebral anomalies of several vertebral bodies, other skeletal anomalies, single kidney, cryptorchidism, strabismus, pulmonic stenosis, and severe intellectual disability with almost no expressive language currently. His younger brother, Patient 3 is a 17 month-old male with submucous cleft palate, imperforate anus, craniofacial dysmorphia, pectus excavatum, scoliosis, fusion of cervical vertebral bodies with rib anomalies, hydronephrosis, VSD, and motor and language delays. Chromosome microarray analysis showed multiple segments of absence of heterozygosity (total 81 Mb), consistent with consanguinity. Quad-exome sequencing of both brothers and parents showed a homozygous variant: NM_019026.4:c.616C>T (p.Arg206*). These cases expand TMCO1 deficiency beyond the initial ethnic groups reported. Additionally, duplicated digits and periventricular leukomalacia with agenesis of the hippocampal commissure in Patient 1, profound SNHL in Patient 2, and imperforate anus in Patient 3, have not been previously reported in the *TMCO1*-related phenotypic spectrum. To date, only truncating *TMCO1* variants have been reported in association with disease. These mutations are distributed among different regions of the *TMCO1* gene, appear to mediate significant syndromic features and are now shown to be present across multiple ethnic groups worldwide.

1150W

Robinow syndrome in an infant with multiple anomalies due to *DVL3* mutation: A lesson in detailed clues. *B. Keena[1], C. Murali[1], P. Liu[2], E.H. Zackai[1].* 1) Clinical Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Baylor College of Medicine, Houston, TX.

We wish to report a patient with Robinow syndrome due to a DVL3 mutation and emphasize the findings in an infant. Recent discovery of mutations in the *DVL3* gene resulting in autosomal dominant Robinow syndrome was reported by Janson White et al (Am J Hum Gen 98, 2016). The authors reported 5 patients ranging in age from 14 years to 38 years who were identified with *DVL3* mutations. These 5 patients were part of a total cohort of 34 patients with clinical diagnoses of Robinow syndrome who all were studied for known Robinow genes-*DVL1*, *WNT5A*, and *ROR2* along with *DVL3* which was a candidate gene at that time. We report a case of an infant who died at 3 months of age with multiple anomalies in 2012. The female infant was transferred to our hospital at 1 month of age for management of multiple anomalies-hypoplastic left heart, omphalocele, cleft lip and palate, choanal atresia, club feet, anorectal malformation, hypertelorism, and broad thumbs and great toes. Microarray testing was normal. X-rays showed bifid terminal phalanges in fingers 1 and 2 bilaterally. The baby died at 3 months of age from worsening apnea and cardiac complications without a specific diagnosis being made. Whole exome sequencing was initially performed while this patient was alive and did not identify a molecular etiology for her findings. Upon recent review, a heterozygous *de novo* mutation in *DVL3* (c.1592delC) was reported consistent with a diagnosis of Robinow syndrome which is thought to explain the patient's findings. Robinow syndrome is characterized by frontal bossing, hypertelorism, cleft lip/palate, mesomelia, genital hypoplasia, and distinctive facial features. Omphalocele, heart defects, and broad thumbs and toes have been reported. The finding of bifid terminal phalanges, which is distinctive and rare, has been reported in Robinow syndrome in the older literature (Kelly, T et al. Am J Dis Child, 129, 1975 and Giedion et al. Helf paediat 20, 1975). In retrospect, this could have been a major clue to this diagnosis in the patient's lifetime. This finding can be very useful in a child with multiple anomalies to lead one to a diagnosis.

1151T

Frontonasal dysplasia with facial skin polyps: From Pai syndrome to differential diagnosis through a series of 25 patients. *D. Lehalle[1,2], AL. Bruel[1], M. Assoum[1], P. Vabres[1], J. Amiel[3], G. Baujat[3], B. Bessieres[4], S. Bigoni[5], L. Burglen[6], G. Captier[7], R. Dard[8], P. Edery[9,10], D. Genevieve[11], A. Goldenberg[12], D. Heron[13], M. Holder-Espinasse[14,15], D. Lederer[16], S. Grotto[12], S. Marlin[3], M. Rossi[9,10], D. Sanlaville[9,10], P. Saugier-Veber[12], S. Triau[17], MI. Valenzuela Palafoll[18], C. Vincent Delorme[14], E. Zivi[19], C. Thauvin-Robinet[1,2], J. Thevenon[1], L. Faivre[1,2], P. Callier[1].* 1) Equipe GAD, INSERM LNC UMR 1231, Faculté de Médecine, Université de Bourgogne Franche-Comté, Dijon, Dijon, France; 2) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est, Centre Hospitalier Universitaire Dijon, 21079 Dijon, France; 3) Service de Génétique, INSERM U781, Hôpital Necker-Enfants Malades, Institut Imagine, University Sorbonne-Paris-Cité, Paris, France; 4) Unite d'embryofoetopathologie, Service d'Histologie-Embryologie-Cytogénétique, Hôpital Necker – Enfants Malades, APHP, Paris; 5) UOL of Medical Genetics, Ferrara Hospital University, Ferrara, Italy; 6) Département de Génétique et Centre de Référence « malformations et maladies congénitales du cervelet», AP-HP, Hôpital Trousseau, Paris, France; 7) Service de chirurgie orthopédique et plastique pédiatrique, Hôpital Lapeyronie, CHU Montpellier, Montpellier, France; 8) Service de Cytogénétique, Centre Hospitalier Intercommunal de Poissy Saint-Germain-en-Laye, 10 rue du Champ Gaillard, F-78303 Poissy, France; 9) Service de génétique et Centre de Référence des Anomalies du développement de la région Auvergne-Rhône-Alpes, CHU de Lyon, France; 10) Centre de Recherche en Neurosciences de Lyon, INSERM U1028 CNRS UMR 5292, UCB Lyon 1; 11) Service de génétique clinique, Département de génétique médicale, maladies rares et médecine personnalisée, Université Montpellier, Unité Inserm U1183, Montpellier, France; 12) Department of Genetics, Rouen University Hospital, Normandy Centre for Genomic and Personalized Medicine, F-76000 Rouen, France; 13) AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Génétique, 75013, Paris, France. Centre de Référence "déficiences intellectuelles de causes rares", 75013 Paris, France. Groupe de Recherche Clinique (GRC) "déficience intellectuelle et autisme" UPMC, 75013 Paris, France.INSERM,; 14) CHU Lille, Department of Clinical Genetics, F-59000 Lille, France; 15) Clinical Genetics department, Guy's hospital, London, UK; 16) Center for Human Genetics, Institut de Pathologie et Génétique (I.P.G.), Gosselies, Belgium; 17) Service de Foetopathologie, CHU Angers, France; 18) Àrea de Genètica Clínica i Malalties Minoritàries, Hospital Vall d'Hebron, Barcelona, Spain; 19) The Raphael Recanati Genetics Institute, Rabin Medical Center, Petah Tikva, Israel.

Frontonasal dysplasias (FND) are rare facial malformations caused by abnormal development of the frontonasal process-derived structures. Several entities have been described, characterized by hypertelorism, median cleft and nasal anomalies. The association with facial skin polyps leads to the diagnosis of Pai, oculoauriculofrontonasal syndromes (PS, OAFNS), encephalocraniocutaneous lipomatosis (ECCL), or Sakoda complex (SC). In this work, we aim to better define the phenotypic spectrum of PS, OAFNS, ECCL and SC, discuss overlapping features and differential diagnosis for these syndromes, and identify the molecular basis of PS. We clinically reviewed 25 individuals referred for a syndromic frontonasal polyp. We performed trio or solo whole exome sequencing (WES) on blood-derived DNA from 10 individuals and whole genome sequencing (WGS) in one individual. Typical or atypical PS was clinically confirmed in respectively 9 (36%) and 2 individuals (8%). OAFNS was diagnosed in 9 individuals (36%). Syndrome overlaps were discussed in 5 patients: 3 (12%) met the criteria for both PS and OAFNS; one (4%) for PS and ECCL; one for PS and SC. We identified 3 *de novo* heterozygous variants in genes from the TGFβ pathway (*TGFBRAP1*, *PCSK7*) in PS patients, thus 12% of the series, and one inherited *GLI2* variant in a PS-ECCL patient. Based on the phenotypic overlap between these entities, we suggest that OAFNS, ECCL and SC should either be considered as differential diagnosis for Pai syndrome or part of the same spectrum. The identification of molecular bases will be the next step toward a better delineation of these entities.

1152F

Treacher Collins syndrome 1: Extreme manifestations. *C. McDougall, A.R. Barone, N. Khalek, E. Schindewolf, L. Conlin, E.H. Zackai.* Children's Hospital of Philadelphia, Philadelphia, PA.

Treacher Collins syndrome (TCS) is the most common and well known mandibulofacial dysostosis caused by mutations in at least 3 genes involved in pre-rRna transcription (*TCOF1*, *POLR1D,* and *POLR1C*). Here we present a case with extreme findings who, on whole exome sequencing (WES), was found to have a heterozygous pathogenic mutation in *TCOF1* (c.4357_4360delGAAA;p.Glu1453fs). No association between severity of the TCS phenotype and the location within the *TCOF1* gene has been observed. An infant female was born at 39 weeks gestation to a 25 year old G1P0->1 mother. Facial abnormalities were noted prenatally including cleft lip and palate, bilateral facial clefts, hypertelorism, possible colobomas, and micrognathia. Examination after delivery noted marked hypertelorism, a small right eye, midline cleft lip and palate, severe bilateral facial clefting, incompletely formed maxilla, and retromicrognathia. The patient was intubated and it was unclear if any external ear tissue was present. Extremities were normal. Ophthalmologic examination revealed bilateral chorioretinal colobomas with lack of normal optic nerve tissue as well as small palpebral fissure on the right with phimosis of the eyelids. Echocardiogram found a VSD and PDA. Given the respiratory failure and severity of the infant's anomalies including likely blindness and deafness, family elected comfort care and the infant expired. Microarray and WES were sent, with WES results as above. There have been two other extreme cases in the literature (Bauer et al., 2013. Am J Med Genet; Writzl et al., 2008. Am J Med Genet) who had molecular testing and were each found to have a heterozygous *de novo* pathogenic mutation in *TCOF1* (c.435 5_4356ins14;p.1456Thrfs*18 and c.4321delG;p.E1441fs, respectively). These variants all fall within a region of *TCOF1* where many insertions and deletions are reported, emphasizing the lack of a genotype-phenotype correlation for this disorder. Several hypotheses were considered to explain the extreme manifestation, including some promoter effect on the gene, some position effect on the gene, a polymorphism elsewhere in the gene, a point mutation elsewhere in the gene, a polymorphism in another gene, or a point mutation in another gene such as *POLR1C* or *POLR1D* (Bauer et al). The fact that this individual had WES and a normal chromosomal microarray has eliminated a rare pathogenic mutation (copy number or SNV) in the other Treacher Collins genes as possible explanations.

1153W

Triple diagnosis by whole exome sequencing. *Z. Yüksel[1], J. Balazs[1], M. Calvo[1], A. Megarbane[2], O. Brandau[1], P. Bauer[1,3], A. Rolfs[1,4].* 1) Centogene AG, Rostock, Germany; 2) Institut Jerome Lejeune, Paris, France; 3) Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany; 4) Albrecht-Kossel-Institute, University of Rostock, Rostock, Germany.

Introduction: Whole exome sequencing (WES) was performed only for the index in order to determine the possible genetic cause of symptoms for a 37-year-old Lebanese patient whose parents are consanguineous (2nd degree cousins). The index presented with hydrocephalus, delayed motor and language development, intellectual disability, abnormal gait, macrocephaly (56 cm), microphthalmos, teeth abnormality, early teeth loss, malformation of heart and great vessels, dysmorphic facial features, failure to thrive, short stature, undescended testis and micropenis. **Methods:** WES was performed on Illumina Platform for the index (Nextera Rapid Capture Exome Kit, ~95% of target bases were covered at 10x or higher). All disease-causing variants reported in HGMD®, in ClinVar or in CentoMD® (class 1), as well as all variants with minor allele frequency (MAF) of less than 1% in ExAc database were considered. Evaluation was focused on exons and intron boundaries +/-20 bps. All relevant inheritance patterns were considered and the provided family history and clinical information were used to evaluate eventually identified variants. Only variants related to the phenotype were reported after confirmation by Sanger sequencing. **Results:** Three variants associated with the patient's phenotype were detected. The first variant c.893del, classified as likely pathogenic, was detected in a homozygous state in the *THOC6* gene (NM_024339.3). The *THOC6* gene is associated with Beaulieu-Boycott-Innes syndrome (OMIM #613680), an autosomal recessive neurodevelopmental disorder characterized by delayed development, moderate intellectual disability, and dysmorphic facial features. The second variant classified as likely pathogenic in the *PTCH2* gene (NM_003738.4:c.528del) is associated with autosomal dominant Basal cell nevus syndrome (OMIM #109400), a genodermatosis characterized by multiple early-onset basal cell carcinoma (BCC), odontogenic keratocysts and skeletal abnormalities. The third variant classified as pathogenic in the *EDAR* gene (NM_022336.3:c.486del), is associated with autosomal dominant Ectodermal dysplasia 10A, hypohidrotic/hair/nail type (OMIM #129490). **Conclusion:** Using WES as a diagnostic tool, three different diagnoses have been possibly confirmed for this patient with a complex phenotype.

1154T

Ovotesticular DSD associated with a pathogenic mutation in the BMP15 gene. *S. Albanyan, D. Chitayat.* The Hospital for Sick Children, toronto, ON, Canada.

Ovotesticular disorders of sex development (DSD) are rare disorders of sex development that are characterized by the presence of both well-differentiated ovarian and testicular tissue co-existing in the same individual. A wide spectrum of atypical internal and external genitalia has been reported and the etiology in most cases is not known. We report a patient who presented with DSD including perineal hypospadias, left sided ovotestis with fallopian tube and a hemi uterus and a right sided testes). Chromosome analysis was 46 XX and microarray analysis was normal and female. Whole exome sequencing identified a missense variant in the *BMP15* gene - c.226C>T in a heterozygous state. The bone morphogenetic protein (BMP) family is part of the transforming growth factor-beta superfamily which includes large families of growth and differentiation factors. These proteins are synthesized as prepropeptides, cleaved, and then processed into dimeric proteins. In 1998 Dube et al. identified an additional member of the BMP family, *BMP15*, mapped to Xp11.22. The *BMP15* variants have been previously implicated in premature ovarian failure (POF) and ovarian dysgenesis can be seen as the most severe form of POF. There have been mouse models that revealed that *bmp15* mutants could result from a failure of primary sex determination or from a failure to maintain female sex during the juvenile and/or adult stages and in zebra fish, Bmp15 was found to have no role in either primary or secondary female sex determination or differentiation. However, the role of *BMP15* in human gonadogenesis and sex development has not been determined. To our best knowledge this is the first reported case of ovotesticular DSD associated with *BMP15* variant pointing to the possibility that it may have a major role in female gonadogenesis with its variant being associated with testicular determination in 46,XX chromosome sex. Our study shows again the importance of whole exome sequencing in identifying the molecular etiology of disorders of sex development.

1155F

A genotype-first approach identifies genes contributing to variable phenotypic presentations in a Smith-Magenis Syndrome cohort. *C. Zhang[1], C.M.B. Carvalho[1], C.R. Beck[1], Z.C. Akdemir[1], P. Liu[1], B. Yuan[1], M. Withers[1], S.N. Jhangian[2], D.M. Muzny[2], R.A. Gibbs[2], J.R. Lupski[1,2,3,4].* 1) Department of Molecular and Human Genetics, BCM, Houston, TX; 2) Human Genome Sequencing Center, BCM, Houston, TX; 3) Department of Pediatrics, BCM, Houston, TX; 4) Texas Children's Hospital, Houston, TX.

Smith-Magenis syndrome (SMS) is a developmental disorder caused by either a deletion or a pathogenic loss-of-function mutation in the dosage sensitive *Retinoic acid-induced 1(RAI1)* gene. The majority of patients (~70-80%) with SMS carry a recurrent 3.6 Mb deletion of 17p11.2. The core clinical phenotypes of SMS include global developmental delay, cognitive impairment, craniofacial differences, and skeletal features. Some SMS phenotypes, such as seizures, sensorineural deafness, and scoliosis, affect < 70% of patients. Remarkably, these less frequent phenotypes are usually observed in patients with 17p11.2 deletions rather than *RAI1* variants. Therefore, we hypothesize that deletions unmask recessive mutations (including both common variants and rare variant alleles) in genes mapping proximal to *RAI1*. To test this hypothesis, we performed Whole Exome Sequencing (WES) in 28 patients carrying recurrent SMS deletions. We identified variants in the 53 protein-coding genes in this region based on inheritance models and database annotation; 21 genes were observed to contain variants after filtering. We performed a genotype-first approach by selecting genes with more than two patients presenting hemizygous variants and then evaluated these individuals for shared clinical phenotypes. Our preliminary analyses identified three genes (*FAM83G*, *DRC3* and *MYO15A*) for which a recessive model could potentially explain variable expressivity of certain clinical phenotypes. *FAM83G* functions in noncanonical bone morphogenetic protein signaling. Patients carrying variants in *FAM83G* have both heart and bone development abnormalities. In addition, FAM83G targets ASNS, which is an asparagine synthetase. Accumulation of aspartate and glutamate in brain may result in increased excitability and seizure activity, which was found in patients. *DRC3* is a key component of the nexin-dynein regulatory complex, which may be associated with defects in ciliary motility; remarkably, patients with *DRC3* mutation had ophthalmological and hearing deficiencies. *MYO15A* is a well-known deafness gene, and resides in the DFNB3 locus for recessive deafness. Aggregated analysis also showed variants in both *MYO15A* and *DRC3* in two patients (BAB1198 and BAB1361) who share a majority of the clinical phenotypes[CB1] . In summary, our data provide evidence for recessive variants in three genes that may act either individually or together to cause some of the less frequent clinical manifestations in SMS patients.

1156W

A novel *FBXO28* frameshift in a patient with the predominant features of 1q41-q42 deletion syndrome: A case for haploinsufficiency and primary pathogenicity. *C.D. Balak[1,2], N. Belnap[1,2], K. Ramsey[1,2], A. Siniard[1,2], S. Szelinger[1,2], T. Izatt[1,2], M.E. Parker[4,5], R. Richholt[1,2], M. Naymik[1,2], M. De Both[1,2], W. Jepsen[1,2], I. Piras[1,2], S. Rangasamy[1,2], I. Schrauwen[1,2,6], D.W. Craig[1,2,3], V. Narayanan[1,2], M. Huentelman[1,2].* 1) Center for Rare Childhood Disorders (C4RCD), Translational Genomics Research Institute, Phoenix, AZ; 2) Neurogenomics, Translational Genomics Research Institute, Phoenix, AZ; 3) Department of Translational Genomics, Keck School of Medicine of USC, Los Angeles, CA; 4) Department of Physical Therapy, Texas State University, San Marcos, TX; 5) U.R. Our Hope, Rare Disorder Organization, Austin, TX; 6) Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Chromosome 1q41-q42 deletions have recently been associated with a distinguishable neurodevelopmental disorder in early childhood (OMIM 612530). Features of this deletion syndrome span a small neurodevelopmental spectrum, likely depending on the inclusion or exclusion of specific genes in the 1q41-q42 deletion. Despite this, a predominant phenotype has emerged with core features of developmental delay (DD), intellectual disability (ID), epilepsy and distinct facial features including a depressed nasal bridge, widely-spaced teeth and gingival hyperplasia. In addition, it is also frequently accompanied with CT/MRI brain abnormalities, growth retardation, nail hypoplasia, mixed tone and limb/digit anomalies or contractures. Less frequently, congenital anomalies including diaphragmatic hernias, palate and cardiac defects have been seen. There has been significant interest in mapping the smallest region of overlap (SRO) for this syndrome to identify the pathological gene(s). To date, no definitive gene(s) have been directly attributed to the predominant 1q41-q42 microdeletion syndrome (MDS) phenotype as even the smallest deletions encompass at least 6 genes. Nevertheless recent reports have proposed *DISP1*, *TP53BP2* and particularly *FBXO28* to play roles in pathogenesis through studies of SRO and single gene murine models. Here we present a single-gene, loss-of-function (LoF) event in a 3 year-old female with global DD, ID, seizures, distinct facial features and other traits highly reflective of the overriding 1q41-q42 MDS phenotype. Through whole exome sequencing (WES) of the family trio, we identified a *de novo*, heterozygous *FBXO28* c.972_973delACinsG (p.Arg325GlufsX3) 2 base-pair event in the female proband. The resulting frameshift and premature stop gain three amino acids downstream has not been reported in any genomic reference database and is the first report of a single-gene LoF event in the 1q41-q42 SRO causing the hallmark features of the syndrome's predominant phenotype. These new findings provide strong foundational evidence supporting *FBXO28*'s role as a novel monogenic neurodevelopmental disease gene and further establishes it as the causative gene contributing to the ID, seizure and dysmorphic phenotype previously proposed in 1q41-q42 MDS. In light of this evidence we suggest the name *FBXO28* Haploinsufficiency Syndrome in describing this predominant phenotype in combination with heterozygous *FBXO28* loss of function.

1157T

Exome sequencing in a family with autosomal recessive amelogenesis imperfecta. *G. Chavarria-Soley[1,3], D. Ugalde-Araya[1,3], A. Avila-Aguirre[1,3], C. Mairena-Acuna[1,3], CT. Thiel[2], S. Uebe[2], H. Raventos[1,3], AB. Ekici[2], A. Reis[2].* 1) Escuela de Biologia, Universidad de Costa Rica , San José, Costa Rica, Costa Rica; 2) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany; 3) Centro de Investigacion en Biología Celular y Molecular, Universidad de Costa Rica.

Amelogenesis imperfecta (AI) is a clinically and genetically heterogeneous group of diseases characterized by defects in enamel formation. A non-consanguineous Costa Rican family with hypoplastic amelogenesis imperfecta was identified, which apparently follows an autosomal recessive pattern of inheritance. Initially, four genes that have been reported to harbor disease-causing variants were excluded by Sanger sequencing: *ENAM*, *KLK4*, *MMP20*, and *WDR72*. As a next step, whole exome sequencing (WES) was performed for one affected individual, one healthy sibling, and their healthy parents in search for variants that segregate with the disease according to autosomal recessive inheritance. No homozygous variants were found in the affected individual. On the other hand, five candidate genes were identified for which the affected individual - but not his healthy sibling - is compound heterozygous, and each parent is heterozygous for one of the variants. Sequencing of these candidate genes in the rest of the family, as well as CNV analysis, are ongoing. It is worth mentioning that after WES, a search for variants in the affected individual was performed for other genes that have been reported to have disease-causing variants for autosomal recessive AI: *FAM20A*, *AMBN*, *ITGB6*, *SLC24A4*, and *C4orf26*. None of these genes had variants that segregated with the disease in the family. However, it was found that the affected individual is heterozygous for a nonsense variant, p.C552X, in *ITGB6*. A second mutation in the gene was not present. For this gene as well, a further analysis in order to detect a possible CNV at the second allele is ongoing.

1158F

De novo coding and noncoding variants in novel disease genes account for a significant fraction of isolated and complex congenital diaphragmatic hernia. H. Qi[1], L. Yu[2], A. Kitaygorodsky[1], J. Wynn[2], G. Aspelund[3], F. High[4], M. Longoni[4], P. Donahoe[4], F. Lim[5], T. Crombleholme[5], R. Cusick[6], K. Azarow[6], M. Danko[7], D. Chung[7], B. Warner[8], G. Mychaliska[9], D. Potoka[10], A. Wagner[11], M. ElFiky[12], Y. Shen[1], W. Chung[2]. 1) Departments of System Biology and Biomedical Informatics, Columbia University Medical Center, New York, NY; 2) Department of Pediatrics, Columbia University Medical Center, New York, NY; 3) Department of Surgery, Columbia University Medical Center, New York, NY; 4) Pediatric Surgical Research Laboratories, Massachusetts General Hospital, Boston, MA; 5) Cincinnati Children's Hospital, Cincinnati, OH; 6) Children's Hospital & Medical Center of Omaha, University of Nebraska College of Medicine, Omaha, NE; 7) Monroe Carell Jr. Children's Hospital, Vanderbilt University Medical Center, Nashville, TN; 8) Washington University, St. Louis Children's Hospital, St. Louis, MO; 9) University of Michigan, CS Mott Children's Hospital, Ann Arbor, MI; 10) Children's Hospital of Pittsburgh, Pittsburgh, PA; 11) Medical College of Wisconsin, Milwaukee, WI; 12) Department of Pediatric Surgery, Faculty of Medicine, Cairo University, Cairo, Egypt.

Congenital diaphragmatic hernia (CDH) is one of the most common and often lethal birth defects, characterized by the incomplete formation of the diaphragm. CDH can occur as an isolated defect or in association with additional anomalies. Previous genetic studies of CDH using exome sequencing (WES) support a significant enrichment of coding *de novo* variants in a large number of risk genes, especially in complex cases. To further investigate the genetic architecture of CDH, we performed WES in 196 case-parents trios and whole genome sequencing (WGS) in 195 trios (28 with both WES and WGS), including 133 complex and 230 isolated cases. We observed a significant overall enrichment of likely-gene-disrupting (LGD) (enrichment=1.7, p-value=3e-4) and missense (enrichment=1.3, p-value=1e-5) *de novo* variants. In total, LGD and missense *de novo* variants explain about 25% of both complex and isolated CDH cases, which is larger than previous estimates from a smaller cohort and older sequencing approaches. Interestingly, isolated and complex cases have distinct patterns of LGD and missense variants burden. LGD burden in complex cases are all carried by genes highly expressed in developing diaphragm (mouse E11.5 pleuroperitoneal fold), but in isolated cases, the burden is spread across genes with a broader range of expression in the developing diaphragm. In complex cases, deleterious missense (d-mis, as predicted by CADD or metaSVM) burden is mostly located in haploinsufficient genes ("HIS", defined by ExAC pLI), whereas in isolated cases, d-mis variants in non-haploinsufficient (non-HIS) genes explain about twice as many cases as d-mis in HIS genes. This suggests that isolated CDH cases are more likely to be caused by mutations in non-HIS genes that are dominant negative or oligogenic compared to complex cases. We identified 4 LGD or d-mis *de novo* variants in *MYRF*, a significant association (p-value=2e-10) based on comparison with background mutation rate. *MYRF* is a transcription factor intolerant of mutations (ExAC pLI=1) and highly expressed in developing diaphragm and heart. We observed 2 genes with 2 LGD or d-mis variants: *TUBGCP6* and *WT1*. Finally, 6 genes carrying *de novo* LGD or d-mis variants are involved in post-transcriptional regulation through 3'UTR. In WGS data, we observed a moderate enrichment of noncoding *de novo* variants in 3'UTR regions of genes highly expressed in developing diaphragm, supporting a novel role of 3'UTR dysregulation in CDH.

1159W

Clinical and molecular phenotypes of Coffin-Siris syndrome among UAE population. F. Al Ali, M. Khalifa, F. Bastaki. Dubai Health Authority , Dubai , United Arab Emirates.

Coffin-Siris syndrome (CSS) is a genetically heterogeneous intellectual disability/dysmorphic features syndrome. So far 8 genes have been associated with CSS; *AIRD1B* is most prevalent with no obvious phenotype/genotype correlation. Since its first description in 1970, a significant number of additional and different findings have been reported expanding the phenotype from the "unusual facies, absent fifth finger- and toenails, and mental deficiency" original prototype. CSS is inherited in an autosomal dominant fashion, and all are *de Novo* with one exception of a mother child inheritance. Whole exome sequencing screening of Emirati children with DD/intellectual disability identified 8 patients; 5 males and 3 females in 6 families with CSS. Attached table shows the clinical and molecular characteristics of these patients. *AIRD1B* variants still is the most prevalent present in 5 patients followed by *AIRD1A* in 2 and *SMARCB1 in 1*. All variants were novel and *de novo* 2/8 patients have nails changes, one of them had the characteristic feature of CSS. 6/8 have 'classical/type A' facial features; that include bushy eyebrows, broad nose, broad nasal tip, thick and fleshy lip. Cognitive and developmental delay existed in all cases. Several previously unreported findings as hepatomegaly (3/8), optic atrophy (1/8), transient neonatal hypocalcemia (1/8), and un-ossified fibroma (1/8) were seen among these patients. In addition, these families have the only reported gondal mosaicism associated with CSS where three affected siblings born to unaffected parents..

Family	1	2	3	4	4	4	5	6
Diagnosis	CSS3	CSS2	CSS1	CSS1	CSS1	CSS1	CSS1	CSS2
Alternative title	MRD15	MRD14	MRD12	MRD12	MRD12	MRD12	MRD12	MRD14
Gene	SMARCB1	ARID1A	ARID1B	ARID1B	ARID1B	ARID1B	ARID1B	ARID1A
DNA change	c.110G>A	c.6697C>T	c.4720+1G>A	c.4357C>T	c.4357C>T	c.4357C>T	c.644C>G	c.145G>T
Protein Change	Arg37His	Arg2233Trp	Splice site	Gln1453*	Gln1453*	Gln1453*	Pro215Arg	Glu49*
Parents	-	-	-	-	-	-	-	-
Face type	B	B	A	A	A	A	A	A
5th finger/toe	+							
Microcephaly	-	-	-	-	+	-	-	-
Short stature	-	-	-	+	+	+	+	-
Frequent infections				+	+	+		
Hirsutism	-	-	+	+	+	+		
Sparse scalp hair								
CVS	-	+	-	+	+	+	-	+
Feeding/GI	+	+	+					+
GU/Renal	-	+						
Brain/Cranial	+	+	+	+	+	-	+	+
Vision	-	+/Corneal dystrophy		+	+	-	+/Optic atrophy	
Hearing	-	+						
Other	Macrocornea	Postaxial polydactyly Riedel's lobe	Hypocalcemia Hepatomegaly Natal pads	Asymmetrical face	Asymmetrical face	Asymmetrical face Premature closure coronal / lambdoid	Unossified fibroma	Hepatomegaly High AFP, ferritin, alanine,

1160T

Twelve novel *GUSB* mutations and genotype-phenotype correlation in mucopolysaccharidosis VII (MPS VII) patients. *Q. Abu Ali[1], W. Song[1], S. Daugherty[1], A. Schatz[1], H. Lau[2], C. Dias da Costa[3], T. Levade[4], G. Touati[5], D. Marsden[1], C. Haller[1].* 1) Ultragenyx Pharmaceutical Inc., Novato, CA; 2) New York University Lagone Medical Center, New York, NY; 3) Centro Hospitalar Lisboa Norte EPE, Lisbon, Portugal; 4) Institut Fédératif de Biologie, CHU Toulouse, France; 5) Hôpital des Enfants, CHU Toulouse, France.

MPS VII is a clinically heterogeneous, ultra-rare, autosomal recessive lysosomal disease in which patients are deficient in beta-glucuronidase enzyme (GUS). Nearly fifty unique MPS VII-causing mutations in *GUSB* gene have been reported. In this study, we present twelve novel *GUSB* mutations and provide potential genotype-phenotype correlation. Twenty-three MPS VII patients were enrolled in early access programs and clinical studies investigating recombinant human GUS, vestronidase alfa (UX003), as a potential enzyme replacement therapy (ERT). *GUSB* testing results were available on 74% of patients (17/23). 65% (11/17) were male, 35% (6/17) female; 35% (6/17) were Hispanic, 29% (5/17) Caucasian, 18% (3/17) Asian, and 6% (1/17) each for African American, native Brazilian and unknown ethnicity. 65% of patients (11/17) were unrelated, in addition to three sets of non-twin siblings. All tested patients harbored at least two mutations. Overall, twenty different *GUSB* mutations were identified, all being missense. 60% (12/20) were novel mutations (c.1A>G, p.G97A, p.E141A, p.H142Y, p.W288L, p.N379S, p.M430T, p.S485F, p.V496L, p.G512R, p.W587S, p.I608M). 59% of patients (10/17), including two of three sibling pairs, were heterozygous for at least one novel mutation, and 30% of these patients (3/10) were compound heterozygous for two novel mutations. 82% of patients (14/17) harbored at least one previously published mutation. To assess potential genotype-phenotype correlation, an individual clinical assessment of disease severity at baseline was done, considering dysmorphic, clinical, diagnostic and functional assessments including walking ability and patient reported outcomes. All patients (4/4) heterozygous for these novel mutations: c.1A>G, p.G512R, p.W288L, p.S485F, p.V496L, p.W587S and p.I608M exhibited an overall severe phenotype. All patients (4/4) heterozygous for these novel mutations: p.M430T, p.G97A, p.E141A and p.H142Y exhibited an overall mild phenotype. One sibling pair harbored p.N379S novel mutation, and overall phenotype was mild for one sibling but severe for the other. Our results add twelve novel *GUSB* mutations to previously reported genotype data, and provide further evidence that genetic heterogeneity could contribute to MPS VII clinical heterogeneity. Genotype-phenotype correlation will benefit from future studies aimed at understanding possible role of various genotypes in potential response to investigational ERT in MPS VII patients.

1161F

Clinical exome data analysis and novel variant identification for MPS VII disease in GUSB gene. *A. Bhattacherjee[1], P. Ranganath[2], A. Dalal[1].* 1) The Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India; 2) Nizam's Institute of Medical Sciences, Hyderabad, Telangana, India.

The deficiency of β-glucoronidase enzyme due to mutations in *GUSB* gene (OMIM*611499), leads to Mucopolysaccharidosis Type VII (MPS VII) (OMIM#253220), which is an autosomal recessive lysosomal storage disease characterized by highly variable phenotype, ranging from severe lethal hydrops fetalis to mild forms with survival into adulthood. At least 55 mutations in the *GUSB* gene have been found to cause MPS VII of which most are single nucleotide changes which reduce or eliminate the function of β-glucuronidase. We report a case of 26 years old female, at 18 weeks of pregnancy with nonimmune fetal hydrops. This was the second conceptus of a third degree consanguineous couple with a previous history of a female fetus terminated at 20 weeks with fetal hydrops. Enzyme analysis in amniotic fluid culture showed β-glucoronidase level of 0 nmol/hr/mg (Control – 84, normal: 75 – 650 nmol/hr/mg) suggestive of MPS VII. All other investigations including fetal karyotype, MLPA for α – thalassemia and TORCH profile were normal. *GUSB* has large number of pseudogenes and hence molecular analysis by Sanger sequencing is very difficult. DNA sample was extracted from the amniotic fluid (AF) with the consent of the patient and the whole exome sequencing of the DNA was performed using Illumina HiSeq platform. Raw data was analysed using BWA-GATK pipeline and reads were aligned to the human reference genome GRCh37/hg19 to generate the variant call file (VCF) in the downstream analysis. Annotation was done using Annovar. A homozygous missense mutation in exon 5 of the *GUSB* gene (NM_000181:c.893C>T; Depth: 48X) that results in amino acid change from Alanine to Valine at the 298th position (p.Ala298Val) was identified. This variant has not been reported as a normal variant in public databases like 1000 Genomes, Exome Variant Server (EVS), Exome Aggregation Consortium (ExAC), dbSNP, ClinVar and HGMD. *In-silico* pathogenicity prediction programs like Combined Annotation Dependent Depletion (CADD phred score 24.8) and Mutation taster predicted it to be disease causing and is assumed to be the cause of MPS VII in this fetus. This report signifies the importance of next generation sequencing technologies for analysis of patients with diseases with pseudogenes where Sanger sequencing is difficult.

1162W

Mutations in the condensin II component NCAPG2 cause autosomal recessive neurodevelopmental syndrome. *T.N. Khan[1], K. Khan[1], H. Reynolds[1,2], A. Sadeghpour[1], M.T. McDonald[3], R.N. Goldberg[3], W.B. Gallentine[3], N. Katsanis[1], E.E Davis[1], Task Force for Neonatal Genomics.* 1) Duke Center for Human Disease Modeling, Duke University, Durham, NC 27701; 2) Furman University, Greenville, SC 29613; 3) Department of Pediatrics, Duke University, Durham, NC 27710.

Neurodevelopmental disorders (NDDs) are clinically and genetically heterogeneous, and the number of known NDD-associated genes is rapidly expanding with the application of whole-exome sequencing in patients. Here we report on two unrelated pediatric cases presenting with overlapping central nervous system, ocular, renal and digit abnormalities. Both individuals had non-diagnostic chromosomal microarray results; female DM516 was array negative, and female DM074 harbored a maternally-inherited heterozygous *NPHP1* deletion. We performed trio-based whole exome sequencing in both families and, after variant filtration, we identified rare missense variants in *NCAPG2* segregating with disease in an autosomal recessive pattern of inheritance. NCAPG2 plays an important role in chromosomal condensation and segregation during cell division. To investigate the functional relevance of *NCAPG2* to NDDs and the pathogenicity of missense variants identified in patients, we conducted morpholino-based suppression of this gene's sole zebrafish ortholog. The morphant zebrafish displayed patient-relevant phenotypes such as microcephaly and renal anomalies. These abnormalities were rescued by co-injecting the morpholino with wild type human *NCAPG2* mRNA, demonstrating morpholino specificity. Coinjection of morpholino with *NCAPG2* mRNA harboring the patients' specific missense variants did not rescue the head size and renal anomalies as efficiently as wild type *NCAPG2* mRNA, suggesting loss of function of *NCAPG2* as a plausible disease mechanism. Additionally, CRISPR-Cas9 editing of *ncapg2* in zebrafish recapitulated the microcephaly and renal anomalies. We also observed increases in apoptosis and in cell proliferation, as well as gross structural abnormalities of the central nervous system in *ncapg2* zebrafish morphants and CRISPR F0 mutants supporting further a role for *ncapg2* in clinical phenotype. Finally, we tested whether sensitization at the *NPHP1* locus, the most common contributor of juvenile nephronophthisis, could explain a more severe renal phenotype in DM074. *In vivo* co-suppression of *nphp1* and *ncapg2* in zebrafish resulted in significantly more dysplastic renal tubules in comparison to either gene model alone. Together, our genetic findings and subsequent functional modeling in zebrafish suggest that loss of function of *NCAPG2* results in a neurodevelopmental syndrome and highlight how multi-gene interactions can exacerbate endophenotypes in rare syndromes.

1163T

Frequency of germline pathogenic variation in *NF1* and eight other RASopathy genes in the Exome Aggregation Consortium (ExAC) database: A pilot study. *A. Pemov, K. Jung, D.R. Stewart.* NIH/NCI, Bethesda, MD.

ExAC is the largest public database of human variation, comprising high-quality variant calls across 60,706 human exomes. Importantly, the curators of the database "have made every effort to exclude individuals with severe pediatric diseases from the ExAC data set," which makes the database a valuable resource for studying variation in Mendelian disease genes in unaffected populations. The RASopathies are a group of inherited disorders caused by mutations in predominantly RAS-MAP kinase pathway genes. We downloaded all coding, non-synonymous germline variants in nine genes: *NF1, HRAS, KRAS, NRAS, BRAF, RAF1, SPRED1, MAP2K1* and *MAP2K2* from the database (excluding 7,601 The Cancer Genome Atlas (TCGA) samples). There were 1,188 mutations in the nine genes (range: 36 (*NRAS*) to 498 (*NF1*), of which almost half were found only once in the dataset and 76% had frequency below 0.01%. We then analyzed the variants using InterVar (ACMG/AMP 2015 guideline) method and phenotype-genotype databases to predict the pathogenicity of the variants. As expected, the majority of the variants were classified as "benign," "likely benign," and "variants with unknown significance." We also compared the ExAC variants with the variation annotated in the Human Gene Mutation Database (HGMD). In the nine genes, we found 60 variants present in both HGMD and non-TCGA ExAC. Forty-one variants were found in *NF1* alone, of which nine were loss-of-function or canonical splice-site mutations. After excluding variants for which the evidence of pathogenicity was less convincing (labeled "DM?" in HGMD), there were 28 disease-causing mutations in *NF1*. Assuming that there was one *NF1* mutation per individual, this observation implies that in ExAC the frequency of *NF1* variants associated with neurofibromatosis type 1, is $28/(60,706_{Total}-7,601_{TCGA})=0.00053$ (~1/1,900), which is within the range of published NF1 prevalence in the general population . There were eight "pathogenic" or "likely pathogenic" mutations in four genes (*H-, K-, NRAS* and *BRAF*), all of them were rare, with frequencies ranging $8.2 \times 10^{-6} – 2.5 \times 10^{-5}$. **Conclusions**: In this pilot study of a large exome database, we found that prevalence of pathogenic variation in *NF1, BRAF, H-, K-* and *NRAS* is close to published estimates. Future work focuses on phenotype spectrum and penetrance of NF1 and other RASopathies in large, exome-sequenced populations linked to an electronic medical record.

1164F

Enhancement of hepatic autophagy increases ureagenesis and protects against hyperammonemia in a mouse model of ornithine transcarbamylase deficiency and other models of secondary hyperammonemia.
L. Soria[1], G. Allegri[2], D. Melck[3], N. Pastore[4], P. Annunziata[1], D. Paris[3], E. Polishchuk[1], E. Nusco[1], B. Thöny[2], A. Motta[3], J. Häberle[2], A. Ballabio[1,4,5], N. Brunetti-Pierri[1,5]. 1) Telethon Institute of Genetics and Medicine, Pozzuoli, Naples, Italy; 2) Division of Metabolism, University Children's Hospital Zurich and Children's Research Center, Zurich, Switzerland; 3) Institute of Biomolecular Chemistry, National Research Council, Pozzuoli, Italy; 4) Department of Molecular and Human Genetics; Baylor College of Medicine; Houston, TX, USA; 5) Department of Translational Medicine, Federico II University, Naples, Italy.

Ureagenesis is the main pathway for waste nitrogen removal in mammals. Ammonia crosses the blood-brain barrier and is neurotoxic at high concentration. Systemic ammonia is elevated in patients with inherited or acquired impairments of ammonia detoxification, such as urea cycle disorders, organic acidemias, or acute and chronic liver diseases. Regardless from its etiology, hyperammonemia may result in irreversible brain damage if not treated early and thoroughly. Current treatments are often inadequate. In this study, we show that hepatic autophagy is triggered in an mTOR-dependent manner by hyperammonemia and deficiency of autophagy impairs ammonia handling. Rapamycin is a drug already used in humans that is known to activate autophagy. Rapamycin reduced blood ammonia in C57BL/6 wild-type mice fed with ammonia enriched diet. Moreover, rapamycin treatment of abnormal skin and hair (*spf[ash]*) mice that have markedly reduced ornithine transcarbamylase (OTC) activity increased ^{15}N-labeled urea enrichment and normalized orotic acid concentration in whole blood/dried blood spots. In conclusion, these data show that autophagy is an important mechanism for ammonia detoxification and its enhancement has potential for therapy of both primary and secondary causes of hyperammonemia.

1165W

Search for the mutation causing the ThoracoAbdominal Syndrome (TAS), an X-linked dominant disorder. *P. Majdalani[1,2], R. Parvari[1,2].* 1) The Shraga Segal Department of Microbiology,Immunology & Genetics; Faculty of Health Sciences (A.B.), Ben-Gurion University of the Negev, Beer-Sheva, Israel; 2) National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer Sheva, Israel.

Introduction: The Thoracoabdominal Syndrome (TAS) is a very rare congenital X-linked dominant disorder presented only in one Jewish North African family. The features of the syndrome are ventral herniae (midline abdominal wall defect), antero-lateral diaphragmatic herniae (manifested almost exclusively in affected males), hypoplastic lung, and associated cardiac anomalies. The aim of this study is to identify and characterize the mutation causing TAS. Materials and Methods: The initial search for the mutation causing the syndrome was done by linkage analysis on the X chromosome using 27 individuals of the TAS family. Genome sequencing of two of the family members was performed on the identified TAS interval on chromosome X. Results: The TAS interval was localized to Chr:Xq27.1 in an interval of 1.06 Mb. Three genes and two miRNA are encoded in the interval. No mutation causing variants were identified in these sequences. Large deletions, insertions and translocations were negated by the sequence. Therefore, further search for indel variants in the non-coding DNA was pursued. Variations presenting in more than 1 person in the population databases and in Short Tandem Repeats (STR) regions were excluded, leaving 4 variations that are further characterized for the possibility of causing the TAS phenotype. Conclusion: Further analysis using the Chromosome Conformation Capture-on-Chip (4C-seq) technology will be held on the mutation causing variant for determining long-range chromatin interactions and the gene that may be affected.

1166T

Germline mutations associated with polycomb repressive complex 2 cause Weaver syndrome. *E. Imagawa[1], K. Higashimoto[2], Y. Sakai[3], C. Numakura[4], N. Okamoto[5], S. Matsunaga[6], A. Ryo[6], Y. Sato[7], M. Sanefuji[3], K. Ihara[8], Y. Takada[9], G. Nishimura[10], H. Saitsu[11], T. Mizuguchi[1], S. Miyatake[1], M. Nakashima[1,11], N. Miyake[1], H. Soejima[2], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga, Japan; 3) Department of Pediatrics, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 4) Department of Pediatrics, Yamagata University School of Medicine, Yamagata, Japan; 5) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 6) Department of Microbiology, Yokohama City University School of Medicine, Yokohama, Japan; 7) Department of Molecular Biology, Yokohama City University School of Medicine, Yokohama, Japan; 8) Department of Pediatrics, Faculty of Medicine, Oita University, Yufu, Japan; 9) Department of Pediatrics, Japanese Red Cross Fukuoka Hospital, Fukuoka, Japan; 10) Department of Pediatric Imaging, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 11) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan.

Weaver syndrome (WS; MIM# 277590) is a rare congenital overgrowth disorder characterized by pre/postnatal overgrowth, accelerated osseous maturation, characteristic craniofacial features, intellectual disability, limb anomalies and susceptibility to cancer. To date, germline heterozygous mutations in *EZH2* (enhancer of zeste homolog 2; NM_004456.4) and *EED* (embryonic ectoderm development; NM_003797.3) have been reported in WS. EZH2 and EED are core components of the polycomb repressive complex 2 (PRC2), which play a role in regulating epigenetic gene repression by catalyzing trimethylation of lysine 27 of histone H3 (H3K27me3). Here, we analyzed three probands with clinically suspected WS using whole-exome sequencing and identified three mutations: a 25.4-kb deletion partially involving *EZH2* and *CUL1*, a missense mutation (c.707G>C, p.Arg236Thr) in *EED* and a missense mutation (c.1829A>T, p.Glu610Val) in *SUZ12* (suppressor of zeste 12 homolog; NM_015355.2) inherited from her father with a mosaic mutation. SUZ12 is another component of PRC2, but germline mutations in *SUZ12* have never been reported in humans. To investigate the functional effects of the *EED* and *SUZ12* missense mutations, we performed rescue experiments to evaluate the H3K27me3 levels using the Tet-on system for 293 T-REx cells. Our results showed that the EED or SUZ12 mutants failed to rescue and produced only similarly low levels of H3K27me3 as the mock transfections. These data indicate that loss-of-function mutations of PRC2 components are an important cause of WS. Interestingly, *SUZ12* at 17q11.2 is mapped approximately 560 kb downstream of *NF1*, whose mutations cause neurofibromatosis type 1 (NF1; MIM# 162200). 5-10% of NF1 patients exhibit 1.2–1.4 Mb microdeletions encompassing *NF1* and *SUZ12*. The clinical phenotypes of patients with the microdeletion are significantly different from those with *NF1* point mutations: overgrowth, learning disabilities and WS-like features (macrocephaly, prominent forehead, hypertelorism long philtrum, micrognathia, and large ears) have been often described in *NF1* microdeletion patients. Therefore, *SUZ12* deletions may possibly contribute to the unique phenotypes in *NF1* microdeletion.

1167F

Functional analysis of a novel *TUBA1A* gene mutation associated with cerebral dysgenesis and cleft palate. *R. Peretz[1], S. Chung[4], A. Derrick[4], M. Rees[4], R. Stratton[2], C. Wang[3].* 1) Department of Pediatrics, Driscoll Children's Hospital, Corpus Christi, TX; 2) Department of Medical Genetics, Driscoll Children's Hospital, Corpus Christi, TX; 3) Department of Neurology, Driscoll Children's Hospital, Corpus Christi, TX; 4) Neurology and Molecular Neuroscience Research, Institute of Life Science, Swansea University Medical School, South Wales, UK.

Tubulinopathies are a group of autosomal dominant brain malformations due to structural defects in microtubules caused by mutations in genes encoding tubulin subunits. Characteristic features of tubulinopathies include cortical dysgenesis - such as lissencephaly, pachygyria, and polymicrogyria - and malformations of the corpus callosum and basal ganglia. We report a newborn with severe lissencephaly, polymicrogyria, agenesis of the corpus callosum and cerebellum, and a unique extra-cranial manifestation of cleft palate. Genetic screening with a lissencephaly gene panel revealed a novel heterozygous missense mutation in the *alpha-tubulin 1A (TUBA1A)* gene: c.1169G>C, p.Arg-390Pro (R390P). This single amino acid substitution of a positively charged arginine with a neutrally-charged proline strongly suggests reconfiguration of the protein structure at this locus. Thus, we conducted FLAG-tagged site-directed mutagenesis of the wild-type TUBA1A construct to create a TUBA1A p.Arg-390Pro construct. Immunocytochemistry in HEK cells, in vitro polymerization assays, and protein expression studies were utilized to analyze the functional consequences of this novel mutation on alpha-tubulin. This study expands the clinical phenotypes of tubulinopathies and provides insight into the function of the protein domain spanning amino acid Arg390 of alpha-tubulin.

1168W

A prospective study of natural history, physiology, and biochemistry of propionic acidemia. *O.A. Shchelochkov, J. Sloan, S. Ferry, C. Van Ryzin, J. Gagne, I. Manoli, C.P. Venditti.* NHGRI/NIH, Bethesda, MD.

Propionic acidemia (PA) has a heterogenous spectrum of laboratory abnormalities and multiorgan pathology. To define the genetic, laboratory, and clinical variables in PA, we initiated a prospective study of the natural history, physiology, microbiome and biochemistry of PA (ClinicalTrials.gov Identifier: NCT02890342). During protocol visits, patients receive clinical, nutritional, imaging, and laboratory evaluations, with biospecimen collection and banking. Since fall 2016, we have evaluated 8 PA patients, ages 3 – 52 years (4 females). Molecular confirmation was made in 7/8 (4 had mutations in *PCCA* and 3 in *PCCB*). Three patients had short stature (z-score ≤ 2), 3 patients were overweight or obese, one patient was underweight. On EKG, 4 patients had non-specific repolarization abnormalities and 2 patients had a QTcB interval greater than the age- and sex-specific cutoff. All had a normal ejection fraction (range 54-65%). Cystatin C-based glomerular filtration rate ranged from 43 to 103 mL/min/1.73 m^2. In 5/6 patients, dual-energy X-ray absorptiometry showed bone density z-score ≤ 2. Hearing loss was present in 4/8 patients. Eye evaluation revealed normal retina in all and subtle optic pallor in 2/8 patients. Epilepsy was reported in 3/8 patients. Four patients completed brain MRI revealing cerebral atrophy and non-specific T2 signals in white matter. Neuro-cognitive studies showed IQ in the normal range in 1/6 patients and moderate to severe intellectual disability in 5/6. Five of 8 patients were diagnosed with autism spectrum disorder. Six of 8 patients received medical foods and 4/8 had a G-tube. One patient underwent a liver transplant for metabolic instability and one received a kidney transplant for end-stage renal disease. To non-invasively assess the whole-body activity of propionyl-CoA carboxylase, we measured 1-^{13}C-propionate oxidation in 7 patients. Patients with a clinically severe form of PA had depressed oxidation (7.04±3.38 % of dose metabolized), while two clinically milder patients were more variable (15.7% and 40.5%) compared to controls (47.7±4.5%). Thus, increased 1-^{13}C-propionate oxidation rate appeared to correlate with a milder clinical course. In summary, our initial experience highlights that all PA patients are at risk for developing multi-organ involvement warranting prospective clinical surveillance and management, and identifies clinical and metabolic domains that could be useful to predict the course and assess new therapies.

1169T

SOPH syndrome: Multisystem disorder with facial dysmorphism, skeletal dysplasia, episodic liver failure, immune dysfunction and intellectual disability. *A. Yadav, R.J. Hopkin, E.K. Schorry.* Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

SOPH syndrome is an AR condition with skeletal dysplasia, brachydactyly, Pelger-Huet anomaly and episodic liver failure triggered by fever due to biallelic *NBAS* mutations. The *NBAS* protein (NAG) is a subunit in the syntaxin-18 complex implicated in ER/Golgi transport. Missense mutations in *NBAS* result in decreased protein product and render syntaxin-18 complex thermolabile. Hence, febrile illnesses may cause disordered ER/Golgi transport. Our proband presented at 16 years during her 4th episode of acute liver failure associated with acute renal failure. She had three prior episodes of liver failure. She was born at 26 weeks to a non-consanguineous Caucasian couple. She had intellectual disability (IQ 72). She had a triangular face, prominent forehead, pointed chin, long fingers, prominent PIP joints and short stature (Z-score -3). Skeletal survey showed long phalanges, thin long bones, anterior subluxation at C2-C3, right thoracic curvature, gracile ribs and narrow upper chest. NK cell function was decreased. Interestingly, IgG was low and normalized later while IgA was low during recovery. She was followed for failure to thrive & precocious puberty. She had hypothyroidism, insulin resistance and PCOS. Novel compound heterozygous missense mutations in the *NBAS* were found by whole exome sequencing. SOPH syndrome was first described by Maksimova et al (2010) in an isolated Yakut population. Clinical features in 33 patients included short stature, facial dysmorphism, senile appearance, brachydactyly, optic atrophy and Pelger-Huet anomaly. Homozygosity mapping found homozygous missense mutations in *NBAS*. Biallelic mutations in *NBAS* were found causative in infantile onset recurrent acute liver failure in a separate cohort of 11 German individuals by Haack et al (2015). Affected individuals had episodes of acute liver failure triggered by fever, with interval recovery of liver function. They did not have any features of SOPH syndrome. Soon after, Segarra et al reported two patients with biallelic *NBAS* mutations and multisystem disease that included progeroid appearance, recurrent acute liver failure, immune dysfunction, hypogammaglobulinemia, skeletal dysplasia with C1-C2 instability. Our case supports the theory that SOPH syndrome and recurrent episodes of liver failure may be part of the same spectrum and can coexist in some patients. In addition, she has unique features that raise questions for further functional studies of this fascinating condition.

1170F

The genetic architecture of Bardet Biedl Syndrome. *M. Kousi[1], A. Ozanturk[1], S. Frangakis[1], J. Muller[2,3], A. Sadeghpour[1], R.A. Lewis[4,5], H. Dollfus[2], E.E. Davis[1], N. Katsanis[1].* 1) Center for Human Disease Modeling, Duke University School of Medicine, Durham, NC; 2) Laboratoire de Génétique Médicale, Institut de Génétique Médicale d'Alsace, INSERM U1112, Université de Strasbourg, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Strasbourg, France; 3) Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 5) Department of Ophthalmology, Baylor College of Medicine, Houston, Texas, USA.

The importance of genetic background on driver mutations is well established. However, understanding how such background interacts with Mendelian loci remains challenging. Bardet-Biedl syndrome (BBS) has been a useful model to study such phenomena. Mutations in >20 genes are necessary for pathogenesis, while numerous BBS patients have also been reported who, in addition to recessive drivers, bear *bona fide* pathogenic alleles in other BBS genes. Here, through sequencing and array analysis, we detected all rare alleles impacting the coding regions 17 BBS genes in two cohorts of BBS patients (201 families). We stratify the distribution of recessive mutations and report a new candidate BBS locus, *CEP76*. Further, we show that in a cohort of 74 families with diagnosed recessive alleles, there is significant enrichment of rare alleles (excluding diagnostic recessive variants) compared to controls (n= 380), with > 40% of BBS patients carrying 1-4 additional rare variants in BBS genes. This observation persisted when we restricted our analysis to likely pathogenic variants; was replicated in a second cohort of 101 recessive families; and enabled the generation of a mutational distribution network for BBS across 175 families. Importantly, this enrichment was restricted to BBS genes. We found no differences in an equivalent-sized set of genes mutated in primary ciliary dyskinesia, nor did we see signal when considering other functional particles in the ciliary apparatus. Taken together, our data inform the mutational architecture of BBS and intimate specificity in the network, where the effect of *trans* variation is enriched in BBS genes, but not the overall ciliary proteome. We propose that such architecture is unlikely to be unique to BBS and that focused analysis of first-order interactors in other pleiotropic genetic disorders will potentially reveal similar enrichment and distribution.

1171W

Clinical and genetic characteristics of seven patients with Floating-Harbor syndrome. *P. Castro[1], R. Honjo[1], G. Lopes[1], L. Albano[1], S. Sugayama[1], D. Bertola[1], T. Homma[2], B. Freire[2], A. Jorge[2], C. Kim[1].* 1) Genetics Clinic Unit, Child Institute, University of São Paulo School of Medicine, São Paulo, Brazil; 2) Developmental Endocrinology Unit (LIM/42), University of São Paulo School of Medicine, São Paulo, Brazil.

The Floating-Harbor syndrome (FHS) is a rare condition, with only about 50 confirmed individuals to date. It is characterized by a particular gestalt, with typical facial dysmorphisms, short stature with delayed bone age and speech delay. Our study included seven patients assisted at the Genetics and Endocrinology Services in Hospital das Clinicas, São Paulo - Brazil, between 1992 and 2017, with clinical and molecular diagnosis of FHS. Analysis of the clinical characteristics of the patients was carried out by a systematic survey of their medical records. The seven patients were 2 females and 5 males, with current ages ranging from 5-29 years old (mean 14.5y). All of them had proportionate short stature (mean SD: -3.14) and delayed bone age greater than 2SD. Six had characteristic facies, with triangular face, deep eyes, thin lips and wide mouth. The typical nose, narrow at the root and broad at the tip, with low-hanging columella and smooth philtrum was observed in five patients, though all of them had a somehow prominent nose. Intellectual disability and language delay were also universal findings ranging from severe cognitive impairment to isolated speech delay. Only three patients presented seizures. Motor development was normal in five patients and only slightly delayed in the remaining two. Microcephaly was reported in four patients. Recurrent ear infections, another common trait associated to FHS, was reported in only three patients. Hirsutism was present in three patients. Other structural anomalies, such as heart defects and cleft lip/palate were observed in isolated patients. Molecular testing revealed known pathogenic variants in *SRCAP in all, except one case, which presented a novel variant*. All mutations were located in exon 34, three in c.7330C>T and three in c.7303C>T, which lead to premature stop codons: p.Arg2444* and p.Arg2435*. These mutations were already reported in literature. The last patient had a novel mutation, a *de novo* 1-bp duplication at nucleotide 7262 causing a frameshift predicted to result in a premature termination codon (p.R2421fs). Overall, the clinical characteristics of our patients were similar to those previously reported cases. (SUPPORT: CNPQ).

1172T

An integrated clinical program and crowdsourcing strategy for genomic sequencing and Mendelian disease-gene discovery. *A. Haghighi[1,2,3,4,5], C. Cassa[2,4], D. Vuzman[2,4], A. Toth-Petrozy[2], C. MacRae[3], C. Seidman[1,3,4], S. Sunyaev[2,4], R. Maas[2], J. Krier[2], Brigham Genomic Medicine.* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Broad Institutes of Harvard and MIT, Cambridge, MA; 5) Partners HealthCare Laboratory for Molecular Medicine, Cambridge, MA.

Mendelian gene discovery is not only potentially beneficial to individuals with these rare disorders, but it also provides insight into more common forms of similar disease phenotypes. Here, we describe an integrated clinical program using whole exome and whole genome sequencing (WES/WGS) for Mendelian disease gene discovery as well as cutting edge science in genome analytics (including genome and exome sequencing) and biological models (mouse and zebrafish mutants, and human iPS cells) to: (1) discover the genetic cause of undiagnosed monogenic diseases, and (2) strategize improved or novel treatment for patients. This program includes specific case ascertainment parameters to select cases not solved by conventional clinical sequencing, a WES/WGS computational analysis pipeline that is optimized for Mendelian gene discovery with callers tuned to specific inheritance modes, an interdisciplinary crowdsourcing strategy for genomic sequence analysis, matchmaking for additional cases, and integration of the clinical findings regarding gene causality with the clinical management plan. Clinicians and research scientists review the clinical and genomic findings and to contribute to discussion about each open case, channeling the clinical and scientific expertise that resides within the local and extended biomedical communities into solving these challenging genetic cases. This efficient platform has promoted education and engagement of faculty and fellows throughout our affiliated academic medical centers and has served as a genomic sequencing re-analysis core for the Harvard Undiagnosed Disease Network clinical site and the NIDCR FaceBase Consortium. Examples of (published) BGM high impact discoveries of disease-associated genes (with different inheritance modes; i.e., autosomal dominant and recessive, X-linked and de novo mutations), include PIEZO2 (where we identified that gain-of-function missense mutations cause a subtype of distal arthrogryposis, a musculo-skeletal contracture and respiratory disease), LOX (in which mutations cause familial aortic dissection), WISP3 (in precocious arthritis) and CHST11 (in T cell lymphoma with limb abnormalities). This integrated monogenic disease gene discovery model could be readily adopted by other academic medical centers, and effectively leverages existing clinical and research expertise in the task of diagnosing and treating these debilitating conditions.

1173F

Insights from applying ClinGen framework in evaluating the clinical validity of 30 published studies of gene-disease relationship of monogenic disorders. *R. Attali, O. Farchy, A. Rafaeli, S. Tzur.* Genomic Research Department, Emedgene Technologies, Tel Aviv, Israel.

The Clinical Genome Resource (ClinGen) has recently developed a framework for evaluating the clinical validity of gene-disease relationship in monogenic disorders. Using this semi-quantitative measurement framework, a scientist can evaluate the relevant genetic and experimental evidence for supporting a certain gene-disease relationship, and to calculate the strength of evidence based on qualitative classifications: "Definitive", "Strong", "Moderate", "Limited", "No Reported Evidence", or "Conflicting Evidence". In an attempt to assay this framework, we evaluated 30 studies (in journals with impact factor greater than 5) that described new gene-disease connections. The aim of our work was twofold: to evaluate the usability of the ClinGen framework and to estimate its power when applied to a single first gene-disease article. Overall, we confirm the usability and the importance of this framework approach in clinical validation of new gene-disease connections. The evaluation process is well detailed and accurate and therefore can be used as a powerful source for gene-disease validity assessment by multiple users. Potential drawbacks are the frequent inconsistencies between curators that will often require ClinGen expert decision, and the technical calculation of the segregation Lod-score for non-experts. Interestingly, in our review 40% of the evaluated articles were scored with "Strong" strength of evidences, whereas 42% were "moderate", and only 18% with "Limited" strength. This tends to show the high standards demanded by most editors and peer-reviewers. The generalized use of this framework will indescribably benefit scientists in building their research evidence, and it can also be used as a standard evaluation in peer reviewed journals for submitters and publishers. Particularly this framework could help defining a strict threshold for the recognition of a new gene-disease connection in databases such as OMIM.

1174W

Evaluating the evidence available for associating genes of unknown significance (GUS) with disease phenotypes: Review of 100 studies. *S. Tzur, N. Mizrahi, E. Feldman, R. Attali.* emedgene, Tel Aviv, Israel.

A large portion of unresolved exome-analysis cases in rare disease research are explained by mutations in genes of unknown significance (GUS). While various types of evidence are available for most human genes (biochemical function, protein interaction, expression, animal model, pathways, and gene family), it is unclear how often these can be used as indirect supportive data for the establishment of a new gene-disease relationship. We reviewed 100 recent publications from journals with an impact factor greater than 5 that described the discovery of new genes causing monogenic diseases. We listed the types of evidence that were available for each gene prior to the publication, and evaluated whether they could be used as indirect supporting evidence for connecting the GUS with disease phenotypes (i.e a shared pathway with another gene causing disease, or an animal knockout model resulted with similar phenotype). We found that in 90% of the studies, biochemical function data was available for the gene under investigation. Most interestingly, in 73% of these cases the data could be used as supportive indirect evidence for the new gene-disease suggested relationship. Additional categories were also examined: Pathways (80% availability, in 76% as supportive), animal model (75% availability, in 76% as supportive), gene family (54% availability, in 50% as supportive), protein interactions (45% availability, in 70% as supportive). On average for each case, 4.3 different types of evidence were available, 2 were supportive for establishing the new gene-disease connection. This study quantifies for the first time how prior available evidence can contribute to the identification of new gene-disease association. These data demonstrate how basic scientific evidence is highly required to establish a gene-disease relationship, and pinpoints the importance of availability of basic molecular research data as well as knockout animal projects. The implementation of these insights in rare disease research protocols would improve the ability to obtain evidence needed for a new gene-disease relationship.

1175T

Oral cavity findings in *A2ML1*-related otitis media. *R.L.P. Santos-Cortez[1,2,3], M. Steritz[1], J.P. Magno[4], M. Pedro[3], S.M. Lagrana-Villagracia[3], T.K.L. Yarza[3,4], M.L.C. Tantoco[3,4], A.L. Chan[3,4], M.R.T. Reyes-Quintos[3,4,5], E.G.dV. Llanes[3,4], A.T.R. Santos[6], R.A.P. Santos[6], R.A.Q. Rosanes[7], C.M. Chiong[3,4,5].* 1) Department of Otolaryngology, University of Colorado School of Medicine, Aurora, CO; 2) Center for Children's Surgery, Colorado Children's Hospital, Aurora, CO; 3) Philippine National Ear Institute, University of the Philippines Manila - National Institutes of Health; 4) Philippine General Hospital, Manila, Philippines; 5) Newborn Hearing Screening Reference Center, University of the Philippines Manila - National Institutes of Health; 6) Nordhoff Craniofacial Center, Manila, Philippines; 7) College of Dentistry, University of the Philippines Manila.

Otitis media is the top reason for health care visits and antibiotic use among children. Previously the *A2ML1* variant c.2478_2485dupGGCTAAAT (p.Ser-829Trpfs*9) was identified to confer increased susceptibility (odds ratio 3.7) to early-onset otitis media of various types (chronic, acute, effusive, healed) in an indigenous Filipino population with a 50% otitis media prevalence. This variant is absent in public variant databases and in Filipino controls. The *A2ML1* gene encodes alpha-2-macroglobulin-like protein, which is highly similar to serine protease inhibitor alpha-2-macroglobulin. In *A2ML1* variant carriers with chronic otitis media, middle ears have increased relative abundance of *Fusobacterium*. In the head and neck, *F. nucleatum* is a common pathogen for gingivitis and periodontitis and has clonal similarity in nasopharynx and saliva. *F. nucleatum* produces serine protease fusolisin and its growth is inhibited by protease inhibitors. We hypothesize that the oral cavity is a potential source of pathogens for the middle ear. The oral cavities of 52 indigenous Filipinos were screened for dental caries, gum bleeding, gingivitis and periodontitis. Salivary DNA samples were Sanger-sequenced for the *A2ML1* variant. Gingivitis increases risk for otitis media (odds ratio 6.4, 95%CI: 1.3, 42.5, $p=0.009$). There is no direct relation between the *A2ML1* variant and oral cavity status. However by multivariate regression analysis, both the *A2ML1* variant ($p=0.02$) and gingivitis ($p=0.002$) increase risk for otitis media. Age and sex were not significant in the regression model. Our dental screening suggests that the oral cavity is a potential source of middle ear pathogens, likely using the nasopharynx as the main route of spread. The lack of protease inhibition due to the *A2ML1* variant is hypothesized to alter the activity of bacteria which rely on serine proteases for growth. RT-PCR showed that the shorter *A2ML1* isoform is expressed in salivary gland. This study will also describe salivary RNA expression according to *A2ML1* genotype.

1176F

Deep phenotyping for patients by patients: A lay-friendly version of the Human Phenotype Ontology. *M.A. Haendel[1,7], N. Vasilevsky[1], M.E. Engelstad[1], E.D. Foster[9], T. Groza[8], J. McMurry[1], C.J. Mungall[6], C. Brownstein[5,7], I. Holm[5,7], C. Boerkoel[2], K. Rageth[2], M. Moore[2], P. Robinson[3], S. Köhler[4].* 1) Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR; 2) Sanford Health Imagenetics, Sioux Falls, SD; 3) Jackson Laboratory of Genomic Medicine, CT; 4) Charite Hospital, Berlin, Germany; 5) Boston Children's Hospital, Boston, MA; 6) Lawrence Berkeley National Laboratory, Berkeley, CA; 7) Undiagnosed Disease Network; 8) Garvan Institute, Sydney, Australia; 9) Indiana University School of Medicine, Indianapolis, IN.

The Human Phenotype Ontology (HPO) has become the de facto standard representation of clinical "deep phenotype" data for computational comparison of abnormalities and for use in genetic disease diagnostics. The HPO enables non-exact matching of sets of phenotypic features (phenotype profile) against known diseases, other patients, and model organisms, and algorithms for inclusion of phenotype profiles. The algorithms based on HPO have been implemented into many variant prioritization tools and are used by the 100,000 Genomes project, the NIH Undiagnosed Diseases Program/Network, and thousands of other clinics, labs, tools, and databases. Patients themselves are an eager and untapped source of accurate information about symptoms and phenotypes - some of which may go unnoticed by the clinician. However, medical terminology is often perplexing to patients, making it difficult to use resources like the HPO. Since patient self-phenotyping is a critical complement to deep phenotyping performed by clinicians, we have created a 'layperson' translation of the HPO. We systematically reviewed all 12,000+ terms in the HPO and added layperson synonyms with approximately 60% being translatable into layperson terminology. We evaluated the quality of the lay-HPO to ensure lay profiles could be effectively used by variant prioritization tools such as Exomiser. Preliminary analyses suggest that the lay-HPO has the features required to be useful in a diagnostic setting, in that lay terms are: a) sufficiently specific and, b) well-represented in our disease-to-phenotype database that is utilized by the aforementioned tools for differential diagnostics. This patient friendly version of the HPO uses the same infrastructure as the primary HPO, so patient-generated phenotyping data can readily be combined with clinical phenotyping data to improve variant prioritization. New patient-centered tools are being developed to help patients to assist clinicians in creating robust computational phenotype profiles to improve molecular diagnostic rates and be active participants in their diagnostic odysseys. Finally, such tools can also assist in creation of patient generated phenotypic profiles for sharing in patient registries, forums, and on the Web for cohort definitions and community formation.

1177W

Sporadic, isolated Fanconi syndrome due to a mutation of EHHADH. *E.G. Seaby[1,2], D. Bunyan[3], R.D. Gilbert[2], S. Ennis[1].* 1) Genomic Informatics, University Hospital Southampton, Southampton, United Kingdom; 2) Southampton Childrens Hospital, University Hospital Southampton, United Kingdom; 3) Wessex Regional Genetics Laboratory, Salisbury, UK.

Background: Hereditary Fanconi syndrome has been described a handful of times in the literature, displaying variable Mendelian patterns of inheritance. Some cases segregate as autosomal dominant and may or may not be associated with renal failure, whilst other kindreds show recessive and X-linked recessive inheritance. In 2001, Lichter-Konecki et al identified the locus (chromosome 15q15.3) in a kindred with autosomal dominant Fanconi syndrome. In 2014, Klootwijk et al found a causal mutation in *EHHADH*, a gene involved in peroxisomal beta oxidation resulting in dominant Fanconi syndrome with preserved renal function in a black family. *Case Diagnosis/Treatment*: We applied whole exome sequencing technology to a Caucasian boy presenting with sporadic, idiopathic Fanconi syndrome. We identified a *de novo* mutation in *EHHADH*, the same variant as previously described by Klootwijk et al. This mutation has been observed to cause the protein encoded by *EHHADH* to mislocalise to the mitochondria, where it interferes with mitochondrial beta oxidation, the primary source of energy in renal proximal tubular cells. Aberrant mitochondrial beta oxidation starves the renal tubule cells of energy required to maintain electrochemical gradients, essential for solute reabsorption. *Conclusion*: Our finding describes the first Caucasian patient with a *de novo* mutation in *EHHADH* causing isolated Fanconi syndrome.

1178T

A combinatorial approach for the selection of novel bioactive peptides. *R.R. Handley, SN. Lawrence, K. Ly, K. Lehnert, RG. Snell.* School of Biological Sciences, University of Auckland, Auckland, New Zealand.

The rapid discovery of rare mutations is revealing an enormous number of drug targets. As a result there is a requirement for almost bespoke molecular tools for understanding the pathophysiology of these conditions and ultimately also bespoke therapies. Our aim is to develop an inexpensive method for generating and screening novel bioactive compounds, using combinatorial randomised sequence generation by degenerate primer PCR amplification. The goal is that the approach will be applicable in any molecular genetics laboratory with particular disease expertise. As proof of concept we developed and tested this method in a modified toxin/antitoxin plasmid cloning system, where the plasmid contains a type II bacterial toxin 'cell death control' sequence *ccdB,* immediately downstream of its *ccdA* antitoxin sequence. To generate a diverse library for screening, PCR with degenerate primers was used to synthesise random DNA sequences in *ccdA*, with products subsequently incorporated as part of the antitoxin sequence of the plasmid. Sequence analyses revealed massive diversity of the libraries as expected. Functionality of modified toxin/antitoxin peptides was tested using a simple live/dead screen of transformed *E.coli*. Small scale sequencing of 16 selected colonies which grew revealed 8 clones containing novel *ccdA* modifications that may have enabled antitoxin function to be maintained. Large scale sequencing of further positive clones is underway. Overall, initial testing indicates that this simple, combinatorial method can generate very complex libraries and may function as an effective source for target specific, bioactive compounds. We are now in the process of identifying peptides that inactivate huntingtin (*HTT*) toxicity in *E.coli* and which may be useful for investigating Huntington's disease.

1179F

Mouse models of human disease: How mouse model data can provide mechanistic insight into human disease etiology and development of therapeutics. *M. Tomczuk, J.T. Eppig, C.L. Smith, the Mouse Genome Informatics Staff and Software Team.* Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME.

Model organisms greatly aid in understanding the cause and development of human disease. Mice provide an excellent system for investigating rare disorders in which patient data may be lacking, for understanding congenital diseases as the mouse embryo is easily manipulated, and for designing and evaluating targeted therapies. The Mouse Genome Informatics (MGI, *www.informatics.jax.org*) resource integrates mouse phenotype data from multiple sources such as published research, large scale mouse mutagenesis projects, the International Mouse Phenotyping Consortium's (IMPC) genome-wide phenotypic pipeline, and the newly added embryonic lethal data from Deciphering the Mechanisms of Developmental Disorders (DMDD) consortium, to facilitate links between known mouse phenotypes and human disease symptoms. Congenital diaphragmatic hernia (CDH) is a condition that affects 1:3,000 births and results in 30-50% mortality and abnormal embryogenesis. Only two human genes, *GATA4* and *ZFPM2*, are so far implicated in CDH and causes are largely unknown. We will show how MGI can be utilized to gain a deeper understanding of the pathogenesis of this defect by studying available mouse models and using these models to identify possible new candidate genes. A search for 'diaphragmatic hernia' displays all genetic mouse mutants with this phenotype. The results indicate 52 mutant alleles in 38 different genes that show this defect. The new Disease Ontology (DO) Browser, a standardized human disease vocabulary, with links to MesH, OMIM, and Orphanet, shows that 18 mutant alleles in 11 genes are reported to model CDH. Searching by extra phenotypes specific to signs observed in patients can help to identify genes responsible for different aspects of the condition. Links to specific alleles provide molecular descriptions about the mutation and genotypes associated with detailed descriptions of the mutant phenotypes. Embryonic abnormalities of these mutants can be analyzed to better understand how CDH develops in humans. Based on the most relevant phenotypic similarities between patient symptoms and phenotypes in mutant mice, a list of possible candidate genes can be generated. Detailed pages for each gene provide comprehensive information about the gene of interest, including biochemical function and process, sub-cellular localization, expression data, gene interactions, polymorphism, and gene homology that can aid in understanding disease pathology and guide discovery of therapeutics.

1180W

Challenges in translating pharmacogenetic variants from Illumina's Multi-ethnic Genotyping array into clinical practice. *N. Rafaels[1], K. Crooks[1,2], T. Phang[1], T. Brunetti[1], C.R. Gignoux[1,3], Y.M. Lee[1,4], A. Monte[1,5], C.L. Aquilante[1,4,6], R.A. Mathias[1,7], K.C. Barnes[1].* 1) Colorado Center for Personalized Medicine, School of Medicine, University of Colorado, Aurora, CO; 2) Department of Pathology, School of Medicine, University of Colorado, Aurora, CO; 3) Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO; 4) Skaggs School of Pharmacy and Pharmaceutical Sciences, Department of Pharmaceutical Sciences, University of Colorado, Aurora, CO; 5) Department of Emergency Medicine, School of Medicine, University of Colorado, Aurora, CO; 6) Skaggs School of Pharmacy and Pharmaceutical Sciences, Department of Clinical Pharmacy, University of Colorado, Aurora, CO; 7) Johns Hopkins Asthma and Allergy Center, School of Medicine, Johns Hopkins University.

   The incorporation of pharmacogenetics into the clinical setting is becoming increasingly feasible with the availability of more affordable genetic sequencing technologies; however, many challenges still remain for clinical implementation. The Colorado Center for Personalized Medicine is using Illumina's Multi-Ethnic Genotyping Array (MEGA-Ex) with up to 300,000 variants of custom content to meet this challenge. Scraping databases from the Clinical Pharmacogenetics Implementation Consortium (CPIC) and PharmGKB, the base chip contains 2,778 of 3,274 (84.9%) of known pharmacogenetic variants, yet only 98 of 334 (29.3%) of known, clinically actionable, CPIC Level A variants. (The more recently released Illumina Global Screening Array contains 2,472 (75.5%) and 81 (24.3%) of known pharmacogenetics variants and CPIC Level A variants, respectively.) Work still needs to be done in public databases to trace 85 of 496 missing known pharmacogentic variants and 5 of 236 missing CPIC Level A variants to rsID and base pair position. In addition, validation of HLA, copy number variant (CNV) and compound heterozygote calling, as well as design of indels, remain challenges (e.g., although 150 of 157 known *CYP2D6* variants can be identified, 21 variants are indels, 107 of 153 * alleles contain multiple SNPs, and 15 * alleles require CNV calling). Current GWAS array technology cannot provide all solutions for clinical pharmacogenetics implementation; however, with the ability to add custom content, it can provide an affordable solution to capture up to 444 of 479 (92.7%) of all CPIC Level A * alleles.

1181T

The ET-HPN highlights a potential ALS-related disease cluster. *B.E. Graham[1], C. Darabos[2], M. Huang[3], L.J. Muglia[4], J.H. Moore[5], S.M. Williams[6].* 1) Systems Biology and Bioinformatics, Case Western Reserve University, Cleveland, OH. 44106, U.S.A; 2) Research Computing, Academic & Campus Technology Services, Dartmouth College, Hanover, NH 03755, U.S.A; 3) Department of Genetics, The Geisel School of Medicine at Dartmouth College, Hanover, NH 03755, U.S.A; 4) Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children's Hospital Medical Center and March of Dimes Prematurity Research Center Ohio Collaborative, Cincinnati, OH 45229, U.S.A; 5) Institute for Biomedical Informatics, Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, U.S.A; 6) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106, U.S.A.

The identification of genes related to human phenotypes is important to understanding disease pathophysiology, but for complex disease this can be a difficult endeavor. We previously developed the ET-HPN, a method of filtering Human Phenotype networks based on the differential distribution of disease between populations and three-way comparisons of the possibly associated allele frequencies. Here we investigated Type 2 diabetes mellitus (T2DM) as an index disease, using ET to filter the HPN. Our sample populations were YRI, GIH and CEU, based on T2DM global prevalences. Our resulting network contained several T2DM related phenotypes, including Glucose and Glyco-sylated Hemoglobin A (HbA1c). Unexpectedly, we uncovered a link between Glucose, HbA1c and Amyotrophic Lateral Sclerosis (ALS). We confirmed, by a literature search, that T2DB is inversely related to ALS, but the underlying mechanism is unknown. However, our network results identified a series of common genes, and we found that T2DM is less common in the ALS population, with a better ALS prognosis and later onset in the presence of T2DM. In our present study, we also used a proteomic, pathway based analysis, with a focus on ALS in context of T2DM. We narrowed the genes examined to those shared by ALS and Glucose or HbA1c. Our examination uncovered three genes associated with KEGG pathways. *GRIN2B* is in the Huntington's (HD), ALS and Alzheimer's (AD) disease pathways. *NDUFA4* is in the AD, HD and Parkinson's (PD) pathways. *RYR3* is in the AD pathway. We also examined molecular, cellular and biological function of the gene products as defined by Gene Ontology and found a Neuron cell-cell adhesion function involving *CNTN4*, *NCAM2* and *NRXN3*. This cluster of T2DM, ALS, AD, PD and HD is very interesting, as T2DM is associated with HD, AD and PD, in addition to the aforementioned inverse relationship to ASL. Also, ALS and HD seem to occur together more than by chance. ALS shares disease features with HD, AD and PD as well as common GWAS hits. Dementia is rare in ALS, but ALS is more severe when dementia is present. All of this together suggests that these five conditions form a disease comorbidity cluster based on pleiotropic gene action. Some links were already known, but not this particular cluster. This study shows that the ET-HPN is a powerful tool for the rapid identification of links and comorbidity of disease and possibly predictive of novel disease clusters.

1182F

The system biology concept applied to a secondary analysis of Body Mass Index Genome Wide Association Study (BMI GWAS) data. *E. Cirillo[1], K. Watanabe[2], M. Adriaens[3], L. Parnell[4], S. Coort[1], C. Evelo[1,3].* 1) Department of Bioinformatics – BiGCaT, Maastricht University, Maastricht, The Netherlands; 2) Department of Complex Trait Genetics, VU University Amsterdam, Amsterdam, The Netherlands; 3) Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands; 4) Agricultural Research Service, USDA, Jean Mayer-USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, USA.

The interpretation of GWAS results is crucial to understand the influence of SNPs in diseases. For instance, identifying genetic loci affected by actionable SNPs can guide functional genomics experimentation and drug target testing. Retrieving and combining the information available to describe the SNPs functionality and then place it in the larger context of a biological process, is a current challenge. Currently, many tools arrange different types of annotation related to variants, genes and pathways, but no single tool exists that provides all the diverse types of information simultaneously. Moreover, the output of multiple tools need to be harmonized and interpreted by the researchers. We present a secondary analysis of BMI GWAS results using several types of data, with the scope to combine different levels of complexity based on SNPs, genes and pathways information, towards the understanding of the obese condition. We performed FUMA to define 77 genomic risk loci containing 10,367 SNPs which are in Linkage Disequilibrium to one of the independent significant SNPs at r^2 0.6. All the SNPs were mapped to genes, if they were closer than 10kb regardless of ANNOVAR annotation. We obtained 182 genes related to 89 coding SNPs, and 10,279 non-coding SNPs. This gene set was analyzed using pathway enrichment analysis and network analysis using the WikiPathways collection. The pathway role was evaluated in relation to both the effect of the coding variants on the genes that they present, and in the context of obesity with the literature support. In addition, all the non-coding variants were described using four online tools (FUMA, HaploREg, rVarBase and GWAVA) from which we considered only the epigenetics information related to chromatin state, histone modification and transcription factor binding sites. This output was then combined with eQTLs analysis based on GTEx data related to: adipose tissue, pancreas, liver, skeletal muscle, brain and thyroid. The preliminary results related to pathway, showed processes such as: energy metabolism and brain functionality that are in line with the pathophysiology of the obesity. However, further analysis is currently underway to elucidate which genes per tissue are regulated by the non-coding SNPs. Those genes will allow the identification of tissue specific pathways. With these results we aim to present an overview of tissue specific processes in which genes affected by BMI GWAS variants can influence the obesity condition.

1183W

Causal gene prediction from type 2 diabetes susceptibility loci through integration of genetic association and functional annotation data. *J. Fernandez-Tajes[1], J. Torres[1], M. van de Bunt[1,2], S. Thomsen[3], W. Rayner[1,3,4], M. Thurner[1], K. Lage[5,6], K. Gaulton[1,7], M. McCarthy[1,3], DIAGRAM consortium, the T2D-GENES consortium, the GoT2D consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Global Research Informatics, Novo Nordisk A/S, Måløv, Denmark; 3) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom, Hinxton, United Kingdom; 4) Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Genome Campus; 5) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 6) Department of Surgery, Massachusetts General Hospital, Boston, MA, US; 7) Department of Genetics, Stanford University, Stanford, CA, USA.

GWAS studies have identified ~100 loci influencing type 2 diabetes (T2D) risk. However, as most associated variants are non-coding, the genes through which causal variants act are mostly unknown. We aimed to identify the genes through which these risk variants act and protein interaction networks (PPi) in which these genes reside through integration of large-scale genetic and genomic annotation data. We developed a scheme to assign causal 'candidacy' scores to genes and applied it to all 1,511 genes within 500kb of 93 known T2D signals. This scheme integrated data from (a) fine-mapped localization of causal variants using Bayesian credible sets, (b) regulatory annotations from ENCODE and T2D-relevant tissues (eg pancreatic islets, adipose), (c) annotations linking regulatory elements to target genes (eg *cis* eQTLs), (d) coding variant T2D associations from exome sequence and array data and (e) biological annotations of gene function. We then normalized the resulting scores across all genes at each locus by partitioning the posterior probability of candidacy. This approach generated high candidacy scores for many established causal genes (eg *SLC30A8, MTNR1B*), and highlighted novel candidates (eg *NKX6-3* at *ANK1* locus). We then used these scores as seeds to build a PPi network, by solving an instance of a rooted weighted price-collecting Steiner tree problem, where the roots were the genes in a 500kb window region from the loci described in Wood et al (2017) and associated to both acute and peak insulin response. This approach allowed us to guide the network modeling to a likely islet-function specific network. The resulting network was significantly enriched for T2D association in DIAGRAM imputed to 1000 Genomes (Scott et al. 2017) (PASCAL, Fisher $p=5x10^{-5}$) after excluding GWAS genes from the enrichment. The most important steiner nodes (proteins that connect GWAS loci genes) in terms of both network connectivity and link to "causal" genes, were *YWHAG*, *SMAD4* and *EP300*, all with independent molecular evidence for a role in T2D. The top significant GO processes enriched in this T2D-specific network included endocrine pancreas development, glucose homeostasis, positive regulation of autophagy, and regulation of insulin secretion. These results suggest a high degree of functional inter-connectivity among causal genes influencing T2D. They provide a resource for experimental validation of causal gene targets at established loci and discovery of novel risk genes.

1184T

LabWAS: A catalog of real-world associations between genetic variants and lab values. *J.A. Goldstein, L.A. Bastarache, P. Speltz, A. Gifford, D.M. Roden, J.C. Denny.* Vanderbilt University Medical Center, Nashville, TN.

Laboratory testing underlies much of modern medicine. Genome-wide association studies (GWAS) on individual lab tests have been used to understand the biology of diseases in which those labs are abnormal. Using 24600 Caucasian adults with genotyping and linked electronic health record (EHR), including labs, we performed GWAS for 110 common lab tests. We summarized multiple measures based on the mean, median, first, minimum or maximum lab values. We also tested whether the patient ever had an abnormal high or low value. We compared our findings to previously reported associations. We replicated 354 of 462 testable associations (77%) previously identified in the GWAS catalog at $p<0.05$. We replicated 60% of associations previously reported a single time, which increased to 100% when associations were previously reported at least 5 times. Probability of replication was also related to prior published P-value. We also identified 1216 putative novel associations for 61 labs at $p<5E-8$. For example, rs34856868 in *BTBD8*, previously associated with Crohn's disease, was associated with a 72 cell/uL increase in monocyte count ($p = 3.6E-13$). The largest number of new associations were identified using labs that have few previous significant findings from GWAS, such as eosinophil counts and oxygen saturation. We also identified variants associated with well-studied labs, such as blood glucose. Many of these blood glucose-associated variants were in genes in the HLA region known to be associated with Type I Diabetes Mellitus (T1D), including *BTNL2, HLA-DQA1*, and *HLA-DQB1*. The novelty of these associations likely reflects exclusion of patients with disease (such as T1D) from prior studies. We then tested whether SNPs associated with lab values were also associated with disease using phenome-wide association studies (PheWAS). We identified 329 lab:SNP:disease associations. Some, such as those noted above between glucose, HLA SNPs, and T1D were expected. Others, such as the association between basophil counts, rs35534614 in *CACNA1S*, and lower gastrointestinal congenital anomalies are surprising and may point to new disease biology. In a real-world cohort, we showed a high degree of replication of known associations and report novel SNPs associations with lab results. Some of these novel findings may occur exclusively in the context of disease. We are developing a public repository of all results at http://www.phewascatalog.org/labwas.

1185F

Genome-wide prediction of susceptibility loci for complex diseases based on regulatory data. *T. Yang, S. Dong, Y. Guo, S. Yao, Y. Chen, M. He, X. Chen, J. Chen.* School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

Abstract Genome-wide association studies (GWASs) is an effective strategy to identify susceptibility loci for human complex diseases. However, missing heritability is still a big problem. Most GWASs SNPs are located in non-coding regions, Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics projects have provided rich resources in interpreting their biological functions. In this study, we developed a pipeline to identify novel susceptibility loci for complex diseases based on the interpretation of the functional features for known disease-associated variants with machine learning. We applied our pipeline to predict novel susceptibility SNPs for type 2 diabetes (T2D). The predicted SNPs could explain heritability beyond that explained by GWAS-associated SNPs. Functional annotation by expression quantitative trait loci (eQTL) analyses showed that the target genes of the predicted SNPs were significantly enriched in T2D related pathways. Our results suggest that combining previous GWASs results and regulatory data could identify additional functional susceptibility SNPs for complex diseases. We hope our pipeline could help to identify novel susceptibility loci for complex diseases and solve the missing heritability problem.

1186W

The Type 2 Diabetes Knowledge Portal: Clearing a path from genetic associations to disease biology. *B. Alexander, M. Costanzo, M. Duby, C. Gilbert, T. Green, M. von Grotthus, R. Koesterer, P. Singh, K. Yuksel, J. Flannick.* Medical Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA.

Genetic research on type 2 diabetes (T2D) and complex traits in general has reached an inflection point where the central challenge has shifted from discovering new loci to understanding disease biology. However, major barriers separate the statistical experts who analyze genetics data and the biologists who might profitably learn from them. Here, we demonstrate the value of a new data sharing initiative, the T2D Knowledge Portal (T2DKP) produced by the Accelerating Medicines Partnership in T2D, for breaking down these barriers. The T2DKP draws from a harmonized database of 39 million variants, 48 traits, and 11 million epigenetic annotations and provides a freely available website (http://type2diabetesgenetics.org). This resource provides a wide variety of interactive tools, including custom association analysis with dynamic sample filtering, covariate specification, and conditional analysis, while enabling sophisticated analytics of sensitive genetic data and ensuring patient privacy. Using this framework, known results that previously took consortium-level coordination can be identified in seconds. A search of the MTNR1B gene displays an integrated table of all variants associated across all traits, notably fasting glucose and T2D, as well as an interactive "LocusZoom" plot of the genomic loci. Displaying credible sets, rather than associations, immediately prioritizes a single variant as likely causal for the signal, with further evidence for a regulatory effect in pancreatic islets based on epigenomic results integrated within the display; these results replicate the results of a 2015 genetic analysis that required over 100 authors (Gaulton KJ et al., 2015, Nat Genet. 47:141). Additionally, investigators without previous access to these data can for the first time use them in their own analyses. A study that characterized the impact of variants in HNF1A (Najmi LA et al. (2017), Diabetes 66:335) was able to validate the significant association of functionally important variants based on lookups in the T2DKP. Similarly, a study that identified associations between ANGPTL4 and coronary artery disease (Dewey FE et al. (2016), N Engl J Med. 374:1123) was able to query the T2DKP for T2D associations, suggesting a novel effect on T2D from this gene. These results provide a platform that removes a major barrier impeding genetic data sharing and demonstrate the potential scientific value of providing broad access to genetic data.   .

1187T

Retrospective electronic medical record analysis to identify patients at risk of hypophosphatasia. *C. Peroutka[1], M. Marzinke[4], N. Parikh[2], A. Alade[1,2], J. McGready[3], K. Schulze[2,3], J. Hoover-Fong[1,2].* 1) McKusick Nathans Inst of Gen Med, Johns Hopkins, Baltimore, MD; 2) Greenberg Center for Skeletal Dysplasias, Johns Hopkins, Baltimore, MD; 3) Bloomberg Sch of Public Health, Johns Hopkins, Baltimore, MD; 4) Clinical Pharmacology Analytical Laboratory, Johns Hopkins Hospital, Baltimore, MD.

   Introduction: Hypophosphatasia (HPP) is a pleiotropic metabolic condition caused by variants in tissue-nonspecific alkaline phosphatase (*TN-SALP/ALPL*), manifesting as perinatal lethal or throughout life with fractures, early tooth loss, osteoporosis, seizures, chronic pain, fatigue and/or respiratory compromise. Low serum alkaline phosphatase (AP) is a key biochemical indicator of HPP, but is often unrecognized as pathologic. Given a broad disease spectrum and under-recognized indicator, HPP is likely underdiagnosed. Misdiagnosis as idiopathic osteoporosis may lead to treatment with bisphosphonates, which counteracts available asfotase alpha enzyme replacement therapy for HPP. We conducted a risk prevalence study for HPP by query of a large, hospital-based electronic medical record (EMR), demonstrating that EMR query may be used to identify rare disease. **Methods:** The Johns Hopkins (JH) IRB approved this protocol under a waiver of consent. JH patients from all clinical sites with $\geq$ 1 AP level reported from Jan 2013-Dec 2015 were compiled by patient medical record number (MRN). Site and ICD-9 codes were merged to MRNs with $\geq$ 1 low AP based on age- and sex-specific AP norms. MRNs with hematologic, oncological, drug and/or transplant codes associated with low AP were excluded. Remaining charts were prioritized for review. **Results:** There were 983,458 AP levels from 156,459 patients, with 45,233 (22,175 patients $\leq$ 19yo) low AP levels from 11,730 patients. Exclusion by ICD-9 and clinical codes yielded 642 patients for review, with data from 45 patients presented. Medical history is suggestive of HPP in 28(62%) patients: 3 of 45(6%) with adult premature tooth loss, 4(8%) excessive caries, 6(13%) nephrolithiasis, 10(22%) muscle pain, and 27(60%) chronic pain. 6(13%) have osteoporosis; 2 received bisphosphonates and 1 received a PTH analog. Multiple specialists provide care: 5(11%) endocrinology, 8(17%) orthopedics, 4(8%) rheumatology, 3(6%) genetics, and 2(2%) pain management. HPP was not on the differential for patients with low AP, though 5(11%) had molecular testing for other conditions. **Conclusions:** This study provides evidence that EMR query of a large patient population can identify those who warrant further investigation for rare, treatable genetic conditions like HPP. Following chart review, we will obtain IRB approval to contact patients at risk of HPP to offer genetic counseling, molecular testing, and treatment to those affected.

1188F

Evaluation of genetic sequencing to improve newborn screening for VLCAD disease. *B. Cai[1], R. Jain[1], D. Lee[1], T. Bergquist[1], C. Bock[1], V. Pejaver[1], R. Gallagher[5], P. Kwok[2], J. Puck[3], S.D. Mooney[1], B. Currier[4].* 1) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA; 2) Institute of Human Genetics, University of California, San Francisco, California, CA, USA; Cardiovascular Research Institute, University of California, San Francisco, CA, USA; Department of Dermatology, University of California, San Francisco, CA, USA; 3) Department of Pediatrics, Division of Allergy, Immunology, and Blood and Marrow Transplantation, University of California San Francisco, San Francisco, CA, USA; 4) Genetic Disease Screening Program, California Department of Public Health, Richmond, CA, USA; 5) Department of Pediatrics, Medical Genetics, University of California San Francisco, San Francisco, CA, USA.

   Background: Very-long-chain acyl-CoA dehydrogenase deficiency (VLCAD) is currently screened for as part of newborn screening programs in all 50 states in the US. We assessed the incorporation of genetic sequencing of the *ACADVL* gene to assess the overall utility of incorporating genetic information by comparing screening true positives (affected= TP) and screening false positives (unaffected but screened positive= FP) samples. **Methods:** The screening results within the study population were refined via Mayo Clinic Collaborative Laboratory Integrated Reports (CLIR) Tool Runner using the tools for VLCAD and VLCAD(het), i.e. heterozygote VLCAD carriers. We performed exome sequencing on 600 individuals and applied a computational pipeline to identify pathogenic variants on the ACADVL gene. This study population included the 16 TP and 45 FP for VLCAD identified by the California Newborn Screening Program, which were resolved as not disease during follow-up by metabolic clinics. **Results:** The results are summarized in the table below. In the 16 TP (affected) individuals, 11 had two potentially pathogenic variants identified and the other 5 each had one variant identified. Review of case history in these 5 suggested that that at least 2 were heterozygous carriers. Of the 44 FP, 3 had two variants identified, 31 had one variant identified and 13 had no variants identified. In the clinical follow-up, of the 14 cases called indeterminate by the CLIR tools, 3 were identified as VLCAD. Of these indeterminate cases, 2 had two pathogenic variants, and 1 had one, 6 were clinically identified as carriers, 5 had 1 pathogenic variant, and 1 had none. 3 were clinically identified as "No Disorder", 2 had no variants, and 1 had one. **Conclusion:** This initial study suggests that sequencing ACADVL together with the CLIR-based tools can significantly refine newborn screening for VLCAD. Table: Comparison of clinical resolution, CLIR refinement of screening results, and sequencing results among VLCAD screen positive cases in California.

NBS resolution for VLCAD Screen Positive	CLIR Dual Scatter Plot Evaluation	Number of Cases	# of ACADVL Variants		
			Two	One	None
True Positive (TP)	VLCAD	10	9	1	0
	VLCAD heterozygote (carrier)	3	0	3 (*)	0
	Indeterminate	3	2	1	0
False Positive (FP)	VLCAD	3	1	0	2
	VLCAD heterozygote (carrier)	30	2	21	7
	Indeterminate	11	0	8	3

(*) Two of these cases were reclassified as carriers (i.e. FP screening tests) following several years of uneventful clinical follow-up by metabolic specialists.

1189W

Using automatic adipose measures from electronic health record based imaging data for discovery. *E.D.K. Cha[1], Y. Veturi[1], C. Agarwal[2], M. Arbabshirani[1], S.A. Pendergrass[1] on behalf of the DiscovEHR collaboration.* 1) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA; 2) Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL.

Using advanced methods for extracting measures from imaging data within electronic health records offers an opportunity for new research, such as investigating the impact of subcutaneous and visceral adipose tissue levels on risk factors and exploring genetic associations with those traits. The location and type of adipose tissue is an important factor in metabolic risk for diseases such as diabetes, hyperlipidemia, hypertension, and atherosclerosis. Visceral adipose tissue (VAT) levels are believed to have a more negative impact on metabolic outcomes over subcutaneous adipose tissue (SAT) levels. Evaluating visceral and subcutaneous fat levels has been pursued through computed tomography (CT). There is a unique opportunity to use existing abdominal CT imaging data for large-scale research without necessitating the enrollment of individuals in a new imaging study. As proof of principle, we pulled ~2000 CT studies from Geisinger Health System PACS and used "greedy snakes", a method in computer vision, to quantify VAT, SAT, total adipose tissue (TAT) and adipose-ratio volumes. After controlling for age and sex, we performed logistic regression analyses to evaluate these adipose measures for association with more than 1000 disease diagnoses (using ICD-9 codes to define cases and controls for each diagnosis). While SAT was more strongly associated with some obesity-related ICD-9 codes (e.g. localized osteoarthritis), we found a number of significantly stronger associations between VAT and metabolic syndrome diagnoses such as type II diabetes (p=1.5E-58; 95% CI =[2.85,2.87]) and obstructive sleep apnea (p=7.7E-37; 95% CI=[2.32,3.16]), which shows the increased impact of VAT on these traits. We also performed linear regression with 35 continuous clinical lab measures, finding stronger associations and consistent effect direction for VAT with lipid levels and white blood cell counts. The same tests stratified by sex revealed stronger associations for many diagnoses and clinical lab measures for women than men, suggestive of an increased impact of VAT in women. Finally, we performed a GWAS between adipose measures and genetic variants identified some suggestive associations for VAT and SAT measures. Our future work will include increasing the sample size for a larger study, including common and rare variant genetic associations with these adipose measures to further elucidate their impact on disease risk, comorbidities and the underlying genetic architecture.

1190T

***In-silico* characterization of cell-type composition in adipose tissue: Implications for 'omic analyses and associations to adiposity measures.** *C.A. Glastonbury[1,2], J.S. El-Sayed Moustafa[1], A.C. Alves[1], K.S. Small[1].* 1) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 2) Big Data Institute, University of Oxford, Oxford, UK.

Adipose tissue is the largest endocrine organ in the body and has a role in the development of cardio-metabolic disorders. Adipose tissue is comprised of a heterogenous array of cell types which can differentially affect disease phenotypes. Cellular heterogeneity is a well established confounder of 'omic analyses in blood, but has rarely been taken into account in analysis of solid-tissue transcriptomes. We used CIBERSORT to estimate the relative proportions of four cell types (adipocytes, macrophages, CD4+ t-cells, Micro-Vascular Endothelial Cells (MVEC)) in bulk subcutaneous adipose RNAseq from TwinsUK volunteers (N =766) and GTEx post-mortem donors (N=326). We confirmed deconvolution accuracy *via* simulations and visual consistency with GTEx histology images. Cell-type composition was markedly different between the two cohorts (adipocyte, macrophage, MVEC medians: TwinsUK =0.82, 0.02, 0.15, GTEx = 0.62, 0.08, 0.30), reflecting differences in the collection protocols, biopsy location (abdomen vs leg) and cohorts. In TwinsUK, cell-type proportion, in particular macrophage proportion, was correlated with BMI (R^2= 0.22, *P* =2×10^{-8}), with even stronger associations to DXA-derived measurements of body-fat distribution including visceral fat (R^2= 0.29, *P* =4×10^{-15}) and android/gynoid ratio (R^2=0.36, *P* =2×10^{-16}), suggesting macrophage infiltration plays a prominent role in body-fat distribution. By contrast, cell-type proportion was not associated with age (R^2 = -0.02). Heritability of adipocyte, macrophage and MVEC proportion was 17%, 30% and 21% respectively. We assessed the effect of cell-type composition on a range of 'omic analyses. Expression levels of 43% of protein-coding genes were associated with BMI (FDR 5%); 20% of these were no longer significant after adjusting for cell-type proportion, indicating a substantial portion of BMI-expression associations are independent of cell-type proportion. WCGNA co-expression networks were heavily associated with cell-type proportion, including the MEMN module and macrophage proportion (R=0.68, *P* =2×10^{-16}). Adjusting for cell-type proportion in *cis*-eQTL analysis achieved a modest increase in eQTL yield (2.3%), but no difference after correction for latent factors. G x Cell Type Proportion interaction models identified 20 cell-type specific eQTLs (12 macrophage, 10 MVEC, 4 adipocyte), demonstrating the potential of *in silico* deconvolution of bulk tissue to identify cell-type restricted regulatory variants.

1191F

Non-additive SNP effect discovery from GWAS with LAMP. *J. Sese[1,2], A. Terada[3,1,4], M. Koido[2], M. Akiyama[2], M. Kubo[5], Y. Kamatani[2,6].* 1) Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology; 2) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences; 3) PRESTO, Japan Science and Technology Agency; 4) Dept. of Computational Biology and Medical Science, The University of Tokyo; 5) RIKEN Center for Integrative Medical Sciences; 6) Center for Genomic Medicine, Kyoto University Graduate School of Medicine.

   Genome-wide association studies (GWASs) have uncovered thousands of disease related SNPs, and recent very large-scale GWASs unveiled the importance of additive effects among the SNPs to understand the mutations related to especially common diseases. These analyses also showed the existence of remaining missing heritability with the traits. One possibility to explain the gap is to consider non-additive effects among multiple SNPs. However, only a few non-additive effects can be found because of the lack of the statistically valid method to detect the effects. Therefore, we developed a statistical method to detect the combinatorial effect of multiple features, called Limitless Arity Multiple-testing Procedure (LAMP) [PNAS 2013], and made a new software named LAMPLINK which incorporates LAMP in widely-used statistical genetic software PLINK [Bioinformatics 2016]. We here show the results of the application of LAMPLINK to a part of the BioBank Japan Project data including 200,000 patients with 47 target diseases. The results revealed the candidate novel non-additive combinations of SNPs related to diseases including Graves' diseases and rheumatoid arthritis. Furthermore, the joint analysis workflow of LASSO with LAMPLINK confirmed that the SNP combinations cannot be explained by the additive effect of SNPs.

1192W

The grid-interpolation algorithm: A novel approach for fast and efficient mixed model analysis of high-dimensional phenotype data. *J.R. O'Connell.* Program for Personalized and Genomic Medicine, University of Maryland, Baltimore, Baltimore, MD.

   High-dimensional phenotype data represents a new and significant analytical and computational challenge for emerging large scale multi-omics and imaging datasets. Two examples of high-dimensional phenotype data include epigenome-wide analysis (EWA) of a complex trait using the 850,000 probes on the Illumina EPIC methylation chip and voxel-wise analysis of cognitive function using 100,000 voxels from Human Connectome project brain images. Using regression variable names we call the analysis of many dependent variables the big Y problem and many independent variables the big X problem. An example of big X is genome-wide association analysis (GWA) of sequence data. Both the above big Y examples require mixed model methods to incorporate random effects to account for genetic relatedness and/or batch effects. Many optimized mixed model algorithms and software exist for efficient GWA analysis, but to date no such tools exist for mixed model EWA analysis. Current options are extremely inefficient: run a separate GWA mixed model for each phenotype. We present a simple solution to the big Y mixed model problem that is thousands of times faster than the current single-phenotype approach and can scale to samples sizes in the tens of thousands. The algorithm, called the grid-interpolation algorithm, relies on evaluating the mixed model likelihood on a set of fixed parameter values, called a grid, then interpolating the likelihood function using Lagrange polynomials to find the maximum value. The computational cost of the grid step is of order N^3, where N is the sample size, then grid likelihood evaluation is of order $P*N^2$, where P is the number of phenotypes. The single-phenotype approach has cost of order $P*N^3$. To examine performance, we simulated 100,000 phenotypes (random normal deviates) on 8000 subjects who had available age, sex, lipid traits, and a genetic relationship matrix. The runtime for a single lipid trait against 100,000 phenotypes was under 30 minutes, while the single-phenotype analysis was estimated to take 25 days based on the runtime for 5000 phenotypes. Our software supports both cluster and cloud computing, outputs both maximum likelihood (ML) and restricted ML (REML) estimates, standard errors and p-values of all fixed and random effects, and can output a file of the residuals for each phenotype for input to other analyses. We will also demonstrate extensions of our algorithm to big Y analysis of genetic data and multiple random effects.

1193T

Identification of genes alternatively spliced in HIV-infected CD4+ T-cells.
*S. Han, S. Byun, Y. Lee *. Biomedical Informatics, University of Utah, Salt Lake City, UT., Select a Country.

Introduction: The most effective treatment for patients infected with human immunodeficiency virus 1 (HIV-1) is the complete elimination of T-cells harboring the latent form of the virus. However, specific targeting of the latently infected T-cells is a major obstacle in such treatment. **Motivation**: Alternative splicing (AS), using different combinations of exons, is a major mechanism for gene regulation and protein diversity. AS patterns (exon skipping, alternative splice sites) are known to occur in specific genes related to T-cell immune processing (e.g. CD44, CD45, CTLA4). However, there is no study yet that has systematically examined differential exon skipping events in HIV-infected T-cells. We hypothesized that immune-related genes have differentially expressed exons in HIV-infected T-cells. We analyzed alternative splicing in RNA-seq data to discover immune-related genes that produce differentially skipped exons between HIV-infected and non-infected T-cells. **Methods:** We downloaded RNA-seq data from the Sequence Read Archive (run IDs SRR5071107-SRR5071122), which consisted of *in vitro* HIV-infected (n=4) and non-infected (n=12) T-cells. Raw data was mapped onto the human reference genome (GRCH37.75) using Tophat. The R package *IMAS* was used to identify differentially skipped exons. Gene Ontology (GO) term enrichment was analyzed using the program GSEA to determine the functional roles of genes. In addition, we searched PubMed for studies documenting genes with roles relating to transmembrane protein and T-cell function. **Results:** We identified 145 genes with differentially expressed exons. Among these genes, the term "Immune Processing Process" was significantly enriched (*p*= 1.75e-8, *q*=7.75e-5, GO:0002376), represented by 19 genes (e.g. NCK1, STK11, CTSL). Thirty-one of the AS genes are known transmembrane proteins, of which 14 are related to immune function (e.g. FCGR1B, FCGR3A, LILRB2). In addition, we found 20 genes that are related to T-cell function (e.g., ITK, CD200, TMEM176B). **Conclusion:** We identified a number of genes with roles relating to immune and T-cell function which produce differentially expressed exons in HIV-infected T-cells. These findings suggest that an assay for alternative splicing may contribute to improved detection of latently infected T-cells.

1194F

Depression and mental illness affect pre and post multiple sclerosis diagnosis. *C. Gardner[1], M. Davis[1], J. Denny[2].* 1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Multiple sclerosis (MS), is an autoimmune disease resulting in the demyelination of the myelin sheath in the central nervous system (CNS). Currently, there are approximately 2.3 million people worldwide affected by MS, with an increased incidence in those with European heritage. The disease course for each individual case of MS is unique and unpredictable with a wide array of symptoms and disabilities following diagnosis. As the disease progresses, the quality of life the individual decreases, often leading to depression. Studies have shown that up to 50% of those diagnosed with MS will suffer from depression or anxiety in addition to the physical symptoms and disabilities. Evidence recently suggests that there may be a stronger relationship between MS and depression and anxiety in which having depression and anxiety increases risk of MS. We investigated this through data extracted from de-identified electronic medical records (EMRs) and genetic data from the Bio-VU database from Vanderbilt University Medical Center, based on ICD9 and ICD 10 codes for MS, depression and anxiety. We then created four groups (one control group, and three case groups) from the extracted individuals based on criterion of diagnosis dates of both diseases. Controls were defined as individuals diagnosed with MS and never diagnosed with depression or anxiety (n=746). The three case groups were defined as 1) having MS and being diagnosed with depression or anxiety (n=118), 2) having MS and being diagnosed with depression or anxiety pre-MS diagnosis (n=78), and 3) being diagnosed with depression or anxiety post-MS diagnosis (n=40). We[MD1] then preformed a logistic regression for each of the case groups against the control group while adjusting for age and gender to identify genetic variants that may be associated with depression or anxiety in MS patients. We saw no SNPs that reached genome-wide significance for any of the analyses, but some results were borderline significant. Seven intronic SNPs across three genes (ANKRD55, RPS6KA2, WDY4) showed association with depression or anxiety in either pre- or post-MS diagnosis (p<1 x 10⁻⁴), but not in the other (p>0.05).

1195W

A haplotype assembly workflow for HLA and KIR typing from next-generation sequencing data. *S. Tian[1], H. YAN[1], S. McDonnell[1], W. Ding[2], E. Klee[1], S. Baheti[1], N. Ihrke[3], M. Kalmbach[3], J. Olson[4], J. Cerhan[4], S. Slager[1].* 1) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 2) Division of Hematology/Medical Oncology, Department of Internal Medicine, Mayo Clinic, Rochester, MN 55905, USA; 3) Division of Research and Education Support Systems, Department of Information Technology, Mayo Clinic, Rochester, MN 55905, USA; 4) Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA.

The human leukocyte antigen (HLA) is a super-locus of genes that encode for cell-surface proteins that regulate the immune system. HLA and the Killer Immunoglobulin-like Receptors (KIR) region it interacts with play critical roles in organ transplant and disease. Determining gene sequences (typing) is crucial for defining immune epitope repertoires in cancer immunotherapy and for predicting adverse drug effects. Classical HLA typing with serology- or PCR-based methods is labor-intensive, time-consuming, and ambiguity-prone. From next-generation sequencing data, the two well-performed algorithms Optitype and Polysolver can only type class I (HLA-A, -B, and -C) but not class II (HLA-DRB1, -DQA1, and -DQB1) genes that are more complex. In fact, results from several bioinfomatics algorithms showed considerable discordance compared to the Sanger sequencing-based typing, the current gold standard. At the four-digit resolution, the overall prediction accuracy in typing HLA classes I and II ranges from 7 to 66%, 26 to 77% and 20 to 81%, respectively, for whole-genome (WGS), whole-exome (WES), and RNA sequencing data in the 1000 Genomes Project. These studies highlight the need to develop novel analytical systems for more accurate typing in HLA and other complex regions. We previously developed SGVar, a *de novo* assembly workflow for haplotype-based variant discovery from WGS and WES, which is particularly sensitive to the HLA region. Here, we aimed to improve typing accuracy and success rate by utilizing SGVar assembled haplotigs. Specifically, we optimized the selection of reads to be used in assembly and the allele inference from haplotig-reference alignments. We selected 13 WGS (2x250bp) and 24 WES (2x100bp) data sets from 1000 Genomes Project that had the gold standard PCR-verified HLA genotype information. Our pipeline achieved an accuracy of 92.3% and a success rate of 94.7% for class I and 90.4% accuracy and success for class II in WGS. For WES, the accuracy was 87.3 and 86.4% and the success rate was 87.9 and 86.4% for class 1 and II genes, respectively. We are currently using the pipeline to type the HLA and KIR genes from 1000 WGS in Mayo Clinic biobank samples. The allele frequency information gained from such a large collection of controls will be particularly useful for the discovery of immune and autoimmune risk alleles in disease susceptibility studies.

1196T

A high-throughput deep sequencing approach for CRISPR off-target assessment in therapeutic genome editing applications. *A. Ajetunmobi[1], A. Hendel[2], V. Dillard[1], R. Doyle[1], N. Humphryes-Kirilov[1].* 1) Clinical Services, Desktop Genetics, London, London, United Kingdom; 2) Bar-Ilan University, Ramat Gan, Israel.

Presented is a multiplexed deep sequencing approach for investigating off-target effects in CRISPR genome-editing experiments. This approach involves the design of Tru-Seq Custom Amplicon (TSCA) primers for multiplexed target amplification and deep sequencing of predicted off-target sites. Specificities of sgRNAs were assessed in K562 myelogenous leukaemia cells and CD34+ human pluripotent stem cells. 3000 amplicons across 24 samples were pooled for parallel sequencing on a single Mi-Seq run. Results show limited cross-hybridization across TSCA primers indicating high specificity to the target loci. Over 80% of amplicons achieve uniform coverage, negating target region bias across the 125 predicted sites. CRISPR editing events were detected at the predicted on-target site as well as several predicted off-target sites. Mutations in the population were detected to a sensitivity of 0.1% minor allele frequency. This approach presents a scalable option for discovering, validating, and assessing the off-target profile of genome-editing modifications.

1197F

High throughput BCR sequencing of the thymus and blood B cell repertoire in myasthenia gravis patients pre and post thymectomy. *R. Jiang[1], E. Dimalanta[2], R.J. Nowak[3], S.H. Kleinstein[4,5], K.C. O'Connor[3].* 1) Immunobiology, Yale University, New Haven, CT, USA; 2) New England Biolabs, Inc. Ipswich, MA USA; 3) Department of Neurology, Yale School of Medicine, New Haven, CT, USA; 4) Interdepartmental Program in Computational Biology and Bioinformatics, Yale University. New Haven, CT, USA; 5) Departments of Pathology and Immunobiology, Yale School of Medicine, New Haven, CT, USA.

Thymectomy along with glucocorticoid treatment has long been part of the standard of care for nonthymomatous myasthenia gravis (MG); thymic hyperplasia has been demonstrated on pathology in the vast majority of patients with MG, particularly those with circulating anti-AChR antibodies. Transplantation of thymic biopsies from patients with proven thymic hyperplasia in the context of MG is known to elicit an MG phenotype in immunodeficient mice (SCID) associated with the production of anti-AChR antibodies. Furthermore, previous studies have demonstrated correlative evidence between the efficacy of thymectomy and the presence of tertiary lymphoid structures in the thymus. While previous studies have used low throughput techniques to attempt to study the role of the thymic B cell compartment in MG, we sought to evaluate the heavy chain repertoire of both the thymic and blood compartments from thymectomy patients at a high depth of sequencing. Using samples from patients that had been treated by thymectomy at Yale-New Haven Hospital, we investigated both their thymic and blood repertoires using high-throughput heavy chain B cell receptor (BCR) sequencing. We applied a method developed for the accurate reconstruction of VDJ constructs from RNA extracted from B cells. Unique molecular identifier (UMI) sequencing combined with novel computational techniques was also applied to resolve sequencing error and to maximize the number of sequences that could be extracted from these samples. Our results demonstrate significant abnormalities in the thymic compartment of these patients relative to the blood as well as shifts in the repertoire pre and post thymectomy. We also attempted to identify similarities across thymic samples from these patients both in terms of broad repertoire level analysis and more closely in terms of the structural identities of the heavy chains representing the repertoires. This study better defines these repertoire abnormalities with a goal towards offering new paradigms for the earlier diagnosis and more definitive prognosis of MG.

1198W

PheWAS and permutation analyses indicated involvement of the *CLEC16A* locus in immune-related phenotypes. *M.E. March[1], F.D. Mentch[1], K. Thomas[1], C. Hou[1], C.E. Kim[1], P.M.A. Sleiman[1,2], H. Hakonarson[1,2].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

The *CLEC16A* locus has been implicated as a region of interest in multiple genome-wide association studies examining autoimmune-, immunodeficient-, or inflammation-related diseases. Associated conditions include (but are not limited to) type 1 diabetes, combined variable immunodeficiency, asthma, atopic dermatitis, multiple sclerosis, and systemic lupus erythematosus. To investigate the potential role of *CLEC16A* in other phenotypes, an approach was developed using phenome-wide association studies (PheWAS). The genotypes for the entire PheWAS cohort from the Center for Applied Genomics at CHOP were imputed against the Haplotype Research Consortium reference panel, and the approximately 2,200 LD-pruned SNPs from the *CLEC16A* region on chromosome 16 were extracted as allele doses. The allele doses were run through our standard PheWAS pipeline, and suggestive results (p < 1E-5) were extracted and analyzed. Twenty-five immunity- and infection-related phenotypes were identified in the suggestive results. To determine if this accumulation of phenotypes represented an enrichment, the SNPs from the *CLEC16A* region were permuted into multiple hundreds of collections. Each permuted collection was run through the PheWAS, and the results were scanned for occurrences of the 25 identified phenotypes. The distribution of occurrences in the permutations indicated that the 25 phenotypes found with the original genotypes is a significant enrichment. A second permutation analysis performed with an irrelevant gene of similar size to *CLEC16A* did not uncover a similar enrichment, suggesting that our PheWAS analysis of the *CLEC16A* locus provides additional evidence for the role of *CLEC16A* in immune-related phenotypes, several of which are novel.

1199T

Understanding chronic fatigue syndrome using immune cell specific RNA-seq in a time series after an exercise perturbation. *P. Comella, N.D. Beckmann, A. Cohain, G.E. Hoffman, E.E. Schadt.* Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.

Chronic Fatigue Syndrome (CFS) is a disease resulting in extreme fatigue without any known underlying medical condition. CFS often presents in patients following a viral infection and patients may appear to have slightly impaired immune systems. With no positive diagnostic, CFS is often misunderstood and misdiagnosed. A pilot clinical study was established to give further insight into the immunological role of CFS. This pilot study included 15 CFS patients and 15 age and sex matched controls undergoing cardiopulmonary exercise testing (CPET). Blood was extracted at four time points - immediately before, 1 day after, 2 days after and 3 days after. Five immune cell types were sorted from whole blood collected from each participant, and RNA was extracted and sequenced for each cell type, as well as whole blood, at each time point. Additionally, each participant underwent extensive clinical surveys and questionnaires. To discover genetic drivers of CFS, we generated differential gene expression (DE) signatures, constructed gene co-expression networks to identify coherent modules of co-regulated genes, and then carried out gene ontology enrichments on the signatures and modules. Using all cell types, we were able to identify B cells as the most CFS informative cell type, with the DE signature and network modules constructed from the B cell data containing the most disease specific information across all time-points. We identified components of DE signatures that were specific to a given time point, with some time points having larger DE signatures than others, suggesting such time points may be more informative in further studies of CFS.

1200F

Finding NEMO: *De novo* mutation detection in the IKBKG gene. *Z. Deng[1], A. Almeida de Jesus[2], R. Goldbach-Mansky[2], R. Siegel[1], E. Hanson[1].* 1) NIAMS, National Institutes of Health, Bethesda, MD; 2) NIAID, National Institutes of Health, Bethesda, MD.

NEMO (NF-kB Essential Modulator, encoded by the *IKBKG* gene) is an important regulator of host defense and inflammation. Mutations in the NEMO gene have been shown to cause a range of human diseases including embryonic lethality, Incontinentia Pigmenti in females, and Anhidrotic Ectodermal Dysplasia with Immune Deficiency. Detecting mutations in *IKBKG*, therefore, is of paramount clinical importance, however this has been difficult due to its location in a segmental duplication region on human chromosome Xq28. The duplication covers *IKBKG* after intron 2 and extends 36kb with near perfect sequence identity. Standard variant callers such as GATK and SAMTOOLS among others are ill-suited for mutation detection in duplicated regions because NGS reads have ambiguous mapping locations. We have developed a bioinformatics method to identify potential *de novo* mutations in *IKBKG* from whole genome sequencing data. We first extract reads in the duplicated regions from whole genome BAM files which are then converted to pileup format to identify variants. Next, a series of parameters including coverage and allele fraction are tested and optimized to filter out false variants. Lastly, potential mutations are manually inspected using IGV. We applied this method to screen a cohort of 69 patients with suspected monogenic autoinflammatory disease and their 123 unaffected parental controls. We identified two *de novo* mutations at the splice sites for exon 5. One mutation is chrX 153,788,776 T>G that disrupts the +2 consensus donor splice site (GT to GG) within intron 6. Another mutation is chrX 153,788,779 G>A, 3 bp away from the first mutation at +5 splice site. Both mutations are predicted to affect exon 5 splicing. We confirmed exon 5 skipping in mRNA from both patients and protein expression of a NEMO isoform (NEMO-Δex5). Both patients identified by this method are phenotypically similar to a previously studied NEMO-Δex5 patient. In conclusion, we have developed a bioinformatic method capable of mutation detection in highly similar segmental duplication regions that led to the identification of 2 *de novo* mutations in *IKBKG*. As there are more than 25 disease genes that, like *IKBKG*, are located in so-called NGS dead zones, our method may also enable detection of mutation in these other genes.

1201W

Exome sequencing by NimbleGen kit is not suitable for *SOX2* and *SOX3* molecular screening due to bald spot. *A.F.F. Benedetti[1], J.M. Silva[1], J.L.O. Madeira[1], Q. Ma[2], J.Z. Li[2], A.B. Ozel[2], B.B. Mendonca[1], I.J.P. Arnhold[1], S.A. Camper[2], A.C. Tahira[3], L.R.S. Carvalho[1].* 1) Unit of Developmental Endocrinology, Laboratory of Hormones and Molecular Genetics/LIM42. Clinics Hospital, University of Sao Paulo Medical School (HC-FMUSP), Sao Paulo, Sao Paulo, Brazil; 2) Department of Human Genetics, University of Michigan, Ann Arbor, USA; 3) Medical Investigation Laboratory from Institute of Psychiatry/LIM23. University of Sao Paulo Medical School (FMUSP), Sao Paulo, Sao Paulo, Brazil.

  SOX2 and *SOX3* genes are members of the SOX (SRY-related HMG-box) family of transcription factors involved in the pituitary development. Mutations in both genes were previously associated with congenital idiopathic hypopituitarism (CIH). The aim of this study was to determine *SOX2* and *SOX3* coverage in the whole exome sequencing (WES) for variants screening in a cohort of twenty-six patients (15 males) with CIH. Genomic DNA was isolated from blood leukocytes by standard techniques and then used to produce exome captured sequencing library by NimbleGen SeqCap EZ v3.0 kit, following manufacturer's protocol. Paired-end exome-sequencing was done using the Illumina HiSeq2000 System. Raw reads were aligned to the 1000 Genomes Phase 1 reference and mapped to GRCh37 using BWA v0.5.9. Single nucleotide variants, insertions and deletions were called using GATK Haplotype Caller v3.3. Variants were annotated using ANNOVAR. Lastly, the coverage of exomic regions corresponding to *SOX2* and *SOX3* genes were explored using the software Integrative Genomics Viewer (IGV), as well as the coverage suite from bedTools (coverageBed). The GC-rich single exon of both *SOX2* and *SOX3* were amplified by polymerase chain reaction using in-house designed primers. Amplicons were sequenced using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit and placed on the ABI PRISM 3100 sequencer. No variants in *SOX2* and *SOX3* were identified by WES. However, *SOX3* poly alanine tract and other parts of the coding region in both genes displayed low coverage, with less than a minimum of twenty reads per base. *SOX3* analysis by Sanger sequencing revealed the allelic variant p.P103T (c.307G→T, rs201101913) in heterozygous state in three patients. In silico analysis and variant population frequency was performed using the prediction sites Mutation Assessor, Mutation Taster, Polyphen2, Human Splicing Finder 3.0, ExAC, and 1000g and the variant was classified according to the American College of Genetics and Genomics (ACMG) as benign. It was previously demonstrate that WES by NimbleGen capture system showed a poor coverage of genes with high GC contents, such as *SOX2* and *SOX3*. Despite the rarity of *SOX2* and *SOX3* mutations, it is important for them to be screened since defects in these genes were associated to CIH. Thus, the screening of these genes in patients with CIH should be done by Sanger sequencing instead of using NimbleGen.

1202T

Hunting rare variants in chronic kidney disease. *S.R. Cameron-Christie[1], E. Groopman[2], C. Wolock[3], J. Fleckner[1], R. March[1], A. Platt[1], D. Goldstein[3], A. Gharavi[2].* 1) Centre for Genomics Research, Astrazeneca LTD, Melbourn, Royston, United Kingdom; 2) Department of Medicine, Columbia University, 635 W 165th St, New York, U.S.A; 3) Institute for Genomic Medicine, Hammer Health Sciences, 701 W 168th St, New York, U.S.A.

  Whole-exome sequencing (WES) has been successfully used to identify rare or private variants that dramatically increase the risk for common or complex disorders, helping to bridge the gap between the highly-penetrant mutations identified in families with Mendelian disorders and the common, low-risk polymorphisms that can be detected by genome-wide association studies. Case-control assessment of the gene-level burden of rare variants of large effect have provided insight into the molecular basis of complex disorders such as amyotrophic lateral sclerosis, inflammatory bowel disease, and idiopathic pulmonary fibrosis. However, to our knowledge, this framework has yet to be applied to the study of chronic kidney disease (CKD). To this aim, we performed a gene-based collapsing analysis using WES data from 2,710 CKD cases and 12,218 healthy or non-CKD controls. CKD cases included adults recruited from the nephrology clinics at Columbia University Medical Center (CUMC) (n=1,569), and the Astrazeneca-AURORA study (n=1141). The cohort was comprised largely of adults and was ethnically diverse (29.5% of patients were of self-declared non-white ancestry). Patients presented common clinical phenotypes, including glomerular, cystic, diabetic, and hypertensive kidney disease, making it broadly representative of all-cause CKD. We first excluded sites that had unbalanced sequence coverage between cases and controls, then used collapsing analysis to identify genes enriched in cases for rare variants predicted to be protein-damaging. We ran analyses under dominant and recessive inheritance models and using different allele frequencies, functional categories of variants, and with or without exclusion of diabetic and hypertensive phenotypes. We used principal component analysis to identify population stratification within and between the cases and controls, and performed ancestry-stratified and whole-population analyses. Preliminary analyses demonstrate enrichment in cases with concordant phenotypes for rare, deleterious variants in multiple genes known to cause Mendelian forms of nephropathy, such as *COL4A5* and *TRPC6*. In addition to known genes for nephropathy, the analysis yielded novel candidate genes for further investigation. Our findings support the potential of WES to elucidate the genetic architecture of CKD, inform understanding of its molecular pathogenesis and help facilitate the development of novel, targeted diagnostics and therapeutics.

1203F

BREATH: An open-access database and website of normal human and mouse lung development generated by the Molecular Atlas of Lung Development Program (LungMAP). *R.F. Clark[1], N.C. Gaddis[1], J.L. Levy[1], M. Duparc[1], H. Pan[1], G.P. Page[1], J.A. Whitsett[2], S.M. Palmer[3], LungMAP Data Coordinating Center.* 1) RTI International, Research Triangle Park, NC; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC.

The lung is a complex organ with high cellular heterogeneity, and research is needed to define the interactive gene networks and dynamic crosstalk among multiple cell types that coordinate normal lung development. Significant knowledge gaps exist in the understanding of lung development from late fetal to perinatal stages and through early childhood, a critical period when the diverse lung cells go through terminal differentiation and maturation and when alveoli form. Coordination of information from disparate assays and experimental approaches is critically needed to help the pulmonary research community uncover the molecular determinants of lung development. The goal of LungMAP is to build an open-access resource providing a comprehensive molecular atlas of late-stage lung development in humans and mice, making otherwise dispersed data, reagents, and protocols freely available to the research community. We have created the Bioinformatics REsource ATlas for the Healthy lung (BREATH) database, applying novel data management and bioinformatics approaches to manage high-throughput multidimensional experimental data. BREATH is built on a triplestore database backbone, integrated with anatomical ontologies for lung development, and provides access to novel web-based tools for the analysis and visualization of data generated by the four Research Centers and Human Tissue Core of the LungMAP Consortium. The current version of BREATH contains confocal immunofluorescence images, *in situ* hybridization images, histological images, nano-DESI, and 3D Vibra-SSIM confocal and uCT images of developing mouse and human lungs at several time points, as well as single-cell RNAseq, multi-cell RNAseq, microRNA, MeDIP-seq, proteomic, lipidomic, and metabolomic data from mouse and human lung cells. The LungMAP website (www.lungmap.net) provides an entry portal to the BREATH database and tools for exploring and interacting with the lung images and omics data. Future versions of BREATH will incorporate additional data types and novel tools for cross-datatype analysis. A better understanding of the basic molecular pathways that regulate normal lung development will enable development of innovative therapies that advance treatment of lung injury repair and regeneration.

1204W

An unbiased, genetic-data-driven benchmarking strategy for gene and variant prioritization algorithms. *R. Fine[1,2,3], T.H. Pers[4,5], Y.H. Hsu[1,2,3], J.N. Hirschhorn[1,2,3].* 1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 5) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark.

Determining the causal genes or variants from genome-wide association studies (GWAS) loci remains a daunting challenge. Many algorithms have been developed to prioritize causal genes/variants within associated loci, but benchmarking the relative success of these methods is not straightforward. A common approach is to assess whether known, "gold-standard" disease-related genes are preferentially prioritized. However, this method relies heavily on prior knowledge of disease etiology and is biased toward better-studied genes in well-characterized biological pathways. A better benchmarking approach would rely only on unbiased data, such as the genetic results themselves, rather than drawing on alternative sources of information whose relevance and comprehensiveness are uncertain. To address this need, we have developed Benchmarker, a leave-one-chromosome-out approach to perform a more robust benchmarking that uses the GWAS data as a gold standard: the GWAS serves as its own, unbiased control. From a GWAS data set, Benchmarker (1) removes the results from all SNPs from one chromosome, (2) applies the gene/variant prioritization method to the remaining GWAS data, and (3) determines if the prioritized variants or the SNPs in prioritized genes on the withheld chromosome are enriched for low p-values in the original GWAS. Benchmarker also generates a permutation-based p-value for the comparison of any two prioritization strategies. We applied our method to DEPICT, an algorithm which prioritizes genes from GWAS results based on shared biological functions, and show that across several well-powered GWASes, DEPICT's gene prioritization performs substantially better than chance. As a proof of principle, we used Benchmarker to show that DEPICT strongly outperforms a prioritization based on gene length only (p<0.002). We also compared DEPICT's gene prioritization performance based on either (1) shared gene set membership or (2) shared patterns of tissue expression. We found that using gene set membership was more effective than tissue expression for height and schizophrenia, while the reverse was true for inflammatory bowel disease and rheumatoid arthritis. Benchmarker can be used to fine-tune existing prioritization strategies, rigorously compare two different prioritization strategies, and compare the efficiency of prioritization based on different data types (e.g. GWAS vs ExomeChip). It thus provides a useful framework for future development of these algorithms.

1205T

The gene expression signature associated with rheumatoid arthritis is altered during pregnancy. *D. Jawaheer[1,8,11], D. Goin[1,2], M. Smed[3], L. Pachter[2,4], E. Purdom[2], J.L. Nelson[5,6], B. Ottesen[3], J. Olsen[7,8], M. Hetland[9,10], V. Zoffmann[3,10], H. Kjærgaard[3].* 1) UCSF Benioff Children's Hospital Oakland Research Institute, Oakland, CA, USA; 2) University of California, Berkeley, Berkeley, CA, USA; 3) Juliane Marie Centeret, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; 4) Caltech, Pasadena, CA, USA; 5) Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 6) University of Washington, Seattle, WA, USA; 7) University of California, Los Angeles, Los Angeles, CA, USA; 8) Aarhus University, Aarhus, Denmark; 9) DANBIO Registry and Copenhagen Centre for Arthritis Research, Rigshospitalet, Glostrup, Denmark; 10) University of Copenhagen, Copenhagen, Denmark; 11) University of California San Francisco, San Francisco, CA, USA.

Goal: To identify gene expression signatures associated with rheumatoid arthritis (RA) before pregnancy and to determine whether the signatures are altered during pregnancy, when disease activity naturally improves or worsens. **Methods:** Global gene expression profiles before pregnancy (T0) and at the third trimester (T3) were generated from 11 women with RA and 5 healthy women using RNA sequencing (RNA-seq). Disease activity scores were used to assess improvement or worsening of RA during pregnancy. Pseudo-alignment and quantification of transcripts were performed using kallisto. Differential expression (DE) analysis of normalized gene-level counts was performed using edgeR comparing (a) each RA subset (RA$_{improved}$ or RA$_{worsened}$) vs. healthy women at each time-point, and (b) T3 vs T0 profiles within each group of women. Functional enrichment analysis was performed using WebGestalt. **Results:** Among the women with RA, 8 improved during pregnancy and 3 worsened. In the RA$_{improved}$ vs healthy analysis at T0, 94 genes had expression patterns associated with RA. These were enriched in various immune-related gene ontology processes. At T3, when disease activity improved, the majority of these genes (79 of 94) had expression patterns that were no longer associated with RA. These included several genes that have been previously reported as part of RA expression signatures. Their loss of DE at T3 was due to changes in expression (from T0 to T3) in RA and/or healthy women such that expression profiles became comparable between the 2 groups by T3. When the RA$_{worsened}$ group was compared to healthy women, 83 genes were differentially expressed at T0. There was little overlap (13 genes) between these baseline signature genes and those identified among RA$_{improved}$ women. Further, only 38 of the 83 genes (46%) lost their DE by T3 when RA worsened. Interestingly, an additional set of 240 genes became differentially expressed at T3 as might be expected with increasing differences in clinical phenotype between the RA$_{worsened}$ and healthy groups as RA worsened. **Conclusions:** In our pilot data, when disease activity improved during pregnancy, expression profiles of numerous RA signature genes became comparable between RA and healthy women. On the other hand, when disease activity worsened, there was an increase in differentially expressed genes during pregnancy.

1206F

The accuracy of whole genome sequencing and the Platypus variant caller in identifying genetic variation within the structurally challenging epidermal differentiation complex. *C. Malley[1], M. Boorgula[2], S. Chavan[2], N. Rafaels[2], D. Leung[3], K. Barnes[2], R. Mathias[1].* 1) Division of Allergy and Clinical Immunology, Johns Hopkins University, Baltimore, MD; 2) Colorado Center for Personalized Medicine, University of Colorado, Department of Medicine, Aurora, CO; 3) Pediatric Allergy and Clinical Immunology, National Jewish Health, Denver, CO.

Rationale: Segmental duplications and tandem repeats in the Epidermal Differentiation Complex (EDC; chr1:151972910-153642037) present challenges to sequencing approaches. Within the Atopic Dermatitis Research Network (ADRN), whole genome sequencing (WGS) is available on ~800 subjects. Given the implications of the EDC in development and severity of atopic dermatitis (AD), increased genotype resolution leveraging WGS will aid the search for disease-associated genes. Here, we apply the Platypus variant caller to integrate the calling of structural variants and single nucleotide polymorphisms within a single framework, and evaluate the accuracy of variant calling within the EDC. **Methods:** Three sets of genetic data were available on N=799 European American subjects from the ADRN: 1) 30X WGS generated on the Illumina HiSeq; 2) the OMNI 2.5 GWAS array; and 3) TaqMan genotypes on a set of four variants in the Filaggrin (FLG) gene of prior clinical interest (R501X, 2282del4, R2447X, and S3247X). The Platypus variant caller was run on all subjects to create multi-sample variant call files (VCF) containing both SNPs and indels. Only sites that passed the variant caller flags and genotypes with coverage and/or read quality >7 were retained for analysis. All platypus SNP calls were compared to the OMNI and TaqMan genotype data to determine concordance. This similar pipeline was run on three additional gene regions without known segmental duplication and tandem repeat issues of equal length as the EDC for comparison. **Results:** There were 15,636 SNPs passing QC filters in the EDC with Platypus with mean total coverage per variant of 30,971 (range: 337-46,696). We found a mean of 99% (range: 98-100%) genotype concordance between Platypus and OMNI, from a subset of 850 non-monomorphic transition overlapping SNPs called in both the OMNI and WGS data. Additionally, there was 98% (787 of 799) genotype concordance with TaqMan called FLG variants. Overall coverage and variant calling statistics were comparable to the three other gene regions. **Conclusions:** The Platypus variant caller provides high quality variant calls from whole genome sequencing data rife with structural variation, which has led to unreliable or impossible calls in the past. Platypus-called genotypes are highly concordant with older-generation genotyping array and qPCR data. We are confident in the quality of the software's joint calling for generating squared-off multi-sample VCFs of both SNPs and indels.

1207W

Somatic mutation hunting: The search for the genetic architecture of linear localized scleroderma. *R.G. Higgins[1,2], M. Theiler[1,3], A. Smith[1,3], R. Wälchli[1,3], L. Weibel[1,3], A.A Navarini[1,2].* 1) Department of Dermatology, UniversitätsSpital Zürich, Switzerland; 2) Department of Dermatology, University of Zurich, Switzerland; 3) Kinderspital Zürich, Zürich, Switzerland.

   Linear localized scleroderma (LLS) is a rare connective tissue disorder (2.7/100'000 per year) characterised by chronic inflammation and massive accumulation of collagen. This results in hardening and thickening of the lesion leading to the affected areas to cave in from atrophy. The sharply delimited and linear lesions, can affect patients throughout the body and more rarely the face in the clinical subtypes *en coup de sabre* and Parry-Romberg syndrome. This leads to terrible disfigurement. The disease affects mostly children and is limited in treatment options, which are most often unsatisfactory. Very little is understood about the condition in terms of genetic and clinical aetiology. There is evidence that LLS might be based on genetic alterations in affected tissues. Blaschko's lines are the patterns of cell migration and proliferation during embryological development. Multiple skin conditions have been shown to follow Blaschko's lines including LLS. Several of these diseases have been demonstrated to be caused by genetic factors such as a *de novo* somatic mutation causing a cutaneous mosaicism. Here we test the hypothesis that LLS is caused by a somatic genetic mutation. Blood and affected skin taken from 19 confirmed LLS patients was Whole Exome Sequenced (WES). Library preparation and hybridization was performed using the SureSelectXT Reagent kits (Agilent Technologies). Sequence analysis was performed as per GATKv3.5. Somatic mutations were called using 3 somatic callers. All SNPs and indels were analysed and no rare and damaging mutation was found in common. 5 patients had damaging germline SNP mutations in *PRSS3*, a gene involved in metabolism and the immune system. No minor allelic frequency data is available for any of the *PRSS3* mutations found so comparison with internal controls has shown 2 of 5 to be truly rare. An analysis of the 3 somatic callers revealed that an increase in read depth (DP) allowed for a more sensitive detection of mutations at low allelic fraction. To this end 4 patients with similar lesions were chosen for deep sequencing (DP of >300). However, somatic analysis of these samples revealed no suitable causative mutation. CGH was performed on 3 patients to find large scale chromosomal aberrations too large to be detected by WES. Samples were compared with 3000 control karyotypes. No rare aberrations were found in common between patients. Taken together, our analysis revealed the absence of genetic mosaicism in LLS.

1208T

***In silico* screening for potentially aberrant novel essential splice sites (PANESS) created by single nucleotide variants (SNVs) in the human genome.** *C.C. Bruels[1], T. Mendoza[2], J. Khan[2], H.M. Reddy[1], E.A. Estrella[3], M. Lek[4,5,7,8], B.B. Cummings[4,5,6], E. Valkanas[4,5], D.G. MacArthur[4,5], L.M. Kunkel[3], F. Modave[2], P.B. Kang[1,9,10].* 1) Division of Pediatric Neurology, Department of Pediatrics, University of Florida College of Medicine, Gainesville, FL; 2) Department of Health Outcomes & Policy, University of Florida College of Medicine, Gainesville, FL; 3) Division of Genetics & Genomics, Boston Children's Hospital and Harvard Medical School, Boston, MA; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 5) Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 6) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; 7) School of Paediatrics and Child Health, University of Sydney, Sydney, Australia; 8) Institute for Neuroscience and Muscle Research, Kids Research Institute, The Children's Hospital at Westmead, Sydney, Australia; 9) Department of Molecular Genetics & Microbiology, and Department of Neurology, University of Florida College of Medicine, Gainesville, FL; 10) Genetics Institute and Myology Institute, University of Florida, Gainesville, FL.

   The muscular dystrophies are inherited, progressive muscle disorders characterized by destruction and wasting of muscle tissues. Although genetic mechanisms for numerous muscular dystrophies are well defined, many affected patients remain undiagnosed. Whole exome (WES) and whole genome (WGS) sequencing are commonly used to discover pathogenic mutations underlying rare genetic disorders. WES of limb girdle muscular dystrophies (LGMDs) typically identifies pathogenic mutations in approximately 40% of cases. Although exons remain the most common sites of causative mutations, recent studies are increasingly focusing on the identification and characterization of pathogenic mutations in regulatory regions, including canonical RNA splice sites. *In silico* tools currently available for identifying RNA splicing effects examine variants only in or near canonical splice sites at intron/exon boundaries. Whole transcriptome sequencing (RNA-seq) can identify splice sites created outside these boundaries; however, not all cases are amenable to this approach. *In silico* evaluation of existing WES/WGS data can identify potentially aberrant novel essential splice sites (PANESS). We developed a pipeline that uses the ANNOVAR software tool for data file conversion and annotation, and uses code written in Python v.2.7.6 to evaluate the potential effect of SNVs. We used the ExAC v.0.3.1 dataset of human protein-coding genetic variants to validate the pipeline. We applied the validated pipeline to sequence data from two families with uncharacterized neuromuscular disorders. The first family has several individuals affected with LGMD. The pipeline identified 189 low frequency (≤ 0.001) variants that formed PANESS on the chromosome containing a previously identified linkage peak, of which 13 variants were within the actual linkage peak. The proband in the second family is affected with myotonia congenita. The pipeline identified 536 variants that created PANESS in 480 genes, 18 of which are involved in other neuromuscular disorders. Further analysis is ongoing to identify the causative mutations in both families. Although specifically designed to assess SNVs that might be involved in neuromuscular disorders, the pipeline is broadly applicable to any disease category. The list of affected genes can be used in pathway, gene ontology, and other *in silico* analyses to prioritize the list of candidate variants prior to laboratory characterization of the effect of the variants.

1209F

Computational prediction and molecular validation of novel therapeutic targets for potent splicing modulators. *D. Gao[1,2,3], E. Morini[1,2], M. Salani[1,2], C. Montgomery[1,2], S. Erdin[1,2,3], M.E. Talkowski[1,2,3,4,5], S.A. Slaugenhaupt[1,2,5].* 1) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 4) Departments of Psychiatry and Pathology, Massachusetts General Hospital, Boston, MA; 5) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA.

 While Mendelian genetics has seen remarkable advances in recent years in the discovery of novel genes and pathogenic mechanisms, therapeutic development remains incremental. One strategy is to discover novel targets for known compounds, and therapeutic targeting of splicing defects is one area in which progress has been made. Here, we applied a combination of computational and molecular approaches to identify such targets. We demonstrate that the pathogenic mechanism underlying familial dysautonomia (FD), namely skipping of exon 20 in the *IKBKAP* gene, is partially rescued by the plant cytokinin kinetin, which improves exon inclusion and increases protein both *in vitro* and *in vivo*. We have developed other kinetin-like splicing modulator compounds (SMCs) to improve drug efficiency and pharmacokinetic properties. However, to what extent SMCs influence splicing across the transcriptome is unclear. We thus conducted a transcriptome-wide *in silico* analysis to identify the sequence signatures that determine response to treatment and to extend the therapeutic usefulness of SMCs. Six independent human fibroblast cell lines from healthy individuals were treated with kinetin and a more potent SMC, BPN-15477. From RNA-seq of these, we investigated splicing by assessing all plausible exon 'triplets' (three consecutive exons) genome-wide. For each triplet, we calculated the Percentage of Splice-in (PSI, ψ), or the frequency of inclusion of the middle exon. This analysis revealed 97 events where ψ changed in response to kinetin and 242 events in response to BPN-15477. We then trained a machine learning approach based on genomic feature enrichment within ±100bp of the middle exons with the most significant ψ changes and with no ψ changes. We found that fourteen sequence features were sufficient to distinguish these two groups of middle exons (ROC=0.89). We are using these signatures to predict which disease-causing splice mutations will likely respond to treatment. Our study is an example of coupling machine learning strategies with classic transcriptome analysis to determine potential genomic drug targets.

1210W

Alterations of transcriptome landscaping in head trauma-related human brain disorders. *H. Cho[1,2], T. Stein[5,6,7], J. Lee[5,6], N. Kowall[5,6], A. McKee[5,6,7], J. Shin[1,3], H. Ryu[4,5,6], J. Seo[1,2,3].* 1) Genomic Medicine Institute, Medical Research Center, Seoul National University, Seoul, 110-799, South Korea; 2) Department of Biomedical Sciences, Biomedical Science Building 117, Seoul, South Korea; 3) Life Science Institute, Macrogen Inc., Seoul, 153-781, Korea; 4) Center for Neuromedicine, Brain Science Institute, Korea Institute of Science and Technology, Seoul, 02792, Korea; 5) VA Boston Healthcare System, Boston, MA, 02132, USA; 6) Boston University Alzheimer's Disease Center and Department of Neurology, Boston University School of Medicine, Boston, MA, 02215, USA; 7) Boston University Chronic Traumatic Encephalopathy Center and Department of Pathology, Boston University School of Medicine, Boston, MA, 02215, USA.

 Chronic Traumatic Encephalopathy (CTE) is a neurodegenerative disease that is associated with traumatic brain injury. CTE and Alzheimer's disease (AD) are known to share similar neuropathological features, which mainly consist of the presence of neurofibrillary tangles and hyperphosphorylated tau, but they have generally been described as distinct entities. To better understand the neuropathological mechanism of CTE and AD, we conducted transcriptome sequencing analysis of post-mortem human brain tissue. Moreover, in order to characterize unique or common transcriptome signatures among head trauma-related disorders, we also included postmortem brains with chronic traumatic encephalomyelopathy (CTEM), an amyotrophic lateral sclerosis (ALS)-like disorder, and CTE/AD in the RNA sequencing analysis. Using the weighted gene co-expression network analysis (WGCNA) and principal component analysis (PCA), we found a clear distinction between CTE and CTEM and as well as similar distributions of CTE and AD. Our systematic and comprehensive analysis serves as a framework to reveal important transcriptome signatures that will be useful for the diagnosis of disease progression and the design of therapeutic strategy in head trauma-related neurodegenerative disorders.

1211T

NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS)
Genomics Database. *E. Greenfest-Allen[2,4], P. Gangadharan[2,4], A. Kuzma[1,2], L. Qu[1,2], O. Valladares[1,2], Y.Y. Leung[1,2,3], A. Naj[5], G.D. Schellenberg[1,3], C.J. Stoeckert Jr.[2,4], L. Wang[1,2,3].* 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine; 2) Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine; 3) Institute on Aging, University of Pennsylvania Perelman School of Medicine; 4) Department of Genetics, University of Pennsylvania Perelman School of Medicine; 5) Department of Epidemiology, University of Pennsylvania Perelman School of Medicine.

   Statement of Purpose: NIAGADS is a national genetics data repository that facilitates access of genotypic data to qualified investigators for the study of late-onset Alzheimer's disease (AD) and related neuropathologies. The GenomicsDB is a searchable annotation resource that provides public access to genome-wide association study (GWAS) summary statistics datasets reposited at NIAGADS. These data are linked to variant and gene annotations and functional genomics datasets, allowing AD researchers to easily identify and interpret interesting genomic regions via detailed record pages, interactive search strategies, and a genome browser.   **Methods:** The GenomicsDB is powered by the Genomics Unified Schema (GUS), a relational database system comprising a modular schema capturing sequence data, functional genomics data, and rich descriptions of methodology and study design. GUS uses OBO Foundry ontologies to annotate all captured datasets and associated meta-data, facilitating data harmonization and integration. The GUS system underlying the NIAGADS GenomicsDB is served via an open-source PostgreSQL 9.5 database, optimized for big-data storage and querying. The website is built using the Strategies-WDK system, which provides both a back-end web development kit and a front-end "strategy"-based graphical search interface for functional genomics databases. "Strategies" are interactive graphical workflows that enable sophisticated data-mining of the GWAS summary statistics datasets via filtering, Boolean logic, and comparisons to sequence annotations or uploaded user-data via colocation queries. Strategy results can be saved and shared with other researchers. Searches can be made for specific record types (e.g., variants, genes, genomic regions) and then combined or transformed (e.g., variants to genes) using the Strategies interface. The NIAGADS GenomicsDB also provides a genome browser, built on the GMOD:JBrowse framework, allowing users to visually inspect NIAGADS reposited tracks and compare against personal data.   **Results**: As of June 2017, the NIAGADS Genomics Database (https://www.niagads.org/genomics) makes available summary statistics from eight NIAGADS GWAS accessions (35 datasets; >150 million variant annotations) for interactive exploration and data-mining. With a newly redesigned search interface and comprehensive record pages linking the summary statistics to variants and genes, the GenomicsDB is a rich resource and valuable tool for AD research.

1212F

Integrated causal network analysis of genomic and epigenomic data.
Z.X. Hu[1,2], N. Lin[2], P.P. Wang[1], Y. Zhu[3], J.Y. Zhao[3], D.A. Bennett[4], L. Jin[1], M.M. Xiong[2]. 1) School of Life Science, Fudan University, Shanghai, China; 2) Department of Biostatistic, Health Science Center at Houston, TX, US; 3) Division of Genetic Epidemiology at the College of Public Health and Health Professions and the College of Medicine at the University of Florida; 4) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA.

   The generation of large multiscale omic data requires new methods for integration with phenotypic data in terms of breadth and depth. Many extant approaches perform analyses individually and separately. The traditional analytic platforms for analyzing biological datasets use association or correlation analysis. However, systematic omic data analysis needs to uncover causal relationships among various molecular components, which form complex biological systems. How to identify the causal relationship inherent in multiscale omics and clinical phenotypes and perform integrated analysis of WGS, other omics data, and clinical data is a key issue to the analysis. Deep analyses of high dimensional and heterogeneous types of correlated omic and clinical data pose huge challenges. We develop a unified analytical framework for systematic causal decomposition through novel statistical methods of trans-omic networks integrating heterogeneous genomic, environmental, RNA-seq, DNA methylation and phenotypic data into multilayer networks underlying disease and health. The proposed method was applied to genetic and epigenetic studies of five diseases: hypertension, obesity, type 2 diabetes, Alzheimer's disease (AD), and Lewy Body disease to infer causal genotype-environment-expression-phenotype-disease networks**.** The inferred network consists of **58,207 nodes** and **192,939 edges**. We found that **960 Genes** were directly connected to phenotype nodes, **1501 genes** to the disease nodes in the causal networks and that **395 genes** and **2088 methylated genes** were connected to **609 gene expression nodes**. **14,368 paths** were identified from these genes to diseases. Several notable features emerge. 1) Several genes such as *HLA-DRB6, HLA-DQA1 / HLA-DQB1, NFKB1and POU3F2 (Brn2)* that play an essential role in AD via expression and methylation causal networks were identified. 2) **Half of the edges** in the signaling pathways were consistent with the pathway structure in the KEGG pathway database. 3) Several diseases appear to share common disease risks, pathways and genes. For example, the *HLA-DRB5* gene, its expression, and methylation, all had direct effects on AD, hypertension, and obesity, and AD share gene *CREBBP*, and hypertension, obesity and Lewy body disease share gene *KMT2C*. 4) Causal omic network analysis can identify the causal pathways from gene to disease via environment, gene expression and phenotype (risk factors) that was supported by literatures.

1213W

FGWAS: Functional Genome Wide Association Analysis. *C. Huang[1], P. Thompson[2], Y. Wang[3], J. Zhang[1], D. Kong[4], R. Colen[5], R. Knickmeyer[6], H. Zhu[1].* 1) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX and University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 2) Imaging Genetics Center, Stevens Institute for Neuroimaging and Informatics, University of Southern California, Marina del Rey, CA, USA; 3) School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA; 4) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 5) Department of Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 6) Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

   Introduction: Functional phenotypes (e.g., subcortical surface representation), which commonly arise in imaging genetic studies, have been used to detect putative genes for complexly inherited neuropsychiatric and neurodegenerative disorders. However, existing statistical methods largely ignore the functional features (e.g., functional smoothness and correlation). Methods and Results: The aim of this paper is to develop a functional genome-wide association analysis (FGWAS) framework to efficiently carry out whole-genome analyses of functional phenotypes. FGWAS consists of three components: a multivariate varying coefficient model, a global sure independence screening procedure, and a test procedure. Compared with the standard multivariate regression model, the multivariate varying coefficient model explicitly models the functional features of functional phenotypes through the integration of smooth coefficient functions and functional principal component analysis. Statistically, compared with existing methods for genome-wide association studies (GWAS), FGWAS can substantially boost the detection power for discovering important genetic variants influencing brain structure and function. Simulation studies show that FGWAS outperforms existing GWAS methods for searching sparse signals in an extremely large search space, while controlling for the family-wise error rate. We have successfully applied FGWAS to large-scale analysis of data from the Alzheimer's Disease Neuroimaging Initiative for 708 subjects, 30,000 vertices on the left and right hippocampal surfaces, and 501,584 SNPs. Conclusion: Three key advantages of FGWAS have been discovered: (i) the spatial correlation structure of imaging data and variability of multiple phenotypic measurements considered in the multivariate varying coefficient model; (ii) much lower computational complexity compared to standard functional GWAS, and (iii) a parallel computing strategy that makes FGWAS feasible for super large-scale genetic data. Our FGWAS is a valuable statistical toolbox for fast, large-scale imaging genetic analysis.

1214T

NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS): Update 2017. *H. Lin[1,2], O. Valladares[1,2], A. Kuzma[1,2], E. Greenfest-Allen[2,4], P. Gangadharan[1,2], D. Childress[1,2], R. Cweibel[1,2,3], C. Zhong[1,2], Y. Zhao[1,2], L. Qu[1,2], Y. Leung[1,2,3], A. Naj[1,5], C. Stoeckert Jr.[1,2,4], G. Schellenberg[1,3], L. Wang[1,2,3].* 1) Penn Neurodegeneration Genomics Center, University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Institute on Aging, University of Pennsylvania, Philadelphia, PA; 4) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 5) Department of Epidemiology, University of Pennsylvania, Philadelphia, PA.

  Statement of Purpose: NIAGADS is a national genetics data repository that facilitates access of genotypic data to qualified investigators for the study of the genetics of late-onset Alzheimer's disease (AD) and other neurological diseases. Collaborations with large consortia such as the Alzheimer's Disease Genetics Consortium (ADGC) and the Alzheimer's Disease Genome Sequencing Project (ADSP) allow NIAGADS to lead the effort in managing large AD datasets that can be easily accessed by the research community and utilized to their fullest extent. **Methods:** Since 2012, NIAGADS has been supported by National Institute on Aging (NIA) under a cooperative agreement (U24 AG041689). All data derived from NIA funded AD genetics studies are expected to be deposited in NIAGADS or another NIA approved site. NIAGADS has partnered with the database of Genotypes and Phenotypes (dbGaP) and the Sequencing Read Archive (SRA) in this effort. NIAGADS is continually developing new databases to facilitate neurodegenerative research. The redesigned NIAGADS Genomics Database is a searchable annotation resource that links published AD studies to AD-relevant sequence features and genome-wide annotations. Based on INQuery developed by CNDR at Penn, the redeveloped genotype/phenotype database allows qualified users to log in and filter for subjects with a particular phenotype and genotype and see if there is a DNA sample available at the National Cell Repository for AD (NCRAD). **Results:** As of June 2017, NIAGADS houses 44 datasets with >55,000 samples and over 24 billion genotypes. With the completion of the Discovery Phase of ADSP, qualified investigators can retrieve sequencing data with ease and flexibility using the ADSP website and data portal (collaboration with dbGaP/SRA). The ADSP project has completed whole-exome sequencing of 10,939 subjects and whole-genome sequencing of 4,036 subjects; raw data as well as quality controlled VCF files for the Discovery Phase are available through the ADSP portal and dbGaP. **Conclusions:** NIAGADS is a rich resource for AD researchers. Datasets, guidelines, and new features are available on our website at https://www.niagads.org.

1215F

The Alzheimer's Disease Sequencing Project (ADSP) data update 2017.

Y. Zhao[1], A.B. Kuzma[1], K.M. Faber[3], W.J. Salerno[6], Y.Y. Leung[1], L.B. Cant-well[1], M. Feolo[4], A. Stine[4], N. Gupta[5], R. Fulton[7], O. Valladares[1], R. Cweibel[1], E. Applebaum[7], S.H. Choi[2], K.L. Hamilton-Nelson[11], H.-J. Lin[1], D. Muzny[6], L. Qu[1], D. Reyes[9], J. Waligorski[7], J. Farrell[12], A.C. Naj[1], J.C. Bis[8], A.L. DeSte-fano[2], S. Seshadri[10], E. Boerwinkle[6], G.D. Schellenberg[1], T.M. Foroud[3], L.-S. Wang[1], Alzheimer's Disease Sequencing Project Data Flow Work Group. 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) School of Public Health, Boston University, Boston, MA; 3) Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana; 4) National Center for Biotechnology Information, National Institutes of health, Bethesda MD; 5) Broad Institute of Harvard and MIT, Cambridge, MA; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 7) The Genome Institute, Washington Univ., St. Louis, MO; 8) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 9) Taub Institute for Research on Alzheimer's Disease, Columbia University, New York; 10) Department of Neurology, Boston University School of Medicine, Boston, MA; 11) John P Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 12) School of Medicine, Boston University, Boston, MA.

Statement of Purpose. The Alzheimer's Disease Sequencing Project (ADSP) was established in 2012 as a key initiative to meet the goals of the National Alzheimer's Project ACT (NAPA) to prevent and effectively treat Alzheimer's disease (AD) by 2025. Developed jointly by National Institute on Aging (NIA) and the National Human Genome Research Institute (NHGRI), the aims of the ADSP are to: 1) identify protective genomic variants in older adults at risk for AD, 2) identify new risk variants among AD cases, and 3) examine these factors in multi-ethnic populations to identify therapeutic targets for disease prevention. **Methods.** The ADSP Data Flow Work Group (DFWG) and the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) support data production, sharing and management for the ADSP, and facilitation of data access to the community. Samples are contributed by the AD Genetics Consortium (ADGC) and Cohorts for Heart and Aging Research in Genomic Epidemiology consortium (CHARGE). Sample plating and shipping is coordinated by the National Cell Repository for AD (NCRAD). Whole-exome (WES) and whole-genome (WGS) sequencing data were generated by three NHGRI funded Sequencing Centers and quality controlled by the ADSP. The DFWG maintains the ADSP web site, providing study design, cohort information, news releases, and application instructions for ADSP access. The ADSP Data Portal is a collaboration with the Database of Genotypes and Phenotypes (dbGaP), Sequence Read Archive (SRA), and NIAGADS. The portal allows users to explore the ADSP project data archived at dbGaP with NIH iTrust user authentication. Approved investigators identify data and download files using a customizable filtering system and check-out cart function. **Results.** The ADSP has completed whole-exome sequencing of 10,939 unrelated cases and controls, whole-genome sequencing of 892 members from 159 families, and whole-genome sequencing of 3,144 unrelated cases and controls. All newly sequenced genomes are being processed on GRCh38 by the Genome Center for AD (GCAD) and are expected to be available in late 2017. Accompanying each sample are phenotypes that were harmonized according to ADSP protocols. **Conclusion.** The DFWG provides support to all ADSP Work Groups for data related issues, coordinates with dbGaP for data transfers, and reviews, posts, and notifies members of new results generated by other Work Groups. Find additional information at the ADSP website (www.niagads.org/adsp). .

1216W

Personalized and cell-specific pathway score computations from risk alleles and regulatory information in 2370 subjects with multiple sclerosis.

L.R. Madireddy[1], K. Kim[1], S.J. Caillier[1], S.L. Hauser[1], J.R. Oksenberg[1], S.E. Baranzini[1,2]. 1) Department of Neurology, Weill Institute for Neurosciences, University of California San Francisco, San Francisco, USA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco, USA.

More than a decade of successful GWA studies on Multiple Sclerosis (MS) has revealed close to 200 genetic loci associated with MS. However, progress on functional characterization, cell type specificity, and prioritization of genes within each of these loci has been very limited. To address these biologically important questions, we developed a novel bioinformatics and network based pipeline where the most functionally associated gene per region and cell type are identified, leveraging multiple sources of evidence including ENCODE and Roadmap Epigenomics Project. Gene regulatory data were scraped from RegulomeDB website for SNPs that passed a p-value threshold of 10^{-4} in a large meta-analysis with ~ 47,000 patients and for the SNPs tagging them. We then grouped experiments into 10 cell type buckets of potential interest for MS. In addition, we classified histone modifications into 3 broad categories representing transcriptional promotion/quiescence/repression. We computed normalized weights for each SNP based on number of times the SNP is shown to be associated with a particular regulatory activity in that particular cell type. These weights were summed across SNPs to compute sum of weighted weights per gene per region across cell types. We ascribed a gene to a region per cell type based on the gene that has highest absolute sum of weighted weights score in that cell type in that region. We next overlapped the gene regulatory potential with an existing protein interaction network, and performed a topological analysis from the resulting sub networks at each cell type. We observed a significant enrichment of interactions across prioritized genes at various cell types compared to the number of connections from randomly generated networks. Finally, we computed individual regulatory burden scores across genes and cell types for 2782 samples from UCSF and observed individualized cell-specific association patterns. This method is relevant for prioritization of genes within a region, and for individualized risk stratification, where networks of associated genes in different cell types can be potentially used to inform therapeutic approaches or disease management strategies. This novel approach will open new avenues to integrate past and future GWA studies with publicly available regulatory information and shines more light on the functional aspects of disease etiology and on the use of individual genetic scores to discover associations with additional phenotypes.

1217T

Omics-based machine learning modeling of monogenic neurological diseases. *J.A. Botía[1,2], S. Guelfi[1], K. D'sa[1,3], J. Vandrovcova[1], H. Houlden[1], J. Hardy[1], M. Ryten[1,3].* 1) Molecular Neuroscience, University College London, London, United Kingdom; 2) Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, Spain; 3) Medical and Molecular Genetics, King's College London, United Kingdom.

Dysfunction of the central and peripheral nervous system is a feature of around 50% of all rare diseases. Yet we know relatively little about why some genes give rise to human brain diseases, whereas others do not. In order to address this question we used machine learning (ML) to model genes associated with neurological diseases from their omics-based features. These features, which are used as predictors, included: i) gene-based measures of mutational constraint generated by ExAC (e.g. probability of intolerance to loss of function or missense variants); ii) gene complexity measures derived from GENCODE (e.g. gene length, and number of transcripts per gene); and iii) tissue-specific gene expression and co-expression metrics generated using GTEx V6 gene expression data covering 47 human tissues (e.g. adjacency for global hubbyness and module membership for local connectivity). We used the Genomics England Panel App to obtain expertly curated lists of genes associated with a range of neurological and neurodevelopmental disorders. Non-disease genes, were defined as all genes not currently known to be linked to any disease. We developed a bootstrapped decision tree model (BDTM) of n=200 trees for each disease gene panel containing >10 genes (n=23). This led to the construction of 23 classifiers, one for each disease type, with an additional classifier, which used all genes associated with any neurological or neurodevelopmental disorder. Using this approach the BDTM created using all neuro-genes predicted 276 new neuro-genes (threshold of 99% agreement on trees within the BDTM). Interestingly, using the Mouse Phenotype Database, we found that this gene list was significantly enriched for genes associated with abnormal neuron morphology (p-value = 1.39×10^{-6}) and abnormal synaptic transmission (p-value = 1.79×10^{-6}) amongst other phenotypes. Furthermore, using DisGeNET, a database of genotype-phenotype associations including those generated through genome-wide association studies, we identified 14 disease-specific enrichment signals including addiction related disorders (p-value = 4.53×10^{-13}), glioblastoma (p-value = 1.11×10^{-7}), autism spectrum disorders (p-value 2.63×10^{-4}), schizophrenia (p-value = 3.56×10^{-4}) and Alzheimer's disease (p-value = 7.08×10^{-3}). Thus, we demonstrate both the predictive and explanatory power of ML-based models in the context of neurological and neurodevelopmental diseases.

1218F

Identification of tolerated reading-frame changes induced by stop-lost and frame-shift variants in Alzheimer's disease. *M. Butkiewicz, J.L. Haines, W.S. Bush for the Alzheimer s Disease Sequencing Project.* Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Background: With increasing life expectancy, Late Onset Alzheimer's disease (LOAD) is affecting an expanding population of elderly individuals and is a growing sociological and financial burden on society. Genetic variants that contribute to a 'change-of-function' character for gene transcripts are of interest with respect to disease pathogenesis. Here, we aim to identify tolerated reading-frame changes within a cohort of well-aged individuals, and to then prioritize variants within a cohort of late onset Alzheimer's disease. **Methods:** We analyzed a whole-genome sequencing (WGS) dataset, comprised of 578 individuals from 111 European Caucasian and Caribbean Hispanic families, from the Alzheimer's Disease Sequencing Project (ADSP), an initiative to identify genetic variation associated with LOAD risk. As a control, we accessed WGS called variants from the Wellderly study, a cohort of 511 successfully aged individuals, older than 80 years, without significant medical conditions. Both datasets were analyzed by COCOS, a recently published plugin for the Ensembl variant effect predictor (VEP), which provides information about the altered reading frame and the translated protein sequence changes resulting from stop-lost variants and small insertions and deletions (InDels). We determined whether these variant-induced reading-frame changes would affect functional domains within the translated protein sequence using the Pfam database, as well as sequence similarity compared to the unaltered protein sequence. **Results:** Applying the COCOS VEP plugin, we identified reading-frame altering variants in the ADSP (9,593 InDels, 352 stop-lost) and the Wellderly (11,450 InDels, 428 stop-lost) datasets. An overlap of 986 InDels was observed for the ADSP (986/9,593=10%) when compared to the Wellderly (986/11,450=9%). 55% (542/986) of the overlapping ADSP InDels induced reading-frame changes leading to a disruption in at least one Pfam domain, and exhibited an average sequence similarity of 16% compared to the original protein sequence. Also, 83 stop-lost variants were overlapping between the Wellderly (83/428=19%) and ADSP (83/352=24%). Overlapping InDels and stop-lost variants can be seen as a reduction in number of variants to test with respect to AD. **Conclusions:** Considering the Wellderly cohort as a source for seemingly more tolerated variation, the overlap of reading-frame changing variants between the ADSP and Wellderly cohorts suggests a reduced significance to AD risk.

1219W

Missense variant interpretation based on mutational burden at analogous amino acid positions across gene family members. *E. Perez-Palma[1], P. May[2], P. Nürnberg[1], M. Daly[3,4,5], A. Palotie[3,4,5,6], D. Lal[1,3,4,5].* 1) Cologne Center for Genomics, University of Cologne, Cologne, NRW, Germany; 2) Luxembourg Centre for Systems Biomedicine, University Luxembourg, Luxembourg; 3) Massachusetts General Hospital, Boston, MA, USA; 4) Harvard Medical School, Boston, MA, USA; 5) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 6) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.

Missense variants are the most frequent disease-associated mutations, but also the most challenging to interpret. Current *in silico* methods are imperfect in predicting their functional consequences and potential disease trajectories. Ultra large exome aggregation approaches such as gnomAD have allowed to identify and score genes and exons intolerant for missense variants. However, these do often not show high enough accuracy and the number of variants is still too low to calculate intolerance with amino acid resolution. We have recently shown that paralogous conserved amino acids are more likely than non-conserved residues to give risk to disease when mutated. In the present work, we propose a novel framework using the total amount of variants of the whole gene family to calculate gene-specific local constrained scores with amino acid resolution. We hypothesize that paralogous amino acid positions across the whole gene family are essential when never mutated in individuals from the general population. Gene families and paralog definition were retrieved from the maximum likelihood phylogenetic gene trees provided by Ensembl. For each gene, the longest CCDS isoform was considered. Non-coding genes and genes without a HGNC symbol were excluded. All alignments were performed using the multiple sequence comparison by log-expectation (MUSCLE) software. Mutations from the general population were extracted from the genome Aggregation Database (gnomAD, version 2.0), neurodevelopmental disorders (NDDs) mutations were extracted from ClinVar database (March 15, 2017). In total, we generated 2,871 gene families with at least 80% identity in their amino acid sequence comprising a total of 9,991 genes. We were able to map 2,219,811 and 97,391 missense mutations from gnomAD and ClinVar, respectively, into these alignments. Gene-specific local constrained scores were generated using a sliding window approach evaluating the proportion of gnomAD variants found among the alignments of each gene family. Our approach allowed to identify mutation-constrained paralog domains enriched for *de novo* variants from 11,000 trios with NDDs and ClinVar variants. Our derived gene-specific scores identify sites in proteins which are never mutated in the general population at the same position across proteins from the same gene family. In particular, in large gene families, such as ion channels, our approach will facilitate variant interpretation and identify functionally essential domains.

1220T

Evaluation of basic massive parallel sequencing parameters in relation to true/false positivity's findings of rare variants from an isolated population from South-Eastern Moravia in the Czech Republic with high incidence of Parkinsonism. *R. Vodicka[1], R. Vrtel[1], K. Kolarikova[1], M. Prochazka[1], K. Mensikova[2], P. Kanovsky[2].* 1) Department of Medical Genetics , University Hospital, Olomouc, Czech Republic; 2) 2Department of Neurology Faculty of Medicine and Dentistry, Palacky University, University Hospital, Olomouc, Czech Republic.

Introduction: Massive Parallel Sequencing (MPS) in 16 genes known to be associated with Parkinsonism including coding DNA, intron/exon boundaries and UTRs loci was used to find rare variants in 30 patients and 12 healthy controls from an isolated population of South-Eastern Moravia in the Czech Republic. Epidemiological data show significantly increased prevalence of Parkinsonism (2.9%). Aim of the study: To evaluate true/false positivity ratio in relation to the basic MPS sequencing parameters (coverage, type of mutation – SNV/INDEL, percentage of rare variants in case of heterozygozity, +/- strand bias and length of homopolymers) in the 16 genes associated with Parkinsonism (ADH1C, ATP13A2, EIF4G1, FBXO7, GBA + GBAP1, GIGYF2, HTRA2, LRRK2, MAPT, PARK2, PARK7, PINK1, PLA2G6, SNCA, UCHL1 and VPS35). Study was supported by MH CZ – DRO (FNOI, 00098892). Method: Final filtered out rare variants were obtained from the Ion Torrent platform with workflow as following: Torrent Suite - Base calling and BAM mapping; IonReporter - Variant calling and rare variant filtering. True from false positivity findings were distinguished by Sanger confirmation sequencing. Results: In total, there were found 36 rare variants (MAF< 1 %) from which 50 % of them were confirmed as true positive. In case of SNV, the probability of false positivity is 11.7 % while in INDEL the false positivity proportion is 84 %. Very interesting indicator of true positivity could be high correlation in strand biases of reference and rare variants in heterozygous findings. Variants were than filtered using following parameters: AQ>20; Read coverage >10; MAF<0,01; SIFT: 0 - 0,05 and/or PolyPhen-2: 0,2 -1. The most important and Sanger confirmed variants: three missense mutations were found in LRRK2 gene: rs33995883 in 6/0 patients/control (p/c), rs33958906 in 1/1p/c, rs781737269 in 3/0p/c, one missense mutation in MAPT gene rs63750072 in 6/1p/c and one mutation in HTRA2 gene rs72470545 in 3/1p/c. Conclusion: Our findings could contribute: 1) to the improvement of NGS data analysis with respect to false positivity from true positivity variant calling recognition 2) to further understanding of molecular pathogenesis of Parkinsonism 3) to the creation of clinically applicable diagnostic procedure.

1221F

A statistical inference framework to improve functional prediction of missense variants in neurodevelopmental disorders. *J. Du[1], E. Pérez-Palma[1], P. May[2], P. Nuernberg[1,3,4], M. Daly[5,6,7], A. Palotie[5,6,7,8], D. Lal[1,5,6,7].* 1) Cologne Center for Genomics, Cologne, NRW, Germany; 2) Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg; 3) Cologne Excellence Cluster on Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, Cologne, NRW, Germany; 4) Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, NRW, Germany; 5) Massachusetts General Hospital, Boston, MA, USA; 6) Harvard Medical School, Boston, MA, USA; 7) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 8) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.

   Many *in silico* methods have been recently developed to predict the effect of missense variants. However, these prediction algorithms do not accurately distinguish between functional and neutral missense variants. In addition, each gene has a different evolutionary and population constraint and the distribution of variant scores is gene-specific. Current pathogenicity prediction methods score each variant without informing whether the observed variant score is similar to or significantly different from the expected score in individuals of the general population. The inability to evaluate whether the observed variant score is exceptional minimizes the interpretation power of the available methods. Therefore, in this study, we present a statistical inference framework to guide functional prediction of missense variants using various variant prediction tools. Methods and Results: The genome Aggregation Database (gnomAD, version 2.0), encompassing 4,483,391 missense variants in 138,632 individuals, was used as a reference set for genetic variants in the general population. All missense variants were annotated to the corresponding canonical transcripts. Established methods for variant interpretation such as SIFT, GERP, CADD and EIGEN were employed together with novel prediction tools developed in our group. Based on these measures we developed a one-sided testing framework that scores missense variants of interest and calculates the likelihood for observing variants with such scores in the gnomAD reference set. Currently, the framework is applied to *de novo* missense variants identified in our cohort (N = 11k) of patients with neurodevelopmental disorders and disease databases (including ClinVar and HGMD). Our approach provides a novel statistical framework for leveraging large population data to easily identify exceptional disease variant scores across a large range of variant prediction tools. The method will allow ranking thousands of variants from a whole exome data set of an individual patient to bring only a few highly significantly different variants to our attention. The framework is implemented in Python and will be freely available as an online tool.

1222W

Assessment of lesion-associated gene and variant pathogenicity in focal human epilepsies. *L.M. Neupert[1], P. May[2], K. Kobow[3], M. Nothnagel[1], P. Nürnberg[1], T. Freiman[4], P. Harter[4], K.M. Klein[4], Y. Weber[5], M. Daly[6,7,8], A. Palotie[6,7,8,9], I. Blümcke[3], D. Lal[1,6,7,8].* 1) Cologne Center for Genomics, Cologne, NRW, Germany; 2) Luxembourg Centre for Systems Biomedicine, University Luxembourg, Luxembourg; 3) University Hospital Erlangen, Germany; 4) Epilepsy Center Frankfurt Rhine-Main, Germany; 5) Hertie Institute for Clinical Brain Research, Germany; 6) Massachusetts General Hospital, Boston, MA, USA; 7) Harvard Medical School, Boston, MA, USA; 8) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 9) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.

   Germline and brain-specific somatic variants have been reported as an underlying cause in patients with epilepsy-associated neuropathologies, including focal cortical dysplasias (FCDs) and long-term epilepsy associated tumors (LEAT). However, evaluation of identified neuropathology-associated variants in genetic screens is complex since not all observed variants contribute to the etiology of neuropathologies not even in genuinely disease-associated genes. Here, we critically reevaluated the pathogenicity of 12 previously published disease-related genes and of 88 neuropathology-associated missense variants listed in the PubMed and ClinVar databases. We (1) assessed the evolutionary gene constraint using the *pLI* and the missense *z* score, (2) used the latest American College of Medical Genetics and Genomics (ACMG) guidelines, and (3) performed bioinformatic variant pathogenicity prediction analyses using PolyPhen-2, CADD, and GERP. Constraint analysis classified only seven out of 12 genes to be likely disease-associated. Furthermore, 35 (40%) of 88 neuropathology-associated missense variants were classified as being of unknown significance (VUS) and 53 (60%) as being likely pathogenic (LPII). Pathogenicity prediction yielded a clear discrimination between LPII variants and a milder discrimination for VUS compared with rare variant scores from individuals present in the Genome Aggregation Database (gnomAD). In summary, our results demonstrate that interpretation of variants associated with neuropathologies is complex while the application of current ACMG guidelines including bioinformatic pathogenicity prediction can help improving variant evaluation. Furthermore, we will augment this set of literature-identified variants at the conference by results from our variant screen using self-generated deep sequencing data in >150 candidate genes in >50 patients not yet analyzed.

1223T

Accurate identification of de novo structural variants in a trio using a reference agnostic, rapidly queryable format to reduce the proportion of unsolved cases. *S.N. Shekar[1], L. Herta[1], W.J. Salerno[2], A.C. English[1,2], C.A. Brownstein[4,5], J. Gonzalez-Heydrich[4], A. Mangubat[1], J. Bruestle[1], E. Boerwinkle[1,3], R.A. Gibbs[1].* 1) Fabric Genomics, Oakland, CA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 3) Human Genetics Center, University of Texas Health Science Center, Houston, Texas, USA; 4) Boston Children's Hospital, Boston, MA; 5) The Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA.

Detecting structural genetic changes (SVs) that are de novo could reduce the proportion of unsolved cases. Currently, SV callers can exhibit upwards of a 50% false discovery rate (English et al, 2015) with ambiguity in the breakpoint coordinates. This makes it difficult to compare calls between the proband and the parents. Here, we detect de novo variants using the Biograph Analysis Format (BAF), a method of indexing NGS data that is reference agnostic and rapidly searchable. To show consistency of calls using this Format, we confirm SV identified by the caller Pindel using overlap assembly in a Ashkenazi Jewish Trio from the Personal Genome Project. Of 1,195 calls that showed evidence in the reads of at least one individual, all except for 25 (2.1%) were consistent with mendelian inheritance. In all cases that the variant was called in multiple samples, the variant was reported exactly the same. For variants that did not follow mendelian inheritance, there was insufficient coverage in regions flanking the breakpoint in at least one individual. In a further case where the proband and both parents genomes were sequenced at 30x coverage using an Illumina HiSeq, we used Anchored Assembly on the proband sample, a caller that uses read overlap assembly to detect variants (English et al, 2015). We called 2,383 genetic changes that were either an insertion or deletion (> 50 base pairs) in the proband. To confirm whether there was evidence of each of these variants in the parents, we search for the count of reads that contain the 50mer that is unique to the breakpoint junction. Of the 2,383 variants, 98 showed, prima facie, evidence of being de novo. 18 of these variants were heterozygous in the proband. Although there was no evidence for the variant in either parent, at least one parent showed a drop in coverage. Of the remaining 80 variants that were homozygous in the proband, the variant was present in at least one parent with the other parent either having no coverage at that location (68 variants) or low reference (12 variants, < 9 reads). Overall, this is suggestive that these variants are more likely to be due to a lack of coverage than true de novo variants. This analysis was completed in 18 hours. Although, we didn't confirm a true de novo SV, calling with a low FDR and rapid search of read data can be used to detect and rule out candidate de novo variants quickly.

1224F

A single-cell transcriptome atlas of mouse hypothalamus: Identification and characterization of high-fat diet and GLP-1 responding arcuate nucleus cell types. *P.N. Timshel, M.A. Bentsen, J. Lee, T.H. Pers.* Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark.

Background: A key challenge for gaining biological insights from genetic associations is to identify etiological cell types and pathways. Genome-wide association studies (GWAS) have identified numerous genetic variants associated with body mass index (BMI). Functional studies using gene expression data of genes in BMI-associated genetic loci have emphasized the role of the brain in predisposition to obesity. However, these studies have been conducted at the brain area level as opposed to the cell-type level. Key appetite regulating brain nuclei reside within the hypothalamus. The arcuate nucleus is one of the key areas and importantly, it has recently been shown that an approved weight-loss drug, liraglutide (a GLP-1 analogue), acts in the ARC. Yet our understanding of the molecular mechanisms behind the effects of GLP-1 analogue treatment remains incomplete. **Data**: We used droplet-based single-cell RNA sequencing (10x Genomics protocol) to construct transcriptional atlases of mouse ARC cell types from four groups of C57BL/6J mice: 1) diet-induced obese (DIO) fed ad libitum, 2) DIO treated with 400 mg/kg liraglutide for three weeks, 3) pair-fed vehicle control and 4) chow fed control. In total, we profiled > 60,000 single-cell transcriptomes. **Methods**: We developed and applied a computational method to integrate single-cell transcriptomic data with GWAS summary statistics, to prioritize cell types and cell-type specific pathways. Our method uses matrix factorization models to decompose the biological components and cellular heterogeneity captured in our ARC cell atlases. **Results**: Liraglutide treated mice lost 17±2% weight from baseline, pair-fed lost 10±2%, whereas DIO mice fed ad libitum lost 2±1% and chow fed mice gained 1±1%. We compared cell types across each group of mice, and identified GLP-1 and high-fat diet responsive cell types. Using our computational method and BMI GWAS data, we prioritized BMI-relevant cell types and pathways. **Conclusion**: Single-cell transcriptomic profiling of the ARC enable identification of brain cell types responding obesity relevant physiological regimes. Finally, we show that integrating genetic data with single-cell transcriptomics is a powerful way of gaining cell-type and pathway etiological insights for human complex traits.

1225W

The GCAD workflow for processing 5000 whole genomes and 11,000 whole exomes from the Alzheimer's Disease Sequencing Project using Amazon cloud. *Y.-F. Chou[1,2], Y.Y. Leung[1,2], O. Valladares[1,2], A.B. Kuzma[1,2], L. Cantwell[1], L. Qu[1,2], H.-J. Lin[1,2], P. Gangadharan[1,2], Y. Zhao[1,2], J. Malamon[1], A.D. Sequencing Project (ADSP)[1], A. Naj[1], W.J. Salerno[3], G.D. Schellenberg[1,2], L.-S. Wang[1,2].* 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Late-Onset Alzheimer's Disease (AD) is a devastating disease that will affect 16 million Americans by 2050. Established in 2016, the Genome Center for Alzheimer's Disease (GCAD) aims to coordinate the integration and meta-analysis of all available AD relevant genetic data with the goal of identifying AD risk and eventual therapeutic targets. It involves the collaboration of diverse scientists within the Alzheimer's disease sequencing project (ADSP) and GCAD.   With the vision to minimize data heterogeneity and to efficiently process thousands of genomes, GCAD has built an Amazon cloud-based pipeline that is 1) fully designed and optimized for large scale production of WGS/WES data; 2) interactive with a tracking database that can display a variety of quality metrics instantly; and 3) developed in coordination with other NIH programs including TOPMed and the CCDG.   This pipeline takes an input FASTQ/BAM file and re-maps to the hg38 genome assembly using BWA. Supplementary alignments are retained for downstream SV optimization. It then performs duplicate marking, base recalibration, indel-realignment and quality score binning using GATK 3.7. Next, it performs SNV genotype calling using GATK. All individual gVCFs from WES and WGS are scaled and aggregated to a project level file by joint genotype call. The final output contains all possible sites for every genotype position, allowing for easy comparison across genomes from other sequencing projects.   Together with the pipeline, a live statistics page is built and supported by a database which stores >100 quality metrics on each sequenced sample. These include information on how the sequencing is performed, processing time, metrics on BAM quality and genome-wide statistics of the called SNPs and indels. On average, the workflow can process a complete genome in <30 hours. The pipeline is available through bitbucket (https://bitbucket.org/GCAD/SNV_pipeline/) and as an Amazon image (ami-e6df35f0).   To date, we have processed 5,000 WGS and 11,000 WES from 26 different cohorts of European, Caribbean Hispanic, and African American descent, which are deeply sequenced across 5 sequencing centers at mean depth 30X (WGS) and 75% of targeted regions at >20x coverage for WES. From the WGS data, 21M novel SNPs were discovered across these sequenced individuals. Among the discovered, 23% were singletons present in a single individual. 39.6M SNPs has reached 99.8% sensitivity tranche threshold, of which 42% of these are novel.

1226T

Neuron-squared (N²): An industrial-scale iPSC disease-modeling project for neuropsychiatric disorders. *R. Randhawa, A. Kaykas, B. Bilican, K. Worringer, C. Ye, B. Altshuler, T. Kulkarni, K. Chan, S. Kommineni, D. Fletcher, B. Henry, K. Theriault, D. Ho.* Neuroscience, Novartis Institutes for BioMedical Research, Cambridge, MA.

The majority of neuroscience drug discovery in the last 3 decades has yielded very few new drugs. One of the major challenges we face is rodent models of neuropsychiatric and neurodegenerative disorders have shown limited success in identifying compounds that are effective in clinical trials. Traditionally, industry has relied on transformed or genetically-engineered heterologous systems to develop cellular assays for drug discovery. Stem cells offer distinct opportunities to establish genetically and functionally faithful neurological disease models, potentially bridging a significant gap in new target identification, validation and drug discovery. However, a major challenge in the field is establishing standardized differentiation protocols that yield human neurons with disease-relevant phenotypes at a truly industrial scale, in both quantity and screening format.   To this end, we established a fully automated human pluripotent stem cell (PSC) maintenance and excitatory cortical neuronal differentiation platform that enables parallel phenotyping of many different lines at once. We have generated over 100 iPSC lines from normal and psychiatric disease patients using non-integrating reprogramming technology. We are currently running these lines through automated differentiation platform and collecting genomic & phenotypic data. We have subjected all patients to whole genome-sequencing and are collecting epigenomic and transcriptomic data at the pluripotency, neuronal precursor and neuronal phases of differentiation. We are collecting high content imaging data on the neurons to capture neuronal morphology and synaptic formation and FDSS/calcium imaging as a proxy for network activity. This large data set will allow us to look at reproducibility and technical noise at unprecedented levels and allow us to correlate the clinical phenotypes of patients, genotype, transcriptional networks and cellular phenotype. Further analysis of this large dataset may yield new targets for psychiatric disease and/or cellular phenotypes that could be used for high-throughput screening. This human disease-modeling platform is being integrated into Novartis' lead discovery pipeline to identify new targets, molecules, and to shed light on the cellular aspects of human neuronal biology.

1227F

Towards translating genetic findings of polygenic diseases to personalized drug development: Proof-of-concept study for drug combinations to target multiple genes. *I.S. Vlachos[1,2,3], B. Kaskow[1,2,3], G. Wong[1,2,3], P. de Jager[1,2,3], W. Elyaman[1,2,3], N.A. Patsopoulos[1,2,3].* 1) Ann Romney Center for Neurological Diseases, Brigham & Women's Hospital, Boston, MA, USA; 2) Harvard Medical School, Boston, MA, USA; 3) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

   Background One of the main promises of genetics is the discovery of new targets for drug discovery. However, two main limitations can arise: i) the vast majority of disease-associated loci fall in non-coding parts of the genome, ii) there are hundreds of genes that can be prioritized in common diseases. Here, we perform a proof-of-concept study by integrating personalized genetic information, expression data, and drug-screen signature. We create and experimentally test a framework to identify compound combinations that can target multiple disease-related genes, given an individual's genome. **Methods** We perused a cohort of 215 healthy subjects to identify pairs of individuals that are extreme homozygotes for multiple MS-associated variants. We used expression data of CD4+ naïve T cells from the same individuals, to prioritize pairs bearing MS variants that were *cis*-eQTLs for one or more genes. Having pairs of individuals, that are expected to have extreme gene expression for the given set of genes, we leveraged the Connectivity Map (CMAP) data to identify combinations of compounds that could reverse the MS-specific gene expression signature. We isolated CD4+ naïve T cells from the selected pairs and treated them with the prioritized compounds. We used SCRB-Seq, a low-input library preparation to measure genome-wide gene expression.   **Results** We prioritized 2 pairs of individuals (n=4) for a total of 3 MS-associated variants that were cis-eQTLs for 4 genes. Perusing the CMAP data, we prioritized 13 chemical compounds, combinations of which were predicted to reverse the MS-specific gene signatures. We sequenced CD4+ T naïve cells, from the above individuals, treated for the 13 compounds and 6 pair-wise combinations. We also included untreated and DMSO-only CD4+ naïve T cells for comparisons. We report the success to replicate the CMAP data and the ability to predict the effect of the drug combinations given their single-drug effect. Finally, we report the extent to which the by disease-specific genetic variation controls the response of gene expression to the compounds treatment.   **Discussion** We present a comprehensive methodology that aims to develop a genetic-guided and personalized combinatorial drug regimen to simultaneously target several disease related genes. We discuss feasibility, issues, and results and we elaborate on next steps towards created personalized drug prioritization approaches.

1228W

SV[2]: Accurate structural variation genotyping and de novo mutation from whole genomes. *D. Antaki[1,2,3,4], W.M. Brandler[1,2,3], J. Sebat[1,2,3].* 1) Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, CA; 2) Department of Psychiatry, University of California San Diego, La Jolla, CA; 3) Department of Cellular and Molecular Medicine and Pediatrics, University of California San Diego, La Jolla, CA; 4) Biomedical Sciences Graduate Program, University of California San Diego, La Jolla, CA.

   Structural Variation (SV) is a major contributor to human genetic variation with 13% of the human genome defined as structurally variable, and is also implicated in a variety of human diseases. In particular, *de novo* structural mutations (those in offspring and not in parents) contribute significant risk for idiopathic autism and intellectual disability. Therefore, family-based studies of rare and *de novo* SVs have become a widely implemented approach for gene discovery in neurodevelopmental disorders. However, SV detection from short-read whole genome sequencing is error prone, presenting significant challenges for population or family-based studies of disease. Here we present SV[2] (support-vector structural-variation genotyper), a machine-learning algorithm for genotyping deletions and duplications from paired-end sequencing data. SV[2] rapidly integrates variant calls from multiple structural variant discovery algorithms into a unified call set with high genotyping accuracy and capability to detect *de novo* mutations. As a proof of concept, SV[2] was implemented in a genetic study of autism (3,169 individuals, 829 families), where 123 *de novo* deletions and duplications were resolved, implicating novel autism gene candidates. Likewise, SV[2] was pivotal in our novel discovery of over-transmission of rare paternally-inherited SVs that disrupt promoters or UTRs of haploinsufficient genes to autistic children and not to their typically-developing siblings (P= 0.0013). In conclusion, the utility of SV[2]'s accurate genotyping benefits any structural variation association study, making SV[2] a palpable solution for genetic association and *de novo* mutation discovery. SV[2] is an open-source software available for download at http://github.com/dantaki/SV2.

1229T

Novel pathway transcriptomics method greatly increases detection of molecular pathways associated with the trait. *C. Chatzinakos[1], A. Docherty[2], V. Vladimirov[1], T. Bigdeli[3], T. Webb[1], K. Kendler[1], S. Bacanu[1].* 1) Virginia Commonwealth University (VCU), Richmond, VA; 2) University of Utah School of Medicine, Utah; 3) SUNY Downstate Medical Center, University Hospital of Brooklyn, New York.

Genetic signal detection in genome-wide association studies (GWAS) is improved by pooling information from multiple single nucleotide polymorphism (SNP), e.g. across genes or pathways. Because many genes influence trait via gene expression, it is of interest to combine information from Quantitative Trait Loci (eQTLs) in a gene or genes in the same pathway. Such type of methods, denoted as transcriptomics, already exists for gene analysis. They use eQTLs to infer the association between trait and predicted expression of gene under study. Among the many transcriptomics based methods/software, our group proposed **G**ene-level **J**oint **A**nalysis of functional SNPs in **C**osmopolitan **C**ohorts (JEPEGMIX). However, due to the $O(n^2)$ computational burden for computing linkage disequilibrium (LD) between numerous (n) SNPs across large regions (even chromosome arms) transcriptomics methods are not yet applicable to arbitrarily large pathways/gene sets. To overcome this obstacle, we propose JEPEGMIX2, a novel $O(n)$ transcriptomics method, which 1) computes LD for gene statistics and 2) uses LD and GWAS summary statistics to rapidly test for the association between trait and expression of genes even in the largest pathways. It first computes chromosome arm pathway χ^2 tests as Mahalanobis statistics of Z-scores for genes in the pathway and chromosome arm. Subsequently, due to variants on different chromosome arms being quasi-independent, JEPEGMIX2 pathway χ^2 statistic and degrees of freedom are simply the sum of their chromosome arm counterparts. Simulations indicate that JEPEGMIX2 controls the type I error at or below nominal rates both for i) the entire set of genes/pathways and each gene/pathway separately. Being written in C++, JEPEGMIX2 takes less than 3.5 hours to compute all gene- and pathway-level statistics. To underline its potential for greatly increasing the power to uncover genetic signals over existing (non-transcriptomics) pathway methods, we applied JEPEGMIX2 to summary statistics from several large meta-analyses. For instance, while current methods found only nominally significant pathways for Psychiatric Genetics Consortium Schizophrenia phase 2, JEPEGMIX2 uncovered around one hundred pathways that were statistically significant after adjusting for multiple testing. Surprisingly, most such significant pathways seem not to be mostly involved in the activity of the central nervous system, but rather in the well-functioning/maintenance of cell/entire organism.

1230F

Gravity: A tool to analyze genetic variants in individuals using gene networks. *F. Cliquet[1,2,3,4], A. Mathieu[1,2,3,4], T. Kergrohen[1,2,3,4], T. Rolland[1,2,3,4], R. Delorme[1,2,3,4,5], B. Schwikowski[4], T. Bourgeron[1,2,3,4].* 1) Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; 2) CNRS UMR 3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) University Paris Diderot, Sorbonne Paris Cité, Paris, France; 4) C3BI, Institut Pasteur, Paris, France; 5) Department of Child and Adolescent Psychiatry, Robert Debré Hospital, APHP, Paris, France.

For many complex traits such as psychiatric disorders or cancer, patients display a highly heterogeneous genetic architecture. In some cases of autism spectrum disorders (ASD), a single *de novo* deleterious mutation is causative, but in the vast majority, a combination of multiple common and rare variants increases disease risk (Bourgeron, T. *Nat Rev Neurosci* 2015). Even within families with more than one affected child, causative mutations can be different from one sibling to another (Yuen, R.K. *Nat Med* 2015). The causative variants are usually detected by progressive filtering of sequencing data, but practical means for the rapid inspection of multiple variants potentially affecting different genes (multiple hits) in functional networks are missing. Here, we present GRAVITY http://gravity.pasteur.fr/, a tool that allows efficient visualization and analysis of all exonic variants, including single nucleotide polymorphisms and copy-number variants, by mapping them onto a protein-protein interaction or co-expression network (Rolland, T. *Cell* 2014). Designed as an app for Cytoscape and highly customizable, it facilitates data analyses; either specifically oriented towards pathways known to be involved in the phenotype (e.g., the glutamatergic synapse pathway in ASD, or sets of candidate genes previously associated with cancer), or by exploring the whole exome. GRAVITY allows regular trio analysis (a patient with his parents) to identify *de novo* mutations, recessive mutations, or Mendelian errors, as well as the visualization of the mutation burden on a gene or a group of genes for an individual. For cancer research, it allows the analysis of the somatic mutations together with the constitutive genetic background of the patient. Finally, it provides customizable options for the identification in a group of patients of all individuals carrying mutations in a given gene or cellular pathway, thereby assisting clinicians to build and validate hypotheses. GRAVITY was developed using Gemini, a flexible, widely used, database to store genetic data that can include pedigree information and potentially more than one phenotype. GRAVITY enables researchers and clinicians to visualize and analyze multiple hits directly from sequencing data, towards a better understanding of how complex mutation profiles translate into diversity and severity of ASD symptoms in patients. GRAVITY is easily adaptable to other neurodevelopmental disorders and other complex human diseases.

1231W

GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. *M. Maciukiewicz[1], V.S Marshe[1,2], A.-C. Hauschild[3,4,5], J.A. Foster[5,6], S. Rotzinger[5,7], J.L. Kennedy[1,2,7], S.H. Kennedy[2,5,7,8], D.J Müller[1,2,7], J. Geraci[9].* 1) Pharmacogenetic Research Clinic, Center for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, Faculty of Medicine, University of Toronto, Ontario, Canada; 3) IBM Life Sciences Discovery Centre, Princess Margaret Cancer Centre, 101 College Street, Toronto Ontario M5G 1L7, Canada; 4) Department of Computer Science, University of Toronto, 6 King's College Road, Toronto Ontario M5S 3H5, Canada; 5) University Health Network, Toronto, Ontario Canada; 6) Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, Ontario, Canada; 7) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 8) Department of Psychiatry, St. Michael's Hospital, Toronto, Ontario, Canada; 9) Department of Molecular Medicine, Queen's University, Kingston, Ontario, Canada.

Background: Major depressive disorder (MDD) is one of the most prevalent psychiatric disorders and is commonly treated with antidepressant drugs. However, large variability is observed in terms of response to antidepressants. Machine learning (ML) models may be useful to predict treatment outcomes. **Patients and Methods**: A sample of 186 MDD patients received treatment with duloxetine for up to 8 weeks were categorized as "responders" based on a MADRS change >50% from baseline; or "remitters" based on a MADRS score ≤10 at end point. The initial dataset (N=186) was randomly divided into training and test sets (70% and 30% respectively), resulting in ten different training and test set pairs. We performed genome-wide logistic regression to identify potentially significant variants related to duloxetine response/remission and extracted the most promising predictors using LASSO regression. Subsequently, classification-regression trees (CRT) and support vector machines (SVM) were applied to construct models, using ten-fold, repeated cross-validation. **Results:** For the response, LASSO regression suggested twenty SNPs for prediction models. Calculated models achieved accuracy of 63.43% for CRT and 78.93% for SVM. For the remission, LASSO filtering suggested five SNPs. Best models were characterized by an accuracy of 65.45% with a sensitivity of 69.31% and specificity of 61.43%. **Discussion and Conclusions**: In this study, the potential of using GWAS data to predict duloxetine outcomes was examined using ML models. These models managed to capture a fraction of responders and remitters (i.e. moderate sensitivity), but failed to filter out non-responders and non-remitters (i.e. low specificity). Inclusion of additional non-genetic variables to create integrated models may help improve prediction results.

1232T

The Registry of Candidate Regulatory Elements: Integrating human and mouse epigenomic data to fine map and annotate genetic variants. *J. Moore, M. Purcaro, H. Pratt, Z. Weng.* Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA.

Genome wide association studies (GWAS) and whole genome sequencing projects have identified thousands of genetic variants associated with human disease. However, understanding how these variants contribute to disease remains a challenge as the majority are in non-coding regions of the genome. To address this problem, we created the Registry of candidate Regulatory Elements (cREs), a collection of putative regulatory regions across human and mouse. We curated this registry by integrating DNase hypersensitive sites with H3K27ac, H3K4me3 and CTCF ChIP-seq signals. We incorporated data from 620 human and 138 mouse cell/tissue types, utilizing experiments from the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics projects. In total, the registry contains 1.3M human cREs and 430k mouse cREs each annotated with promoter-like, enhancer-like, or CTCF-only signatures. We validated our approach using complementary data such as transcription factor ChIP-seq peaks and experimentally validated enhancers from the VISTA database. These cREs are publically available on our interactive database SCREEN (Search Candidate Regulatory Regions by ENCODE) at www.screen.umassmed.edu. To aid in variant annotation and fine mapping, we integrated SCREEN with variants from the NHGRI-EBI GWAS catalog. SCREEN allows users to simultaneously investigate human and mouse regulatory regions, enabling them to (1) prioritize variants in orthologous cREs (i.e. cREs conserved across human and mouse) and (2) survey data from a greater number of cell types. To demonstrate SCREEN's utility, we analyzed variants associated with neuropsychiatric disorders. By comparing cRE activity with corresponding RNA-seq data across human and mouse, we identified over 50 potentially causal cREs. The majority of these cREs have promoter-like or enhancer-like signatures, several of which are primarily active during brain development. For example, *AGAP1* contains an intronic cRE that overlaps a schizophrenia associated variant. Due to the limited scope of fetal brain data, we are unable to annotate this cRE using only human datasets. By incorporating mouse data, we observe that the orthologous cRE has enhancer-like signatures only in embryonic brain tissue suggesting that this cRE regulates AGAP1 only during brain development. In the future, we can apply this same methodology to fine map and annotate variants associated with other diseases, particularly those associated with human development.

1233F

Bayesian multivariate analysis of RNA sequencing data to identify brain-specific protein-protein interactions. *S. Muller[1,2], T. Li[1,2], A. Kim[1,2], C. Hartl[3], E. Malolepsza[1,2], K. Lage[1,2].* 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute, Cambridge, MA; 3) University California Los Angeles,- CA.

 Human protein-protein interaction (PPI) studies have empowered the understanding of biological mechanisms involved in diseases. Although the human interactome is providing useful information on the mediation of biological processes, it is very sparse and likely to be specific to a tissue or cell type. Undercovering the biological processes leading to a brain disease through PPI networks rely therefore on the amount of information available on human brain proteins, which is poor due to the difficulty of uniformization of experiments across tissues. To try to tackle this gap in literature, we hypothesize that the physical PPI in the brain can be learnt from tissue-specific RNA sequencing data. In this study, we propose a new algorithm based on multivariate Bayes theory that aims at inferring a significant differential expression network that is specific to a tissue (e.g. the brain). The algorithm prioritizes genes that have a significant higher or lower co-expression in a given tissue. It assumes that a physical interaction between two proteins in a specific tissue is the consequence of the whole genes expression signature in that given tissue. This type of multivariate inference, particularly suited for small sample size studies, insures that the model learns the putative regulatory effects between the expression of the genes and how, as a whole, gene expression in a given tissue will modulate the protein interactions. From tissue specific RNA sequencing data, we identified a set of 55 brain-specific genes with a significant differential expression network. The expression patterns observed in the network are reproducible in other datasets (in vitro, neuronal-induced cells, in-house RNA sequencing) and in vivo. The obtained network is highly enriched for physical PPI (p=0.003). The characterization of the network reveals significant lower and higher expression clusters of interacting genes both at the RNA and protein levels. Globally, we show that genes in the predicted network are more translated in the brain than in other tissues. Finally, we demonstrated that brain related diseases are significantly more associated to genes in the predicted network than random. The predicted gene network accumulates evidences that it is underlying biological mechanisms specific to the brain. Further in vitro biological validation is planned.

1234W

Tissue-specific gene expression inference. *K. Vervier, J.J. Michaelson.* Psychiatry, University of Iowa, Iowa City, IA.

 Gene expression varies by tissue, and projects like Gene-Tissue Expression (GTEx), have enabled the systematic quantification of such variation. Wheeler *et al.* (2016) estimated tissue-specific gene expression with the use of "local heritability" only, noting that current sample sizes in GTEx do not allow for the integration of distal heritability. In this study, we propose to filter predictor variants based on a tissue-specific functional score (Vervier et al., 2017) and then train predictive models on a large dataset from whole blood samples (Battle et al., 2014). Reducing the number of functional input variables also lowers the features-to-observations ratio, thus enabling the use of both local and distal information. Transfer learning is used to extrapolate from whole blood to other tissues, such as human brain and heart. We demonstrate that tissue-specific predictive models are more consistent at identifying disease-related genes than tissue-agnostic methods such as PrediXcan (Gamazon et al., 2015) and SLINGER (Vervier et al., 2016). On the WTCCC bipolar disorder case-control cohort, differentially expressed genes predicted by a brain-specific model show a significant enrichment in calcium channel complexes, which are known players in bipolar disorder. Notably, this finding was not observed using generic, tissue-agnostic methods.

1235T

Augmenting multi-ethnic image signals to enhance schizophrenia prediction. *H. Qin[1], Y. Li[1], Y.-P. Wang[1,2].* 1) Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, New Orleans, LA 70112, USA; 2) Department of Biomedical Engineering, Tulane University School of Science and Engineering, New Orleans, LA 70118, USA.

Identifying predictive brain voxels is a critical, but currently difficult, task in multi-ethnic biomedical image studies of schizophrenia. Susceptible voxels are highly correlated with adjacent voxels and can interact with latent environmental factors. Classical voxel wise method (e.g., LIMMA), exploits mean heterogeneity between case and control groups only, assuming dispersion homoscedasticity. The dogmatic assumption of homoscedasticity severely violates the strong intra-voxel correlation and latent interaction effects. As such, we proposed a double moderate Welch test (DMWT) to integrate the mean and second order heterogeneities of a voxel after calibrating the mean and dispersion effects of confounders in a generalized linear mixed model. By extensive simulations, the proposed method properly controlled type I error rates and appeared strikingly more powerful than mean heterogeneity test. The novel method displayed particular utility in analyzing biomedical fMRI data of 104 multi-ethnic controls and 80 multi-ethnic schizophrenia cases. The DMWT and the LIMMA identified 4 and 3 significant voxels, respectively. The four voxels identified by the DMWT displayed 33% prediction accuracy over those three identified by the LIMMA; and the improvement proved significant (p = 0.016). In summary, we have paved an effective avenue to integrate voxel wise high-order informative heterogeneities while adjusting for heteroscedastic background confounders. Our results have demonstrated the tremendous potential gain of explicitly handling higher-order heterogeneities in multiethnic image data. This work was supported by Carol Lavin Bernick Faculty Grant (632119), Tulane's Committee on Research fellowship (600890), and Innovative Programs Hub (I2PH) Grants Award of Tulane (632037).

1236F

A statistical framework of mapping risk genes from *de novo* mutations in whole-genome sequencing studies. *Y. Liu[1], E. Cicek[2,3], Y. Liang[3], J. Li[4,6,10], R. Muhle[7,8,9], N. Knoblauch[5], M. Krenzer[9], Y. Mei[4], Y. Jiang[6,10], E. Geller[7,9], Z. Li[6], I. Ionita-laza[11], J. Wu[4,6], K. Xia[10], J. Noonan[7,9], Z. Sun[4,6], X. He[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Computer Engineering Department, Bilkent University, Ankara, Turkey; 3) Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA; 4) Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing, China; 5) Committee on Genetics, Genomics and Systems Biology, The University of Chicago, Chicago, IL; 6) Institute of Genomic Medicine, Wenzhou Medical University, Wenzhou, China; 7) Department of Genetics, Yale School of Medicine, New Haven, Connecticut; 8) Child Study Center, Yale Medicine, New Haven, Connecticut; 9) Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, Connecticut; 10) The State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China; 11) Department of Biostatistics, Columbia University, New York, New York.

Analyzing *de novo* mutations (DNMs) from sequencing data of nuclear families has identified risk genes of a variety of complex diseases, especially neurodevelopmental and psychiatric disorders. Such efforts have been mostly focused on mutations targeting protein-coding sequences. Non-coding regions harbor important regulatory information and play important roles in human diseases, as evidenced by the finding from genome-wide association studies (GWAS) that most disease loci are located in non-coding sequences. Extending the DNM-based gene mapping approach to non-coding regions is thus an important direction to advancing our understanding of complex psychiatric disorders. Unlike coding sequences, however, we do not have a "genetic code" that allows us to easily predict the functional consequence of noncoding mutations, and as a result, how to interpret DNM data from whole genomes remains a major challenge. In this study, we developed a statistical framework for analyzing DNMs from whole-genome sequencing (WGS) data of parent-child trios. Built upon on our earlier TADA model for DNM analysis, the new method allows us to combine coding and non-coding mutations at the gene level to detect risk genes. It incorporates various genomic annotations in the model, such as conservation scores and epigenomic marks. The model learns from data which annotations are informative of pathogenic mutations, and utilizes such knowledge to better identify causative non-coding mutations. An additional benefit of the model is that it enables meta-analysis that combines multiple DNM studies, while adjusting for study-specific effects on mutation rates (e.g. difference of sequencing depth). Application of this new method, which we name TADA-Annotations, to WGS data of ~300 autism family trios (combining five studies) increases the power of disease gene mapping and predicts a number of new genes with possible roles in ASD.

1237W

Comparison of different approaches to detect CNV from SNP genotyping array and whole-exome sequencing. *B. Chaumette[1], Q. He[2], G. Huguet[3], D. Spiegelman[1], A. Dionne-Laporte[1], P. Dion[1], G.A. Rouleau[1], L. Xiong[2].* 1) Neurology and Neurosurgery, McGill University, Montréal, Quebec, Canada; 2) Centre de recherche, Institut universitaire en santé mentale de Montréal, Université de Montréal, Montréal, Quebec, Canada; 3) CHU Sainte-Justine, Université de Montréal, Montréal, Quebec, Canada.

The detection of Copy Number Variants (CNV) from large datasets remains challenging due to variable sensitivity and specificity offered by different programs and approaches. While comparative studies have been published, those comparing CNV calls from whole-genome genotyping data (SNP chip) and whole-exome sequencing data (WES) are scarce. We present a comparison between CNV calls from SNP chip and from WES in a large dataset (n=243 individuals). We illustrate this comparison with the specific results from a multiplex family with autism spectrum disorder (ASD). ASD is a neurodevelopmental condition defined by altered social interaction and communication, restricted interests and repetitive behaviors. The heritability is estimated to 80% and some pathogenic CNV have been identified. First, we used three programs - PennCNV, QuantiSNP and cnvPartition - to detect CNV from SNP chip. Second, we used the WES data from the overlapped samples to detect CNV using two programs - Conifer and XHMM. Concordance between programs and datasets were determined using CNVision program. The multiplex family design was useful to detect true positives called by each program, based on the familial segregation. In our genotyping dataset, 16.5% of the CNV were found by the three algorithms with a comparable concordance rate for duplication and deletion. Adjusting the quality control parameters (quality score, minimal number of SNP) increased this concordance to 40%. In the WES dataset, Conifer called 72.8% of the CNV that were also called by XHMM but the inverse gives 7.8% concordance. Only 2.3% of the CNV were concordant between WES and SNP genotyping data. As an example, in our SNP chip data, we found a 11q22.1 deletion encompassing the *CNTN5* gene and segregating in all three affected male siblings in a multiplex family. This deletion is inherited from the unaffected father and shared by one sibling with learning disabilities. This deletion in *CNTN5* gene has been previously identify as a strong candidate in autism and could be considered as a positive control. Interestingly, the deletion was impossible to identify in the WES data because the region was out of the capture-kit targets. This comparison, as well as this example, sheds light on the importance of a complementary approach from different datasets to efficiently detect CNV. Our design helped us to determine the sensitivity and the specificity of the different methodologies and to optimize the parameters of each program.

1238T

Exomerate: A machine-learning approach to identify high-confidence CNVs from exome sequencing data. *V. Pounraja, G. Jayakar, M. Jensen, N. Kelkar, S. Girirajan.* Department of Biochemistry and Molecular Biology, The Pennsylvania State University, PA.

Most studies that detect copy-number variants (CNVs) from exome-sequencing data employ algorithms that model the sequence read distribution across the genome. Each of these algorithms employ different criteria to filter noise from signal, with almost no consensus between the CNV calls. Their performances are often estimated by either validating only a subset of predictions or comparing them against calls made by an alternate algorithm. Therefore, an objective method that borrows strengths from multiple callers is necessary to address these issues. We propose a machine-learning framework, *Exomerate*, that learns to segregate true CNVs from false positives based on the ability of individual callers. We first tested the feasibility of the proposed framework with two sets of independent samples, 134 samples from the 1000 Genome Project (1KG) and 503 samples from the Simons Variation in Individuals Project (SVIP), for which biological validations were available. We analyzed both sets of samples using five different exome CNV callers, including Conifer, CODEX, Canoes, Exomedepth and XHMM, and obtained 69,254 (1KG) and 143,619 (SVIP) CNV predictions. Each prediction was then labeled either 'True' or 'False' based on the presence of overlap with at least one biologically validated CNV. Since the likelihood of CNVs vary across the genome based on the regional context, we chose to leverage GC content, mappability, exome probe counts and aggregate RVIS (Residual Variation Intolerance Score) as features for our algorithm. We supplemented CNV predictions made by individual exome CNV callers with these variables, and trained random forest classifiers to assign prediction scores to each call based on its likelihood of being a true CNV. The models built independently on both datasets attained substantially high precision rates (65-70%) and recall rates (35-40%). Our framework assigned high confidence even to those true calls lacking consensus across multiple exome CNV callers that would have otherwise been discarded. Our results indicate that a dynamic framework with a small set of biologically validated CNVs could help differentiate true CNVs from false positives from exome sequencing data. Our framework addresses several limitations of existing exome CNV callers and potentially helps making cautious and yet meaningful interpretations of the identified CNVs.

1239F

Improving pathogenicity prediction of structural variation in neurodevelopmental disorders: A machine learning approach. *P.T. Tandon, O.S. Shanta, D.A. Antaki, W.B. Brandler, M.K. Kleiber, O.H. Hong, J.S. Sebat.* University of California, San Diego, San Diego, CA.

   Structural Variants (SVs) consist of changes in the copy number or arrangement of DNA sequences (> 50 bp) in the genome and include deletions, duplications, and inversions. Gene-disrupting SVs contribute to elevated risk for a variety of psychiatric disorders including Autism Spectrum Disorders (ASD) and schizophrenia (SCZ) (Malhotra and Sebat 2012; Walsh et al. 2008). Despite continued success in identifying SVs that confer risk, current analytical approaches have been limited to investigations of the genome-wide burden of SVs or the enrichment of large gene-disrupting SVs at individual loci. Limited progress has been made in identifying smaller causal variants within individual genes or in non-coding regions of the genome that may be pathogenic. This study seeks to develop more accurate pathogenicity prediction and functional interpretation tools for SVs using a machine learning approach. The study develops a machine learning pipeline to fuse multiple biological functional annotations into a single pathogenicity score for a structural variant to assess risk for neurodevelopmental disorders. The trained classifier can distinguish deleterious SVs from neutral variants on a variety of data types including next-gen, whole genome sequencing, and clinical microarray data. On ClinVar data, the trained predictor can leverage multiple functional constraint predictors to predict the pathogenicity of an SV accurately for many sizes of SV, even when its ExAC probability of loss of function (pLI) score is on the margin (i.e. pLI=0.5).   The developed SV pathogenicity scoring system is used to investigate the unexplained genetic basis of psychiatric disease in ASD and SCZ. The predictor is more accurate than previous methods for medium to large (10^5-10^6 bp) deletions in the Psychiatric Genomics Consortium (PGC) data, after controlling for known covariates via logistic regression and filtering known loci (Wilcox p < 0.01). Emphasis in our study is placed on exploring the contribution of SVs that overlap non-coding regulatory elements to disease risk. The predictor utilizes fetal brain enhancer and promoter overlaps to improve predictions for non-exonic SVs. We have also seen supporting evidence for a "female protective effect" in SCZ by computing Bayesian person-level risk scores. Female cases display a stronger risk score enrichment over female controls (p=3.605*10^-6, Odds Ratio=1.2381) compared to the males (p=2.8959*10^-4, Odds Ratio=1.0256).

1240W

Biological pathways and drug gene-sets: Analysis and visualization. *H.A. Gaspar[1,2], C. Hübel[1,2], G. Breen[1,2].* 1) King's College London, Institute of Psychiatry, Psychology and Neuroscience, MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, UK; 2) National Institute for Health Research Biomedical Research Centre, South London and Maudsley National Health Service Trust, UK.

   Introduction Here, we use summary statistics from genome-wide association studies (GWAS) to find significant biological pathways, disease pathways, drug targets and drug gene-sets in schizophrenia, BMI, body fat % (BF%), and fat-free mass (FFM). We introduce a workflow encompassing data collection and curation, pathway analysis of drug gene-sets and biological pathways, drug class enrichment analysis, and visualization of "pathway landscapes". **Materials and Methods** We used the latest schizophrenia GWAS from the PGC, and new GWASs of BMI, BF% and FFM based on UK Biobank data. We curated drug/gene interactions from various databases (DGIdb, DSigDB, Ki DB, PHAROS, ChEMBL). Biological pathways were collected from MSigDB (GO and canonical pathways) and disease pathways from the Open Targets platform. The software MAGMA was used to produce gene-wise and pathway-wise associations, and MetaXcan together with GTEx data (Genotype-Tissue Expression project) to predict gene expression levels in different tissues. The enrichment of drug classes was estimated by labelling drugs by ATC code (Anatomical Therapeutic Chemical) and estimating the enrichment using a Wilcoxon-Mann-Whitney test and the AUC (area under the enrichment curve). Pathway landscapes were generated using the Generative Topographic Mapping (GTM), a probabilistic dimensionality reduction algorithm.   **Results** Results show new Bonferroni-significant pathways in schizophrenia. Antipsychotics and antiepileptics are enriched in the latest and largest schizophrenia GWAS from the PGC(1). Different disease pathways are found in BMI and BF% - e.g., "binge eating" and "age at menarche" for BMI. 47 disease pathways and 24 biological pathways are significantly associated with FFM - related to bone, height, etc. Pathway landscapes reveal that many significant pathways overlap in "pathway clusters".   **Conclusion** A comprehensive approach is necessary to investigate drug repurposing opportunities and biological pathways using GWAS results. Two key issues are data availability (drug/target affinities, gene annotations) and sample sizes for GWAS studies. Growing public databases and increasing sample sizes will help us to improve our understanding of the genetic etiology of complex diseases.   (1) Pathways analyses of schizophrenia GWAS focusing on known and novel drug targets. H. A. Gaspar, G. Breen. bioRxiv 091264; doi: https://doi.org/10.1101/091264.

1241T

A paradigm for using human GWAS summary statistics to accurately test gene expression correlation networks derived from model organisms or post-mortem tissues. *S. Bacanu, V. Vladimirov, C. Chatzinakos.* Virginia Commonwealth University, Richmond, VA.

Due to the direct assessment of biologically relevant variables, gene expression experiments are very powerful in detecting genes and pathways associated with traits. Unfortunately, in humans very rarely researchers can assay the tissue of interest, e.g. brain tissues for psychiatric disorders. Nevertheless, such experiments are easily performed in animal models and post-mortem human tissues. However, the main issue for such experiments is their perceived lack of generalizability to their homologues in living human subjects. First, the correlation networks derived from animal models might not be relevant for humans. Second, post-mortem tissue experiments are beset by the gene expression differences/correlations not being causal to disease but occurring due to the medication, changes in nutrition post disease onset, tissue decomposition after death, etc. To validate these potentially high powered but "softer" hypotheses, researchers must confirm them using firmer human data, e.g. summary statistics from Genome-Wide Association Studies (GWAS) for relevant traits. To this end we propose a three-step procedure that is based on our **G**ene-level **J**oint **A**nalysis of functional SNPs in **C**osmopolitan **C**ohorts (JEPEGMIX) transcriptomics method, which uses Quantitative Trait Loci (eQTLs) to infer the association between trait and predicted expression of the gene under study. First, it decomposes the gene correlation matrix into m principal components (PCs), which can include all PCS or only the most "important" ones. Second, it uses GWAS summary statistics to obtain a novel, fast Z-score statistic that tests for the significance of the association between trait and linear combination of expression for human orthologous genes associated with each PC. Third, it combines the Z-scores for each of the PCs in an overall χ^2_m pathway test. Simulations indicate that our method is fast and controls the false positive rates at nominal rates. We applied JEPEGMIX to summary statistics from large meta-analyses for several psychiatric traits, e.g. schizophrenia and bipolar disorders. As softer hypotheses needing validation we used gene expression correlation networks derived from postmortem brain expression studies involving tissues donated by the Stanley Medical Research Institute from normal subjects and subjects with schizophrenia and bipolar disorders. More than half of the PCs associated with these networks were found to be significant after validation using GWAS statistics.

1242F Unprogrammed presentation number

1243W

Ultra-accurate complex disorder prediction: Case study of neurodevelopmental disorders. *L. Huynh[1], F. Hormozdiari[1,2,3].* 1) Genome Center, UC Davis, Davis, CA; 2) MIND institute, UC Davis; 3) Department of Biochemistry and Molecular Medicine, UC Davis.

Early prediction of complex disorders (e.g., autism and other neurodevelopmental disorders) is one of the fundamental goals of precision medicine and personalized genomics. An early prediction of these disorders can have a significant impact on increasing the effectiveness of interventions and treatments in improving the prognosis and, in many cases, enhancing the quality of life in the affected patients. Considering the genetic heritability of neurodevelopmental disorders, we are proposing a novel framework for utilizing rare coding genetic variation for early prediction of these disorders. We provide a novel formulation for the Ultra-Accurate Disorder Prediction (UADP) problem and develop a novel combinatorial framework for solving this problem. The primary goal of this novel framework, denoted as Odin (Oracle for DIsorder predictioN), is to make an accurate prediction for a subset of affected cases while having virtually zero false positive predictions for unaffected samples. Note that in the Odin framework we will take advantage of the available functional information (e.g., pairwise coexpression of genes during brain development) to increase the prediction power beyond genes with recurrent variants. Application of our method accurately recovers an additional 8% of autism cases with LoF variants not disrupting recurrent mutated genes in the training set. Note that our approach has less than 0.5% false positive prediction based on our analysis of unaffected controls. We also extended the Odin method to predict genes that severe variants will potentially cause neurodevelopmental disorder with high probability. Utilizing this approach, we also predicted a set of 391 genes that severe variants in these genes can cause autism or other neurodevelopmental disorders. These 391 genes were not only significantly enriched in known autism genes (10-fold enrichment, p<0.001), but also in reported *de novo* variants in autism probands. Furthermore, we observed that the genes which were predicted can also help in dissecting some of the phenotypes and predicting potentially more severe subtypes of autism. Finally, we observed that the probands which were accurately predicted by Odin had significantly smaller male to female gap than expected. Odin is publicly available at https://github.com/HormozdiariLab/Odin. .

1244T

Sample-specific background correction leveraging vast historical patient cohort maximizes sensitivity of noninvasive prenatal screening. *D. Muzzey, J. Tratner, K.M. D'Auria, K.E. Kaseniit, D.H. Hite, G.J. Hogan, C.B. Oyolu, C. Haverty, K.R. Haas.* Counsyl, South San Francisco, CA.

Background: Noninvasive prenatal screening (NIPS) using whole-genome sequencing (WGS) identifies fetal chromosomal aneuploidies by measuring larger-than-expected deviations in the depth of sequenced cell-free DNA. A key determinant of test quality, therefore, is an accurate expectation of the disomic depth distribution, such that significant deviations are properly detected. Early implementations of the WGS method calculated expected disomic depth in a simple "one-dimensional" manner, using either the average depth of other "reference" chromosomes in the same sample or the average of the single chromosome of interest across the other 50-100 samples in the batch. These approaches have two drawbacks: (1) limited background observations (e.g., only 23 chromosomes), and (2) insufficient accounting for systematic biases—e.g., GC bias and total depth—between samples and chromosomes. Here we present a strategy for determining an accurate expected disomic depth that is multidimensional, accounts for sources of bias, uses all chromosomes, and utilizes tens of thousands of samples. **Methods:** WGS data from 34,000 anonymized clinical NIPS samples was featurized by three values: the depth, the GC-bias profile, and the chromosome-bin-depth medians. Applying K-means clustering to the resulting dataset yielded 200 "background clusters", each with 500 samples. Each newly sequenced sample was next assigned to the closest background cluster. Aneuploidy calling used the chromosome's z-score, which reflects the extent of depth deviation from the background cluster. A regression model leveraging all chromosomes and all samples of the background cluster predicted the expected disomic depth for the chromosome of interest in the given sample. Sensitivity was calculated by developing a modeling framework that uses (1) the fetal-fraction distribution and (2) the relationship between z-scores and fetal fraction to predict the proportion of aneuploid samples expected to be called positive. **Results & Conclusions:** Our algorithm appreciably increased sensitivity relative to standard algorithms, notably yielding a two-fold reduction in the expected false-negative rate for chromosomes 13, 18, and 21. Though our work showcases sample-specific background correction for NIPS, we expect that harnessing multidimensional data from thousands of historical samples will improve other large-scale clinical genomic assays as well.

1245F

Identification of homozygous deletions from exome sequencing data.
M. Kohda[1,2], T. Hirata[2], T. Fushimi[3], H. Harashima[4], K. Murayama[3], A. Ohtake[4], Y. Okazaki[1,2]. 1) Intractable Disease Research Center, Juntendo University, Bunkyo-ku, Tokyo, Japan; 2) Research Center for Genomic Medicine, Saitama Medical University, Saitama, Japan; 3) Department of Metabolism, Chiba Children's Hospital, Chiba, Japan; 4) Department of Pediatrics, Saitama Medical University, Saitama, Japan.

Exome sequencing are routinely used to identify single nucleotide variations and small insertion/deletions. In addition to this, many tools have been developed to detect copy number variations (CNVs) using depth of coverage data. Most tools work efficiently for finding of relatively large and heterozygous CNVs. There are few tools, however, for detecting relatively small (single or several exons) and homozygous deletions. On the other hand, variations of loss of function (LoF) has attracted attention in the field of disease genetics and population genetics, because it's interesting to decipher the relationship between biallelic LoF and human phenotypes. In this study, we report newly developed tool (named HomDel) that identify putative homozygous deletion regions from exome sequencing data through its depth of coverage. We tested HomDel for the detection of homozygous deletions in our 284 cases with suspected mitochondrial diseases. HomDel could identify multi-exonic deletions in known cases, and also could detect putative homozygous deletion in the *ATAD3* gene cluster region where harboring segmental duplications. The homozygous deletion approach has a potential to find complete LoF gene from patient's exomes. Especially, in patients with recessive traits, this strategy could identify missing causative genes by normal exome analysis pipelines. We will present our approach and evaluations at the meeting.

1246W

Comparison of pipelines and databases for detection and annotation of mitochondrial variants from whole-exome sequencing data. *J.C. Tsai[1,2], B. Chaumette[1,2], A. Dionne-Laporte[2], Q. He[1], G.A. Rouleau[2], L. Xiong[1,2].* 1) Centre de Recherche, Institut Universitaire en Santé Mentale de Montréal, Université de Montréal; 2) Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada.

Introduction: Detecting mitochondrial (MT) variants in human diseases has garnered interest in recent years. In the last decade, Whole-Exome Sequencing (WES) has become a research standard due to its decreasing cost, availability and manageable data output. While panels are designed to capture and sequence the entirety of an individual's nuclear exons, off-target regions, like the MT genome, are also captured based on sequence similarity to the capture probes. Bioinformatic tools can be used to analyze the off-target sequences thus providing a unique and additional data source for investigations. Our primary objectives are to investigate the best bioinformatic pipeline and technology platform to detect, evaluate, annotate and validate the mitochondrial variants from off-target WES captures, as well as to explore the implication of MT variants in various cohorts with neuropsychiatric disorders. Methods: Two alignment procedures were tested: alignment with BWA to a reference MT genome and inputing the cleaned exomic capture into two programs: MitoSeek and MToolbox. These pipelines were chosen for their ability to align WES data to the MT genome and were used to detect MT variants with various filtering criteria. A literature comparative screening was conducted to determine the best reference databases to be implemented by ANNOVAR. Input parameters were manipulated to ensure comparable results. In our pilot study, WES data from 243 individuals suffering from major psychiatric disorders was used for testing these pipelines. Calls between pipelines were compared using GATK tools. Interesting variants will be further confirmed by re-sequencing. Preliminary Results: Both tools align WES data to a reference MT genome without discrepancy. A total of 601 variant calls were made in 243 WES data, 251 have associated dbSNP IDs and are considered as true positive. Further investigation of the validity and the utility of MT variant calls from WES data are underway. Conclusions: There are clear advantages and drawbacks for each tool, and features unique to each program. We herein present a comprehensive pipeline to detect and annotate mitochondrial variants for various disease cohorts where mitochondrial variants are likely to play a role. The long-term objective of this project is to develop and implement the "best-tool" from this pilot study to a larger WES dataset (4000+ exomes with neuropsychiatric disease phenotypes) to identify potential disease associated MT variants.

1247T

Assessment of the performance of splicing predictors at non-canonical intronic sites and implications for variant classification. *D.H. Tran, H.P. Kang, R. Mar-Heyming, S. Guha, 4.* Counsyl, South San Francisco, CA.

   Genetic variants that disrupt or alter normal splicing are important factors underlying many genetic disorders. The appropriate clinical classification of these variants represents a challenge that could be assisted by accurate *in silico* predictors and comprehensive population-genetic information. Here we leveraged the Counsyl variant database to evaluate the general features of intronic-variant interpretation, and assessed the performance of *in silico* predictors. We focused on "non-canonical" intronic variants, those located more than 2bp from a splice junction rather than the nearly invariant "canonical sites". 30,646 curated, intronic SNPs alleles comprised our initial dataset. When we compared the allele frequencies to Exome Aggregation consortium (ExAC) frequencies, our internal allele frequencies correlated with those in ExAC (r-squared of 0.8739, p<.00001), suggesting our sample cohort is concordant with an ostensibly general population. 13,132 intronic variants resided within 20bp of the splice junction. 1326 variants at canonical sites were excluded from further analyses. Among the non-canonical variants: 58 alleles were classified as deleterious/likely deleterious by literature/database curation of patient cases and/or functional studies, 476 alleles were classified benign/likely benign, and 9539 alleles were designated as variants of unknown significance (VUS). Comparing ExAC allele frequencies for this cohort, non-canonical pathogenic variants averaged an allele frequency of 0.019±0.00036% (n=38), benign variants 18.53±0.85% (n=471), and VUS variants 0.10±0.0013% (n=9539). Next we partitioned sites into 5' and 3' intronic regions, and analyzed the performance of MaxENT, NNsplice, and SpliceSiteFinder predictions. MaxENT had the best overall performance, as estimated by the area under the receiver-operator characteristic curves (5' AUC = 0.95, n = 322; 3' = 0.86, n = 267), followed by NNsplice (5' AUC = 0.95, n = 302; 3' = 0.68, n = 222), and SSF (5' AUC = 0.88, n = 328; 3' = 0.77, n = 266). The overall sensitivity and specificity for these tools were higher for non-canonical intronic variants in the 5' region than variants in the 3' region. In summary, our analyses show that outside of the canonical sites, these tools, while providing a useful estimation of potential splicing defects, need additional lines of evidence for accurate classification.

1248F

Chromosomal sequencing analysis: Assessing the performance of copy number variant analysis by next generation sequencing compared to traditional clinical microarray analysis. *S.P. Strom, W. Scaringe, C.Y. Lee, M. Li, H. Yen, Y. Xue, Y. Shen, H. Gao.* Fulgent Diagnostics, Temple City, CA.

   INTRODUCTION: Chromosomal Sequencing Analysis (CSA) is the detection of structural variants using Next Generation Sequencing (NGS). We have developed a highly sensitive CSA pipeline using a combination of low-pass whole genome NGS and capture based targeted NGS. This screen, when combined with orthogonal confirmation enables high performance CSA with high sensitivity and specificity. In this study, we compare the types of genetic alterations clinically validated as detectable using this method in comparison with the gold standard clinical approach, Clinical Microarray Analysis (CMA). METHODS: Whole genome sequencing was performed with an average coverage under 1x and large copy number variants (CNV) were called using CNVgenome™. Target capture NGS was performed and exon-level CNV were called using CNVexon™. Results for were analyzed by at least two licensed molecular geneticists and compared with known positive and negative controls previously tested by CMA or other standard methods. RESULTS: CNV were detected with sensitivity and pre-confirmation specificity ranging from >99% and >99.9%, respectively, for CNV affecting multiple exons or larger regions down to 97% and 94% for CNV affecting only two exons. CONCLUSIONS: Chromosomal Sequencing Analysis has a higher resolution than SNP-based CMA, with the ability to detect much smaller deletions and duplications, and has equivalent ability to detect microdeletions and microduplications >100kb. CMA offers better resolution at detecting absence of heterozygosity (AOH). Combined with the benefit of potential reflex testing to sequence variant analysis, this test may be a resource-efficient alternative front-line test. Turn-around time for large CNVs is competitive with CMA (estimated at 2 weeks). Deeper whole genome sequencing and additional bioinformatics tools are required to detect copy neutral structural variants such as balanced translocations, and is not yet validated.

1249W

A Bayesian network approach for *de novo* variant calling and its application on rare Mendelian disorders. *A. Toth-Petroczy[1], A.K. Mohanty[1], L. Francioli[3,4], C.A. Cassa[1], N. Carmichael[1], D. Vuzman[1], R. Maas[1], J. Krier[1], S. Sunyaev[1,2], Brigham Genomic Medicine.* 1) Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 2) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA; 3) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) Massachusetts General Hospital, Boston, MA, USA.

A growing number of studies focus on the analysis of *de novo* mutations detected by sequencing of parent-offspring trios. *De novo* mutations are a predominant cause of sporadic dominant monogenic diseases and play a significant role in genetics of complex neuropsychiatric disorders. *De novo* mutation studies inform population genetics models and shed light on biology of DNA replication and repair. However, there is room for improvement with regards to the accuracy of *de novo* mutation calling. We designed the novoCaller variant calling algorithm to detect *de novo* mutations in trio sequencing data. novoCaller is a Bayesian method that uses the information from both read-level data in the pedigree and unrelated samples. The method was extensively tested using large trio sequencing studies. We also applied the algorithm to 22 cases of suspected rare Mendelian disorders including an affected proband and unaffected parents as part of the Brigham Genomic Medicine gene discovery initiative. The caller performed with 98% specificity and 91% sensitivity on a set of validated variants. Its application resulted in a significant reduction in the resources required for manual inspection and experimental validation of the calls. Further, four *de novo* variants were found in known genes associated with rare disorders leading to a genetic diagnosis for the probands. novoCaller is available for the biomedical community to further aid diagnostics of rare genetic disorders. .

1250T

MiRMed: Therapeutic recommendation and drug repositioning framework for rare and common diseases using micro-RNA signatures. *K. Shameer[1], B.S. Glicksberg[1], B. Readhead[1], K.W. Johnson[1], M.A. Badgeley[1], M.S. Tomlinson[1], A. Laganà[1], C. Becker[1], K.K. Yadav[2], B.A. Kidd[1], P. Mukherjee[3], A.K. Tewari[2], R. Bhattacharya[3], J.T. Dudley[1].* 1) Biomedical Informatics, Genetics and Genomic Sciences, Institute of Next Generation Healthcare, Mount Sinai Health System, New York, NY, USA; 2) Department of Urology, Mount Sinai Health System, New York, NY, USA; 3) Peggy and Charles Stephenson Cancer Center, University of Oklahoma Health Science Center, Oklahoma, OK, USA.

Micro-RNAs (miRNAs) regulate gene expression of target genes and act as key players of pathophysiology pathways of various diseases. Over the last decade, miRNAs were used as therapeutic agents and has shown great promise in altering tumor progression, immune responses, and heart failure. Oligonucleotides based anti-miR have shown therapeutic efficacy by inhibiting the function of miRNAs in the setting of viremia, cancer, and autoimmune diseases. Gene knockout studies have shown that for disorders such as cancer, perturbing targets of multiple miRNAs have a higher therapeutic impact than targets of a single miRNA. Irrespective of experimental evidence, the delivery of miRNAs for therapeutic purposes is still a challenge. To address this problem, we present MiRMed - a computational drug discovery framework that uses vast compendium of published miRNA-disease signatures integrated with systematic drug repositioning data from RepurposeDB (http://repurposedb. dudleylab.org) to find compounds based on miR-disease signatures. MirMed uses meta-expression signature of miRNAs associated with a disease and targets the consensus sub-network of the target genes to find drugs that can impart the synergistic or antagonistic effect as miRNAs. To illustrate the utility of the framework, we have applied MirMed to 240 diseases covering various genetic diseases, including cancers, cardiovascular diseases, infectious diseases, and autoimmune diseases. Systematic analyses of the dataset of 240 disease signatures with 423 miRNAs and 11, 551 target genes reveal core biological functions, pathway cross-talks, and functional modules driving miRNA-based gene regulation. Using MirMed, we have identified therapeutic recommendations and drug repositioning opportunities for common cancers (ovarian cancer and prostate cancer), rare types of cancer (Waldenström Macroglobulinemia and Sézary syndrome), cardiovascular diseases (hypertrophic cardiomyopathy and heart failure), autoimmune diseases (systemic lupus erythematosus and Sjogren's Syndrome) infectious diseases (sepsis and lupus vulgaris) and genetic diseases (fragile X syndrome, Gerstmann-Straussler-Scheinker Disease, and Multiple Hamartoma Syndrome). Pre-clinical studies, experimental and electronic medical based validations combined with clinical trials using the lead compounds identified using MirMed would further improve drug development for several conditions without currently available treatments.

1251F

Examining age, tissue, and genetic effects on RNA splicing with allele-specific resolution in a diverse mouse population. *D.A. Skelly, N. Raghupathy, K. Choi, A. Srivastava, G.A. Churchill.* The Jackson Laboratory, Bar Harbor, ME.

Alternative splicing is a ubiquitous feature of gene expression that generates a wide variety of transcript isoforms from a relatively small number of protein-coding genes. Variation in splicing is increasingly being recognized as a contributor to complex traits, including human disease. We used RNA-Seq to examine splicing patterns associated with genotype, age, tissue, and sex in a diverse multi-parent mouse population that provides powerful and precise dissection of the genetic contribution to splicing variation. We used the RNA-Seq data to reconstruct the diploid genomes of individual mice and obtained estimates of allele-specific splice junction usage by aggregating transcript abundance across shared haplotypes. Differences in splice junction usage are widespread, and we identify numerous examples of age-, tissue-, and sex-specific splicing patterns. We mapped quantitative trait loci that drive variation in splice junction usage (sQTL). We found that sQTL are common, primarily local, and concordant across tissues. We examined conserved donor and acceptor splice site sequences but found few examples of variants that could explain the sQTL. This observation suggests that the regulation of splicing is genetically complex and may be dependent on as yet uncharacterized DNA elements. Our results highlight the power of multi-parent populations for genetic analysis of molecular phenotypes and shed new light on the role of genetic diversity as a driver of variation in transcript structure.

1252W

Exomic variants of an elderly cohort of Brazilians: ABraOM database.
M.S. Naslavsky[1,2], G.L. Yamamoto[1,3], T.F. Almeida[1], S.A.M. Ezquina[1], D.Y. Sunaga[1], N. Pho[4], D. Bozoklian[1], T.O.M. Sandberg[1], L.A. Brito[1], M. Lazar[1], D.V. Bernardo[5], E. Amaro Jr[2,6], Y.A.O. Duarte[7,8], M.L. Lebrão[7], M.R. Passos-Bueno[1], M. Zatz[1]. 1) Human Genome and Stem Cell Research Center, University of Sao Paulo, Sao Paulo, SP, Brazil; 2) Hospital Israelita Albert Einstein, São Paulo, SP, Brazil; 3) Department of Clinical Genetics, Children's Hospital, Medical School, University of São Paulo, São Paulo, SP, Brazil; 4) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States of America; 5) Laboratório de Estudos em Antropologia Biológica, Bioarqueologia e Evolução Humana. Instituto de Ciências Humanas e da Informação. Universidade Federal do Rio Grande, Rio Grande, RS, Brazil; 6) Radiology Institute, Medical School, University of São Paulo, São Paulo, SP, Brazil; 7) Department of Epidemiology, Public Health School, University of São Paulo, São Paulo, SP, Brazil; 8) School of Nursing, University of São Paulo, São Paulo, SP, Brazil.

We hereby present the SABE609, a collection of exomic variants from 609 elderly from a São Paulo city census-based cohort deposited in ABraOM (Online Archive of Brazilian Mutations at abraom.ib.usp.br), a web-based publicly available database. Age-related phenotypes and their underlying biology are a focus of research worldwide and components driving healthy aging are largely unknown. Comprehensive repositories of phenotypic and genotypic information are essential for variant interpretation by allele frequency filtering. Elderly individuals are less likely to harbor pathogenic mutations for early and adult-onset diseases, therefore such variant databases are of great interest. São Paulo population is highly admixed, with parental populations from Europe, Africa and Asia, underrepresented in major public databases. Among the over 2.3 million variants, 1,282,008 are high-confidence calls. We identified more than 200 thousand variants absent from major public databases. 9,791 variants with potential loss of function consequence were identified, with an average of 300 mutations per individual. Pathogenic variants on clinically relevant genes (ACMG) were found in 1.15% of the individuals. These observations illustrate the relevance of collecting data from different populations. Datasets of census-based of aged individuals are extremely valuable to improve the interpretation of variant pathogenicity.

1253T

Age-related changes in abundance of extracellular RNA in human serum.
D.F. Dluzen[1], S. De[2], W.H. Wood[2], K.G. Becker[2], A.Z. Zonderman[1], T. Tanaka[3], L. Ferrucci[3], N. Noren Hooten[1], M.K. Evans[1]. 1) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, NIH, Baltimore, MD; 2) Laboratory of Genetics and Genomics, National Institute on Aging, NIH, Baltimore, MD; 3) Translational Gerontology Branch, National Institute on Aging, NIH, Baltimore, MD.

Extracellular RNAs (exRNAs) are cell-free RNAs found in human serum and other bodily fluids and can be used as biomarkers for disease. Profiling exRNAs in a readily-accessible biofluid such as serum may help identify biomarkers of healthy aging or age-related disease. However, previous studies are limited either by profiling one type of RNA biotype or by RNA size (small or large). As a result, there has never been a single comprehensive analysis of total exRNA profiles in the context of healthy aging. We hypothesized that age-related changes in exRNAs are important for normal aging and age-related degenerative processes. Here, we sequenced both large and small RNAs from human serum under one comprehensive and novel protocol. Total RNA was isolated from serum from 13 young (30-35 yrs) and 10 old (80-85 yrs) healthy, African American females from the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) and the Baltimore Longitudinal Study of Aging (BLSA) studies. RNA-seq was performed on the Ion Torrent Proton and reads were aligned to the human genome (v19) using separate scripts to identify linear, micro-, and circular RNAs. There was an average of 8.4 million reads per individual. We identified over 30,000 unique RNA transcripts, of which almost 9,000 were shared between young and old. We sequenced varying RNA transcript biotypes, including: messenger, micro-, long non-coding, ribosomal, small nuclear and nucleolar, transfer-, Y-, and circular. The total number of unique transcripts varied by individual. We found that transcript abundance of total small nuclear RNAs significantly decreased with age and abundance of mitochondrial transfer-RNAs increased with age significantly. We identified 142 linear transcripts in at least 90% of our samples, 38 of which were present in every individual. There were significant age-related changes in serum abundance for 10 of these transcripts: *HBB, MT-TL1, SNORD69, RNY4P8, MTRNR2L9, MT-TM, RNU2-59P, MTRNR2L6, GNL3,* and *MT-ND1*. We identified 454 unique miRNAs throughout the entire cohort and observed age-dependent changes in serum abundance for 11 of them, including; miRs -320a, -3653, -145-5p, -6087, -425, -1307-5p, 3607, -301a-5p, -191-5p, -2110, and -30d-5p. Our novel approach provides evidence of an extensive, age-related exRNA transcriptome present in human serum.

1254F

Mixture models reveal multiple positional bias types in RNA-seq data and lead to accurate transcript concentration estimates. *J. Aleksejeva, A. Tuerk, G. Wiktorin, S. Güler.* Lexogen, Vienna, Austria.

Accuracy of transcript quantification with RNA-Seq is negatively affected by positional fragment bias. Mix^2 (rd. "mixquare") is a transcript quantification method which uses a mixture of probability distributions to model and thereby neutralize the effects of positional fragment bias. The parameters of Mix^2 are trained by Expectation Maximization resulting in simultaneous transcript abundance and bias estimates. We compare Mix^2 to Cufflinks, RSEM, eXpress and PennSeq; state-of-the-art quantification methods implementing some form of bias correction. On four synthetic biases we show that the accuracy of Mix^2 overall exceeds the accuracy of the other methods and that its bias estimates converge to the correct solution. We further evaluate Mix^2 on real RNA-Seq data from the Microarray and Sequencing Quality Control (MAQC, SEQC) Consortia as well as the ABRF NGS study. On MAQC data, Mix^2 achieves improved correlation to qPCR measurements with a relative increase in R^2 between 4% and 50%. Mix^2 also yields repeatable concentration estimates across technical replicates with a relative increase in R^2 between 8% and 47% and reduced standard deviation across the full concentration range. We further observe more accurate detection of differential expression with a relative increase in true positives between 74% and 378% for 5% false positives. In addition, Mix^2 reveals 5 dominant biases in MAQC data deviating from the common assumption of a uniform fragment distribution. On SEQC data, Mix^2 yields higher consistency between measured and predicted concentration ratios. A relative error of 20% or less is obtained for 51% of transcripts by Mix^2, 40% of transcripts by Cufflinks and RSEM and 30% by eXpress. Titration order consistency is correct for 47% of transcripts for Mix^2, 41% for Cufflinks and RSEM and 34% for eXpress. We, further, observe improved repeatability across laboratory sites with a relative increase in R^2 between 8% and 44% and reduced standard deviation. On the ABRF data we study repeatability under different types of RNA degradation and for libraries generated with either poly(A) enrichment or rRNA depletion. We ran Mix^2 and Cufflinks and found an average increase in R^2 of 28.23% for Mix^2 over Cufflinks. When correlating FPKM and qPCR fold changes between UHR and HBR for variable conditions, Mix^2 achieved an average increase in R^2 of 48.36% leading to improved ROC curves in our classification experiments of differentially expressed isoforms.

1255W

A novel clustering model for droplet-based single cell transcriptomic data. *W. Chen[1,2], Z. Sun[1], T. Wang[2], K. Deng[3], R. Lafyatis[4], M. Hu[5].* 1) Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 2) Division of Pulmonary Medicine, Allergy and Immunology; Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA; 3) Center for Statistical Science, Tsinghua University, Beijing, China; 4) Department of Medicine, Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 5) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio.

   Single cell transcriptome sequencing (scRNA-Seq) has become a revolutionary tool to study cellular and molecular processes at single cell resolution. Among existing technologies, the recently developed droplet-based platform enables efficient parallel processing of thousands of single cells efficiently. Despite the technology advances, statistical methods and computational tools are still lacking for analyzing droplet-based scRNA-Seq data. Model-based approaches for clustering large-scale single cell transcriptomic data are not well studied. We developed DIMM-SC, a Dirichlet Mixture Model for clustering droplet-based Single Cell transcriptomic data. This approach explicitly models Unique Molecular Identifier (UMI) count data from scRNA-Seq experiments and characterizes variations across different cell clusters via a Dirichlet mixture prior. An expectation-maximization algorithm is used for parameter inference. We further extended the model to incorporate heterogeneity across multiple individuals. We performed comprehensive simulations to evaluate DIMM-SC and compared it with existing clustering methods such as K-means, CellTree and Seurat. In addition, we analyzed public scRNA-Seq datasets with known cluster labels and in-house scRNA-Seq datasets from a study of systemic sclerosis with prior biological knowledge to benchmark and validate DIMM-SC. Both simulation studies and real data applications demonstrated that overall, DIMM-SC achieves substantially improved clustering accuracy and much lower clustering variability compared to other existing clustering methods. More importantly, as a model-based approach, DIMM-SC is able to quantify the clustering uncertainty for each single cell, facilitating rigorous statistical inference and biological interpretations. DIMM-SC has been implemented in a user-friendly R package with a detailed tutorial available on www. pitt.edu/~wec47/singlecell.html.

1256T

CookHLA: Accurate, efficient, and memory-efficient HLA imputation. *S. Cook[1,2], X. Jia[3], P. Bakker[4,5], S. Raychaudhuri[6,7,8,9], B. Han[1].* 1) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea; 2) Department of Biomedical Science and Engineering, University of Ulsan College of Medicine, Seoul, Republic of Korea; 3) Department of Neurology, University of California–San Francisco, San Francisco, CA, USA; 4) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands; 6) Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, University of Manchester, Manchester, UK; 7) NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester Foundation Trust and University of Manchester, Manchester Academy of Health Sciences, Manchester, UK; 8) Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 9) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

   Motivation: Human leukocyte antigen (HLA) genes in the major histocompatibility complex (MHC) region influence many disease phenotypes. However, in order to thoroughly investigate MHC and identify disease-driving HLA variations, HLA typing in a large number of individuals is required, which can be costly. To overcome this challenge, we previously developed an HLA imputation method called SNP2HLA, which helps investigators to impute HLA alleles and amino acid residues using the intergenic SNP markers. To date, SNP2HLA has been widely used in the community to define disease-driving HLA alleles and amino acids for many disease phenotypes. However, SNP2HLA has several limitations; (1) The imputations are often inaccurate, (2) The computational cost is prohibitively high for a large reference panel (N>10,000), and (3) The memory requirement is high for large sample. **Results:** In this study, we developed a new HLA imputation method, CookHLA. CookHLA differs from previous approaches in that it learns the pseudo genetic distance between the binary markers and utilizes this information for imputation. This strategy dramatically improved the imputation accuracy; in the simulation experiment where we used a European reference panel to impute HLA of HapMap CEU data and measured the average 4-digit imputation accuracy of 6 HLA genes, CookHLA superbly achieved 97.4% accuracy, whereas SNP2HLA and HIBAG achieved 91.2% and 93.1% accuracy, respectively. A wide range of different simulation experiments including the Asian data analysis showed similar relative performances between the methods. CookHLA is more computationally efficient than any other competing methods, consumes minimal memory, and supports multi-core computation. Utilizing the largest-to-date reference panel (Chinese, N=10,689) takes only 3 hours of computation time and 3Gb memory in CookHLA, even when using a single core. .

1257F

A novel Word2vec based tool to estimate semantic similarity of genes by using gene ontology terms. *D. Duong[1], J. Li[2,3], E. Eskin[1,3].* 1) Computer Science, University of California, Los Angeles; 2) Statistics, University of California, Los Angeles; 3) Human Genetics, University of California, Los Angeles.

The Gene Ontology (GO) database contains GO terms that describe biological functions of genes and proteins. A GO term contains one or two sentences describing a biological aspect. One application of the GO database is the comparison of two genes or two proteins by first comparing semantic similarity of the GO terms that annotate them. Previous methods for this task have relied on the fact that GO terms are organized into a tree structure. In this old paradigm, the locations of two GO terms in the tree dictate their similarity score. In this paper, we introduce a new solution to the problem of comparing two GO terms. Our method uses natural language processing (NLP) and does not rely on the GO tree. In our approach, we use the Word2vec model to compare two words. Using this model as the key building-block, we compare two sentences, and definitions of two GO terms. Because a gene or protein is annotated by a set of GO terms, we can apply our method to compare two genes or two proteins. In addition, we also train the Doc2vec model using the descriptions of genes from ncbi.nlm.nih.gov, so that the gene names not seen in open accessed Pubmed articles are converted to vectors. These vectorized gene names are used to directly measure semantic similarity between genes. We validate our method in two ways. In the first experiment, we measure the similarity of genes in the same regulatory pathways. In the second experiment, we test the model's ability to differentiate a true protein-protein network from a randomly generated network. Our results are equivalent to those of previous methods which depend on the GO tree. This gives promise to the development of NLP methods in comparing GO terms.

1258W

Systematic assessment of joint genotyping. *U. Evani, R. Musunuri, A. Abhyankar.* New York Genome Center, New York, NY.

Today we routinely see designing of projects that involve sequencing of tens of thousands of individuals. To obtain better sensitivity and specificity in variant calling, a joint genotyping strategy is generally recommended. This strategy typically involves multiple steps to be run on a large cohort of samples. Although this strategy has been used in a lot of large-scale projects, to date we have not seen a study that systematically characterizes the benefits or limitations of the joint genotyping strategies. The process is computationally intensive and beyond the capacity of most research labs, so we think such an assessment would be beneficial to the community. To that end, we performed whole genome sequencing on HapMap sample NA12878 on Illumina HiSeqX platform and pre-processed the data using GATK best practices recommendations. The resulting gVCF was jointly genotyped along with gVCFs from other samples sequenced and analyzed using the same protocol. NA12878 was jointly genotyped with these additional samples in batch sizes ranging from 2 to 4500 samples. After joint genotyping, we analyzed the resulting variant calls for NA12878 in each batch to, (1) measure the change in sensitivity and specificity of variant calling in comparison to confident callset provided by Illumina as part of Platinum genomes effort, (2) evaluate the effect of joint genotyping batch sizes on the individual genotype calls and the variant parameters that affect the called genotype. We also analyzed the effect of large batch sizes on genotyping extremely rare variants.

1259T

Improving sequence read mapping and allele calling for the polymorphic *PRDM9* gene using a reference graph approach. *H. Gibling[1,2], A. Ang Houle[1,2], J.T. Simpson[1,3], P. Awadalla[1,2].* 1) Informatics, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Computer Science, University of Toronto, Toronto, Ontario, Canada.

Mapping high-throughput sequencing reads to a haploid linear reference genome is useful for identifying small variants, but is problematic for regions of high genetic polymorphism, particularly in repetitive loci. One consequence of this is a mapping bias where reads with variant alleles might not map to the reference genome. This leads to the underestimation of non-reference variants, resulting in a limited estimate of human genetic diversity. The polymorphic gene *PRDM9* is difficult to map with current methods due to its minisatellite-like zinc finger (ZnF) repeats. Typically only expressed during meiosis, PRDM9 is involved in recruiting machinery that initiates double-stranded DNA breaks during recombination. The ZnF repeats that determine its DNA-binding sites show extensive allelic variation across different populations, with more alleles observed in Africans than in Europeans. Rare alleles that diverge greatly from the reference allele are more likely to be subject to mapping biases, and additional Sanger sequencing is currently necessary to accurately genotype samples. To overcome the challenges with mapping high-throughput sequencing reads from the *PRDM9* locus, we use a sequence graph to create a flexible population-based reference, incorporating ZnF repeat arrays from 36 known *PRDM9* alleles. We used the graph to align sequencing reads from the 1000 Genomes Project. Many reads from samples of both African and European populations had higher alignment scores compared to when mapped to the GRCh37 reference. Additionally, a greater proportion of reads had improved alignment scores in the African samples than in the European samples, suggesting our reference graph captures more of the *PRDM9* genetic diversity present in the African samples than does the GRCh37 reference. We are developing a Hidden Markov Model approach to call alleles from sequencing reads aligned to our *PRDM9* graph, and are validating it with simulated data. We are then calling alleles for samples from the Pan Cancer Analysis of Whole Genomes (PCAWG) project. Our lab previously found an association between rare *PRDM9* alleles and pediatric acute lymphoblastic leukemia, and is currently investigating aberrant *PRDM9* expression in several PCAWG cancer samples. This allele-calling approach will help elucidate the relationship between PRDM9 and some cancers, and will provide a resource for genotyping samples that diverge from the reference allele.

1260F

Comprehensive benchmarking of error correction methods for next generation sequencing via unique molecular identifiers. *B.L. Hill[1], S. Mangul[1], I. Mandric[2], R. Littman[1], D. Yao[1], H. Yang[1], K. Hsieh[1], L. Chen[1], P. Ingle[1], A. Nguyen[1], S. Gill[1], N. Wu[3], R. Sun[1], J. Schroeder[4], P. Skums[2], A. Zelikovsky[2], E. Eskin[1].* 1) University of California Los Angeles, Los Angeles, CA; 2) Georgia State University, Atlanta, GA; 3) The Scripps Research Institute, La Jolla, CA; 4) Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria, Australia.

Error correction is an important computational technique that promises to deliver highly accurate sequencing calls and improve the results of next-generation sequencing (NGS) analysis. While errors in data sets are a concern for any NGS-based application, they mostly affect applications that use variants at frequencies similar to the frequencies of errors. Currently, a plethora of error correction methods based on different computational approaches is available. The optimal choice of error correction method for users of NGS platforms is often unclear, because we lack standardized comparison platforms and 'gold standard' benchmark data sets. For this reason, NGS-based applications cannot adopt today's most accurate and efficient error correction techniques. In this paper, we provide the first comprehensive assessment of error-correction algorithms (n=15) based on high-quality sequencing data derived from heterogeneous populations, the most challenging population on which to apply error correction methods. We use two case studies; one consists of a population of viral genomes, and the other consists of a population of T cell receptor clonotypes. Such heterogeneous populations of distinct, but closely related, clonotypes pose a serious challenge to error-correction algorithms due to the comparable level of artificial sequencing noise (i.e., errors) and true low-frequency genetic intra-population diversity. We used the UMI-based high-fidelity sequencing protocol (safe-SeqS) to eliminate errors from the sequencing data. Using raw reads and high fidelity reads obtained with the protocol, our approach provides an accurate and robust baseline for performing realistic evaluation of error correction on sequenced genomes. The most challenging reads to perform error correction on are derived from heterogeneous samples composed of highly similar, yet distinct, genomic variants, such as viral and T cell populations. We apply different methods to reads from such communities in order to assess whether error correction affects the ability of the assembly method to identify low frequency variants. We present the assessment criteria we used in the study to allow the user to make an informed choice of the most suitable software for specific NGS projects. Scripts for computing the accuracy of error-corrected reads and the 'gold standard' data are publicly available at https://smangul1.github.io/ec/, allowing users to easily benchmark any new error-correction algorithm.

1261W

ART_PacBio: A fast and accurate simulator for PacBio sequencing platforms. *W. Huang[1], L. Niu[2].* 1) NERL, EPA, Cincinnati, OH; 2) Department of Environmental Health, University of Cincinnati, OH.

   In the last couple of years, PacBio SMRT long-read sequencing technology has been drastically increasingly used for *de novo* whole genome sequencing, metagenomics studies, full-length transcription studies, survey large population for structural variants, and many others. Such fast expansion of PacBio sequencing applications is largely due to the drastically improved PacBio sequencing systems with markedly improved base accuracy, substantial higher throughput, and much shorter sequencing time needed. To facilitate biomedical research using PacBio sequencing system, we are developing a new sequencing simulator ART_PacBio to support the latest PacBio sequencing system (e.g., Sequel System), and to support the newest sequencing chemistries (e.g., Sequel 2.0), which currently are not supported by the available simulation tools. ART_PacBio also supports the previous PacBio system PacBio RS II, allowing users to simulate data from the popular old system. In order to generate realistic simulated PacBio data, we evaluated substitution, insertion, and deletion sequencing error rates from multiple latest PacBio datasets. In addition, we developed a new Gamma-based distribution to model raw read length generated from zero-mode waveguides (ZMWs) of PacBio SMRT Cell, and use a Gaussian distribution for the original DNA fragment length or eventual consensus sequence length. Furthermore, ART_PacBio can directly use empirical distributions generated from real PacBio sequencing data for simulation. ART_PacBio, implemented in C++, will be freely available to the public once it is released.

1262T

A high-throughput pipeline for genotyping within primary health care to enable precision prevention. *S. Jai Kumar Ahuja, M. Soe, L. Fricke, J.M. Lubieniecka, S.D. Bennett, K.P. Lubieniecki, C.F. Boerkoel, R. Pyatt, M. Sincan, C-H. Chan, P.F. Cherukuri.* Genomic & Molecular Medicine, Sanford Imagenetics, Sioux Falls, SD.

   Implementation of large-scale genotyping within primary health care for purposes of precision prevention remains problematic in part due to high costs of genetic testing. We hypothesized that SNP chip genotyping using Illumina's Infinium® Global Sequencing Array with 660,000 markers in 24-sample format, provides a cost effective first line genotyping for pharmacogenetics and some disease predisposing alleles deemed medically actionable by the American College of Medical Genetics (ACMG) and Sanford clinicians. To test this hypothesis, we developed a high-throughput laboratory and a set of bioinformatics tools that are tightly integrated with the electronic medical record. In this context, we are able to offer pharmacogenetics and some predisposition testing to all individuals within the health care system at the time of a visit to their primary care physician. We launched this as part of a precision prevention initiative within a rural primary health care setting. We present the process and logic underlying this as well as our initial results on patient response, clinician response, and educational initiatives encouraging uptake of the tool.

1263F

Reducing reference bias in whole-genome variant calling using reference graphs. *M. Kallberg, M. Popovic, G. Rakocevic, M. Suciu, K. Ghose, Y. Li, J. Browning, P. Komar, V. Semenyuk, A. Jain, J. Spencer, D. Kural.* Seven Bridges Genomics, 1 Main St, Cambridge, MA 02142.

The linear reference genome conventionally used when aligning reads from a sequencing experiment represents a single consensus haplotype derived from a small genetically homogenous cohort. Consequently, the range of known genetic variants in human subpopulations cannot be taken into account during sample analysis. This so-called "reference bias" causes alignment to variant-rich genome regions to become increasingly inaccurate because the applicable sample cohort diverges from the reference. Ultimately this reduces the number of aligned reads which leads to decreased sensitivity in detection of sample variation. Here we present a graph-based genomic reference structure capturing both the linear reference and known variants, thereby enabling the simultaneous representation of several common population haplotypes. Using our graph-aware aligner, we demonstrate how utilizing known variations can significantly improve alignment quality while maintaining compute-cost lower than that of standard linear aligners. Further, we evaluate the variants called from the graph alignments through several orthogonal experiments using both simulated and real data. We demonstrate that our graph-based analysis pipeline achieves higher accuracy than standard variant-processing workflows when benchmarked against public variant truth-sets. Finally, we establish that using a reference graph enables detection of common structural variants well-beyond a single read-length in size. Our methods can be readily applied to large sample cohorts and become increasingly accurate as the reference graph is enriched with cohort unique variants, thus enabling the construction of application specific variant-calling pipelines.

1264W

Integrating networks and comparative genomics reveals retroelement proliferation dynamics. *B.A. Knisbacher[1], O. Levy[2], S. Havlin[2], EY. Levanon[1].* 1) The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, Israel; 2) Department of Physics, Bar-Ilan University, Ramat Gan, Israel.

Retroelements (REs) are transposable elements that multiply and spread throughout genomes by a copy-and-paste mechanism. These parasitic elements promote diversity, cause disease and accelerate evolution in virtually all eukaryotes, including humans. Due to their high copy number and similarity, studying their activity and tracking their proliferation dynamics is a challenge. It is particularly complicated to pinpoint the few REs in a genome that are still active in the haystack of degenerate and suppressed elements. Here, we present a computational framework based on RE similarity networks that tracks the path of RE proliferation throughout evolution and highlights the most probable currently active communities and elements. Integrating biological data into the networks (e.g., comparative genomics, RE expression during embryonic development, population data) enhances the ability to track the course and tempo of RE proliferation. We use SVA, the youngest RE family in human, to demonstrate the potential of this approach. SVA insertions are associated with multiple hereditary diseases, but remain poorly characterized. By constructing and comparing SVA networks from human and other hominids, we identify the genomic locus of specific elements that most probably played pivotal roles in SVA proliferation. Our analysis identifies active SVA communities in hominids and predicts the most probable sources of *de novo* insertions in the human population, which promote diversity and genomic disease. Beyond SVA, the application of biological characteristics specific to LTR and non-LTR retrotransposons in other genomes will be discussed. For example, how footprints of DNA editing by APOBECs can enhance network-based retroelement ancestry inference.

1265T

Limits of indel detection using CLC alignment and variant calling. *T. Koganti[1,2], Z. Tu[1,2], A. McDonald[1], P. Lundquist[1], Z. Niu[1,4], D. Babovic-Vuksanovic[1,4], C. Klein[1,3,4], E. Klee[1,2,4].* 1) Department of Laboratory Medicine and Pathology , Mayo clinic , Rochester, MN; 2) Department of Health Sciences, Mayo clinic , Rochester, MN; 3) Department of Neurology, Mayo clinic , Rochester, MN; 4) Department of Clinical Genomics, Mayo clinic , Rochester, MN.

   Introduction: Variant calling is an important step in bioinformatics analysis of patient sequencing data to report pathogenic variants. Most variant calling software packages can detect Single Nucleotide Polymorphisms (SNP) very reliably but the same is not true when it comes to detecting insertions and deletions (indels). In this study, we evaluated the exact indel size at which insertions, deletions and duplications can be detected at 95% accuracy using CLC pipeline on Illumina Hiseq 2500 data. **Methods**: Samples were processed through library preparation and targeted enrichment using SureSelect hybridization methods. Prepared library was sequenced on Illumina Hiseq 2500 instrument using 101 paired-end reads. An NGS pipeline using CLC software was used to generate alignment files. 500 synthetic insertions and deletions of size 1-50 and 500 synthetic duplications of size 1-30 were artificially inserted in the sample BAM file. These BAM files were converted to fastq files and the entire bioinformatics pipeline was repeated to generate new alignment and variant files. These files were scanned for artificially inserted indels and the indels identified were categorized into 1) Indels identified with same size in output; 2) Indels identified at same start location but different indel size; 3) Indels identified in neighboring regions 4) No indels identified. This procedure was repeated for insertions, deletions and duplications. **Results**: We found that insertions of less than 30 base pairs were detected at an accuracy rate of 95%, deletions less than 27 base pairs were detected at the same rate. Even though duplications are a type of insertions, performance of insertions was far better compared to duplications. Duplications less than 18 base pairs were detected at 95% accuracy. Our study not only helps show the limits of detection size of indels for NGS pipeline but also demonstrates that insertions and duplications need to be assessed individually.

1266F

Comparison of short tandem repeat estimation methods with various conditions. *K. Kojima, Y. Kawai, K. Misawa, T. Mimori, M. Nagasaki.* Integrative Genomics, Tohoku University, Sendai, Japan.

   In the estimation of repeat numbers in a short tandem repeat (STR) region from high-throughput sequencing data, two types of strategies are mainly considered: a strategy based on counting repeat patterns included in sequence reads spanning the STR region and a strategy based on estimating the difference between the actual insert size and the insert size inferred from paired-end reads. The quality of sequence alignment is crucial, especially in the former approaches although usual alignment methods have difficulty in STR regions due to insertions and deletions caused by variations of repeat numbers. lobSTR and RepeatSeq were applied to three types of aligned reads: synthetically generated sequencing reads aligned with BWA-MEM, those realigned with STR-realigner, those realigned with ReviSTER, and those realigned with GATK IndelRealigner. From the comparison of root mean squared errors between estimated and true STR region size, the results for the dataset realigned with STR-realigner are better than those for other cases. For real data analysis, we used a real sequencing dataset from Illumina HiSeq 2000 for a parent-offspring trio. RepeatSeq and lobSTR were applied to the sequence reads for these individuals aligned with BWA-MEM, those realigned with STR-realigner, ReviSTER, and GATK IndelRealigner. STR-realigner shows the best performance in terms of consistency of the size of estimated STR regions in Mendelian inheritance. Root mean squared error values were also calculated from the comparison of these estimated results with STR region sizes obtained from high coverage PacBio sequencing data, and the results from the realigned sequencing data with STR-realigner showed the least (the best) root mean squared error value.

1267W

A graph method for population genotyping of structural variants. *P. Krusche[1], E. Dolzhenko[2], N. Johnson[2], M.A. Bekritsky[1], A. Gross[2], B.R. Lajoie[2], V. Rajan[2], Z. Kingsbury[1], S.J. Humphray[1], S.S. Ajay[2], R.J. Taft[2], D.R. Bentley[1], M.A. Eberle[2].* 1) Illumina, Saffron Walden, United Kingdom; 2) llumina Inc., San Diego, CA.

Accurate detection and annotation of structural variants (SVs) is a critical component of clinical variant calling pipelines. A significant challenge with SV calling is that the breakpoints of the same event are often represented differently from sample to sample. Nearby variants and sequence instability around the breakpoints of the SVs can further complicate these issues. As a result, SVs can be difficult to aggregate across samples leading to errors, such as underestimation of variant frequencies. We have developed a joint genotyping method for SVs based on re-aligning sequence reads to breakpoint graphs. Using this method, we are able to evaluate breakpoints uniformly across many samples and genotype SVs jointly in a population. Furthermore, we can optimize our genotyping parameters based on population analysis to improve genotyping accuracy in individual samples. A distinguishing feature of this method is that we can account for repetitive sequence around the breakpoints and produce consistent results across all samples even when small variants occur close to the breakpoint. As a demonstration of this method we have identified 185,395 putative insertions and deletions identified from a variety of sources including individual SV calls made on >2,500 samples sequenced to high depth. We jointly genotyped these variants in a cohort of over 220 WGS samples from the Polaris cohort (available at ENA as PRJEB20654 and PRJEB19931), as well as in the Platinum Genomes pedigree. Of these variants 183,098 were biallelic in at least one of these samples. We are able to confirm the accuracy of 37,878 of the variants that are bi-allelic in the Platinum Genomes pedigree using Mendelian inheritance (plus an additional 100,154 variants that are monomorphic in the pedigree). Additionally, we identified 71,505 common (MAF>5%) variants are not significantly out of Hardy-Weinberg equilibrium (HWE) in the population data. Combined we are able to validate that 70,009 of these variants pass both of these tests indicating that they are accurately genotyped in these samples. We will continue to investigate the remaining rare variants using HWE in 3,000 additional samples. The variant calls for all these samples and the joint alignment and genotyping software will be publicly available for comparison and method development. Further work will identify and solve the problems with the variants that failed our quality requirements and apply this method to more SVs.

1268T

Accurate radiation biodosimetry through automation of metaphase cell image selection and chromosome segmentation. *Y. Li[1], J. Liu[2], B. Shirley[1], R. Wilkins[3], F. Flegal[4], J.H.M. Knoll[1,2], P.K. Rogan[1,2].* 1) CytoGnomix Inc, London, Ontario, Canada; 2) University of Western Ontario, London, Ontario Canada; 3) Health Canada, Ottawa, Ontario, Canada; 4) Canadian Nuclear Laboratories, Chalk River, Ontario, Canada.

The dicentric chromosome (DC) assay is a standardized method that is recommended for determination of biologic radiation exposure[1,2]. Software to fully automate this assay has been developed in our laboratory[3]. This method relies on high quality microscope-derived images of metaphase cells to reduce the rate of false positive (FP) DCs. We present image processing methods to eliminate suboptimal metaphase cell images based on novel quality measures and to reclassify FPs by analyzing their morphological features. A set of chromosome segmentation thresholds selectively filtered out FPs, arising primarily from extended prometaphase chromosomes, sister chromatid separation and chromosome fragmentation. This reduced the number of FPs by 55% and was highly specific to the abnormal structures (≥97.7%). Image segmentation filters selectively remove images with consistently unparsable or incorrectly segmented chromosome morphologies, while image ranking sorts images according to their qualities and enables selection of optimal images in samples. Overall, these methods can eliminate at least half of the FPs detected by manual image review. By processing data to derive calibration curves and to assess samples of unknown exposures with the same image selection models, average dose estimation errors were reduced from 0.6 Gy to 0.3 Gy, without requiring manual review of DCs. During this presentation, we will use our software to demonstrate that metaphase image filtering and object selection constitute a reliable and scalable approach for biodosimetry, resulting in more accurate radiation dose estimates. 1. International Atomic Energy Agency. (2001) Cytogenetic Analysis for Radiation Dose Assessment, a Manual: Technical Reports Series. No. 405, International Atomic Energy Agency, Vienna. 2. International Atomic Energy Agency. (2011) Cytogenetic Dosimetry: Applications in Preparedness for and Response to Radiation Emergencies, International Atomic Energy Agency, Vienna. 3. Rogan, P. K., Li, Y., Wilkins, R. C., Flegal, F. N., and Knoll, J. H. M. (2016) Radiation Dose Estimation by Automated Cytogenetic Biodosimetry, Radiation Protection Dosimetry 172, 207-217. .

1269F

WGSA 07: Updated annotation pipeline for human genome sequencing studies. *X. Liu[1,2], R. Gibbs[3], E. Boerwinkle[1,2,3].* 1) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.

   Large-scale sequencing technologies have made unprecedented progress during the recent years. We are witnessing an exciting era when genetic variants—including single nucleotide variant (SNV) and insertion-deletion variants (indels)—in both healthy human subjects and those with Mendelian or complex diseases can be comprehensively identified by re-sequencing studies. Unfortunately, there is a large gap in the ability to ascertain enormous amounts of variants and the ability to understand their functionality. This gap will grow even wider as we transition from whole exome capture sequencing to whole genome sequencing (WGS). As a first step to narrow this gap, we developed WGS Annotator (WGSA), an annotation pipeline for human genome sequencing studies, to facilitate the functional annotation step of WGS. WGSA supports the annotation of SNVs and indels locally without remote database requests, so it can be scaled up for large studies. Since its first release in 2015, significant updates have been made. Here are a few highlights of the current version WGSA 07: 1. We integrated annotations from 3 popular annotation tools (ANNOVAR, SnpEff and VEP) versus 3 popular gene model databases (Ensembl, RefSeq and UCSC knowGene). As the consequence prediction of a variant often varies depending on the underlying gene model and the annotation tool used, it is suggested to obtain annotations from multiple tools across multiple databases for a complete interpretation of the variants. Here we provide a one-stop resource to achieve that goal. 2. To enable scalability, we precomputed the annotations for all potential human SNVs (a total of 8,584,031,106 and 8,812,967,043 based on human references GRCh37 and GRCh38, respectively) and use them local databases. Multithreading further speeds up the query 3-fold faster.   3. We integrated many annotation resources for both coding and non-coding variants, including functional prediction scores (CADD, DANN, Eigen&Eigen PC, FATHMM-MKL, genoCanyon, fitCons, Funseq&Funseq2, RegulomeDB, etc), population allele frequencies (1000G, ESP6500, ExAC, UK10K, GnomAD, etc), conservation scores, mappability, and annotations from epigenomic projects (ENCODE, Roadmap, FANTOM5, BLUEPRINT, etc). 4. To provide convenience to a broader community, we have built an Amazon Machine Image for running WGSA on the cloud from Amazon Web Services, in addition to a downloadable version.   More information can be found at https://sites.google.com/site/jpopgen/wgsa.

1270W

PoolHap2: Inferring within-host haplotype frequencies from pathogen next-generation sequencing data. *Q. Long[1], J. Wang[2], L. Mak[1], C. Cao[1], K. Ye[3], D. Jeffares[4].* 1) Biochemistry and Molecular Biology, University of Calgary, Calgary, AB, Canada; 2) Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, USA; 3) Department of Automation, Xi'an Jiaotong University, Xi'an, China; 4) Department of Biology, University of York, York, UK.

   There is a burgeoning demand for tools capable of analyzing the composition and evolution of a mixture of microbes within a patient. The emergence of affordable next-generation DNA sequencing allows researchers to explore the genetic variation and dynamics of microbial genomes at increasingly precise scales. However, when microbe DNA of multiple strains is collected from host tissues, the microbes are sequenced together in pools as isolating and culturing single strains of microbes is frequently a difficult and expensive task. We present PoolHap (version 2.0), a tool that makes use of advanced statistical models to accurately infer the identity and frequency of individual strains (haplotypes) from pooled sequencing *in silico*. In clinical settings, it is expected that accurate inference of pathogenic haplotypes in a heterogeneous population will facilitate targeted, strain-specific treatment. Beyond mixed pathogen analysis, this tool has a broad range of applications such as the approximation of single-cell sequencing outcomes using only conventional pooled sequencing protocols.

1271T

LRSim: A linked reads simulator generating insights for better genome partitioning. *R. Luo[1,2], F. Sedlazeck[1], C. Darby[1], S. Kelly[3], M. Schatz[1,2,4]*. 1) Johns Hopkins University, Department of Computer Science, Baltimore, MD; 2) Johns Hopkins University School of Mecine, Center for Computational Biology, Baltimore, MD; 3) New York University School of Medicine, Center for Health Informatics and Bioinformatics, New York, NY; 4) Cold Spring Harbor Laboratory, Simons Center for Quantitative Biology, New York, NY.

Haplotype-resolved or phased genomes are essential for studying allele-specific regulation and expression, and other important genomic features. However, most of the genomes assembled to date are only a single haploid 'consensus' sequence with parental alleles merged arbitrarily. 10X Genomics recently invented a low-cost, labor-efficient library preparation protocol to obtain phased genomes. The protocol ligates adapters to reads which will be sequenced on an Illumina instrument; the adapters allow each read to be traced back to its progenitor molecule. The so-called linked reads, spanning tens to hundreds of kilobases, offer an alternative to long-read sequencing for de novo assembly, haplotype phasing and other applications. However, there is no available simulator, making it difficult to measure their capability or develop new informatics tools for linked reads. Our analysis of 13 real human genome datasets of 10X Chromium linked reads revealed their characteristics of barcodes, molecules and partitions. The first "linked read" simulator we wrote, named LRSim, generates simulated linked reads by emulating the library preparation and sequencing process with highly customizable parameter settings. We compared the simulated results from LRSim to NA12878 to illustrate a high concordance between the simulated and real data. We concluded that from the phasing results of 6 simulated datasets with different mean molecule lengths and a real dataset of NA12878 that if constrained at a certain sequencing depth, the best molecule size to achieve the best phase block size needs to be meticulously chosen. This can be done by wet-lab experiments, but would be more efficient with a simulator in silico. We also performed experiments on 6 simulated A. thaliana datasets with a different number of partitions and demonstrated a substantial degradation in assembly performance with an improper number of partitions, which leads to insufficient cover-age per molecule. Finally, we concluded an appropriate sequencing depth needs to be chosen for different applications and species before sequencing to achieve the best performance out of linked-reads.

1272F

Genetic correlations as a tool for finding new biomarkers for female reproductive health phenotypes. *R. Magi[1], T. Laisk[2,3], A. Salumets[2,3], A. Metspalu[1]*. 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; 3) Competence Centre on Health Technologies, Tartu, Estonia.

Together with the personalised medicine revolution and the continuously decreasing cost of large-scale genotyping, a lot of attention has been focused on genomic biomarkers, which are robust and reliable and in theory could be used for screening, diagnosing and monitoring the activity of a disease, but also for personalised counselling and treatment regimens, or for assessing therapeutic responses. A genetic component has been identified for many common diseases, including those affecting female fertility and well-being. New methodologies have made it possible to analyse the genome-wide genetic correlations across diseases and phenotypes by comparing the association analysis summary statistics. These methods are not only useful for understanding the genetic background and etiology of studied traits but could also serve as a tool for finding new prognostic biomarker candidates. Comparison of the GWAS summary statistics of age at menopause and nuclear magnetic resonance (NMR) metabolites has revealed several NMR metabolites, which are genetically correlated with age at menopause and could be tested as predictive biomarkers. Based on genetic correlation analysis, five metabolites had nominally significant correlations ($p<0.05$) with the age at menopause. Out of these five, concentrations of chylomicrons and extremely large very-low-density lipoprotein (VLDL) particles, as well as the average number of methylene groups in a fatty acid chain have previously been associated with menopausal status, as has been docosahexaenoic acid, corroborating the view that these metabolic biomarkers could be associated with the timing of menopause. Furthermore, the remaining two metabolites from our analysis, the ratio of bis-allylic groups to total fatty acids and average number of double bonds in a fatty acid chain, are novel biomarker candidates for menopause age, and have previously been associated with body weight change. Before testing these novel biomarker candidates in the clinical setting, further in silico testing will be undertaken to verify the predictive properties of potential biomarkers.

1273W

Privacy preserving Fisher's exact test for GWAS. *K. Misawa[1,2], S. Hasegawa[3], K. Hamada[3], K. Chida[3], S. Ogishima[1,2], M. Nagasaki[1,2].* 1) Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan; 2) Graduate School of Medicine, Tohoku University, Miyagi, Japan; 3) NTT Secure Platform Laboratories, Tokyo, Japan.

   A genome-wide association study (GWAS) is a method for identification of genetic variations that are statistically significantly associated with a particular disease. Because a GWAS requires not only genomic data from the participants but also their health condition, it is necessary to keep the data private. Secure multiparty computation (MPC) protocols allow a set of participants to perform calculations privately. In this study, we developed algorithms of Fisher's exact test for a GWAS using MPC, because Fisher's exact test is widely used for GWAS.We conducted Fisher's exact tests for all possible tables with the given total sample size. On the basis of the results of these tests, a decision tree was constructed. We then developed an algorithm utilizing the decision tree by means of MPC. We devised a function, $f(\mathbf{X})$, that always yields a smaller value than the p value of Fisher's exact test, where $\mathbf{X}$ is a 2 × 2 contingency table intended to reduce computational time. We measured the computational time with sample sizes varying between 100 and 4,000. We found that the new algorithm can carry out Fisher's exact test in a shorter period in comparison with the previous algorithm.

1274T

The predictive power of RNA-Seq: Modeling blood chemistry and hematology test results. *M. Naymik, A. Siniard, M. De Both, A. Wolfe, M. Huentelman.* Neurogenomics Division, Translational Genomics Research Institute (TGen), Phoenix, AZ.

   Blood chemistry and hematology tests are common for diagnosing disease, tracking disease progression, or as an assessment of and individual's health. These tests generally involve a venous blood draw of several milliliters in volume. We are interested in developing methods that can measure these quantities using small volumes or alternative bio specimen types. To facilitate this we explored the potential use of RNA signatures as effective correlates. In this experiment we attempted to use simple and multiple linear regression to predict a panel of 59 blood tests using gene expression from 3 different bio specimen types.   Whole blood (1 ml), plasma (1 ml), and dried blood spot (DBS) (30 µl) samples were collected from 50 individuals and RNA sequencing was performed along side the blood panels. Each blood test was examined for correlation with a single gene followed by a more complex analysis where subsets of up to 5 genes were selected to model the blood test results. We considered models with an r-squared greater than 0.7 to have good predictive power. Using simple regression we identified 5 tests that correlated with a single gene in the whole blood including Eosinophils (0.81), Neutrophils (0.8), Lymphocytes (0.79), PSA (0.71), and Monocytes (0.74). In DBS specimens PSA (0.81) was the only test that passed the cutoff. In plasma there were no tests that passed the cutoff. With multiple linear regression 6 tests passed the cutoff using whole blood specimens including Lymphocytes (0.87), Monocytes (0.79), Neutrophils (0.77), Aspartate Aminotransferase (0.74), Cholesterol (0.72), and Eosinophils (0.72). For plasma specimens no models passed the cutoff. For DBS specimens 5 models passed the cutoff including Non HDL Cholesterol (0.84), Eosinophils (0.78), RBC (0.76), Absolute Neutrophils (0.73), and T7 Index (0.73). Finally, we attempted to combine multiple specimen types by using predictors from whole blood and plasma which improved the r-squared over 3 whole blood only models including RDW (0.57 to 0.68), T7 Index (0.61 to 0.65), and MCHC (0.54 to 0.62). Our results suggest good correlates can be found between gene expression and blood panel tests and in the future it may be possible to use RNA expression instead of the current tests for the blood panel. Our results also suggest most of the predictive power lies in the cellular fraction of the whole blood and not in extracellular plasma.

1275F

Quality comparison of whole exome sequencing data by different types of specimen. *K. Park, C. Fung, H. Kwon, C. Park, S. Lee, S. Park, D. Seo.* Macrogen Corp, Rockville, MD.

Nowadays, there are many researches and clinical studies performed using next-generation sequencing (NGS) data to detect genetic variations from samples. As a sequencing service provider, we have many experiences handling different types of specimens and noticed that their qualities of sequencing results are affected from different types of specimens. To examine these types of biases, we investigated mapping rate, percent of non-redundant, and off-target rate with 600 whole exome sequencing (WES) data including 200 gDNA, 200 blood, and 200 saliva samples that passed our initial QC criteria. From these comparisons, we observed off-target rates across all three types of specimens were comparable, but mapping rates and proportions of non-redundant rates in saliva samples showed much lower efficiency than the other two types of samples. Based on this study, we conclude that gDNA or the blood samples will warrant a better quality results of WES than saliva samples.

1276W

ClinGen Pathogenicity Calculator 2.0: New features and lessons learned from the data mining of 2,400 variant interpretations according to ACMG/AMP guidelines. *R. Patel, N. Shah, A. Jackson, P. Pawliczek, S. Paithankar, S. Milosavljevic, S. Plon, A. Milosavljevic.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

ClinGen Pathogenicity Calculator is a web service for assessing pathogenicity of genetic variants (Genome Medicine, 2017, 9:3). The service has been open to the public since October 2015. As of May 2017, the Calculator has more than 600 registered users. During the month of May 2017 more than 2300 page visits were generated by users from 41 countries. Based on the experience gained, Calculator was updated to include several new features to better assist and manage the variant interpretation process. Updates include a reconfigured dashboard that facilitates navigation by gene, interpretation, and phenotype. Prediction charts from key pathogenicity predictors, conservation scores, pooled allele frequency information, and link-outs to two additional sources (VarSome and InterVar) are included to assist the interpretation process. We evaluated usage patterns for the American College of Medical Genetics and Genomics and the Association of Molecular Pathologist (ACMG/AMP) guidelines using both user-contributed interpretations in the Pathogenicity Calculator and those imported from published articles reporting variant interpretations based on the guidelines. From a combined dataset of 2400 interpretations, we derived key statistics such as ACMG/AMP guidelines code usage, upgrades and pathogenicity assessment patterns for variants in 73 genes having > 10 interpretations (BRCA1, BRCA2, ATM and EFHC1 with the largest number of interpretations). Although the Calculator supports customization of variant interpretation guidelines, most users employ the published ACMG/AMP guidelines without any customization. The three most widely used variant interpretation codes on the combined dataset are BS4 (Non-segregation with the disease), PM2 (Absent in population databases) and PP2 (Missense mutation in the gene with a low rate of benign missense variants) ACMG codes. When only user-contributed assessments are considered, the most widely used codes were PP4 (Patient's phenotype or family history highly specific for the gene), PM2 and PP3 (computational evidence supports a deleterious effect). Other analysis results and lessons learned from this evaluation of ACMG/AMP guideline usage patterns to date will be presented during the meeting. The Pathogenicity Calculator is publicly available upon registration at https://calculator.clinicalgenome.org. This work was funded by the NIH National Human Genome Research Institute grants U01 HG007436-01 and U41HG006834.

1277T

***Indexcov*: Whole-genome coverage in <1 second per BAM.** *B.S. Pedersen, A.R. Quinlan.* Human Genetics, University of Utah, Salt Lake City, UT.

Comparing consecutive file offsets indicated in a BAM or CRAM index allows one to estimate the number of alignments per genomic region without parsing the BAM. By leveraging the BAM/CRAM index in this way, as a proxy for coverage, *indexcov* **can analyze a 60X human genome dataset in ~1 second and thousands of samples in minutes**. This gives a coarse (~16KB resolution) whole-genome coverage estimate that, for the most part, closely tracks the actual depth. This represents a substantial improvement in the time required to profile coverage with existing methods and a valuable, early quality-control measure. The output of indexcov is an interactive HTML page that facilitates the **detection of sex chromosome anomalies, large copy-number variants, high-variance samples, batch effects, and other coverage anomalies**. *Indexcov* infers sample sex from coverage on the sex chromosomes and outputs metrics in a simple text files for filtering and custom visualization. In addition to coverage plots for each chromosome, it outputs a summary plot that simplifies finding outlier samples. *indexcov* is available as an executable from: https://github.com/brentp/goleft/releases with source-code available under the MIT license.

1278F

Variants impact on splicing regulatory element determination pipeline. *N. Prodduturi, G. Oliver, Y. Li, E. Klee.* Mayo Clinic, Rochester, MN.

ABSTRACT Background: Misregulation in the splicing component is linked to multiple hereditary diseases[1]. Many disease-causing variants disrupt splicing mechanism which affects splice sites or splicing regulatory sites in humans[2]. Splicing regulatory elements (SREs) are of different types i.e. splice sites, splicing silencers, and splicing enhancers and causal variants can impact these elements. Available DNA-based methods predict variants that affect these regulatory sites but applying them to mRNA level leads to a significant number of false positive events predicted as pathogenic mutations. Existing RNA-based tools also give large numbers of false positive candidate events. Despite available tools, a comprehensive Workflow to identify splice variants is still needed. **Method:** We have developed an integrative rule-based pipeline to annotate cis-acting variants in SREs and splice sites that cause splicing aberrations based on DNA-Seq and RNA-Seq data. BAM file with RNA-seq reads, mutations from RNA-seq and DNA-Seq (variant calls from DNA-Seq are optional), junction read counts and exon and gene level expression data are inputted into the workflow. Optional modules like gene expression sample filtering and variant calling are also included. The pipeline filters the mutations to the local vicinity Junctions, integrates different genomic input data, summarizes them to mutation level and annotates them as variants causing partial or whole exon exclusion or intron retention, and the deletion or insertion of splicing sites using a rule-based method. Where reference expression data exist, the workflow has an option to predict affect of splicing aberration on a relative gene or exon level expression. **Results:** With this workflow, aberrant splice mutations can be detected with high sensitivity and which can be applied to transcriptomics and genomics studies. We have used this workflow to detect disease causal genetic mutations affecting various splicing events in unsolved diagnostic odyssey cases. **References:** Lim, KH; Ferraris, L; Filloux, ME; Raphael, BJ; Fairbrother, WG (2011). "Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes". Proc. Natl. Acad. Sci. USA. 108 (27): 11093–11098. 2.López-Bigas, Núria; Audit, Benjamin; Ouzounis, Christos; Parra, Genís; Guigó, Roderic (2005). "Are splicing mutations the most frequent cause of hereditary disease?". FEBS Letters. 579 (9): 1900–3.

1279W

Working through heterozygous variants in NGS screenings: The DOMINO algorithm allows recognizing potentially dominant genes by machine-learning. *M. Quinodoz[1], B. Royer-Bertrand[1,2], K. Cisarova[1], S.A. Di Gioia[1], A. Superti-Furga[2], C. Rivolta[1,3].* 1) Département de Biologie Computationelle, Université de Lausanne, Lausanne, Lausanne, Switzerland; 2) Division of Genetic Medicine, Lausanne University Hospital (CHUV), Lausanne, Switzerland; 3) Department of Genetics, University of Leicester, Leicester, United Kingdom.

Introduction: In contrast to recessive conditions with biallelic inheritance, identification of dominant (monoallelic) mutations in novel disease genes is more difficult, because of the abundance of benign heterozygous variants that act as massive background noise (typically, in a 400:1 excess ratio with respect to pathogenic mutations). Methods: To reduce this overflow of false positives in Next-Generation Sequencing (NGS) screens, we developed DOMINO, a tool assessing the likelihood for a gene to harbor dominant changes. Unlike commonly-used predictors of pathogenicity, DOMINO takes into consideration features that are proper of genes, rather than of variants. It uses a machine-learning approach to extract discriminant information from a broad array of features, including: genomic data, intra- and interspecies conservation, gene expression, protein-protein interactions, protein structure, etc. DOMINO's iterative architecture includes a massive input of gene-associated properties (432, in total), a training process on 985 disease genes with well-established inheritance patterns, and repeated cross-validation that optimizes its overall discriminant power. Results: When validated on all newly-discovered disease genes published in the *AJHG* and in *Nat Genet* over a 15-month period, the algorithm displays an excellent final performance, with the receiver operating characteristic (ROC) function giving values for the area under the curve (AUC) higher of 0.91 and 0.92 for training and validation respectively. Conclusion: DOMINO is a robust and reliable tool that can predict dominance of candidate disease genes with high sensitivity and specificity, making it a useful complement to any NGS pipeline dealing with the analysis of the morbid human genome.

1280T

Precision phenotyping of sample biopsies of kidney transplants by leveraging public microarray data and cell types enrichment analysis. *D. Rychkov[1,2], T. Sigdel[2], M. Sarwal[2], M. Sirota[1].* 1) Institute for Computational Health Sciences, UCSF, San Francisco, CA; 2) Department of Surgery, UCSF, San Francisco, CA.

Acute Rejection (AR) still remains the driving cause of renal failure within one year after transplantation. The access to publicly available transcriptomic data provides a great opportunity for meta-analysis enlightening new diagnostic and therapeutic strategies. However, the data can have phenotypic annotations that don't truly reflect biological process going on in transplant, since a histological diagnosis can vary among pathologists. To reduce variance in analyses, we developed a new approach for precision phenotyping of allografts. We leverage the publicly available microarray data from the NCBI GEO from 38 datasets of human kidney allograft biopsies. After data quality control and filtering procedures, we ended up with the 1,734 samples, down from 2,973 samples. For each sample, we first used the phenotypes as they are described in the original publications: stable (1,061), chronic allograft nephropathy (CAN) (239) (including interstitial fibrosis/tubular atrophy and calcineurin inhibitor toxicity), and AR (434) (including antibody and T-cell mediated rejection, Mixed, and AR+CAN). Next, by taking advantage of a new tool for cell type enrichment analysis xCell (D. Aran et.al, 2017), we identified unique signatures of 45 cell types (p-value < 0.05) for each phenotype. Adopting this as training data we applied machine learning techniques to carry out sample phenotype prediction. This allows us to precisely select samples based on their biological functionality and thus strengthen the overall gene expression signal for a given phenotype. Also, this new approach can be applied as an independent *in silico* validation for a patient diagnosis and warn about possible future allograft rejection. We reserve three independent longitudinal cohorts with 217 samples (STA (157) and others (60)) for validation of our results. Further, we performed a statistical analysis on re-phenotyped samples to identify a more robust gene expression signature associated with acute rejection resulting in several hundred significant genes. Among them there are up-regulated genes that are involved in the immune response, T- and B-cell activation, and cytokine signaling in immune system. The proposed approach for sample control and diagnostic validation offers a robust analytical design for identification molecular signatures in acute kidney rejection, that can lead to better understanding biological mechanisms of allograft injury and increase the graft survival rate.

1281F

Cloud-based quality measurement of whole-genome cohorts. *W. Salerno[1], O. Krasheninina[1], W. Nasser[1], J. Farek[1], A. Mansfield[1], D. Ames[2], A. Carroll[2], F.J. Sedlazeck[1], Z. Khan[1], G. Metcalf[1], D.M. Muzny[1], E. Boerwinkle[1,3], R.A. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) DNAnexus, Mountain View, CA, 94040; 3) Human Genetics Center, University of Texas Health Science Center, Houston, TX, 77030.

Large-scale human disease sequencing programs such as the Centers for Common Disease Genomics (CCDG), Trans-Omics for Precision Medicine (TOPMed), and Alzheimer's Disease Sequencing Project (ADSP) each have generated whole genome sequence (WGS) data sets derived from tens of thousands of phenotyped individuals. These samples also serve as a resource to the Precision Medicine community, providing the foundation for variant and genome annotation, method and technology development, and sequencing best practices. Cost-optimized processing, management and quality assessment remain technical and logistic challenges. We assessed the data quality of more than 25,000 WGS samples, reflecting multiple sequencing protocols and NIH cohorts, a 20x range in sequence coverage, a variety of multiplexing strategies, three sequencing platforms and the effects primary analysis harmonization. Both sample-level and project-level assessments were performed, including generation of GRCh38-specific variant and genome quality control metrics, structural variant (SV) yield-sensitivity to these metrics, and replicate-based protocol comparisons. All analyses utilized updated workflows deployed on the DNAnexus platform, including an NIH-compliant FASTQ-to-CRAM workflow optimized with Sentieon software, >200 FASTQ and CRAM metrics with the HGSC's AlignStats app, and a multi-method variant calling method parallelized in a single application, Parliament (PMID:25886820). This infrastructure allows for rapid and inexpensive harmonization of other data sets with existing NIH cohorts. These assessments also leverage project-specific features to isolate data bias that can confound downstream analysis. For example, in a multi-institution study, replicate-based SV analysis of 584 WGS samples across 111 families showed strong platform and institution batch effects, and removal of protocol-specific events (49%) did not resolve the heterogeneity. However, a novel PCA replicate harmonization identified a smaller SV set (17%) that did resolve the batch effects when removed: the remaining data properly clustered the replicate samples and reflected the underlying family structure. Taken together, these results and infrastructure describe a scalable model for WGS data harmonization.

1282W

Ultra-rapid detection of identity-by-descent tracts for biobank-scale inference. *R. Shemirani[1], G.M. Belbin[2], E.E. Kenny[2], C.R. Gignoux[3], J.L. Ambite[1].* 1) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Colorado Center for Personalized Medicine, University of Colorado, Aurora, CO.

The identification of identity-by-descent (IBD) tracts is a standard step in many pipelines for genome-wide studies in both population and medical genetics. With novel mega-scale initiatives ongoing from biobanks and nation-wide resources there is an ongoing need for efficient methods to detect these IBD segments for identification of cryptic relatedness, population linkage, and demographic inference. While IBD detection has been a topic of vigorous development, existing algorithms scale poorly, exhibiting emergent quadratic or higher order run time as study size becomes large and IBD segments become commonplace. This makes integration into large studies such as across health systems or national biobank initiatives intractable. Here, we present a novel algorithm for mega-scale IBD detection via partitioning and locality sensitive hashing, allowing close-to-linear time scaling and enabling IBD segment detection genome-wide in studies of hundreds of thousands of samples and beyond. Our algorithm, called I-LASH (Ibd by LocAlity Sensitive Hashing), provides a pipeline-ready approach to IBD detection, uses standard file formats, and contains flexible tuning parameters for arrays or sequencing. By leveraging parallelism and our computationally efficient design I-LASH allows entire mega-scale studies to be analyzed on a standard multi-core machine. We benchmark I-LASH on simulated datasets and across reference datasets including the ethnically diverse >51,000 individuals genotyped in the PAGE-II Consortium. While maintaining similar precision and recall metrics to the standard GERMLINE method, I-LASH provides a >40-fold speedup compared to GERMLINE with 50,000 samples. We anticipate that I-LASH will empower IBD inference for the next generation of medical and population genetic studies.

1283T Unprogrammed presentation number

1284F

Collective feature selection to identify important variables for epistatic interactions. *S. Verma[1,2], A. Lucas[2], X. Zhang[1], Y. Veturi[2], S. Dudek[2], B. Li[1], R. Li[3], D. Kim[2], M. Ritchie[2].* 1) Huck Institute of Life sciences, The Pennsylvania State University, University Park, PA; 2) Biomedical and Translational Bioinformatics Institute, Geisinger Health System, Danville, PA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Machine learning (ML) methods have gained popularity and practicality in identifying linear as well as non-linear effects of genetic variants associated with complex disease traits. Genotype data (used as input variables for ML methods) generally consists of ~100,000 or more common variants (MAF>20%) after quality control. Detection of epistatic interactions for 100K or more variants still remains a challenge, thus leading to the so-called "curse of dimensionality" problem. The efficiency of ML methods can be increased by limiting the number of input features. Thus, it is essential to perform variable selection before searching for epistasis. Many methods have been previously proposed to perform feature selection; I but all methods do not follow "one method fits all" approach. We tested a wide range of parametric, non-parametric, and data mining approaches to select variables with epistatic effects in a simulation study. In our simulated data analysis, we found that data mining approaches like Multifactor Dimensionality Reduction (MDR) performs best in identifying true positives with epistatic as well as additive effects but in other published simulation studies, it is observed that other parametric and nonparametric methods select features with smaller type 1 error rates. Thus, we propose a collective feature selection approach to select variants that are in the "union" of the best performing methods. Since the scoring algorithms for all methods are different, we propose the use of a user-defined percentage of top-ranking features from all methods and then proceed with these collectively selected features for downstream analysis. In this study, we demonstrate in extensive simulation datasets that this collective feature selection method maintains high power to detect the true positive signals in the simulated datasets. We present collective feature selection as a powerful approach in identifying biologically pertinent epistatic features for discovery analysis.

1285W

PipelineDog: A simple and flexible graphic pipeline construction and maintenance tool. *J. Xing[1], A. Zhou[1], Y. Zhang[1], Y. Sun[2].* 1) Dept of Genetics, Human Genetic Institute of New Jersey, Rutgers Univ, Piscataway, NJ; 2) New Jersey Center for Science, Technology & Mathematics, Kean University.

Data manipulations and analyses are essential components of bioinformatics research. An analysis pipeline defines an analysis workflow by specifying the order and configuration of various analysis tools, and their input/output data. Although pipelines can be coded in a scripting language such as BASH or python, such scripts usually suffer from limited readability and poor reusability. For enhanced reusability, a design that allows clear decoupling and easy reassembly of individual steps is needed. Several attempts have been made to address these challenges, such as the Common Workflow Language (CWL) and Workflow Description Language (WDL). However, most existing pipeline management system or workflow engines are too complex for rapid prototyping needs, and a lightweight, user-friendly, and flexible solution is thus desirable. We developed a pipeline construction and maintenance tool, PipelineDog (http://pipeline.dog/), to deliver both simplicity and flexibility. PipelineDog is a comprehensive web-based integrated development environment (IDE) with a modern web graphical user interface (GUI). It offers cross-platform compatibility, project management capabilities, code formatting and error checking functions, and online repositories. It uses an easy-to-read/write script system that encourages code reuse, and with an online repository, it also encourages sharing of pipelines, which enhances analysis reproducibility and accountability. For most users, PipelineDog requires no additional software installation. Overall, this system provides a way to rapidly create and easily manage *ad hoc* pipelines.

1286T

Bayesian multiple eQTL detection with control for population structure and sample relatedness. *B. Zeng, G. Gibson.* Georgia Institute of Technology, Atlanta, GA.

Expression quantitative loci, or eQTL, are widely used as a tool to interpret GWAS results. There is increasing evidence that gene expression is often controlled by multiple cis-eQTLs (Wen et al, 2015; Hormozdiari et al, 2017). A battery of recent methods, frequentist and Bayesian, are being developed to detect such multiple eQTL, and there is some evidence that the latter (CAVIAR, FMQTL, DAP, FINEMAP) have increased power to resolve independent effects of closely linked sites. However, most of the available packages ignore the effects of sample relatedness due to population and family structure. Ignoring these phenomena can be detrimental, as they may lead to false positives or loss of resolution when fine-mapping the multiple causal eQTL variants. Here we describe MeQTLPolyG (*Multiple eQTL detection controlling for PolyGenic background*), a package for identifying and estimating complex eQTL effects while controlling for potential population structure and relatedness. This package first performs sequential conditional analysis to detect independent peak signals, then integrates a genetic relatedness matrix to remove the influence of structure. For each peak signal, a Bayesian-based multiple eQTL detection pipeline is used to estimate the number of causal variants and choose most plausible candidate list. We apply the method to two large blood gene expression data sets, CAGE and Framingham, both of which include relatives and population diversity, and quantify the increased detection of multiple cis-eQTL.

1287F

Robust and accurate estimation of DNA sample contamination agnostic to genetic ancestry. *F. Zhang[1,3], H.M Kang[2,3].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI; 3) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Detecting and estimating DNA sample contamination has been established as a critical step to ensure high quality sequence reads and reliable downstream analysis. Our verifyBamID software tool has been run over 2 million samples to date from multiple sequencing centers. The current method relies on externally provided allele frequency information to estimate estimation of contamination levels, and misspecified allele frequencies result in substantial bias in contamination estimates. Our experiment demonstrates that current methods may fail to screen highly contaminated samples (~10%) at a stringent contamination threshold (~3%) when the genetic ancestry is misspecified. However, correctly specifying population allele frequencies in early stage of sequence analysis is cumbersome or sometimes even impossible for large-scale sequencing centers that simultaneously process samples from multiple studies across diverse populations. We propose a robust statistical method that accurately estimates DNA contamination agnostic to genetic ancestry of the intended or contaminating samples. Our method integrates the estimation of genetic ancestry and DNA contamination in a unified likelihood framework by leveraging individual-specific allele-frequencies projected from reference genotypes onto principal component coordinates. As a result, our method not only estimates contamination accurately, but also estimates genetic ancestry of both intended and contaminating samples. Our experiments show that our method robustly corrects for the bias in both contamination level estimates and genetic ancestry estimates under different scenarios of contamination. In the experiment with *in-silico* contaminated samples constructed from the 1000 genomes samples, our new method resulted best estimation accuracy and least bias compared to the current method across all different settings. With current method, the contamination levels were most severely underestimated for African American samples, and overestimated when intend and contaminating samples originate from different continental populations. In experiment with real dataset, we observed that that 18 of 1,000 deeply sequenced genomes are marked as contaminated with new method at 3% threshold, even though all of them were estimated to have <3% contamination with the current methods. Our method is publicly available at https://github.com/statgen/verifyBamID2.

1288W

Predicting exposure to ionizing radiation by biochemically-inspired genomic machine learning. *J.Z.L. Zhao, E.J. Mucaki, P.K. Rogan.* Biochemistry, University of Western Ontario, London, Ontario, Canada.

Analyzing gene expression in peripheral blood mononuclear cells reveals profiles that predict radiation exposure in humans and mice by logistic regression (PLoS Med. 4:e106; PLoS ONE. 3:e1912). Using biochemically-inspired methods (Mol. Onc. 10:85-100), we derive gene signatures to predict the level of radiation exposure with improved accuracies. DNA repair genes responsive or differentially expressed upon radiation exposure and orthologs highly expressed in species resilient to radiation exposure (n=998) were analyzed by two-sampled t-tests comparing expression in individuals unexposed and exposed to radiation (150-200 cGy: humans or 50-1000 cGy: mice). Significance thresholds for including a gene in developing a signature were adjusted based on radiation dose, from $p < 0.01$ (50 cGy) to < 1E-14 (1000 cGy), equivalent to ~10% of genes. Support Vector Machine (SVM) signatures were derived by backward feature selection (BFS) or minimum-redundancy-maximum-relevance (mRMR) and validated using leave-one-out cross validation (LOOCV) and external datasets. GEO datasets GSE6874 and GSE10640 were used for training and testing. Signatures derived by BFS from the human patients of GSE6874 (n=78) included α) *GADD45A, GTF3A, TNFRSF4, XPC* and β) *ATR, GADD45A, GTF3A, IL2RB, MYC, NEIL2, RBM15, SERPINB1, XPC*, which both distinguished irradiated from unirradiated individuals with 98% sensitivity and 100% specificity in LOOCV. Validating these signatures on the human patients of GSE10640 (n=71) confirms that α and β are both 92% sensitive and, respectively, 94% and 96% specific. mRMR found the 10 "best" genes from the murine samples of GSE10640 (n=104) to create a signature at each radiation dose; several genes were common among signatures. Signature δ (50 cGy) included *PHLDA3, BAX, NBN, CCT3, CDKN1A, CCNG1, POLK, ERCC5, GCDH,* and *RAMP1*. Signature ε (200 cGy) included *PHLDA3, LIMD1, CCT3, BAX, MS4A1, GLIPR2, BLNK, BCAR3, CDKN1A,* and *TFAM*. Signature ζ (1000 cGy) included *CCT3, SUCLG2, EI24, CNBP, PHLDA3, TPST1, HEXB, FEN1, CDKN1A,* and *BLNK*. When validated on the murine samples of GSE6874 (n=14), each signature correctly predicted the exposure status of all mice. Our approach produces signatures with higher accuracies in cross- and external validation datasets than prior logistic regression models, with significantly improved sensitivities in detecting radiation exposure in humans. This will be useful in identifying nearly all radiation-exposed individuals in a mass casualty.

1289T

Leveraging allele-specific expression to refine fine-mapping for eQTL studies. *J. Zou[1], F. Hormozdiari[2], J. Ernst[1], J. Sul[1], E. Eskin[1].* 1) University of California, Los Angeles, Los Angeles, CA; 2) Harvard University, Boston, MA.

Many disease risk loci identified in genome-wide association studies are present in non-coding regions of the genome. It is hypothesized that these variants affect complex traits by acting as expression quantitative trait loci (eQTLs) that influence expression of a nearby gene. This indicates that many causal variants for complex traits are likely to be causal variants for gene expression. Hence, identifying causal variants for gene expression is important for elucidating the genetic basis of not only gene expression but also complex traits. However, detecting those variants is challenging due to complex genetic correlation among variants known as linkage disequilibrium (LD) and the presence of multiple causal variants within a locus. Although several fine-mapping approaches have been developed to overcome these challenges, they may produce large sets of pu- tative causal variants when true causal variants are in high LD with many non-causal variants. In eQTL studies, there is an additional source of information for fine-mapping called allele-specific expression (ASE) that measures imbalance in gene expression due to different alleles. In this work, we develop a novel statistical method that leverages both ASE and eQTL information to detect causal variants that regulate gene expression. We illustrate through simulations and application to the Genotype-Tissue Expression dataset that our method identifies the true causal variants with higher specificity than an approach that uses only eQTL information.

1290F

The role of CTCF and cohesin complex in chromatin looping and higher-order organization of human genome. *D. Plewczynski[1], M. Kadlof[1], P. Szałaj[1], M. Lazniewski[1], M. Sadowski[1], Z. Tang[2], Y. Ruan[2].* 1) Centre of New Technologies, University of Warsaw, Warsaw, Mazovia, Poland; 2) Jackson Laboratory for Genomic Medicine, CT, USA.

The recent genomic and bioimaging insights in the higher order chromatin organisation in human nucleus motivated us to propose novel simulation method aimed at the effective prediction of three-dimensional human genome structure. We applied an advanced long-read Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) experimental strategy [1] combined with computational modelling [2] to comprehensively map higher-order chromosome folding and specific chromatin interactions mediated by CCCTC-binding factor (CTCF), cohesin and RNAPII with haplotype specificity and nucleotide resolution in different human cell lineages [3]. We demonstrate the effectiveness of biophysical modeling in building 3D genome models at multiple levels, including the entire genome, individual chromosomes, and specific segments at megabase (Mb) and kilobase (kb) resolutions of single average and ensemble structures. We find that CTCF/cohesin-mediated interaction anchors serve as structural foci for spatial organization of constitutive genes concordant with CTCF-motif orientation, whereas RNAPII interacts within these structures by selectively drawing cell-type- specific genes towards CTCF-foci for coordinated transcription. Furthermore, we show that haplotype-variants and allelic-interactions have differential effects on chromosome configuration influencing gene expression and may provide mechanistic insights into functions associated with disease susceptibility [1]. 3D-genome simulation suggests a model of chromatin folding around chromosomal axes, where CTCF is involved in defining the interface between condensed and open compartments for structural regulation [2]. Our 3D-genome strategy thus provides unique insights in the topological and thermodynamical mechanisms [4] of human variations and diseases. **References** [1] "CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription" Tang Z, Luo OJ et al., Plewczynski D, Li G, Ruan Y. Cell. 2015 [2] "An integrated 3-Dimensional Genome Modeling Engine for data-driven simulation of spatial genome organization" Szałaj P, et al., Ruan Y, Plewczynski D. Genome Res. 2016 [3] "3D-GNOME: an integrated web service for structural modeling of the 3D genome" Szalaj P, et al., Ruan Y, Plewczynski D. Nucleic Acids Res. 2016 [4] "Free energy based high-resolution modeling of CTCF-mediated chromatin loops for human genome" Dawson W and Plewczynski D (bioRxiv 2017).

1291W

Weighted gene co-expression network analysis using peripheral blood of patients with 22q11.2 deletion syndrome. *A.G. Dantas[1], M.L. Santoro[2], C.A. Kim[3], D.C.Q. Soares[3], V.A. Meloni[1], N.R. Nunes da Silva[1], S.I. Belangero[1,2], G.G. Carvalheira[1], M.I. Melaragno[1].* 1) Department of Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, SP, Brazil; 2) Department of Psychobiology, Universidade Federal de São Paulo, São Paulo, SP, Brazil; 3) Genetics Unit, Universidade de São Paulo, São Paulo, SP, Brazil.

22q11.2 deletion syndrome (22q11.2DS) results from hemizygous deletions of chromosome 22, being the 3 Mb deletion between the LCR22A-D the most frequent one. Even though most patients have the same size deletions, the phenotype is highly variable among individuals and different mechanisms has been described to justify this variability. During the years, the genes responsible for clinical conditions have been delineated from gene expression data sets, where differentially expressed genes are correlated to the phenotype. Recently, a new system biology approach has been developed to analyze gene expression data. The WGCNA (weighted gene co-expression network analysis), a *R* analytical package, identify groups of probes that have similar expression in the sample and produces a set of modules. Each module is a group of genes with strongly shared co-expression relationships. The WGCNA analysis on microarray gene expression data set from 11 22q11.2DS patients and 14 controls identified a total of 50 modules of different sizes in terms of gene number, which were labelled by different colors according to WGCNA package functions, and a gene ontology (GO) enrichment analysis was performed. Eight modules were correlated with the patients' phenotype (correlations ranging from -0.51 to 0.96, p<0.03). The tan module was enriched in neuronal functions and cellular processes such as nucleolus organization and DNA repair; the turquoise module in neuron development, differentiation and neurotransmitter secretion; plum2, in protein modification processes and negative regulation of glutamate metabolic process; white, in uridine and borate transport and regulation of cell motility; thistle1, in cellular processes and cell differentiation; royal blue, in cerebellar cortex morphogenesis, telomere maintenance and organization; magenta, in immune cell response, and red module was enriched in regulation of imprinting and methylation, cell-substrate and cell-matrix adhesion, mesenchymal cell development and GABA signaling pathway. Our data show that the WGCNA is able to indicate the enrichment of GO terms associated to the pathology observed in patients with 22q11.2DS, providing the possibility of searching for genetic modifiers that were not previously observed in gene expression studies. The study of gene networks allowed us to do more than single gene comparisons, but also identify relationships between genes, pathways and phenotypes. Financial support: FAPESP, Brazil.

1292T

Showing your work: Combining genetic variant interpretations with evidence to enable reanalysis and reuse. *B.C. Powell[1], M. Brush[2], M. DiStefano[3], R.R. Freimuth[4], R. Ghosh[5], M.A. Haendel[2], S. Harrison[6], T. Nelson[7], H.L. Rehm[3,6], S. Rynearson[8], N. Strande[1], C. Bizon[9], L. Babb[10], ClinGen Consortium Data Model Working Group and the Monarch Initiative.* 1) Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR; 3) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medcine, Cambridge MA; 4) Health Sciences Research, Mayo Clinic, Rochester MN; 5) Pediatrics-Oncology, Baylor College of Medicine, Houston, TX; 6) Harvard Medical School, Boston MA; 7) Geisinger Autism and Developmental Medicine Institute, Lewisburg PA; 8) Biomedical Informatics, University of Utah, Salt Lake City UT; 9) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill NC; 10) GeneInsight, a SunQuest company, Boston MA.

Determining the clinical relevance of a genetic variant requires detailed evaluation of diverse data. Resources that share simple statements of variant pathogenicity, alone or in combination with unstructured metadata, do not provide sufficient detail for reanalysis in light of new information or for adjudication of discordant interpretations. While the recently published ACMG-AMP guidelines encourage a more systematic assessment of pathogenicity, they fundamentally rely on the synthesis and evaluation of the evidence for pathogenicity in the context of individual criteria. Comparisons of interpretations performed by different laboratories demonstrate that there is still considerable variability in how these criteria may be applied. Further, our knowledge is constantly changing through updating of existing disease and molecular resources, collection of new data and data types, and redefinition of diseases and phenotypes. It is therefore necessary to develop a robust model of evidence that can be leveraged to provide for consistent interpretation of variant pathogenicity and to rapidly incorporate new knowledge in real time. The Clinical Genome resource (ClinGen) and the Monarch Initiative have created a robust informatics foundation for tools to support knowledge integration and interoperability regarding verifiable and updatable interpretations of variant pathogenicity. ClinGen provided domain knowledge in the methods and evidence types used for variant interpretation. This included development of structured representations of over 95 examples documenting the application of the ACMG-AMP guidelines. The Monarch Initiative contributed the Scientific Evidence and Provenance Information Ontology (SEPIO), a generalizable semantic model for linking scientific assertions to underlying lines of evidence and provenance metadata. SEPIO provides a unified and flexible representation of evidence for assertions that supports data integration across diverse biomedical databases. The result is a structured but practical approach to representing variant pathogenicity assertions with evidence. This approach addresses the imperative for structured and granular sharing of variant interpretations and supporting evidence.

1293F

Supervised enhancer prediction with epigenetic pattern recognition and targeted validation across organisms. *M. Gu[1], A. Sethi[1,2], E. Gumusgoz[6], L. Chan[3], K. Yan[1,2], K. Yip[4], J. Rozowsky[1,2], R. Sutton[6], L. Pennacchio[7], M. Gerstein[1,2,5].* 1) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 2) Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 3) School of Medicine, The Chinese University Hong Kong, China; 4) Department of Computer Science, The Chinese University Hong Kong, China; 5) Department of Computer Science, Yale University, New Haven, CT; 6) Department of Internal Medicine, Section of Infectious Diseases, Yale University School of Medicine, New Haven, CT; 7) Functional Genomics Department, Lawrence Berkeley National Laboratory, Berkeley, CA.

Cis-regulatory elements(CREs) are difficult to characterize both experimentally and computationally. With only a few validated mammalian enhancers, it is difficult to properly train statistical models for their identification. Instead, postulated patterns of genomic features were used heuristically for prediction. More recently, a large number of massively parallel assays for characterizing enhancers have been developed. Here, we use them to create shape-matching filters based on enhancer-associated metaprofiles in epigenetic features. We then combine different features with simple, linear models and predict enhancers in a supervised fashion. Using in vitro transduction-based reporter assay in human H1 hESC cell line, as well as in vivo transgenic mouse enhancer assays in 6 different tissues, we show that our model is able to accurately predict enhancers in different species without re-parameterization between cell lines and even between organisms. Finally, we predict enhancers in cell lines with many transcription-factor binding sites. In turn, this highlights distinct differences between the type of binding at enhancers and promoters, enabling the construction of a secondary model discriminating between these two.

1294W

Novel high-resolution multi-ethnic HLA imputation reference panels constructed based on high-coverage whole-genome sequencing data. *Y. Luo[1,2], M. Gutierrez-Arcelus[1,2], A. Metspalu[3], J.G. Wilson[4], S. Kathiresan[5], T. Esko[2,3], S. Raychaudhuri[1,2], the TOPMed project.* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 4) Physiology and Biophysics, University of Mississippi Medical Center; 5) Center for Genomic Medicine, Massachusetts General Hospital.

The human leukocyte antigen (HLA) region harbors genes that are crucial to a wide range of human diseases. However it has been challenging to identify and fine-map HLA associations from GWAS data. HLA imputation approaches have been restricted by limitations in reference panels to European populations and to amino acid (4-digit) resolution. Here, we infer HLA types from large scale whole-genome sequencing (WGS) data, and built the first multi-ethnic high resolution HLA reference panels. These references can be applied to a wide range of worldwide data sets, where HLA imputation was not previously possible. The computational strategies that we present can also be applied in any high coverage WGS data, thereby allowing users to build their own HLA reference panels. First, we present a computational strategy that takes advantage of the recent rapid development of WGS technology to infer a set of alleles that encode the same protein sequence for the peptide binding sequence of an HLA molecule (i.e., 6-digit). We applied this strategy to 2,244 Estonian and 3,027 African-American WGS samples with coverage ≥25x. We observed that the frequencies of inferred HLA alleles are highly consistent with publicly available data within the same population group (Pearson r^2 of 0.96 and 0.98 respectively). Next, we constructed a 4- and 6- digit HLA reference panel of the 5,271 WGS samples, and a 4-digit panel combining these with samples from a perviously published 4-digit panel based on data from 5,225 individuals in the Type 1 Diabetes Genetics Consortium (T1DGC). To evaluate performance of these panels, we imputed HLA alleles with each of them into a multi-ethnic cohort of 2,291 individuals genotyped on the ImmunoChip array. The average imputation r^2 at 4-digit resolution is 0.96 compared to 0.90 using the T1DGC panel. For individuals of African descent, r^2 improved to 0.92 in the combined panel. To further assess the accuracy of the imputed HLA types we analyzed data obtained from 26 individuals by the gold-standard sequencing method. Of 520 alleles tested in the five classical HLA genes (HLA-A, B, C, DQA1, and DRB1), 100%, 96.9% and 94.2%, respectively, were correctly imputed at 2-, 4- and 6-digit resolution. We subsequently demonstrate how imputation and association testing at 4- and 6-digit resolution could facilitate discovery and fine-mapping of HLA association signals, giving specific examples from psoriasis in 50K individuals from the Estonian Biobank.

1295T

Detecting copy number variants in 200,000 individuals: The Department of Veterans Affairs Million Veterans Program (MVP). *M. Li[1,2], Y. Shi[1], S. Pyarajan[1,3] on behalf of the VA Million Veteran Program.* 1) VA Boston Healthcare System, Boston, MA; 2) Division of Nephrology, University of Utah School of Medicine, Salt Lake City, UT; 3) Harvard Medical School, Boston, MA.

Copy number variants (CNVs) are one of the most important genomic sources of variability that have not been comprehensively evaluated for association with complex diseases or human traits in the GWAS era. The Million Veteran Program (MVP), sponsored by the Department of Veterans Affairs, is partnering with Veterans to establish one of the world's largest databases of genomic, clinical, lifestyle and military exposure information. With the goal of enrolling one million participants, MVP has enrolled over 570,000 Veterans as of June 2017. In phase 1 200,845 samples have been genotyped using the Affymetrix Axiom MVP 1.0 Custom array and data quality assessment completed. In order to systematically assess the contribution of CNVs to complex diseases or traits, we aim to develop a map of human CNVs informed by hybridization intensity data generated from the Affymetrix Axiom MVP 1.0 array. CNV calls were obtained using Affymetrix Power Tools and PennCNV software. We "trained" a hidden markov model (hmm) file on 96 HapMap samples genotyped on the MVP 1.0 array, using the available Affymetrix Axiom hmm file as a template in PennCNV. The reference CNV database combined previously identified CNV regions from the HapMap3 project, the 1000 Genomes Project, and the Exome Aggregation Consortium (ExAC). We examined the population frequency of CNVs that are a) within the reference CNVs; b) outside of the reference CNVs. In total, we identified 4,665 autosomal CNVs, 50% of them reside within the reference CNVs. Among 2,334 novel CNVs, 14 are common genic CNVs (≥ 2% frequency) and 2,283 are rare genic CNVs (<0.5% frequency). Our results open the door for new opportunities such as conducting genome-wide association study for CNVs or burden of rare CNVs with different diseases or traits in MVP. We plan to extend the CNV map with genotype data from the entire MVP cohort.

1296F

A graph-based Arab reference genome using whole read overlap assembly. *Y. Mokrab[1], A.C. English[2], J. Bruestle[2], S.N. Shekar[2], K.A.M. Fakhro[1].* 1) Systems Biology and Bioinformatics, Sidra Medical and Research Centre, Doha, Qatar; 2) Fabric Genomics, 1611 Telegraph Avenue, Suite 500, Oakland, CA 94612, USA.

Recent exploration of ancestral populations' genomes has shown that the widely used Human reference genome (GRCh37/GRCh38) lacks the diversity needed to represent populations of specific anthnicities (Fakhro et. al. 2016. Hum Genome Var. 3:16016). By incorporating Single Nucleotide Variants (SNVs) and indels found by alignment-based variant calling into GRCh37, a population-representative reference can be created which increases the proportion of alignable reads from an individual of the same population. Incorporating structural variants (SVs) would further improve alignment. Here, we look at the prevalence of SVs in a cohort of 100 individuals with Qatari ancestry to further enhance the Arab reference genome (Fakhro et. al. 2016. Hum Genome Var. 3:16016) with SVs having a high population frequency. The detection of SVs with short-read data is difficult due to inaccurate breakpoints, unresolved insertion sequences and high false discovery rate (English AC, et. al. 2015. BMC Genomics. 16:286). To characterise SVs in this cohort, we run the BioGraph variant caller - a graph-based caller combining extended BWT and FM index to create a graph structure of the reads while allowing for rapid query of all possible assembly paths. Using this graph-genome representation, BioGraph detects SVs with high sensitivity, low false discovery rate and exact sequence of breakpoints and insertions (Parikh H et al 2016. BMC Genomics. 17:64). In a study of SVs within a family using BioGraph, over 97% of the detected SVs were shown to follow mendelian inheritance (Shekar et al, 2016 unpublished). In a preliminary analysis, the BioGraph variant caller was run on the caucasian HapMap individual NA12878 and two individuals of Arab ancestry (SRR2098185 and SRR2098177). Approximately 4,000 SVs were identified in each of these individuals. Of these, 104 (2.6%) were Arab-specific, identical between the Arab individuals but absent from NA12878. These results suggest this method enables accurate, reproducible SV detection across samples. Here, we will show characteristics of SVs that have a high allele frequency in the Qatari population as well as comparisons of BioGraph results against other graph-based and linear SV detection methods, including long-read single molecule sequencing.

1297W

DUP-OE: A new tool to discovery the origin and expansion of duplication. *X. Zhuang.* HKU, HK, Hong Kong.

Copy number variation (CNV), including deletion and duplication, has been proved to be associated with a lot of human diseases. Researchers have developed lots of tools to detect CNV. Read-Depth, Pair-end and Split-read methods are the widely used priciple for discovering CNV. The detection of deletion is easier than that of duplication, and the concordance of deletion among different tools are quite high. However, there exists high preference in duplication when using different tools, as the derivation of duplication is complex, which includes ectopic recombination, replication slippage and retrotransposition. Morever, the detection of duplication on complex genomic region, such as repeat and GC-content unusual regions might be more challenge. Most of the current tools only give the results of the duplicated locations. However, in most cases, the duplicated genes would not just stand next to the duplication region. On the contrary, it may inserted to other genomic regions. In these cases, if we have not realised where the duplicated gene expanse, we might just focus on the function discovery of the original region. However, the duplicated region might be normal, and it is the expansion region that been disrupt and causing certain kinds of diseases. Based on the above reasons, we develop a new method: DUP-OE, to discovery both the origin and the expansion of duplication. It can be used to detect the CNV based on the whole genome sequence. DUP-OE performs well on both the simulated data and the real clinical samples.

1298T

Illuminating the Druggable Genome (IDG): An NIH Common Fund program. *J. Baker[1], C. Colvis[2], A. Pawlyk[1].* 1) National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, M.D. 20892; 2) National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, M.D. 20892.

The Pilot Phase of the **Illuminating the Druggable Genome (IDG)** was launched as an NIH Common Fund program in 2014. The pilot program supports a Knowledge Management Center (KMC) and a consortium of 7 projects for the Adaptation of Scalable Technologies to Illuminate the Druggable Genome (TechDev). The Pilot Phase was intended to provide feasibility of the approaches to identify and study understudied **G protein-coupled receptors**, **protein kinases**, **ion channels,** and **nuclear receptors**. The Implementation Phase of the IDG is expected to launch in 2018 and run through 2023. The NIH has issued RFAs for **Data and Resource Generation Centers** (RFA-RM-16-026), a **Knowledge Management Center** (RFA-RM-16-024), and a **Resource Dissemination and Outreach Center** (RFA-RM-16-025). Applications were due on March 14th, 2017 and awards will be made in late 2017. The experimental arms of the Implementation Phase of the IDG will focus on understudied members of the **G protein-coupled receptors, ion channels,** and **protein kinases** while informatics efforts will extend across the proteome. Its overall goal will be to generate, aggregate, analyze, and disseminate knowledge and tools around understudied proteins. These new knowledge and tool sets are intended to equip the scientific community, including small businesses and the pharmaceutical industry, with the ability to explore previously understudied biology with the potential to rapidly impact human health. .

1299F

Enhancing discoverability and reusability of 1.2 million human genomic datasets. *M. Corpas, N. Kovalevskaya, R. Shaw, C. Whicher, A. Alexa, F.G.G. Nielsen.* Repositive Ltd, Cambridge, United Kingdom.

Publicly-funded human genome studies require their data to be catalogued and deposited in internationally recognized repositories. As sequencing data generation explodes, a rising number of data repositories are being created to accommodate the increasingly divergent data types and regulated sharing environments. This makes finding and accessing genomic datasets more cumbersome for users, who ultimately may choose to generate their own data rather than reusing the existing data (even if that means duplicating efforts). Repositive (https://repositive.io/?ASHG), a free online portal for finding and accessing human genomic datasets, is a centralized platform currently indexing over 1.2M datasets from 47 data sources. Repositive harbors a community of >2000 users, actively engaged in crowdsourcing the annotation of registered datasets via Repositive's powerful social networking features. By following or favoriting Repositive's datasets, users can track a dataset's user activity and connect with other users interested in the same dataset, hence seeding potential new collaborations. To facilitate the search, datasets are further organized into collections, which aggregate meaningful dataset classifications and specific search criteria according to data types: e.g., population studies, microbiomes, methylomes and other NGS-derived data of human origin. By using the Repositive platform, researchers are able to find all publicly available (restricted or open access) human genomic datasets, acquire a global overview of the genomic data landscape and forge new collaborations, drastically enhancing their dataset discoverability and reusability for further research advancement.

1300W

Semantic web technology accelerates integration of genetic and phenotypic information in biomedical databases. *T. Katayama, S. Kawashima, Y. Yamamoto.* Database Center for Life Science, Kashiwa, Chiba, Japan.

The Database Center for Life Science is a Japanese national institute for the development of technologies on database standardization and integration. We have been developing an integrated genome database, named as TogoGenome, which incorporates genome and gene/protein annotations with the Semantic Web technology. To effectively integrate genetic and phenotypic information in the genome database, we have developed Resource Description Framework (RDF) version of major public biomedical databases such as MedGen, ClinVar and ExAC, where RDF is a standard for data exchange in the Semantic Web which is standardized by the World Wide Web consortium (W3C). While a lot of important biomedical databases are available, integrated use of them is still hard because of lacking in standards for data formats, identifiers and common vocabularies among them. This has laid a heavy burden on both database developers and users working in research institutes and pharmaceutical companies. RDF can effectively resolve the integration issues and also provide the semantics of original data explicitly with the help of ontologies. Once the databases are represented as RDF data, those are interlinked and also linked to external RDF resources. In the case of biomedical data, genetic variation can be seamlessly linked to genomes, transcripts, proteins, diseases, drugs, pathways and literature information. To date, most of major life science databases are already provided or are being developed as RDF datasets such as 1) genome annotations (Ensembl), 2) DNA sequences and annotations (GenBank/ENA/DDBJ and RefSeq), 3) protein sequences and annotations (UniProt), 4) protein structures (PDB), 5) chemical and drug related information (PubChem, ChEMBL, Nikkaji) 6) biological pathways (Reactome), 7) gene expressions (Expression Atlas, RefEx), 8) taxonomy, 9) literature (PubMed, MeSH), and omics data including cancer genomics, toxico genomics, orthologies, epigenetics, proteomics, glycomics along with many biomedical ontologies mainly hosted at the National Center for Biomedical Ontologies (NCBO). To fully utilize those kinds of valuable information for human genetics, here we report our work on RDF versions of the biomedical databases which are going to be integrated to the TogoGenome application. The application and the RDF datasets are publicly available so that biomedical researchers and application developers can make use of them without suffering from incompatibilities.

1301T

Controlled-access databases for genetic and phenotypic human data in Japan. *Y. Kodama[1], A. Fukuda[1], J. Mashima[1], M. Minowa[2], M. Kawashima[2], K. Miyazaki[2], N. Mitsuhashi[2], E. Kaminuma[1], O. Ogasawara[1], K. Okubo[1], Y. Nakamura[1], T. Takagi[1,2].* 1) DDBJ Center, National Institute of Genetics, Mishima, Shizuoka, Japan; 2) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo, Japan.

The Japanese Genotype-phenotype Archive (JGA, http://trace.ddbj.nig.ac.jp/jga) serves as one of centralized repositories providing authorized access to individual-level phenotype and genotype data in collaboration with our partner institute, the National Bioscience Database Center (NBDC, http://humandbs.biosciencedbc.jp) of the Japan Science and Technology Agency. JGA is the Japanese counterpart of the dbGaP (database of genotypes and phenotypes) and EGA (the European Genome-phenome Archive) at NCBI and EBI, respectively. NBDC in Japan established guidelines and policies compliant with the amended Act on the Protection of Personal Information and ethical guidelines fully put into effect on May 30, 2017. The Data Access Committee (DAC) at NBDC reviews and makes decisions about data submission to JGA and usage requests from researchers. JGA accepts metadata in which personal identifiers have been removed by submitters. Acceptable data types include raw and processed data from array-based and next-generation sequencing platforms, clinical images, and phenotype data associated with data samples. Upon receipt, JGA archives the original data files in encrypted form in the database. JGA data are organized in the hierarchical JGA data model based on that of EGA. JGA assigns stable, unique identifiers prefixed by 'JGA' to studies and subsets of metadata from those studies. Approved user can download permitted datasets with secure transfer software after granted by DAC. To enhance collaboration by providing workspace for sharing and analyzing pre-publication data, we have started a new service 'DDBJ Group Cloud (DGC)' since February 2017. As the first use case, we provide 'AMED Genome Sharing Database (AGD)' for sharing pre-publication human genomic data from AMED-funded researches. Because AGD and JGA use the same data model, users can smoothly transfer the AGD data to JGA when their papers are published. Summaries of available studies are listed on the JGA (https://ddbj.nig.ac.jp/jga/viewer/view/studies) and NBDC websites. As of 31 May 2017, 53 studies are available at JGA. In the poster, we present contents, system and recent developments of JGA.

1302F

DASHR 2.0: Database of small non-coding RNAs in normal human tissues and cell types. *P.P. Kuksa[1,2], Y.Y. Leung[1,2], A. Amlie-Wolf[1,2,3], O. Valladares[1,2], L.-S. Wang[1,2,3].* 1) Penn Neurodegeneration Genomics Center, University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA.

DASHR is a database developed at the University of Pennsylvania with the most comprehensive information on ~48000 human small non-coding RNAs (sncRNA) genes and mature sncRNA products in 187 smRNA high-throughput sequencing (smRNA-seq) datasets from multiple public sources. We present the updated DASHR 2.0 database with an improved preprocessing workflow, more samples and tissue types, more sncRNA classes and availability in both hg19 and hg38 genome builds. Beyond small RNA-seq data sets in DASHR 1.0, we have incorporated >150 high quality short total RNA-sequencing and miRNA-sequencing data sets from ENCODE, as well as ~500 single cell small RNA-seq data into DASHR 2.0. The full collection of DASHR 2.0 contains 800 samples across 55 tissues, 33 cell types including 6 stem cells types, and 8 cell lines. In addition to the seven existing small RNA classes (microRNAs, Piwi-interacting RNAs, small nuclear, nucleolar, cytoplasmic RNAs, transfer and ribosomal RNAs), DASHR 2.0 includes other functional sncRNAs such as tRNA-halves, tRF -1, -3, -5, long non-coding RNAs-derived small RNAs, updated annotations for snoRNAs, piRNAs and Y-RNAs. More importantly, we have included all the previously unannotated sncRNAs identified in the datasets described, thus allowing researchers to use our database to perform in silico validation on their own data. All the datasets have been processed using an updated small RNA-seq pipeline (SPAR, http://lisanwanglab.org/SPAR) with additional features, including conservation scores, consistency of mature products and read accumulation relative to the genomic background. These new features, especially the conservation scores should greatly increase the convenience of studying these sncRNAs. All annotations are viewable as tracks on the UCSC genome browser so researchers can integrate with other existing omics data easily. DASHR 2.0 is a valuable resource for researchers to query and characterize their small RNAs of interest. Users can access DASHR 2.0 at http://lisanwanglab.org/DASHR/v2.

1303W

Predicting genetic ancestry for 805,482 patients using clinical data from electronic health records. *N. Tatonetti[1], T. Lorberbaum[1], K. Quinnies[1], R. Vanguri[1], H. Qi[2], W. Chung[2], D. Goldstein[3].* 1) Biomedical Informatics, Columbia University, New York, NY; 2) Pediatrics, Columbia University, New York, NY; 3) Institute for Genomic Medicine, New York, NY.

Genetic ancestry is essential for studies aiming to discover variants with clinical significance. Dimensionality reduction techniques, such as principal components analysis, make shared ancestry in complex populations easily identifiable. Unfortunately, genetic data are not available for most patients in clinical and research settings. Clinical data, like those stored in electronic health records (EHRs), are vast but do not directly capture ancestry information. Here, we present a novel method for inferring genetic ancestry using only variables routinely captured in the EHR. Using 1,161 patients that have both genetic and clinical information at Columbia University/NewYork-Presbyterian Hospital, we trained and validated a machine learning model that uses clinical data to estimate genetic ancestry. We conducted principal components analysis on 3,154 samples from 1000 Genomes and HapMap and applied the learned principal components (PCs) to the genetic data of 1,161 consented patients at Columbia. We then trained a random forest regressor model on demographic and billing code (ICD-10) data from these patients to predict their PC1 and PC2 values; we evaluated model performance using out-of-bag estimates against the genetic reference standard. Top predictive features included the self-reported race/ethnicity information, diseases of the nervous system, and congenital malformations. We found that merging related ICD-10 codes up to the Chapter level (21 condition features and 12 demographic features) achieved the greatest Pearson correlation (0.609) and lowest RMSE (68.29) for PC2 with only a small change in performance for PC1 (Pearson correlation = 0.600, RMSE = 36.66). We also found that the tree ensemble standard deviation was significantly correlated with the absolute error for both PC1 (Pearson correlation = 0.847, P < 1E-100) and PC2 (Pearson correlation = 0.849, P < 1E-100) and therefore could be used to estimate confidence of the predicted ancestry. 805,482 patients had sufficient data to predict their genetic ancestry from the clinical data stored in their EHRs. We applied our trained model to these patients to predict PC1 and PC2 values for patients that otherwise have no genetic data available. Our approach uses a minimal set of commonly collected data to accurately infer genetic ancestry and is a new tool for genetic screening in EHR research.

1304T

Improved phenotype-based computational methods to support diagnosis of genetic disease. *J. Chen[1], H. Xu[2], A. Jegga[1], K. Zhang[1], P. White[1], G. Zhang[1].* 1) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) University of Cincinnati Medical Center, Cincinnati, OH.

Clinical diagnosis of genetic diseases is often difficult because patients usually exhibit many clinical features that are not readily recognizable as known genetic syndromes. The Human Phenotype Ontology (HPO) provides a structured and standardized vocabulary of phenotypic abnormalities of human disease and it has been widely used for differential diagnostics, phenotype-driven analysis of genetic testing results. The main challenges of these approaches include how to efficiently extract the diagnostic information embed in the phenotypic ontology and to make correct and robust diagnosis when the query phenotypes are not specific or imprecise. Here we introduce two new methods to prioritized diagnosis of genetic diseases based on integrated semantic similarity (Method 1) or ontological overlap (Method 2) between the phenotypes expressed by a patient and phenotypes annotated to known diseases. Using simulated and real patients' data, we showed that these two methods were computationally efficient and exhibit better performance than previous methods in correctly prioritizing candidate diseases. We implemented our methods in a freely available web application to aid diagnosis of genetic diseases.

1305F

Quantification of transplant-derived circulating cell-free DNA in absence of a donor genotype. *E. Sharon[1,2,9], H. Shi[8], S. Kharbanda[3], W. Koh[4,5], L.R. Martin[4,5], K.K. Khush[6], H. Valantine[7], J.K. Pritchard[1,2,4,9], I. De Vlaminck[8,9].* 1) Department of Genetics, Stanford University, Stanford, CA, 94305, USA; 2) Department of Biology, Stanford University, Stanford, CA 94305, USA; 3) Pediatric Stem Cell Transplantation, Lucille Packard Children's Hospital, Stanford, California 94304, USA; 4) Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA; 5) Departments of Bioengineering and Applied Physics, Stanford University, Stanford, California 94304, USA; 6) Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; 7) National Institutes of Health, Bethesda, MD 20892, USA; 8) Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA; 9) Corresponding authors.

Quantification of cell-free DNA (cfDNA) in circulating blood derived from a transplanted organ is a powerful approach to monitoring post-transplant injury. Genome transplant dynamics (GTD) quantifies donor-derived cfDNA (dd-cfD-NA) by taking advantage of single-nucleotide polymorphisms (SNPs) distributed across the genome to discriminate donor and recipient DNA molecules. In its current implementation, GTD requires genotyping of both the transplant recipient and donor. However, in practice, donor genotype information is often unavailable. Here, we address this issue by developing an algorithm that estimates dd-cfDNA levels in the absence of a donor genotype. Our algorithm predicts heart and lung allograft rejection with an accuracy that is similar to conventional GTD. We furthermore refined the algorithm to handle closely related recipients and donors, a scenario that is common in bone marrow and kidney transplantation. We show that it is possible to estimate dd-cfDNA in bone marrow transplant patients that are unrelated or that are siblings of the donors, using a hidden Markov model (HMM) of identity-by-descent (IBD) states along the genome. Last, we demonstrate that comparing dd-cfDNA to the proportion of donor DNA in white blood cells can differentiate between relapse and the onset of graft-versus-host disease (GVHD). These methods alleviate some of the barriers to the implementation of GTD, which will further widen its clinical application.

1306W

Copy number variation detection and variant curation improves interpretation of exomes for inborn errors of metabolism. *S.E. Brenner[1], A.N. Adhikari[1], Y. Wang[1], R.C. Gallagher[2], R.L. Nussbaum[5], Y. Zou[1], U. Sunderam[3], J. Shieh[2], F. Chen[2], B.A. Koenig[2], M. Kvale[2], S.D. Mooney[4], R. Srinivasan[3], P.-Y. Kwok[2], J.M. Puck[2], The NBSeq Project.* 1) University of California, Berkeley, CA, USA; 2) University of California, San Francisco, CA, USA; 3) Tata Consultancy Services, Hyderabad, TS, India; 4) University of Washington, Seattle, WA, USA; 5) Invitae, San Francisco, CA, USA.

The NBSeq project is evaluating effectiveness of whole exome sequencing (WES) for detecting inborn errors of metabolism (IEM) for newborn screening (NBS). De-identified archived dried blood spots from MS/MS true positive and false positive cases previously identified in the California NBS were studied. 26 out of 145 affected individuals lacked two rare potentially damaging single nucleotide variants or short indels in genes responsible for their Mendelian disorders. The sensitivity of causal mutation detection in 145 NBSeq exomes varied across disorders; all affected PKU cases were predicted correctly, but several cases of other IEMs were missed. Deeper analysis of the data from these 32 individuals was undertaken to assess sources of discrepancy between sequencing results, MS/MS call, and clinical diagnosis. Copy number variation (CNV) calling tools were evaluated on NBSeq exomes for ability to resolve some of these exome false negatives. CNV tools can both miss CNVs in exomes and report them spuriously. We optimized tools for our data and filtered out genes (PRODH, HCFC1, ETFA) harboring common CNVs (identified from CNV calls on the 1000 genomes project exomes). This identified deletions in the correct genes for 4 of the 32 exome false negatives using XHMM: 2 isovaleric acidemia cases, 1 methylmalonic acidemia case and 1 OTC deficiency case. Our initial pipeline reported any variant in IEM genes with annotation of pathogenic or likely pathogenic by HGMD or ClinVar with allele frequency <5%. However, critical re-assessment of the primary literature for 59 variants with MAF >0.1% found that only 18 were reportable, including 1 dominant variant. The resulting curated inclusion/exclusion variant list reduced both false positives and false negatives, while highlighting the clinical heterogeneity among monogenic IEMs. Incorporation of CNV detection and variant curation into our analysis pipeline improved overall accuracy from 81.32% to 85.16% on the initial 145 affected NBSeq samples. This updated pipeline will be run on additional NBSeq exomes to assess the potential role for WES in NBS. While still not sufficiently specific alone for screening of most IEMs, WES can facilitate timely and more precise case resolution.

1307T

QRank: A novel quantile regression tool for eQTL discovery. *X. Song, G. Li, Z. Zhou, X. Wang, I. Ionita-Laza.* Columbia University, New York, NY.

   MOTIVATION: Over the past decade, there has been a remarkable improvement in our understanding of the role of genetic variation in complex human diseases, especially via genome-wide association studies. However, the underlying molecular mechanisms are still poorly characterized, impending the development of therapeutic interventions. Identifying genetic variants that influence the expression level of a gene, i.e. expression quantitative trait loci (eQTLs), can help us understand how genetic variants influence traits at the molecular level. While most eQTL studies focus on identifying mean effects on gene expression using linear regression, evidence suggests that genetic variation can impact the entire distribution of the expression level. Motivated by the potential higher order associations, several studies investigated variance eQTLs. RESULTS: In this paper, we develop a Quantile Rank-score based test (QRank), which provides an easy way to identify eQTLs that are associated with the conditional quantile functions of gene expression. We have applied the proposed QRank to the Genotype-Tissue Expression project, an international tissue bank for studying the relationship between genetic variation and gene expression in human tissues, and found that the proposed QRank complements the existing methods, and identifies new eQTLs with heterogeneous effects across different quantile levels. Notably, we show that the eQTLs identified by QRank but missed by linear regression are associated with greater enrichment in genome-wide significant SNPs from the GWAS catalog, and are also more likely to be tissue specific than eQTLs identified by linear regression. Intriging by heterogeneous eQTLs on gene expression, we have further extended the QRank approach to maximize the identification of heterogeneous effects.

1308F

A novel set of 96 indices for accurate sample demultiplexing on Illumina platforms. *B. Carrion, J. RosaFigura, S.K. Sandhu, J. Irish, L. Kurihara, T. Harkins, V. Makarov.* Swift Biosciences Inc., Ann Arbor, MI.

   The ability to uniquely and correctly assign sequencing reads to the appropriate sample is essential for all next generation sequencing projects. Recent publications have highlighted sequencing reads being misassigned due to index hopping on the Illumina pattern flow cell platforms. In addition, there is a growing demand for new indices to eliminate errors associated with certain index combinations on Illumina's 2 and 4-channel technologies. These errors are mostly due to insufficient separation of the hamming edit distances (the minimum number of base substitutions required to change one index sequence to the other) among index sequences with equal length. We have observed bleed-through index collisions among Illumina TruSeq HT indices at a frequency up to 1.5%. To remedy these issues, we have developed 96 single (i7) indices with a minimum hamming distance of at least three bases between all 96 indices, allowing proper discrimination of distinct index bases for sample identification during demultiplexing. These 8 nt i7 indices are compatible with Illumina's TruSeq HT i5 index sequences, enabling up to 768 high through-put dual combinations. To ensure there are no bleed-through index collisions, we co-sequenced indexed libraries containing unique non-overlapping inserts on both 2- and 4-channel instruments. Using this method, we were able to identify the presence, the extent, and the source of any observed bleed-through. Sample bleed-through was not observed at rates >0.1%. Also, it has been recently demonstrated that index hopping occurs on patterned flow cells as DNA library strands can have index misassignment due to carryover indexing PCR primers or adapters during the exclusion amplification (ExAmp) reaction, that is part of cluster generation. We have developed both wet lab protocols and unique dual indices (UDIs) to minimize this issue and where it does happen, the misassigned sequencing reads are rendered undetermined, eliminating any sequencing reads from being assigned to the incorrect indexed sample. Using this strategy, there is a minimum of 96 UDI combinations used for sequencing on patterned flow cells. Good laboratory practice suggests using different 96 indices from sequencing run to run to eliminate any possible carry-over, a possible third source for sequencing reads to be misassigned.

1309W

Gene-based tests using imputed genotype dosages showed increased statistical power than using best-guessed genotypes. *M. Hwang, Y. Kim, D. Shin, S. Moon, B. Kim.* Division of Genome Research, Korea National Institute of Health, Cheongju-si, South Korea.

Recently, imputation analysis is widely used in genome-wide association studies to gain increased statistical power and enhanced mapping resolutions, especially for low frequency variants. Association analysis using imputed data is more powerful considering genotype uncertainty by using probabilistic dosages of predicted genotypes. While imputed genotype dosages (DS) have frequently been used for single variant association tests, gene-based association tests often use best-guess genotypes (GT) that is the imputed genotypes with maximum posterior probability. The approach using DS for gene-based tests is limited because 1) only one imputation software, Minimac3, support VCF format with DS as an output, 2) no tool convert the output from other imputation software to VCF format with DS, which is generally used for tools supporting gene-based association tests, and 3) significant amount of time is required for conversion process using in-house scripts. Therefore, the conversion tool and study on advantages in using DS for gene-based tests are warranted. In this study, we developed a software Converting Imputation output to VCF format(CIV) with GT, Genotype Posterior Probabilities (GP), and DS information. And we compared gene-based association results of GT and DS. Imputation analysis was performed using IMPUTE v2 software with 1,000 Genomes project phase 3 data as a reference panel. SKAT-O was used for gene-based association analysis. Association for TCHL was tested using chromosome 1 data of imputed genotypes from about 7,000 genotyped samples (Korea Biobank Array). During conversion process, CIV showed more than approximately 2-3 times faster than qctool v1.4, only support GT and GP output, and in-house python script. In gene-based association tests, DS approach could test about 9.3% more genes than GT approach. In overall, for commonly discovered genes, DS showed slightly increased statistical power over GT. Our study showed that gene-based tests using DS is more powerful than using GT and CIV facilitated the approach using DS followed by imputation analysis in terms of faster conversion time and support for various variables of imputed genotypes.

1310T

User-driven prioritisation of study addition to the NHGRI-EBI Genome-Wide Association Study (GWAS) Catalog. *J.A.L. MacArthur[1], A. Buniello[1], T. Burdett[1], M. Cerezo[1], H. Coley[2], P. Hall[2], L. Harris[1], L.A. Hindorff[3], C. Malangone[1], A. McMahon[1], A. Milano[1], J. Morales[1], D. Welter[1], F. Cunningham[1], P. Flicek[1], H. Parkinson[1].* 1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom; 2) National Human Genome Research Institute (NHGRI), NIH, Bethesda, MD.

The importance of GWAS in identifying associations between genetic loci and traits or diseases is well established. GWAS produce a vast amount of data; we estimate that well over 15 billion variant-phenotype associations have been assayed in published GWAS to date. The GWAS Catalog (www.ebi.ac.uk/gwas) summarises this information into a structured, freely available and user-friendly resource. The Catalog contains the most significant associations, with p-value $<1 \times 10^{-5}$, from all published GWAS assaying at least 100,000 variants genome-wide. As of June 2017 this represents over 40,000 SNP-trait associations covering more than 2,300 unique traits from 2,900 publications. The volume of eligible data continues to grow with an 85% increase in associations from 2015-2016 publications compared to 2013-2014. This trend is likely to continue with many studies now analysing multiple complex phenotypes.Currently, eligible studies are curated into the GWAS Catalog in publication date order. Here, we propose alternative strategies to ensure data with the highest relevance and utility to users is available in the Catalog as quickly as possible. We have identified a series of metrics that can be used to assess the utility of GWAS, including sample size, extent of genome-wide coverage, relevance of trait, execution of a replication stage, inclusion of ancestrally diverse populations and the availability of summary statistics. We have surveyed users of the GWAS Catalog to help us identify which of these metrics are most valued by the community. We are optimising an algorithm to allow us to prioritise curation of GWAS studies according to these criteria, with each study characteristic weighted according to its relative importance. Prioritisation of the Catalog queue according to this algorithm will require triage of each publication as soon as it is identified as eligible. To maintain transparency of the Catalog methods, the queue of eligible GWAS publications will be made publicly accessible. The algorithm also takes into account requests for specific studies by Catalog users and therefore the interface will include the functionality for users to "up-vote" an eligible publication in the queue. These proposed changes to Catalog data acquisition will ensure the maximal value of the data available through the Catalog and sustainability into the future. We welcome feedback from the user community on these improvements and any other changes they would like to see.

1311F

GRASP v3: An updated GWAS catalog and contrast to similar catalogs.
B.A.T. Rodriguez[1], C. Zellmer[1,2], J. Li[1,3], A. Panda[1,4], N. Jenssen[1], J.P. Lien[1], J.D. Eicher[1], A.D. Johnson[1]. 1) Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, Framingham, MA; 2) University of Wisconsin, Madison, WI; 3) The Ohio State University, Columbus, OH; 4) Carnegie Mellon University, Pittsburgh, PA.

Background: Access to the expanding number of GWAS studies is a vital tool to biomedical research, providing opportunity for novel scientific discovery, identification of targets for functional assessment and mechanistic insights into disease biology. Current GWAS catalogs each have unique design features as well as limitations. The goal of GRASP (https://grasp.nhlbi.nih.gov/) is to provide a simple, intuitive means for the scientific community to query a centralized repository of reported SNP associations ($P \leq 0.05$) with human traits, including studies of methylation and expression QTL. Current challenges to developing and maintaining GWAS databases include: (1) data sharing, (2) non-standardized formats, annotation practices across GWAS, (3) human curator study search, review and extraction strategies, and (4) criteria for inclusion. Here we describe efforts toward the release of GRASP v3. **Results:** The v3 update increases the total number of studies to >3,300. To facilitate results sharing summary statistics are now included in the GRASP web portal. To investigate differences among widely used catalogs, we contrast GRASP and the NHGRI-EBI. The latter catalog was downloaded May 5, 2017 and limited to results with clear SNPids. Random studies (n=50) common to both were selected for comparison. At a genome-wide significant threshold (P<=5e-8) there was marked variation between the catalogs with respect to the studies' results. There were 656 SNP-association results unique to GRASP and 8 results unique to NHGRI-EBI. Focusing on 373 SNPs common to both we characterized consistency in associations for the studies. While most were consistent, differences in phenotype nomenclature were frequent. In general, GRASP contained more results for specific sub-phenotypes per SNP (e.g., sex-specific results, disease sub-classification). When comparing P-value associations for identical SNPs/publication/phenotype only a small number of associations (n=14/373) deviated by >1 log(P-value) unit, with others being highly correlated (r=0.99). Sampling of 50 more studies yielded similar results. **Conclusions:** Study exclusion criteria and data extraction methods vary between catalogs and likely account for many differences. However, our results indicate marked heterogeneity across popular GWAS results catalogs and researchers may be best served by querying multiple resources instead of relying on a single one. Further characterization and analysis of GRASP v.3 will be presented.

1312W

Reinterpreting genetic studies of kidney disease with integrated analysis of kidney-cell specific chromatin accessibility and transcriptomics data.
K. Sieber[1], S. Akilesh[2]. 1) Target Sciences, GlaxoSmithKline, King of Prussia, PA; 2) Department of Pathology, University of Washington, Seattle WA.

It is now evident that genetic variants identified by genome wide association studies (GWAS) largely map to non-coding regulatory DNA elements that in turn drive cell-type specific gene transcriptional programs. Regulatory DNA elements are efficiently identified genome-wide by chromatin accessibility assays such as DNase-seq. However, such maps are not available for rare or difficult to isolate cell types, which often play a critical role in human diseases. Specifically, the lack of cell-type specific chromatin-accessibility maps has limited the evaluation of GWAS of complex phenotypes such as chronic kidney disease, which affects ~10% of adults worldwide. Thus far, chromatin accessibility maps have only been generated for kidney tubule cells and endothelial cells and maps for a key component cell of the kidney filter, the glomerular podocyte, have been lacking. Therefore, to address this deficit, we generated high resolution chromatin accessibility (DNase-seq) and gene expression (RNA-seq) data for primary cultures of human podocytes (n=4) and tubular epithelial cells (n=3). Integrated analysis of these chromatin accessibility data with transcriptomic information enables us to map genetic variants associated with kidney disease to specific kidney cell types, a critical step in understanding biology and pathophysiology. Of the 63 replicated kidney loci identified by CKDGen in their 2015 and 2017 publications, 53 of the loci are within 5kb of open chromatin (DNase hypersensitive sites) in our data sets. For example, the *WNT7A* locus has numerous tubule-specific open chromatin regions which correlate with high expression of the *WNT7A* gene in tubule cells. The rs6795744 genetic variant associated with chronic kidney disease maps within 2kb of two tubule-specific DNase-hypersensitive sites. Increased *WNT7A* expression is associated with renal fibrosis in animal models. Therefore, this approach is the first use of cell type specific genomic data to functionally connect genetic variants for kidney disease to specific regulatory DNA elements. Extension of these approaches to other genetic variants is ongoing and will uncover novel therapeutic target genes and pathways affecting specific kidney cell types.

1313T

Genomic multilateration. *K. Kim[1], H. Baik[2], E. Eskin[3,4], B. Han[1].* 1) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Seoul 05505, Republic of Korea; 2) Department of Mathematical Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea; 3) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA 90095, USA; 4) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA.

Multilateration is a surveillance technique used to locate sensor nodes in wireless sensor networks. Using multilateration, the spatial coordinates of an unknown node can be estimated from its distances to reference nodes at known positions. For example, GPS analyzes radio signals received from satellites to pinpoint user position. Our motivation is to implement the concept of multilateration with genomic data. Specifically, we map individual data with K SNPs to a node in K-dimensional space, where each coordinate represents each SNP. In this space, pairwise Euclidean distance represents the genetic distance between two individuals. Using public data such as HapMap as our satellites, we investigate whether the distances to the satellites can uniquely determine the position of an individual in the K-dimensional space. Here, we analytically show that a minimum of K+1 satellites are required, which is not satisfied in many practical situations in which the number of satellites (public data) is smaller than K. However, we show that the distance from an individual to a set of satellites (distance vector) can have wide utility in genetic studies. First, the distance vectors can be used to detect duplicate individuals among multiple association studies in a meta-analysis without transferring genotype data. In a meta-analysis, detecting duplicate individuals is important to avoid spurious associations, but is often difficult because sharing genotype data is prohibited due to privacy issues. Our method can determine duplicate individuals by comparing distance vectors without breaching privacy and can be easily shared among studies. Second, the distance vectors can be used to group individuals into distinct populations. In simulations using the 1000 Genomes Project data as our satellites, our method based on k-means clustering successfully distinguishes HapMap individuals into distinct populations. Third, the distance vectors can be used to map individuals in a 2-D map. Previous studies used principal component analysis (PCA) on genotype data to geographically map individuals. We show that, without genotype data but with distance vectors only, we can approximate the mapping of POPRES data using the 1000 Genomes Project data as the satellites. Our estimated projection resembles the plot based on the PCA result of the corresponding POPRES data and the geographical map of the Europe.

1314F

Estimating incidence of inborn errors of metabolism from the frequency of variants in general population. *I. Mihalek, O. Bodamer.* Boston Children's Hospital, Boston, MA.

Inborn errors of metabolism (IEM) constitute a large class of genetic diseases, specific in that many of them are amenable to (nutritional) treatment, especially if detected early. As the first-line screening moves from metabolite to genetic, it opens new possibilities for the number of diseases that can be screened for, which in turn may bring them to the attention of the medical and nutrition industry - the existence of effective treatment currently being one of the prerequisites for the IEM to be included in the screening panel. Collecting data about the incidence of the IEM on the national level is a difficult process, and not systematically pursued. However, the information about human genetic variation is becoming available (e.g. ExAC database) on the scale that enables us to put a bracket on the expected incidence of all recognized IEMs with monogenic origin. We are collecting that information in a publicly available place (monogenic.org). In this presentation we discuss our current experience in using data integration - modeling of the related protein structure, evolutionary conservation, known disease causing variants, and the genetic variability in the protein coding regions of the genome, stratified across ethnic sub-populations - to put an upper limit on the expected incidence of IEMs. In 75% of the cases for which the information about the IEMs prevalence on the US national level we can make a reasonable prediction on its upper limit. The remaining cases, discussed here, bring to the attention the need to collect the information on a deeper genetic level, perhaps extend the definition of some diseases to include polygenic origin, and hopefully motivate the need for further systematic reporting on the IEMs detected through newborn screening.

1315W

A novel compuational strategy for DNA methylation prediction. *F. Yu, H.W. Deng, S. Hui.* Center for Bioinformatics and Genomics, Department of Global Biostatistics & Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA.

   Whole-genome bisulfite sequencing (WGBS) is the most comprehensive, unbiased method for characterization of genome-wide DNA methylation profiles, but is still too expensive to apply for large-scale studies. On the other hand, the Illumina HumanMethylation450 BeadChip (450K) array is a cost-effective and the most widely used method for DNA methylation analysis. Previous epigenomic studies have generated millions of 450K array data on various samples to study the methylation patterns associated with various diseases/traits. However, the 450K array only measures <2% of all the 28 million CpG sites in the human genome, which considerably limited the scope of the DNA methylation analysis in those studies. In this project, we developed a novel computational strategy to expand the existing 450K data by predicting DNA methylation levels at the unassayed CpG sites based on local DNA methylation patterns and other epigenomic functional annotations. Specifically, we first explored the co-methylation region based on the correlation pattern in a reference WGBS data. Then within each region, we trained a linear regression model of radial basis functions (RBFs) to predict a DNA methylation local score using the local upstream and downstream DNA methylation information. Finally, we incorporated the local score with other functional annotation information in a random forest classifier to predict the DNA methylation levels at unmeasured CpG sites. We found that the proposed prediction strategy using RBFs outperforms some existing methods, which use the methylation value of one nearest upstream and one nearest downstream CpGs as one kind of the input information for prediction, in terms of correlation between the predicted and true methylation values without any significant loss of area under the ROC curve. A potential extension by using the mixture of regression models of RBFs to estimate the local score are expected to improve the performance of our strategy by considering the bimodal distribution of methylation data. By application, this strategy will allow us to expand the existing 450K data and uncover new disease-related epigenetic signals without re-analyzing those samples by WGBS with high cost.

1316T

Scaling workflows for growing microbiome applications. *J. Lai.* Labcyte Inc, San Jose, CA.

   16S ribosomal RNA (rRNA) sequencing has long been a technique employed by microbial ecologists to identify and compare bacteria present within a given sample. It is a well-established method for garnering a sense of organism identity and diversity from complex microbiomes and environments that are difficult to culture or study otherwise. As our understanding of microbiomes grows, we have become increasingly aware of the implications these communities have in areas such as hospital ecology, human health, and metagenomic discovery efforts. Thus, the utility and applications for 16S rRNA sequencing are growing. Unlike PCR-based approaches or capillary sequencing, next-generation sequencing (NGS) has already enabled the analysis of entire microbial communities that are unable to be cultured or studied in the lab. Furthermore, NGS enables the capacity and throughput to perform the method as a routine procedure. The reagent cost, however, can still be prohibitive. We demonstrate that the 16S rRNA amplicon sequencing library preparation procedure for the Illumina platform can be miniaturized 10-fold while providing equivalent amounts of data at equivalent quality levels. This reduces the cost for a technique that is essential to microbiome projects and spurs the technique's feasibility for other projects. As microbiome applications continue to grow, more information is becoming necessary to answer the questions researchers are asking, involving biovars, pathogenicity, and fungal diversity. Here, to address these requirements, we show that whole genome sequencing can be performed at a 10-fold reduced scale to provide much more information and context for microbiome studies than 16S rRNA sequencing, while increasing throughput and reducing costs.

1317F

Deriving disease signatures using data driven appraches on micro-biome. *K. Poorey, L. Whitmore, D. Ray, C. Ball.* Systems Biology, Sandia National Laboratory, Livermore, CA.

Human body comprises 100 trillion microbial cells which make up 90% of total cells. They comprise 90% of our immune system, our digestive system, our second brain. Disruption of microbiome has been correlated to many of the diseases and also in some cases re-establishing a healthy microbial flora is key to restoring normal bodily functions. This puts research and understand-ing of human microbiome highly relevant in perspective of human health and diseases. Although, until recently the research in human microbiome research and its relation to human health have been limited due to lack of tools and resources. With the availability of Human Microbiome Project 1 and 2 (iHMP) data from individuals with disease and healthy gives us an opportunity to apply sophisticated machine learning and data-driven approaches to understand the role of microbiome in human health. Also, can it be used to identify key bio-markers of diseases for early detection and intervention? We have developed a state of the art Metagenomics and Amplicon sequencinG Pipeline (MAGPie) to analyze the NGS data generated from the iHMP project. We combine these datasets with available metadata of individuals to use ensemble machine learning methods and deep learning to understand the dynamics of microbial community and how it changes in diseases. Our ensemble machine learning pipeline included hierarchical clustering approaches to cluster microbes to identify core communities, which reduces the feature space to apply ensemble methods with deep learning to build a predictive model of onset of disease. We find that the reduced feature model reduces the prediction error by 75% in the validation dataset. We apply this methodology on three datasets of Onset of Inflammatory Bowel Disease (IBD), Preterm Birth and, type 2 diabetes. These machine learning models also measure feature importance which is key to identify biomarkers of diseases and health. Which can be used to design a microbial consortium that can be used to counter related health problems. We intend to expand this study to incorporate metabolic modeling using the metatranscriptomics dataset in these systems to model and understand the functional changes happening in diseases.

1318W

Noninvasive reconstruction of fetal methylome by sequencing of ma-ternal plasma DNA. *K. Sun, F.M.F Lun, R.W.K Chiu, Y.M.D Lo, H. Sun.* The Chinese University of Hong Kong, Hong Kong, Hong Kong.

The placental tissue plays key roles during the human pregnancy process and its DNA methylation profile is tightly related to the normal growth and development of the fetus, therefore possesses high clinical value in prenatal diagnosis. The discovery of cell-free fetal DNA in maternal plasma and the fact that these cell-free fetal DNA molecules mainly originate from the placental tis-sue promises the feasibility of noninvasive fetal/placental methylomic analysis using the plasma DNA. In a previous study, our group had demonstrated that the fetal methylome could be assessed through maternal plasma DNA. How-ever, the deduced fetal methylome suffered from low resolution and accuracy thus limited its clinical applications. To this end, in this work we propose a nov-el algorithm, which we named Femer, to reconstruct the fetal methylome by sequencing of plasma DNA. Femer is completely noninvasive and control-free. We demonstrate that Femer could achieve high accuracy as well as high resolution and could correctly recover the Partially Methylated Region (PMD) pattern of the placental tissue, thus provides us a high-quality view of the fetal methylome from the maternal plasma interface. In addition, Femer could also deduce the fetal methylation profile of the CpG islands with extremely high accuracy and demonstrates potential clinical utility in monitoring of the key genes for placental development during the pregnancy process. Our work thus moves one step forward of the noninvasive fetal methylomic analysis using maternal plasma DNA for potential clinical utilities.

1319T

Stargazer: A software tool for calling star alleles from next-generation sequencing data using *CYP2D6* as a model. *S. Lee[1], R. Dalton[2], M.M. Wheeler[1], K. Patterson[1], S. McGee[1], A. Gaedigk[3], T.A. Thornton[4], K.E. Thummel[5], E.L. Woodahl[2], D.A. Nickerson[1].* 1) Department of Genome Sciences, School of Medicine, University of Washington, Seattle, WA, USA; 2) Department of Biomedical and Pharmaceutical Sciences, University of Montana, Missoula, MT, USA; 3) Division of Clinical Pharmacology, Toxicology and Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO, USA; 4) Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA; 5) Department of Pharmaceutics, School of Pharmacy, University of Washington, Seattle, WA, USA.

Many enzymes play a role in pharmacological responses. One important enzyme is Cytochrome P450 2D6 (CYP2D6), which metabolizes nearly 25% of all prescription drugs. Prescription of such medications without the knowledge of a patient's CYP2D6 phenotype status can lead to adverse drug reactions or a loss of efficacy due to inappropriate drug choice and/or dosing. Enzyme activity varies considerably between individuals due to the highly polymorphic nature of the *CYP2D6* gene locus. There are over 100 defined *CYP2D6* haplotypes (designated as star alleles); these are characterized by single nucleotide variants (SNVs), insertion-deletion variants (indels), and structural variants (SVs) that include gene deletions, duplications, and conversions. However, genotype analysis of *CYP2D6* is complex due to a number of reasons including the presence of a highly related non-functional paralog, *CYP2D7*, positioned upstream of *CYP2D6*. We have recently developed a novel tool named Stargazer that uses next-generation DNA sequence reads and a customized computational pipeline to call SNVs, indels, and SVs from targeted or whole genome sequences. To call SVs, Stargazer calculates paralog-specific copy number by leveraging nucleotide differences unique to each paralog. Genotyping of 32 ethnically diverse trios by Stargazer was 98% concordant with data obtained by Sanger sequencing, long-range PCR, TaqMan assays, or a combination of the former. Stargazer accurately called SVs, including *CYP2D6/CYP2D7* hybrids, all of which segregated according to expectations in the trios. In one case, Stargazer detected a duplication event missed by the orthogonal methods. We also assessed the correlation between Stargazer's genotype calls and *in vitro* measurements of CYP2D6 abundance and activity in a large collection of human liver tissue samples (n > 300). To improve interpretability, we translated the samples' genotype information into a standard unit of predicted phenotype called activity score (AS). The AS predicted by Stargazer was strongly correlated with CYP2D6 abundance (r = 0.72) and activity (r = 0.69). This study demonstrates that *CYP2D6* genotype is the most important single factor explaining the large interindividual variability of CYP2D6 abundance and activity. We are currently extending Stargazer to genotype novel *CYP2D6* star alleles and other pharmacogenes.

1320F

Construction of Japanere reference genome and Japanese reference panel of thousands of individuals in Tohoku Medical Megabank Project. *M. Nagasaki[1,2,3], F. Katsuoka[1,2], Y. Kuroki[1,2,4], T. Shibata[1,2], T. Mimori[1,2], Y. Kawai[1,2,3], N. Minegishi[1,2], A. Hozawa[1,2], S. Kuriyama[1,2,5], Y. Suzuki[1,2], H. Kawame[1,2], F. Nagami[1], T. Takai-Igarashi[1], S. Ogishima[1], K. Kojima[1,2,3], K. Misawa[1,2], O. Tanabe[1,2], N. Fuse[1,2,6], H. Tanaka[1], N. Yaegashi[1,2,6], K. Kinoshita[1,3], S. Kure[1,2,6], J. Yasuda[1,2], M. Yamamoto[1,2], TMM Japanese Reference Panel Project.* 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 2) Graduate School of Medicine, Tohoku University, Sendai, Japan; 3) Graduate School of Information Sciences, Tohoku University, Sendai, Japan; 4) National Research Institute for Child Health and Development, Tokyo, Japan; 5) International Research Institute of Disaster Science, Tohoku University, Sendai, Japan; 6) Tohoku University Hospital, Tohoku University, Sendai, Japan.

We have been conducting one of the largest prospective genome cohort studies in Japan. One of our missions is to set up the genome infrastructure in Japan for the promotion of personalized healthcare and genome medicine. Towards this mission, we have developed a whole-genome reference panel consisting of the 1,070 Japanese individuals (1KJPN) with short read Illumina sequencers. With this effort, we have succeeded in cataloging more than 20 million variants. However, at the same time we realized difficulties in cataloging long insertions ranging from one hundred to around ten thousand bases. To overcome this difficulty, we undertook *de novo* assembly of a Japanese genome using a deep long-read sequencer and successfully cataloged ten thousand of long insertions missing in the current international reference assembly. While there are efforts of *de novo* assembly in Korea and China in which one each individual sequence was determined by the similar approach to that employed in our study, we for the first time are able to extensively analyze and characterize these insertions in comparison with more than one thousand international individuals by using the whole-genome sequencing results from our 1KJPN (1,070 Japanese individuals) and 728 individuals from available from international consortium database. Most of the insertions we have identified in this study are shared among populations, such that parts of these sequences are found in the genomes of Denisovan, Neanderthal and chimpanzee. This fact suggests that many common insertions are originated from primitive anthropoids. We also have discovered almost one hundred insertions reside in all international individuals we have analyzed, demonstrating that these insertions are the strongly preferred sequences to complement to the latest international reference assembly. More accurate reference sequence is critically important for current clinical sequencings, given that an increasing number of patients are getting diagnosed by clinical whole-genome or -exome sequencings. In this regard, we provide a decoy Japanese reference genome (decoyJRGv2). The decoyJRGv2 improves the alignment ratio of unmapped sequencing data by more than 10% on average, as well as decreases the miss-called variants in coding regions. With the Japanese reference assembly, we are reconstructing the Japanese reference panel of thousands individuals from short-read sequencers in ToMMo dataset to catalog more accurate variants. .

1321W

SPACE, a tool for dynamic exploration of principal component analyses.
N.D. Berkowitz[1], D.A. Turissini[1], E. Elyashiv[1], K.A. Rand[1], Y. Wang[1], E.L. Hong[1], C.A. Ball[1], K.G. Chahine[2]. 1) Ancestry, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

Principal component analysis (PCA) is frequently used in studies of human genetic diversity to study variability among a set of hundreds of thousands of single nucleotide polymorphisms (SNPs). PCA transforms this highly dimensional data into a new dimensional space, prioritizing the dimensions that capture the greatest variance. While powerful, PCA results can themselves be challenging to interpret, often involving arduous data manipulation and figure generation. We present SPACE, a PCA exploration tool that can be run from a web browser. SPACE, implemented in R using the package Shiny, allows users to dynamically scale, label, subset and transform data with a simple interface. In addition, users can examine metadata for a single sample or a collection of samples solely by pointing and clicking. We have designed SPACE to empower population geneticists to efficiently mine their data and get the most out of their analyses.

1322T

Quickly determining subject ancestries in large datasets using geno-types of dbGaP fingerprint SNPs. *Y. Jin, M. Feolo, S.T. Sherry.* National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

Genotypes and phenotypes of more than a million human individuals have been released by dbGaP. One researcher can now access the genotypes of hundreds of thousands of individuals from different dbGaP studies, but the data across studies is typically obtained using different genotyping and sequencing methods. It is therefore essential, but challenging, to systematically determine the ancestries, or population structures of all subjects in large, combined datasets. Numerous methods have been developed to determine subject ancestries from genotypes, including model-based approaches like STRUCTURE, and model-free, usually PCA-based method such as EIGEN-STRAT. Many methods work well for small or medium sized datasets with low genotype missing rates. However, no existing methods can be applied to large datasets with genotypes collected using different methods, like the datasets combined from many dbGaP studies. At dbGaP, we extract genotypes of a fixed set of 10,000 SNPs from each study, and use these genotypes to find subject overlaps among different studies. We have also developed a method called GRAF (Genetic Relationship and Fingerprinting) to determine subject relationships using these "fingerprint SNPs". In order to determine subject ancestries, we first divide dbGaP subjects into subpopulations based on the self-reported ancestries and calculate the allele frequencies of the fingerprint SNPs for each subpopulation. Then for each subject we calculate the values of a statistical metric that reflect the distances between this subject and every subpopulations, and use these values to predict which subpopulation this subject belongs to, or if this subject is a mixture of several subpopulations. This method, called GRAF-pop, is very quick, running in linear time relative to the number of subjects. It is also accurate, with 99% of the subjects predicted as European, Asian, and African American match the self-reported ancestries. GRAF-pop is also robust, tolerating high genotype missing rates up to 99%. When only 100 of the 10,000 SNPs have genotypes, GRAF-pop can still be used to separate Europeans, Asians and Africans from one another with 95% accuracy. GRAF-pop can be applied to genotype datasets obtained using diverse methods, and the results are comparable between different dbGaP studies, and any other datasets. In this presentation I will describe GRAF-pop in more detail, and how it is used as a QC tool at dbGaP to find errors in self-reported ancestries.

1323F

PubCases: A diagnosis assistant tool for rare diseases based on disease-phenotype associations extracted from published case reports. *T. Fujiwara[1,2], Y. Yamamoto[1], J.D. Kim[1], T. Takagi[2].* 1) Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems; 2) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo.

Recently, next-generation sequencing (NGS)-based diagnostic tests, which particularly employ whole-exome sequencing (WES), have been attempted for diagnosis of patients with rare diseases. Usually, a run of WES identifies more than 30,000 variants, among which tens or hundreds remain after filtering out common variants and prediction of their pathogenicity. For further narrowing down to a handful number of variants to be manually evaluated, a phenotype-driven diagnosis assistant tool such as Phenomizer is often used. Note that the performance of a tool like Phenomizer relies on its underlying database such as Orphanet, and that manually curated databases like Orphanet inherently show a limited coverage. We present PubCases (https://pubcases. dbcls.jp), a new diagnosis assistant tool which employs a text mining-based approach to improve the coverage of Orphanet. More than one million case reports were harvested from PubMed, and they were annotated for concepts of rare diseases and phenotypes. For the annotation, ConceptMapper was used with Orphanet Rare Disease Ontology (ORDO) and Human Phenotype Ontology (HPO). A co-occurrence of a ORDO term and a HPO term within a sentence was treated as a signal for association between them. PubCases exploits the database of disease-phenotype associations sourced both from Orphanet and our result, to assist a diagnostic process: for a set of phenotypes (HPO terms) as input, it suggests a ranking of diseases based on phenotypic similarity. The performance of PubCases was evaluated in comparison with Phenomizer, using the 114 clinical cases retrieved from MyGene[2], with "recall at ranks" (fraction of cases where the correct result appears in top 1, top 5, top 10, and top 100) as the evaluation metric. Orphanet had 6,233 rare diseases, among which 2,323 were associated with HPO terms. Our method could find phenotype-associations for 3,100 rare diseases, of which 52% did not have an association in Orphanet. PubCases showed the recall rate, 0.24@1, 0.30@5, 0.34@10, and 0.67@100, while Phenomizer showed 0.13@1, 0.29@5, 0.36@10, and 0.54@100. The experimental results support that our approach is effective in improving the coverage of the disease-phenotype associations, and PubCases performs better than Phenomizer in terms of average recall, enlightening a promising path to improving the diagnosis rate of NGS-based approaches.

1324W

MVP: A deep neural network method of predicting pathogenicity of missense variants enables novel genetic discoveries. *Y. Shen[1,2], H. Qi[1,3], J.J. Long[3], N. Zhu[1,4], W.K. Chung[4,5], Y. Guan[6].* 1) Department of Systems Biology, Columbia University, New York, NY; 2) Department of Biomedical Informatics, Columbia University, New York, NY; 3) Department of Applied Physics and Applied Mathematics, Columbia University, NY; 4) Department of Pediatrics, Columbia University, New York, NY; 5) Department of Medicine, Columbia University, New York, NY; 6) Departments of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine.

Missense variants are the most abundant type of coding mutations in patients with cancer, birth defects, and neurodevelopmental disorders. However, the phenotypic consequence (pathogenicity) of missense variants are largely unknown, a major barrier in genetic studies and clinical testing. Previously published *in silico* prediction methods have facilitated the interpretation of missense variants, but have limited performance due to a few major issues. First, conventional statistical approaches are sub-optimal in leveraging large amount of training data or many correlated predictors. Second, databases of pathogenic variants are known to have substantial rate of false positives that limit the performance of supervised learning methods. Here we describe a new method (MVP) to leverage large training sets with correlated predictors. We use a deep residual network method in which a hierarchical structure increases model expressiveness without making it excessively large. Besides predictors commonly used in other methods, we included additional predictors at the variant-level (e.g. probability of being in protein interaction interface) and gene-level (e.g. protein complex formation and dosage sensitivity, which can modulate the *pathogenicity* of deleterious missense variants). We obtained large curated pathogenic variant datasets as positives and random rare missense variants from populations as negatives for training. MVP achieved superior area under the receiver operating characteristic curve (AUC) compared with other methods in cross-validation. To benchmark real-world performance, we compiled known cancer somatic mutation hotspots as positives, and random somatic missense as negatives, and showed that MVP had much better AUC. Additionally, we collected germline *de novo* mutations in ~6000 exome sequencing samples from recent studies of developmental disorders. Since such data do not have ground truth, we used the excess of predicted pathogenic missense variants in cases compared to controls as proxies of sensitivity and specificity. We showed MVP achieved 30-100% better sensitivity under a reasonable specificity compared with other methods, and implicated a number of new candidate risk genes for diseases with low reproductive fitness (e.g. *GLRA2*, *MINK1*, *CUL1*, *CAMK1D*, *ROBO2*, *PPP4C*, and *CTNND1*). These results using actual disease associated data suggest that MVP can improve diagnostic yield and power of detecting novel risk genes in genetic studies.

1325T

Search Candidate Regulatory Elements by ENCODE (SCREEN): A web-based tool for visualizing genomic annotations. *H.E. Pratt, M.J. Purcaro, J.E. Moore, Z. Weng.* Program in Bioinformatics and Integrative Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

   The Encyclopedia of DNA Elements (ENCODE) Consortium and Roadmap Epigenomics Consortium have generated thousands of genomic datasets with the goal of annotating the functional elements in the human and mouse genomes. The ENCODE Data Analysis Center has integrated these datasets to create a collection of annotations termed the ENCODE Encyclopedia. The Encyclopedia includes ground level annotations such as DNase hypersensitivity sites (DHSs) and peaks called for each dataset, and integrative level annotations, the core of which form a registry of candidate regulatory elements (cREs) anchored on DHSs in roughly 200 human and 50 mouse cell types. Here we present SCREEN, a web-based tool for searching and visualizing the cREs and their interactions with ground-level annotations. SCREEN is divided into three "apps". The core app is the cRE-centric search, where the user can filter cREs by various characteristics and identify interaction with genomic features such as transcription factor binding sites, SNPs, and topologically-associated domains. The gene-centric app plots RNA-seq data, and, for mouse data, plots differentially-expressed genes and differential cRE activity across cell types and developmental time points. Finally, the SNP-centric GWAS app intersects cREs with SNPs from more than 50 published GWAS studies and presents associated analysis. We have used the Encyclopedia and SCREEN to characterize several novel regulatory elements. For example, SCREEN identified a candidate enhancer in GM12878, a lymphoblastoid cell line, overlapping SNP rs1250568 from a multiple sclerosis GWAS study; ground-level annotations revealed that this candidate enhancer lies in a binding site for transcription factor ELF1, a player in the IL-2 and IL-23 immune pathways which have been previously implicated in multiple sclerosis. SCREEN has also identified several candidate promoters and enhancers which may drive increased Ogn expression during mouse limb development, and several candidate enhancers overlapping SNPs from a schizophrenia GWAS study which may influence transcription factor SP3 binding during human and mouse brain development. Together, the Encyclopedia and SCREEN provide powerful tools for addressing a wide array of biological questions, such as predicting enhancer-gene interactions, comparing gene expression profiles across cell types and developmental time points, and annotating genetic variants and possible functional roles in disease.

1326F

SinCCE (Single-Cell Cluster Ensemble): Cluster ensemble for single-cell RNA-seq data. *Y. Yang[1], R. Huh[2], H. Culpepper[1], Y. Li[1,2].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

   Accurately identifying and characterizing cell type from a mass of heterogeneous cells is a crucial step for downstream analysis of single-cell RNA-seq data. Recently, multiple algorithms have been developed for this purpose. However, different clustering methods, utilizing different characteristics of data for clustering, typically yield different numbers of clusters and different cluster assignment. In this study, we present **Sin**gle-**C**ell **C**luster **E**nsemble (SinCCE) for highly accurate and robust single-cell clustering. SinCCE starts with clustering results from any existing clustering methods and builds one consensus clustering. We performed extensive evaluations across over 20 datasets with number of clusterings ranging from 3 to 14, and number of single cells ranging from 49 to 32,695. Specifically, single cells are first clustered using four state-of-the-art methods, SC3, CIDR, Seurat and t-SNE + k-means; and individual solutions are then combined into one consensus cluster using three hyper-graph-based partitioning algorithms hypergraph partitioning algorithm (HGPA), meta-cluster algorithm (MCLA) and cluster-based similarity partitioning algorithm (CSPA). In our evaluations, SinCCE generates high-quality clustering, in terms of both cluster number and shapes, across various datasets. Moreover, SinCCE is computationally efficient to accommodate large datasets. For example, it takes < 5 minutes to process a data set of 28,733 cells.

1327W

Improving quality of variant calling by integrating whole genome and whole exome sequencing from same samples. *X. Li[1], L. Gauthier[1], A.V. Segré[1], M. Lek[1], E. Gelfand[1], G. Getz[1,2], K. Ardlie[1], D. MacArthur[1,3].* 1) Broad Institute, Cambridge, MA; 2) Department of Pathology, Massachussetts General Hospital, Charlestown, MA, USA; 3) Analytic and Translational Genetics Unit, Massachussetts General Hospital, Boston, MA, USA.

Whole genome sequencing (WGS) and whole exome sequencing (WES) are two important genotyping technologies, which provide high density ascertainment of variants in different/overlapping genomic regions, suitable for various downstream analyses. As the price of Next Generation Sequencing goes down, more and more studies may employ both technologies to get optimal variant calling in coding regions. To fully utilize the sequencing information from both technologies, while reconciling differences of genotype calls among them, we extended the current variant calling framework of Genome Analysis Toolkit (GATK) to enable variant calling and genotyping by integrating WGS and WES data from same samples within a cohort. Instead of merging at the raw sequencing reads level, our method directly merges the Phred genotype likelihood scores, making the integration much easier when GVCFs were generated. To evaluate our method, we generated high confident call sets from WGS and WES data of 147 individuals in GTEx dbGaP release v6, using our latest variant calling and sample and variant quality control tool set we developed. Comparing the integrated WGS and WES call set in exome intervals with WGS or WES alone, we found that our integrative method led to the discovery of an additional ~5,544 variants per sample in exome regions compared to the WES call set (27% of WES calls), while keeping a similar concordance rate with array genotyping technology as the WGS and WES separate call sets. The increase in sensitivity was most pronounced among low frequency variants. Our method also greatly improved genotype quality score (GQ) of on average 80% of WES calls and 0.4% of WGS calls (~16,000 variants per sample). Moreover, we found our method could help reconcile effects of different technologies, such as Illumina HiSeq2000 versus Illumina HiSeqX, and Agilent versus Illumina's Capture Exome (ICE). We further evaluate whether our method might introduce ascertainment bias of variant calling in coding versus noncoding regions, as this may have an effect on genome-wide analyses. In all, our method for integrating WGS and WES calls from the same donor, generated higher sensitivity and quality of variant calls in coding regions, in particular compared to the WES call set, while harmonizing the differences among sequencing technologies.

1328T

The Mega2R suite of R packages: Tools for accessing and processing common genetic data formats in R. *D.E. Weeks[1,2], R.V. Baron[1], J.R. Stickel[1].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

The standalone C++ Mega2 program has been facilitating data-reformatting for linkage and analysis programs since 2000. Support for more analysis programs have been added over time, though this has required software developers keenly knowledgeable in the Mega2 internals. Currently, Mega2 converts data from several different genetic analysis data formats (including PLINK, VCF, BCF, IMPUTE2) into the specific data requirements for over 40 commonly used linkage and association analysis programs (including Mendel, Merlin, Morgan, SHAPEIT, ROADTRIPS, MaCH/minimac3, etc.). Recently, Mega2 has been enhanced to use a SQLite database as an intermediate data representation. Additionally, Mega2 now stores bialleleic genotype data in a highly compressed form, much like that of the GenABEL R facility and the PLINK binary format. Concurrently, the R community and Bioconductor community have developed a variety of genetic analysis programs complimentary to the programs available through Mega2. We have now made it easy to load SQLite3 Mega2 databases directly into R as data frames to use these R facilities. In addition, we have developed C++ functions for R to decompress needed subsets of the genotype data, on the fly, in a memory efficient manner. We created the base Mega2R package that loads and manipulates data frames. We have also created several more R packages that illustrate how to use the data frames as well as perform useful functions: the Mega2pedgene package to run the pedgene R package to carry out gene-based association tests on family data using selected marker subsets, the Mega2VCF package to output the Mega2R data as a VCF file and related files (for phenotype and family data), and the Mega2GenABEL package to convert the data frames into GenABEL R gwaa.data-class objects. The Mega2R program enhances GenABEL since it supports additional input data formats (such as PLINK, VCF and IMPUTE2) not currently supported by GenABEL. The Mega2 program and the Mega2R R package are both open source and are freely available, along with extensive documentation and a 'quick-start' tutorial, from our https://watson.hgen.pitt.edu/register web site. This work was supported by NIH grant R01 GM076667 (PI: Weeks). Earlier contributions to our code base were made by Nandita Mukhopadhyay, Lee Almasy, Mark Schroeder, and William P. Mulvihill.

1329F

An iterative strategy to improve the power of epistasis analysis. *J. Wen, C. Nadzak, X. Shi.* Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC.

Epistasis is an important genetic component in human diseases and quantitative traits whereby non-additive interactions of genes result in varied phenotypic effect. Many studies show that epistasis helps elucidate the genetic foundation for modeling human traits and evolution. However, a major challenge for epistasis analysis arises with the advent of high throughput sequencing and omics data, which the number of features (genetic variants) is usually much larger than sample size and produces an over-saturated model. To address this challenge, we propose an iterative pipeline base on an empirical Bayesian elastic net (EBEN) method that increases the power and accuracy in genomic studies of epistasis. First, we screen each single feature with the phenotype using a simple statistical procedure where features passing a predefined threshold are kept in the model, and then an estimation of the significant features with marginal effect is produced using EBEN. The resulting significant features from EBEN are then used to correct the raw phenotype. Second, a screen on the pairwise epistatic interactions of features is performed using the same statistical procedure and we use EBEN again to estimate the significant pairwise features against the corrected phenotype. Third, we put all the significant features, including the marginal and epistatic effect output by EBEN in Step 2 and 3, into a complete EBEN model to conduct parameter estimation. Significant marginal and epistatic variants are then selected from the complete EBEN model. Monte Carlo simulations verify that the power for epistasis detection is improved using our strategy. We apply our pipeline to a yeast data set and identify the marginal and epistatic effect of SNPs on indoleacetic acid levels, which provides a more comprehensive understanding of the molecular mechanism of yeast fitness via pathway and enrichment analysis.

1330W

Telomere length estimation and analysis on large scale whole-genome sequencing data. *M.A. Taub[1], K. Iyer[2], J. Weinstock[3], L.R. Yanek[4], H.M. Kang[3], T. Blackwell[3], A. Correa[5], S.L. Salzberg[6], D. Vaidya[4], D.M. Becker[4], J.G. Wilson[7], L.C. Becker[4], G. Abecasis[3], A. Reiner[8], R.A. Mathias[4].* 1) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 2) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 3) Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI; 4) The GeneSTAR Program, Johns Hopkins School of Medicine, Baltimore, MD; 5) Department of Medicine, University of Mississippi Medical Center & Jackson Heart Study, Jackson, MS; 6) Departments of Computer Science and Biomedical Engineering, and McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 7) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 8) Department of Epidemiology, University of Washington, Seattle, WA.

Telomere length (TL) is a hallmark of aging, and has been associated with a range of diseases. Variation in TL is linked to exposures such as smoking and stress, and to genetic factors. TL is also considered a measure of "biological age", a concept that disengages the calendar age of an individual from their physical age. While some prior studies have assessed these relationships, the advent of large-scale whole-genome sequencing (WGS) studies in diverse populations has created new resources to expand knowledge of the role of TL in aging and disease.   The optimal method for estimating TL from WGS data is still unknown. In a pilot analysis of 574 samples from the GeneSTAR study, we compared three methods (TelSeq, Computel, and a hexamer counting method), and observed strong correlation among methods (0.94 and above within a batch) but wide differences in computational efficiency (ranging from a few hours to a few days to process one sample, on one core). To date, only small-scale comparisons between WGS-based estimates and low-throughput estimates, such as flow-FISH or qPCR have been done. Preliminary analysis of 19 samples with flow-FISH estimates from two cell types (lymphocytes and granulocytes) show consistent correlation between 0.7 and 0.8 with all methods considered, indicating that computational efficiency is likely to be a more important consideration as all methods perform otherwise similarly to one another.   Samples from the pilot were processed in two batches using two different sequencing centers, to similar depth (~30x). We saw striking differences in the distribution of TL estimates from these batches, with TelSeq estimates on the first batch of samples showing a median of around 5kb (interquartile range 4.2kb to 5.5kb), and the second batch of samples showing a median of around 2.8kb (IQR 2.5kb to 3.1kb). For large scale sequencing studies, with samples likely to be processed in batches, concerns about the ability to combine data generated separately are clearly warranted for these TL estimates. Ongoing work involves estimation on an additional 2500 samples from the Jackson Heart Study through the TOPMed program, for which Southern Blot TL estimates calculated on blood from the same collection point are available for comparison. This will aid in determining the optimal TL estimation method and investigating how to detect and adjust for batch effects to enable effective analyses of the influence of TL on health outcomes.

1331T

1000 Genomes Project data and additional openly consented data resources can be accessed via the International Genome Sample Resource (IGSR). *S. Fairley, P. Harrison, E. Lowy, D. Richardson, I. Streeter, G. Yordanova, H. Zheng-Bradley, L. Clarke, P. Flicek. European Molecular Biology Laboratory. European Bioinformatics Institute, Hinxton, United Kingdom.*

The International Genome Sample Resource (IGSR) provides continued access to the resources created by the 1000 Genomes Project, while also expanding and updating those resources. Resources from the 1000 Genomes Project have been updated to the current reference genome. Data from phase three of the Project has been aligned to GRCh38, an assembly mapping-based "liftover" of the call set has been made available and the alignments to GRCh38 are being used to create a new call set directly on GRCh38. FreeBayes, GATK and Samtools have been used to call on GRCh38, while working toward creation of an integrated call set. New populations have been added to those sampled by 1000 Genomes. An example of this is data from the Gambian Genome Variation Project (GGVP). This project sampled an additional 100 individuals from the same Gambian population that was sampled in the 1000 Genomes Project. Further, GGVP sampled 100 individuals from each of three further Gambian populations, making three new populations and 400 new samples available. These data have also been aligned to GRCh38. New data generated on samples from 1000 Genomes is also made available. In particular for samples included in the work of the Human Genome Structural Variation Consortium (HGSVC), a wealth of data types are available, including PacBio, 10X, TruSeq, Strand-seq, high-coverage PCR-free Illumina data and others. Among the work that IGSR does to facilitate the availability of open genomic data is data coordination work, assistance in ethical review and alignment of sequence data. IGSR is open to establishing new collaborations for projects with openly consented genomic data. As a means of improving access to the data in IGSR, a data portal was added to the project website. This provides a number of ways to view the available data and we continue to develop this to meet users' needs. This, combined with the ability to browse the data in genomic context in Ensembl, provides users with a range of mechanisms for accessing the available data.

1332F

An alignment-based approach for sensitively detecting SVs using optical maps data. *X. Fan[1,2], J. Xu[3], A. Delcher[4], L. Nakhleh[1], F. Yue[3], K. Chen[1,2], Human Genome Structural Variation Consortium. 1) Department of Computer Science, Rice University, Houston, Texas; 2) Department of Bioinformatics and Computational Biology, The University of Texas M.D. Anderson Cancer Center, Houston, Texas; 3) Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, Pennsylvania; 4) Nathans-McKusick Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland.*

Structural variation (SV), although associated with genetic diseases and population diversity, has not been comprehensively studied due to limited sequencing length in widely used Next Generation Sequencing (NGS) data. Emerging new technologies such as the Optimal Maps (OM) generated by Bionano Inc. produce much longer reads (between 100kbp to 250kbp), making it possible to detect SVs in large repetitive regions that have been missed in current studies. However, OM reads lack precise sequence-level information and often contain errors. Current methods attempt to perform whole genome assembly to reduce random errors in individual reads. Such methods, however, require deep coverage and are insensitive to detecting subclonal SVs in heterogeneous cancer genomes. In this abstract, we present a novel alignment-based approach that can sensitively detect somatic SVs in heterogeneous cancer OM data. Starting from the global alignment of Optical Map reads, we develop a dynamic programming algorithm to perform local re-alignment around the breakpoints. This algorithm takes into consideration the three types of errors in Optical Maps (additional and missing labels, sizing error), and thus can effectively improve the alignment accuracy. To enhance specificity, an SV is called only when a cohort of reads supports it, where the cohort size depends on the sequencing coverage, estimated variant allele fraction and tumor purity. We evaluated our algorithm using a public dataset, NA12878, and compared it with two previous approaches. We found that our sensitivity is twice that of the method published in Genetics'16, and 15% greater than that of the Irys program. Our FDR is much lower than Irys when compared against results derived from 10X genomics data. Our unique calls appear to be enriched in repetitive regions, confirming the advantage of our method. We also applied our algorithm to a chronic myeloid leukemia cell line K562. We detected some validated SVs that Irys missed, in addition to the majority of those that Irys detected. In all, our method is the first to combine global and local re-alignment to effectively utilize Optical Maps to detect SV, and thus further improves the detection of SV and, consequently, advances translational research. .

1333W

Leveraging unique molecular identifiers to improve low-frequency variant calling in QIASeq V3 panels. *B. Vilhjalmsson[1], D. Thomsen[1], J. Hill[1], R. Rhode[1], D. Kamaev[1], J. Vesterstrøm[1], T. Rask[1], P. Nielsen[1], M. Christensen[1], J. Nielsen[1], R.G. Madsen[1], M. Hussong[2], S. Rulli[2], L. Husted[1], A. Materna[1], R. Samara[2], M. Simonsen[1], A.-M. Hein[1], Z. Wu[2], E. Lader[2].* 1) QIAGEN, Aarhus, Risskov, Denmark; 2) QIAGEN, Frederick, Maryland, USA.

   Unique molecular indices (UMIs) have in recent years become a standard practice in sequencing. UMIs are random molecular barcodes that are attached to the original DNA or RNA molecules at an early stage before amplification. The benefit of this technology is two-fold: 1) it provides a more accurate snapshot of the original DNA or RNA fragments before amplification; 2) it enables to correct for some PCR and sequencing errors (Xu *et al.*, BMC Genomics 2017).   To take advantage of UMIs, we developed an integrated Sample-to-Insight approach encompassing proprietary chemistry (QIAseq targeted panels) to incorporate UMIs into library construction and three different bioinformatics workflows processing UMI-containing reads in the QIAGEN Biomedical Genomics Workbench (version 4.1) for different data types: one for estimating mRNA expression levels, one for variant detection, and one for detecting fusion gene expression.   All three workflows make use of UMIs to either provide more accurate expression level estimates, or to improve the effective sequencing fidelity and thus variant calling accuracy. This is achieved through two key algorithms, one that identifies UMI groups, and one that merges reads within a UMI group to produce a consensus UMI read. A UMI group consists of reads that both map to the same position, and have an identical UMI, while accounting for errors. A consensus UMI read is obtained by maximizing the likelihood for each base, accumulating statistical evidence across multiple reads within the same UMI group. In practice, this corrects for PCR and sequencing errors, [RS-Q1] thereby increasing the average estimated per-base Q-scores substantially.   We evaluated the benefit of using UMIs, by comparing the three different types of workflows to corresponding workflows where UMIs were ignored. We applied these to several different QIAseq targeted panel datasets and a GIAB (genome in a bottle) dataset for NA12878, where we estimated expression levels, variant frequencies, and empirical Q-scores. We found that UMIs provide a significant increase in the estimated empirical Q-scores, as well as providing more accurate variant frequency estimates overall. These two benefits are particularly relevant for low-frequency variant calling, highlighting UMIs as a promising technology for improving variant detection in cancer applications.

1334T

Multi-sample isoform quantification from RNA-seq for known and novel transcripts. *A.E. Byrnes[1,2], F. Aguet[2], T. Sullivan[2], A. Bloemendal[1,2], J.M. Maller[3], C. Churchhouse[1,2], B.M. Neale[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 3) Genomics plc, Oxford, UK.

   Alternative splicing is critical for the regulation and diversity of most human genes. The ability of a single gene to give rise to several diverse transcripts, and subsequently proteins, has been implicated in a wide range of processes and disorders from brain function to cancer proliferation. However, most RNA-seq transcript quantification methods rely upon existing lists of possible transcripts that may be incorrect or incomplete. Our field has long considered the problems of transcript discovery and quantification separately; however, we advocate evaluating both simultaneously to improve performance. We propose a 2-step, multi-sample method for discovery and quantification of transcript isoforms, both known and novel, from RNA-seq data.   Our method aims, first, to maximize inference on splicing behavior by combining information from all aligned RNA-seq samples to construct a graph representing all possible transcripts, similar to Cufflinks' appraoch, but on all samples pooled together. In graph-building we weight each possible junction between exons by the number of junction reads observed across all samples. We represent each isoform as a possible path through the graph which provides an empirically derived list of transcripts that may be present in our samples and a preliminary estimate of their respective frequencies.   After constructing a list of all possible isoforms, we aim to quantify relative levels of transcription for each isoform in each sample. In order to do so, we conduct a comparison of existing isoform quantification methods, including Cufflinks, Kallisto and RSEM, in addition to our own expectation-maximization (EM) algorithm which explicitly models many of the biases endemic to RNA-seq data by using Alpine. This second step allows us to characterize the isoforms present in any individual and quantify their respective transcription for each sample separately. The isoforms with very low or zero frequency estimates can be pruned from the anti-conservative list in the first step, providing a reliable list of isoforms likely to exist in nature.   We demonstrate robust re-identification of many annotated transcripts, novel identification of additional transcripts not previously identified in Ensembl and several annotated transcripts with no evidence of expression in our data. We will discuss the details of this method and the relative performance of the above methods. Our results have clear implications for future work in alternative splicing.

1335F

PennSeq2: Efficient quantification of isoform-specific gene expression from RNA-seq data using weighted likelihood method. *Y. Hu, M. Li.* Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA.

The emergence of RNA-Seq has provided a powerful tool to study gene regulation at isoform levels. We previously developed PennSeq (Hu et al., 2014 Nuc Acids Res 42(3):e20), a statistical method for isoform-specific gene expression estimation by modeling non-uniform read distribution. However, in addition to the accuracy of estimated isoform expression, efficiency, which measures the degree of estimation uncertainty is also an important factor for downstream analysis. Here we developed PennSeq2, a successor of PennSeq, which employs a weighted likelihood framework to improve estimation efficiency. Briefly, for each read, PennSeq2 assigns a weight based on the number of compatible isoforms to assure that more informative reads (with less compatible isoforms) are given more weight in the likelihood function. PennSeq2 then employs an EM algorithm to estimate isoform-specific gene regression by maximizing the weighted likelihood. Compared to PennSeq, PennSeq2 has several advantages. First, the EM algorithm converges faster than PennSeq. Second, the efficiency of isoform expression estimation is significantly improved, which lays the groundwork for downstream analysis. Third, PennSeq2 yields equally accurate isoform expression estimation compared to PennSeq, even though this weighted likelihood approach theoretically trades bias for efficiency. To verify these advantages, we conducted simulations and compared its performance with PennSeq. We simulated 10 million 76-bp paired-end reads and mapped them to the reference genome using Tophat. Then, isoform relative abundances from 1455 genes were estimated using PennSeq and PennSeq2, respectively. Squared Pearson correlations between the ground truth and estimated relative expression demonstrate good estimation accuracy for both PennSeq (R^2=0.933) and PennSeq2 (R^2=0.926). Meanwhile, compared to PennSeq, on average, PennSeq2 reduces estimation uncertainty by 76% and the number of iterations in EM algorithm by 36%. Additionally, the running time of PennSeq2 is reduced by 10 times compared to PennSeq. These results indicate performance improvement of PennSeq2 without sacrificing estimation accuracy. Currently, we are performing extensive simulations and real RNA-Seq data analysis to verify the benefits of PennSeq2 estimates for differential expression and differential splicing analysis. As estimation efficiency improves, we expect downstream analysis based on PennSeq2 to have improved power.

1336W

VariantFX: An open-source framework for aggregation, visualisation and analysis of Mendelian disease cohort sequence data. *M. Ahmad[1,2], E. Mazaika[1,2], R. Govind[1,2], N. Whiffin[1,2,3], E. Edwards[1,2], B. Weisburd[4,5], M. Solomonson[4,5], A. Harper[2], R. Walsh[1,2], X. Zhang[1,2], A. Wilk[1,2], A. O'Donnell-Luria[4,5], H. Najgebauer[1,2,3], P.J.R. Barton[1,2], S.A. Cook[1,2,3], D. MacArthur[4,5], J.S Ware[1,2,3].* 1) National Heart and Lung Institute, Imperial College London, London, United Kingdom; 2) Royal Brompton and Harefield NHS Foundation Trust, London, United Kingdom; 3) MRC London Institute of Medical Sciences, Imperial College London, London, United Kingdom; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA.

Genomic reference datasets such as gnomAD have been transformative in clinical genome interpretation. However, the integrated presentation and analysis of equivalent data from cases with genetic diseases is challenging. Existing repositories that aggregate single variants do not preserve critical case allele frequency data, and case aggregation is likely best undertaken in a distributed manner by experts in each disease domain. VariantFX provides an open-source web framework comprising a variant database to host aggregated case data, an integrated API, a built-in web front-end for visualisation and quantitative analysis alongside the gnomAD reference dataset. We use Docker as a containerization platform, to create a downloadable resource with multiple complimentary components. Separate containers house each of the three main components of VariantFX; the database, associated API and front-end web scripts. The decoupling of the services into these distinct parts supports easy software upgrade, simplifies debugging, avoids software conflicts and supports adaptability. All the containers are managed by a single *docker-compose.yml* file. The database container hosts sequence data stored within a NoSQL MongoDB, which supports highly efficient queries for unstructured data. The database is shipped with sample seeded data. Users can easily upload their own data which is automatically imported to the database and presented in the front-end via the API. The second container hosts an API, configured through GraphQL, to serve data from the MongoDB database to the front-end and external web services. GraphQL queries have low over-head compared to the traditional REST API, resulting in low bandwidth consumption and flexible front-end coding. The final container hosts the front-end, developed in JavaScript using a React library for efficient rendering of data-rich pages. It uses the React virtual-DOM to render the front-end, providing highly interactive and responsive user experience. VariantFX is portable and can run on all major operating systems with Docker installed. We have provided all the scripts and full instructions on how to download and customise VariantFX on Github (https://github.com/ImperialCardioGenetics/variantfx). Our live implementation of VariantFX combining gnomAD with >10,000 cardiomyopathy cases can be found at http://cardiac.variantfx.org.

1337T

NGS based CNV calling in a clinical diagnostic setting. *D. Becker[1], M. Ziegler[1], Y. Dinçer[1,2], K. Pindl[1], J. Schulz[1], K. Mayer[1], S.H. Eck[1], HG. Klein[1].* 1) Center for Human Genetics and Laboratory Diagnostics (AHC), Martinsried, Germany; 2) Chair of Social Pediatrics, Technical University of Munich, Munich, Germany.

The implementation of Next Generation Sequencing (NGS) in a clinical diagnostic setting opens vast opportunities through the ability to sequence all genes contributing to a certain indication simultaneously at a cost and speed that is superior to traditional sequencing approaches. Furthermore, these data allow the identification of small to medium sized copy number variations (CNVs). While the detection of single nucleotide polymorphisms (SNVs) and small insertions and deletions (InDels) is very straight forward, the reliable detection of CNVs from NGS enrichment data still poses several problems during data analysis. Here we present an approach for CNV calling in a clinical diagnostic setting as part of our multicentric Multiple Integration and Data Annotation Study (MIDAS). To call CNVs with sufficient sensitivity and specificity it seems impossible to rely on only one tool. Therefore, we combined different tools into an analysis pipeline and tried to find a combination of their respective results to optimize our calls. We incorporated ExomeDepth, CoNVaDING, XHMM, Canoes and VisCap in our current pipeline. The primary data in bam file format was produced by CLC Genomics Workbench as part of the primary data analysis. Interval files, GC-content or tables containing the coverage per region for the samples were produced by custom developed R and shell utility scripts. Results came in various formats and had to be normalized and condensed. There are different ways to combine data from the different tools. A union of the calls or a majority vote would improve the sensitivity but in routine diagnostics, specificity was the pressing requirement. The intersection of different tools removed the majority of false positives while true positives were called by all tools and remained in the final call set. Our approach contains a normalization of the scores given by each tool and a workflow to produce the final, scored calls based on unions and intersections of the primary results. Since some tools perform better than others for deletions or duplications (cite Canoes), a pipeline should be able to differentiate and use other combinations of the intermediate results. The assembly of the final calling and scoring mechanism varies therefore between different types of CNVs.

1338F

A generalized non-parametric genotype caller using an EM-like algorithm. *T.A. Benaglia[1], B.S. Carvalho[1,2].* 1) Department of Statistics, State University of Campinas, Campinas, SP, Brazil; 2) Brazilian Institute of Neuroscience and Neurotechnology / BRAINN, Campinas, São Paulo, Brazil.

Current biomedical research makes constant use of high-throughput technologies for genotyping. Microarrays and, more currently, sequencing experiments provide information at millions of variant sites, which are later investigated for association with a phenotype of interest. Prior to this association assessment, genotyping algorithms quantify the allele-specific abundance and classify subjects into genotype clusters. The series of statistical procedures involved in the aforementioned quantification and classification steps is complex and includes strong hypotheses on the distribution of the data. Models for microarray data assume that intensity signals follow an Exponential distribution combined with Gaussian error, while count data from sequencing are seen as results of a Negative-Binomial processes. We propose an alternative genotype caller that does not rely on assumptions regarding the distribution of the data. Our method uses a non-parametric approach combined with an adapted Expectation-Maximization (EM-like) algorithm to determine the membership likelihoods of each subject at every *loci*. Following this strategy, we use the maximum posterior probability to accurately perform genotype calling and consider the use of this statistic as genotype dosage, which can be used later for association studies, providing a strategy to account for genotype errors in data modelling.

1339W

Phenotype-specific information improves prediction of functional impact for noncoding variants. *C. Bodea[1,2], A. Mitchell[1], H. Runz[1], S. Sunyaev[2].* 1) Merck Research Laboratories, Boston, MA; 2) Brigham and Women's Hospital, Boston, MA.

A myriad of noncoding variants and association signals in the human genome are currently awaiting assignment of a function or the identification of causal alleles. Here, we introduce the computational framework PINES (Phenotype-Informed Noncoding Element Scoring) which evaluates the functional impact of noncoding variants in an unsupervised setting by integrating epigenetic annotations from diverse sources in a phenotype-dependent manner. A unique feature of PINES is that analyses may be customized towards genomic annotations of the highest relevance to phenotypes of interest. We show that PINES identifies functional noncoding variation more accurately than methods that do not use phenotype-weighted knowledge. To demonstrate its versatility, we show how PINES can be applied to systematically distinguish variants in enhancer regions from background variation; to delineate high-penetrance noncoding variants leading to Mendelian phenotypes; or to pinpoint epigenetic signals underlying noncoding alleles at GWAS loci. We further show how PINES can complement fine-mapping approaches and predict novel causal alleles from GWAS on Parkinson's and inflammatory bowel disease. Its flexibility and ease of use through a dedicated web portal establish PINES as a powerful new in silico method to prioritize and assign functions to noncoding genetic variants.

1340T

Faster genotype phasing and imputation for large-scale data. *B.L. Browning[1,2], S.R. Browning[2], X. Tian[2].* 1) Department of Medicine, Division of Med Genetics, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA.

We present new methods for genotype phasing and genotype imputation that have high accuracy and fast compute time for large-scale data. Both methods make use of a dynamic model state space for each target sample that is constructed from a set of reference haplotypes that continuously evolves across the genome. The genotype phasing method employs a novel approach that models haplotypes without modeling diplotypes. We compared our new genotype phasing method to Eagle2 (v2.3.2), the current state of the art, on UK Biobank data (n=150,000) for chromosomes 1, 5, 10, 15, and 20. Using current default settings, our phasing method reduced per-chromosome switch error rates and compute times by 23-36% and 23-39% respectively relative to Eagle2. In order to assess the performance of our genotype imputation method with very large reference panels, we made use of recent advances in demographic modeling and simulation methods to simulate 3 million UK European samples. The simulated data incorporate recurrent mutation and have a near-saturated marker density of 1 non-singleton single-nucleotide variant per 5 bases. Relative to Beagle 4.1, the current state of the art for reference panels with millions of samples, our new imputation method reduces compute time by a factor of 9 with no loss in accuracy when imputing 1000 target samples from 3 million reference samples in binary reference (bref) format. Extrapolating to a 3000 Mb genome, we estimate that the compute time for genome-wide imputation from 3M reference samples is < 12 min per sample on a 2.4 GHz compute server with 12 CPU-cores Our new genotype phasing and imputation methods will be incorporated in Beagle version 5 (http://faculty.washington.edu/browning/beagle/beagle.html).

1341F

Fast and easy pipeline for quality control and assurance for GWAS. *T.M. Brunetti[1], M. Campbell[1], N. Rafaels[1], T. Phang[1], M. Boorgula[1], S. Chavan[1], M. Daya[1], A. Shetty[1], R.A. Mathias[1,2], K.C. Barnes[1].* 1) Colorado Center for Personalized Medicine, University of Colorado, Denver, CO; 2) Johns Hopkins Asthma & Allergy Center, School of Medicine, Johns Hopkins University, Baltimore, MD.

Genome-wide association studies (GWAS) tend to have a high degree of variability which can be attributed to both biological and technical confounders. Although biological variation is generally addressed in most GWAS during the downstream analysis of the data, the technical variation needs to be addressed prior to any downstream analysis to ensure significant results are due to true biological differences rather than unwanted technical aberrations. These technical factors can often arise from the improper handling of quality assurance measures within the initial dataset such as poorly performing SNPs and samples. For example, failures can arise due to poor cluster separation, large batch effects, low genotyping rates, and gender discrepancies. For these reasons, we have designed an automated, high-throughput, customizable pipeline implemented in Python, for assessing the quality and filtration of geno-typing data for GWAS. The pipeline has been tested and optimized using the Multi-Ethnic Global Array (MEGA) from Illumina which contains over 1.7 million SNPs. We tested this pipeline across 1,667 samples and produced thorough results with an average run time of ~7 minutes on a standard laptop. The pipe-line outputs a set of deliverables that contain detailed PDF reports and files, as well as PLINK and VCF files populated with metrics for a wide range of anal-yses, including CNV analysis. These files can be used directly in downstream analysis. The pipeline filters out poorly performing SNPs and samples, and reports which metrics the SNPs and samples fail and the value of each failure. In addition, several summary statistics of passing samples and SNPs with both quantitative and visual results are published. Due to its speed and ability to be easily customizable, the user has the ability to change, optimize, and run the pipeline several times or in parallel to determine what thresholds may be most optimal for their experiment, although carefully chosen defaults have been se-lected should the user not want to optimize the pipeline. The automated quality control and assurance pipeline is being implemented at the Colorado Genetics Research Facility and the Colorado Center for Personalized Medicine at the University of Colorado Anschutz Medical Campus.

1342W

An empirical strategy to screen markers on case-control genomic stud-ies. *B.S. Carvalho[1,3], B. Henning[2,3], M.R. Rodrigues[2,3], F. Rossi[2,3], T.A. Benaglia[1], I. Lopes-Cendes[2,3].* 1) Department of Statistics, State University of Campinas, Campinas, SP, Brazil; 2) Department of Medical Genetics, State University of Campinas, Campinas, SP, Brazil; 3) Brazilian Institute of Neuroscience and Neurotechnology, BRAINN, Campinas, SP, Brazil.

The development of Precision Medicine (PM) depends strongly on the statistical comparison of genomic data from control subjects to patients of the phenotype of interest. Next-generation sequencing is one of the most used technologies for this task and, due to its nature of massive parallel sequencing, downstream analysis methodologies should be able to handle thousands, if not millions, of comparisons. The multiple comparison problem is a well-known issue in statistical genetics, as it affects significantly the power to detect association between markers and phenotypes. We present a strategy that combines the latest developments in machine learning with high-through-put sequencing data to identify smaller sets of candidate markers that show evidences for association with the phenotype of interest. More specifically, we call variants in control subjects and epilepsy patients using the Best Practic-es for GATK, after mapping the reads using BWA to the genome reference HG38. Later, we train a random forest (RF) on 70% of our data and optimize the parameters using the remaining 30%. On our dataset, containing 199 healthy individuals and 203 epilepsy patients, we found that the best model for discrimination between case and control statuses in our dataset resulted from a random forest built with 100 decision trees and 16 variables as the maximum depth. On both training and optimizing samples, we obtained MSE=0.24 and AUC=0.63. The error rate was 42% on the training dataset and 38% on the optimizing dataset. Although this strategy shows a low discrimination rate between cases and controls, we use this optimized RF to compute the variable importance for each of the markers used as predictors. Given a threshold for the cumulative percentage of variability explained by the RF, we can then select a subset of markers that are potentially associated to the phenotype and study its overall effect on the response via techniques like enrichment. This subset can be also used by the researcher to generate hypotheses that will be then assessed via specific statistical models. When selecting markers that are responsible for only 10% of the explained variance, we identified 146 markers (out of 297) that were significantly associated with the phenotype at the FDR level of 10%.

1343T

Accurate quantification of allele-specific methylation from genetically diverse population. *K. Choi, A. Srivastava, R. Korstanje, G. Churchill.* The Jackson Laboratory, Bar Harbor, ME.

Quantifying genome-wide hyper and hypo-methylation is crucial for characterizing the regulatory landscape of gene expression. Recent advances in next-generation sequencing technologies, such as reduced representation bisulfite sequencing (RRBS), enable investigators to assess methylation status at single CpG dinucleotide resolution. In RRBS, unmethylated cytosines are converted to uracil, which requires read aligners to efficiently process higher number of mismatches. But, since the performance of Burrows-Wheeler-based aligners quickly degrades as the number of mismatches increases, many existing methods align to four separate three-letter reference genomes — C-to-T and G-to-A converted genomes for both strands. Although this strategy is generally adequate, it does not perform well when investigators seek to quantify allele-specific methylation (ASM). This is because the alignments of each read to four different reference genomes must be post-processed in order to identify the allelic origin of the read. We propose a novel ASM quantification pipeline, **asmtools**, in which we reconstruct individualized diploid genomes by incorporating annotated polymorphisms into the reference genome. Individualized diploid genomes identify the allelic origin of reads automatically and reduce the overload of increased mismatches for read aligner programs. Our method aligns reads to the digested regions of the individualized genomes, efficiently disregarding C-to-T or G-to-A mismatches. This implies our approach does not require the creation of three-letter converted genomes or three-letter converted reads and subsequent post-processing of alignments. Our method also minimizes the information loss caused by discarding multi-mapping reads as we developed a statistical model that disambiguates the origin of each read. Since a probability is assigned to each alignment and consequently to all the mismatches in it, we can evaluate *expected* ratio of methylated probe for each allele of CpG sites. We evaluated the performance of our pipeline using a kidney RRBS data sampled from Diversity Outbred (DO) mice, a genetically diverse population consisting of three different age groups (6, 12, and 18-month) of both sexes. In the QTL analysis, we found thousands of ASM sites are local methylation-QTLs, on which our approach generated higher LOD scores than other existing methods in most cases. These results represent the effectiveness of our pipeline for allele-specific methylome survey.

1344F

Development of a framework for laboratory information tracking in support of research and clinical operations. *B. Craig, B. Tibbils, K. Topper, S. Griffith, L. Watkins, K. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins Genomics (JHG), IGM, JHU-SOM, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR's current genotyping Laboratory Information Management System (LIMS) was designed to handle the tracking of detailed information for large numbers of samples going through specific genotyping assays. This LIMS has over time become difficult to maintain and extend as newer resources and NGS services have emerged. A third party LIMS was selected for NGS applications that addressed the widely different set of tasks and information to be stored for sequencing services. Annual licensing costs and requirements to re-write custom code and upgrade to newer releases to maintain support add to the difficulty of sustaining two LIMS deployments. These deficiencies combined with the expansion of research and clinical genomic services at JHG (jhgenomics.jhmi.edu) have highlighted the need to create a new LIMS to support the breadth of current and future protocols. These rapidly changing laboratory protocols require reliable tracking of large amounts of workflow information including a record of sample movements, identification of robot resources, reagent association and user comment tracking. Here we describe the development of a MySQL database backed Java framework consisting of tools to reduce disruption to lab processing with module hot swapping, limit database schema complexity and increase reusable graphical user interface (GUI) components to flexibly build and extend workflows for any protocol driven technology.

1345W

A new method for analysis of sequencing reads without reference genomes. *H. Dai[1], Y. Guan[1,2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX.

The mapping-to-a-reference-genome based approach to analyze sequencing data has two challenges. One is limited ability to map short sequencing reads associated with INDELs, structural variants and repetitive regions. The other is incompleteness/unavailability and low quality of reference genomes. Here we present Nubeam (nucleotide be a matrix), a method to quantify difference between sets of sequencing reads without reference genomes. In the method, a nucleotide is represented by a matrix, a read is the product of a sequence of matrices, and a number, which is the trace of the product matrix, is assigned to the read. Since matrix multiplication preserves the order, similar sequences are represented by similar numbers. The difference between sets of reads can be quantified by the distance between empirical distributions formed by corresponding numbers. We first used simulations to demonstrate that Nubeam can reconstruct true phylogeny. We simulated sets of reads from trees representing different degrees of sequence divergence and assessed the accuracy of phylogeny reconstruction. We validated that the distance between empirical distributions could measure the genetic distance between samples. We then applied Nubeam on human microbiome project (HMP) WGS metagenomics samples collected from various body habitats. Using mapped reads, dimension reduction of resulted pair-wise distance matrix showed the primary clustering of samples is by body habitats, comparable with the results in published HMP studies. Interestingly, using unmapped reads only, oral and gastrointestinal samples showed less distinct clustering, while skin, nasal and, particularly, urogenital samples displayed much greater variation. Our analysis revealed the contribution of unmapped reads in metagenomics studies. We also applied Nubeam on the developmental transcriptome RNA-seq data of social amoeba *Dictyostelium discoideum*. Dimension reduction of resulted pair-wise distance matrix revealed that developmental transcriptome showed gradual changes interspersed with a few large shifts, indicating the slow change of transcriptome in certain developmental stages followed by rapid change of gene expression when the developmental stages switched. Our analysis recapitulated the findings of the original study. The applicability of Nubeam on the datasets demonstrated the method's efficiency in capturing the characteristic of sequence and ability in investigating the genetic contribution of unmapped reads.

1346T

Gene identifiers: More than just a number. *P. Denny, B. Braschi, K. Gray, R. Seal, S. Tweedie, B. Yates, E. Bruford.* HUGO Gene Nomenclature Committee (HGNC), EMBL-EBI, Hinxton, United Kingdom.

The HUGO Gene Nomenclature Committee (HGNC, www.genenames.org) is the only organisation worldwide assigning standardized gene symbols and names to all genes in the human genome. As human genomics increasingly becomes a key aspect of clinical practise it is critical that everyone knows exactly which gene they are referring to. In this context not only are gene symbols indispensable for everyday communication, stable gene identifiers (IDs) are also very important for cataloguing and datamining. As well as assigning a gene symbol and name the HGNC also creates a unique HGNC ID for every gene, in the format HGNC:#. While specific circumstances can result in a change in gene symbol, e.g. updating an uninformative placeholder symbol when the function of the encoded protein is elucidated (e.g. C14orf169 > RIOX1, ribosomal oxygenase 1), this would not result in a change in the HGNC ID (HGNC:20968) which is intrinsically linked to the gene locus and sequence. The HGNC ID is also a very reliable way to link between other major genomic resources; while these resources also aim to have stable gene IDs, changes for example in genome assembly, gene models and coding status can trigger a new ID being created for the same gene, which would not result in a change in the HGNC ID. Only in the rare circumstances of a gene merge or split would an HGNC ID potentially be altered; even in such a case the previous ID would remain fully traceable and linked to the new ID(s). Links between HGNC IDs and external resources are regularly updated by both automated quality control and manual curation, and HGNC IDs are widely referenced in biomedical resources. As such, we believe HGNC IDs are the most stable and practicable human gene IDs available. We strongly advise authors to reference the HGNC ID in article metadata (or at least once in the abstract/introduction) when publishing on a human gene. This policy is also being advocated by the UK Wellcome Trust funded Transforming Genetic Medicine Initiative (TGMI, www.thetgmi.org/), and we plan to contact journals and publishing houses to promote the utility of HGNC IDs, and model organism database gene IDs in general, in all articles. Not only would using these IDs give readers an immediate link to databases with specific information on the gene(s) in question alongside multiple links to external resources, they would also serve to clarify exactly which species were being discussed, a common ambiguity in many publications.

1347F

Cohort based precise interpretation algorithms of copy number variants detected by exome sequencing on individuals of severe disorders. *X. Dong[1], B. Liu[1], L. Yang[2], H. Wang[2], B. Wu[2], R. Liu[2], H. Chen[2], X. Chen[2], S. Yu[2], F. Xia[3], W. Zhou[1,2], Y. Lu[1,2].* 1) Children's Hospital & Institutes of Biomedical Sci, Fudan University, Shanghai, Shanhai, China; 2) Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defects, The Translational Medicine Center of Children Development and Disease of Fudan University, Shanghai, Shanghai, China; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA.

Background Copy number variations (CNVs) is a major cause of severe genetic disorders. Various algorithms have been developed to solve the estimation strategy or statistical model of CNV detection from WES data. However, lack of clinical phenotype integration or the ability to combine CNV analysis with SNV have made these tools not available for clinical-relevant CNVs evaluation. A more comprehensive variation detection and analysis pipeline is desirable to cover both CNVs and SNVs in clinical practice. Methods We established *a pipeline for clinical NGS-involved CNV detection* (PICNIC) to assist identifying potential disease-causing (PDC) CNVs and provided an *SNV and CNV integrated* (SCI) strategy for combinational CNV and SNV analysis. PICNIC used depth-based algorithm to detect CNV from BAM files, then annotated and filtered candidate variants with gene function, literature annotation and HPO-standardized clinical phenotype. The SNV and CNV Integrated (SCI) strategy could provide combinational potential disease causing (PDC) variants based on gene's inheritance model, zygosity of variants, and clinical phenotype matching details. Results The PICNIC and SCI strategy were applied to assist clinical diagnoses in a large Chinese pediatric cohort with sever genetic disorders, including 861 patients and 1,304 related pedigree samples. Altogether, 16,081 raw CNV calls were analyzed , including 3,610 duplications (DUPs) and 12,471 deletions (DELs). After PICNIC refinement, 1,024 CNVs (669 DELs, 355 DUPs) remained for further SCI analysis. Finally, 70 PDC CNVs (47 DELs, 23 DUPs) were identified, including 52 CNVs larger than 1 Mb (32 DELs, 20 DUPs), 10 CNVs smaller than 1 Mb covering multiple genes (7 DELs, DUPs) and 8 DELs covering single gene. These PDC CNVs were found in 63 patients (7.3%), including 50 SNV-negative cases. Specially, compound heterozygous SNV-CNVs were identified in 7 cases, and 25 out of 28 randomly selected PDC CNVs were confirmed with other test. Altogether, an increase of 6.6% diagnostic rate was achieved by PICNIC and SCI compared with traditional WES-SNV analysis. Conclusion In this study, our pipeline is important for the identification of PDC-CNVs and PDC compound heterozygous (CH) which were not covered in previous WES analysis. The combinatorial analysis could improve the positive diagnosis rate of rare genetic disorders. In the rapid development of clinical genetic testing, both PICNIC and SCI strategy deserve a widespread application.

1348W

Fabric Genomics' Opal variant interpretation platform enables rapid, whole genome analysis turnaround in under an hour. *A.P. Fejes, B. Stade, E.S. Kiruluta, M.G. Reese.* Fabric Genomics, Oakland, CA.

Opal Clinical is well suited for use in clinical applications, supporting delivery of rapid diagnostic outcomes on whole genome data. NGS-based diagnostics are poised to revolutionize the way clinical labs identify and interpret inherited or de novo diseases. However, the long sequencing and processing times make it difficult to incorporate into existing workflows. For diagnostics purposes, a raw sample must be sequenced, aligned to the reference genome, called for variant bases, annotated and interpreted. While large-scale sequencing for 30x whole genomes, suitable for diagnostic purposes, continues to improve, accurate genome annotation and interpretation remains one of the most challenging and time consuming components of the end-to-end process. To solve this issue, Fabric Genomics has designed a platform that performs variant annotation for a whole genome in ~8 minutes. It provides an interface for interpretation that can yield immediate results for many diseases and enables deep genome interrogation and interpretation. The quality of the annotations is ensured through robust regression testing. As each data source is upgraded, or source code is modified, data annotation tests are performed to ensure the accuracy of the annotations is not compromised. This ensures that Fabric Genomics' platform remains both up to date and accurate in its predictions. We discuss here some of the benchmarks of our complete genome annotation platform, as well as some of its key features and design principles. The annotation engine is currently in use widely, including in the UK's 100,000 Genomes Project. It is also in use at Rady's Children Hospital, where they are launching a rapid genome service to return results for babies in the newborn intensive care unit (NICU) and pediatric intensive care unit (PICU) in 24 hours.

1349T

Minerva & Me: Public participation in research: Crowdsourcing for computational phenotyping method development. *M. Ferlaino[1], M. Alvi[2], J. Barret[7], M. Dawson[2], J. Diprose[8], R. Esnouf[8], N. Hallowell[9], G. Baynam[3,4], H. Brunner[5], D.R. FitzPatrick[6], A. Zisserman[2], C. Nellåker[1], Minerva Consortium.*

1) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, UK; 2) Visual Geometry Group, University of Oxford, UK; 3) Genetic Services of Western Australia, King Edward Memorial Hospital, WA, Australia; 4) Western Australian Register of Developmental Anomalies, King Edward Memorial Hospital, WA, Australia; 5) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 6) MRC Human Genetics Unit, IGMM, University of Edinburgh, Western General Hospital, Edinburgh, UK; 7) Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK; 8) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 9) Ethox Centre, Nuffield Department of Population Health, University of Oxford, UK.

The Minerva Consortium is an open academic consortium that has been formed to allow computational phenotyping development together with clinician researchers. The mission is to enable research into the use of image analysis for the diagnosis of diseases, prevent data siloing, and foster further healthy competition between various image phenotyping approaches. There is huge potential for new deep phenotyping and machine learning approaches to have significant impact in health care pathways. However, big data machine learning approaches are reliant on having access to large data sets. Here we present Minerva & Me, which allows families, patients with rare diseases, and the broader public to directly participate in a research project into computational phenotyping. A dynamic consent model allows participants persistent granular control of how their data can be used in research. All data contributed remain under the users' control, including the ability to delete the data. The information to participants has been formulated in an easy to read version to allow active consent as broadly as possible. Our aim is to build on recent developments in using computer vision approaches in the context of rare disease phenotyping. Approaches such as the Clinical Face Phenotype Space have great potential to provide diagnosis decision support, statistical support for causative variants, and to augment current undiagnosed patient matchmaking. All such methods are vulnerable to implicit biases if the underlying data sets for the research are biased in terms of equitable representation of ancestral backgrounds, ages and genders. It is thus critical for the success of deep phenotyping approaches that platforms for collecting data sets from around the world are built. Minerva & Me has been built to have the requisite ethical, data security, and legal frameworks to allow open participation. It has been built to evolve to be able to allow other research projects to add consent and participation questions – and modular to allow additional data inputs in future. To regulate and control such developments an Advisory Board has been assembled with members from patient representative groups Unique, Rare Disease UK, and Rare Voices Australia, clinical genetics, legal, as well as data security experts. Together, the Minerva Consortium will be the means through which computational phenotyping is developed as part of the clinical pathway and ultimately to patient benefit in health care settings.

1350F

Prediction of gain-of-function and loss-of-function mutations using Combined Annotation Dependent Depletion (CADD). *A. Ghosh[1,2], A. Navarini[1,2].*
1) Department of Dermatology, University Hospital Zurich, Zurich, Switzerland; 2) Competence Center Personalized Medicine UZH/ETH, Zurich, Switzerland.

Introduction: Next generation sequencing (NGS) technologies can identify on average 20,000 missense variants per exome, out of which very few can be held responsible for causing the underlying disease. Therefore, it is very important to interpret the functional significance of these variants. A number of in silico bioinformatics tools have been already developed for the similar purpose, out of which CADD, SIFT (Sorting Intolerant From Tolerant) and PolyPhen2 (Polymorphism Phenotyping v2.0) are most commonly used. While SIFT and PolyPhen2 predict the deleteriousness of variants, CADD scores the variants based on a supervised machine learning approach (Support Vector Machine). The aim of this study is to utilize the CADD score to predict whether the variant causes a gain-of-function or loss-of-function at the protein level. Material and Methods: A large number of mutations was extracted from public databases ClinVar and DisGeNET, namely 435 gain-of-function (GOF) and 3,350 loss-of-function (LOF) variants. All of these variants have supporting evidence of their functional significance in terms of literature texts, clinical reports and research projects. Results: Our results show two peaks of mean CADD scores across GOF and LOF variants ($p < 0.05$) indicating a non-random difference. This difference along with the CADD score distribution over all the variants helped us to choose a cut-off score of GOF and LOF variants. The cut-off scores, in combination with either SIFT or PolyPhen2, resulted in correct prediction of 71% GOF and 75% LOF variants. Discussion: Therefore, these CADD cut-off scores can be useful to predict the functional significance of a missense variant. However, they should be used more as an indicator as they need to be interpreted with further evidence on pathogenicity. .

1351W

Efficient pipeline for whole genome simulation and summary statistic calculation with flexible demographic models. *A.L. Gladstein[1], C.D. Quinto-Cortes[2], J.L. Pistorius[3,4], M.A. Skodiak[5], L.K. Gantner[6], A.E. Woerner[7], J.C. Watkins[3,6], M.F. Hammer[3,8].* 1) Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ; 2) National Laboratory of Genomics for Biodiversity, Irapuato, Guanajauto, Mexico; 3) BIO5 Institute, University of Arizona, Tucson, AZ; 4) CyVerse, University of Arizona, Tucson, AZ; 5) School of Information, University of Arizona, Tucson, AZ; 6) Department of Mathematics, University of Arizona, Tucson, AZ; 7) Department of Genetics, University of Arizona, Tucson, AZ; 8) ARL Division of Biotechnology, University of Arizona, Tucson, Arizona.

Approximate Bayesian Computation (ABC) has become a widely used method for demographic inference. In order to use ABC millions of simulations per model and summary statistics of the data are required and small loci are usually used. We have developed a pipeline that utilizes the coalescent simulator MaCS and calculates a variety of genomic summary statistics in Python. In order to reduce storage requirements we bypass all printing by MaCS. We are able to simulate a complex demographic model of 100's of samples of whole chromosomes and calculate summary statistics in a runtime as little as a few seconds and an average of 1hr, with max memory use of 4Gb. This means that we are able to optimally run our pipeline in parallel on high throughput computers, to complete millions of complete chromosome simulations in a few weeks and use little storage. We conveniently read in a user input file containing model parameters and priors, allowing for our pipeline to be easily applied to any demographic model. We provide the option to model ascertainment bias so that ABC can be used on SNP array data. Our pipeline will be publically available on GitHub. We also provide a framework for rigorous testing and debugging in case users would like to add their own functions.

1352T

The European Variation Archive: A repository for short and structural genomic variation. *C.Y. Gonzalez Garcia, G. Saunders, J. Kandasamy, J.M. Mut Lopez, P. Arce Garcia, S. Venkataraman, T. Smith, T. Keane.* Archive Infrastructure, EMBL - European Bioinformatics Institute, Cambridge, United Kingdom.

The European Variation Archive (EVA, https://www.ebi.ac.uk/eva) is a primary open repository for archiving, accessioning, and distributing genomic variation data including single nucleotide variants, short insertions and deletions, and larger structural variants in any species. Since their launch in 2014, the EVA and its sister project the Database of Genetic Variants Archive (DGVa) have archived more than 520 million unique variants across more than 200 studies, containing approximately 300,000 samples across 27 species. A key function of the EVA as a long term data archive is to guarantee that study metadata and submitted VCF files are valid, and hence useful for data analysts. All VCF files are checked using the EVA VCF Validation Suite, which supports all 4.x versions of the VCF specification. We are committed to maintain and improve the VCF specification as part of our collaboration with the Global Alliance for Genomics and Health (GA4GH). The EVA also provides stable identifiers so that discovered variants and alleles can be referenced in publications, cross-linked between databases and integrated with successive reference genome builds. The EVA currently peers with the NCBI-based dbSNP and dbVar databases to form a worldwide network for exchanging and brokering submissions. From Q4 2017, issuing and maintaining locus identifiers will be divided by taxonomy: dbSNP will be responsible for human locus accessioning, and EVA responsible for all non-human loci. The EVA provides a comprehensive view of all these data through its website, mainly through the study browser and the variant browser. The study browser allows users to explore all archived datasets and download them via FTP. The variant browser allows fine-grained queries based on annotation generated using Ensembl's Variant Effect Predictor, protein substitution scores from SIFT and PolyPhen, and allele frequencies. Data can be downloaded at study, file or variant level, in VCF and TSV formats. More recently, we have been working on displaying variant data filtered by the user for a specific annotation version, and implementing a web services API for advanced search of structural variation. Future plans involve improving the integration with other EMBL-EBI services such as the European Genome-phenome Archive (EGA), to support browsing of controlled-access human datasets. All our software is open-source, licensed under Apache 2 and freely available in GitHub (https://github.com/ebivariation/).

1353F

The analysis of negative selection with heterogeneous Mendelian models in coding, as well as noncoding regions for cohorts of diverse undiagnosed diseases. *F. Gu, M. Gordon, T. Markello, S. Macnamara, L. Vlahos, D.A. Adams, W.A. Gahl.* National InstituteOffice of the Clinical Director, NHGRI, and NIH Undiagnosed Diseases Network and Program, NIH, Bethesda MDs of Health, Washington, DC.

   A previous study of negative selection shows excess deleterious de novo variants in the affected cohort compared to healthy control siblings, but that analysis was done solely for the de novo model and for a single disease (autism). To assess the overall fitness landscape of an individual of any given disease, we developed an agnostic diagnostic pipeline that allows high confidence genotyping in low-coverage regions and permits analysis on coding and noncoding regions for various Mendelian models including homozygous recessive, de novo dominant, compound heterozygous recessive, x-linked, deletion with transmission, and novel deletion in very well-sequenced positions. This agnostic diagnostic pipeline presumes nothing about phenotype or molecular causes of these diseases. It was performed forwardly on affected probands, and again backwardly on their healthy siblings reversing the affectedness-statuses. The cohort consists of 96 non-consanguineous quartets or quintets in the Undiagnosed Disease Program of the intramural NIH, all with one affected proband and one or two unaffected siblings. Deleterious appearing candidates were assessed primarily by CADD and population frequency statistics. A combined score for compound heterozygous pairs strongly weighted towards equality in the deleteriousness of the two alleles was used to assess overall deleteriousness of that compound heterozygous pair. The only non-automatic part of the pipeline is the manual curation of the bam files. The analysis showed a higher number of disease-causing variants in the affected probands compared with their healthy controls. This demonstrates power of the current variant analysis relative to chance alone as represented by nonequality of deleterious variants in probands versus unaffected sibling controls and suggests the candidate variants generated from this agnostic genetic analysis are likely to contain the true causal variant.

1354W

Graphtyper: Population-scale genotyping using pangenome graphs.
B. Halldorsson[1,3], H. Eggertsson[1,2], H. Jonsson[1], S. Kristmundsdottir[1,3], E. Hjartarson[1], B. Kehr[1], G. Masson[1], F. Zink[1], A. Jonasdottir[1], A. Jonasdottir[1], I. Jonsdottir[1,4], D. Gudbjartsson[1,2], P. Melsted[1], K. Stefansson[1,4]. 1) deCODE genetics, Reykjavik, IS, Iceland; 2) School of Engineering and Natural Sciences, University of Iceland, Reykjavík, Iceland; 3) School of Science and Engineering, Reykjavik University, Reykjavík, Iceland; 4) Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland.

   A fundamental requisite for genetic studies is an accurate determination of sequence variation. While human genome sequence diversity is increasingly well characterized, there is a need for efficient ways to utilize this knowledge in sequence analysis. Here we present Graphtyper, a publicly available novel algorithm and software for discovering and genotyping sequence variants. Graphtyper realigns short-read sequence data to a pangenome, a variation-aware graph structure that encodes sequence variation within a population by representing possible haplotypes as graph paths. Our results show that Graphtyper is fast, highly scalable, and provides sensitive and accurate genotype calls. Graphtyper genotyped 87.9 million sequence variants in whole-genomes of 28,075 Icelanders using less than 100,000 CPU days, including detailed genotyping of six human leukocyte antigen (HLA) genes. We show that Graphtyper is a valuable tool in characterizing sequence variation in population-scale sequencing studies.

1355T

PathwayMatcher: Direct mapping of omics data to the Reactome pathway knowledgebase. *L.F. Hernández Sánchez[1,4], O. Helgeland[1,3], B. Burger[5,7], H. Barsnes[5,7], S. Johansson[1,4,6], A. Fabregat[2], P.R. Njølstad[1,3], H. Hermjakob[2], M. Vaudel[1,4].* 1) K.G. Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Norway; 2) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; 3) Department of Pediatrics, Haukeland University Hospital, Bergen, Norway; 4) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 5) Proteomics Unit, Department of Biomedicine, University of Bergen, Norway; 6) K.G. Jebsen Centre for Neuropsychiatric Disorders, University of Bergen ,Norway; 7) Computational Biology Unit, University of Bergen, Norway.

Many studies performed in biomedical and clinical sciences is to uncover the biological mechanisms responsible for a phenotype, to provide the means to prevent or cure diseases. Identification of the underlying mechanism remains often elusive, and consequently the cause of the clinical observation cannot be confidently assessed. It is now possible to associate a disease condition with genetic variants, gene expression levels, abundances of proteins, their posttranslational modification states or other biochemical compounds. Here, we present PathwayMatcher, an open source bioinformatics tool to map clinical omics data to the Reactome knowledgebase (reactome.org). Reactome is a free, open-source, open-data, curated and peer-reviewed knowledgebase of biomolecular reactions and pathways. PathwayMatcher is written in Java under the Apache licence 2.0. The input Omics data can be gene names, Single Nucleotide Polymorphisms (SNPs), peptides sequences, protein accessions, and peptides or proteins carrying posttranslational modifications (PTMs). Variant effects are inferred using the Ensembl Variant Effect Predictor (VEP), and protein level data are mapped directly to the Reactome Neo4j graph database installed locally. In the future it will be possible to integrate these new types of mapping to the online Analysis Service of Reactome. PathwayMatcher is available as Docker container image, which is a standalone executable package with all the necessary libraries, including the latest version of VEP and Reactome. It can run offline and does not require the installation of dependencies besides Docker itself. It is possible to integrate PathwayMatcher in genomics bioinformatics pipelines, retrieving biochemical reactions and pathways that can be affected by SNPs and Copy Number Variation (CNV). The functional results can be used to inspect variant distributions among pathways, identify upstream and downstream processes, and design functional assays. In the case of proteomics experiments, it is further possible to map the peptide or protein level results to Reactome, including PTMs, without gene-centric summarization. Results can be visualized in the graphical interface of Reactome to see how the expression patterns of the proteins vary over time or between conditions, allowing visually distinguishing the regions of highest variation and guiding the design of future experiments.

1356F

Development and organization of genomics metadata by ENCODE. *B. Hitz, C. Sloan, U. Baymuradov, E. Chan, C. Davis, I. Gabdank, A. Narayana, K. Onate, J. Hilton, T. Dreszer, J.S. Strattan, F. Tanaka, K. Dalton, J.M. Cherry.* Genetics, Stanford University, Palo Alto, CA., United States.

The goal of the Encyclopedia of DNA Elements (ENCODE) project is to annotate functional regions in the human and mouse genomes. Functional regions include those that code protein-coding or non-coding RNA gene products as well as regions that could have a regulatory role. To this end, the project has surveyed the landscape of the human genome using over 40 high-throughput experimental methods in more than 400 different cell and tissue types, resulting in over 8500 experiments. These datasets have been supplemented with over 4500 complementary assays from other projects; the Roadmap for Epigenomics Consortium, the Genomics of Gene Regulation project, and our sister projects modENCODE and modERN comprising genomics experiments in Drosophila and C. elegans. A description of these datasets, collectively known as metadata, encompasses, but is not limited to, the identification of the experimental method used to generated the data, the sex and age of the donor from whom a tissue sample was taken, and the software used to align the sequencing reads to a reference genome. Defining and organizing the set of metadata that is relevant, informative, and applicable to diverse experimental techniques is challenging. Here we describe how metadata are organized at the ENCODE DCC and define the metadata standard. A metadata standard encompassing key experimental variables as well as software versions, parameters, and quality metrics determined from uniform processing of genomics data is critical to the comparability and reproducibility of the results and findings of these experiments and computational analyses . Understanding the principles and data organization will help improve the accessibility of the ENCODE datasets as well as provide transparency to the data generation processes. This understanding will allow integration of the diverse data within the ENCODE consortium as well as integration with related assays from other large-scale consortium projects and individual labs.

1357W

Compare HLA typing by next generation sequence methods: An example in Taiwan Biobank database. *C. Hsiung[1,2], M. Su[1,2], H. Chu[1,2], C. Shen[1,2].* 1) Taiwan Biobank,Academia Sinica, Taipei, Taipei, Taiwan; 2) Institute of Biomedical Sciences,Academia Sinica, Taipei, Taipei, Taiwan.

Human leukocyte antigens (HLA) genes play an important role in organ transplantation and are associated with autoimmune and infectious diseases. The gene family contains the most polymorphic variation in humans. Current DNA genotyping methods, such as the Sanger sequencing-based typing (SBT) and sequence-specific oligonucleotide (SSO) method which are extensively used in HLA typing, generally result in ambiguous typing results because of the limitations of oligonucleotide probe design or phasing ambiguity of HLA allele assignment. Next Generation Sequencing(NGS) Technology performs massively parallel sequencing, which leads to reconstruct allele sequences from overlapping gene fragments, and to identify the primary structure, providing the region phasing information. Currently, several NGS platforms, such as Ion Torrent, PacBio and whole genome sequence, which use different sequencing chemistry and technologies may be expected to solve the phase ambiguity problem. In order to compare the HLA typing accuracy obtained from different platforms, we conduct a pilot study by using the Taiwan Biobank database. The Taiwan biobank database has collected information from over 80,000 participants and whole-genome sequence has been performed in 1,000 individuals. We randomly selected 20 participants, whose sequences have whole genome-wide sequenced by illumina Hiseq 2500 sequencer, are typed by two different NGS platforms, including Ion Torrent and PacBio, and SBT method. We use SBT method as the gold-standard to estimate the accuracy of HLA tying at 2-digt, 4-digit and a super-high resolution (formerly known as the 8-digit level of resolution). The results show that the accuracy for HLA class I typing at 2-digit and 4-digit resolution, are not significantly difference between these three NGS methods, but the accuracy which is used BWA analysis of whole genome sequence data of the HLA-A typing at 8-digit resolution shows much lower (20%). In HLA class II, the accuracy of which is measured by HLA-DQB1 and HLA-DRB1 typing at 2-digit and 4-digit, is not different between the three NGS methods, but in HLA-DQB1 at 8-digit resolution the accuracy from PacBio analysis pipeline is lower. Our finding suggests that HLA typing at 2-digit and 4-digit resolution by the use the three NGS methods may wildly use in clinical and laboratory application, but at 8-digit Ion Torrent HLA typing is a better method.

1358T

Target Gene Notebook: Connecting genetics and drug discovery through enabling computational and logistical tools. *J. Hutz[1], J. Wierzbowski[1], A. Kirby[3], J. Heymont[1], M. Daly[2], M.P. Reeve[1].* 1) Eisai, 4 Corporate Drive, Andover, MA 01810; 2) Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts General Hospital, Boston, MA; 3) 3309 Bayview Dr, #B, Manhattan Beach, CA 90266.

Genetic studies have generated a surfeit of robust associations in recent years, providing potential insights into the causal mechanisms of many human diseases. While many browsers and databases provide access to unfiltered genome annotations and genetic results, very few of these facilitate the development of therapeutic hypotheses due to limitations in the presentation and interpretation of functional and molecular information. Linking associations to functional biological information is essential to translating genetic insights to drug discovery ; however, there is currently no way to maintain group curation of relevant genetic and functional information and to integrate it with proprietary experimental data. We introduce an e-notebook to facilitate the organization of results and provide freely-available software, Target Gene Notebook, for the purposes of assisting therapeutic target evaluation and creating durable institutional or public knowledge bases. With a gene or locus as the focal point, the software identifies candidate published associations and allows the user to review and store an assessment of each finding. Qualitative assessments are pre-computed via an automated assessment of statistical robustness and the variants' known or predicted functional. These analyses utilize published LD patterns across global populations derived from reference panels and derive 'credible sets' of variants. Public resources of relevance to target evaluation (e.g., variation patterns in ExAC/Gnomad, expression patterns in GTEx) are also presented. In addition, eQTLs as defined from GTEx and other sources are included as associations (complete with credible set definition) for visual assessment of phenotype – eQTL overlaps. The notebook can be extended to include other association modalities (e.g., protein-QTLs) and variant types (e.g. Mendelian disease mutations). Implementation is such that private instances can be shared across groups or sites as needed - potential users include pharma (for tracking knowledge of targets) and disease consortia for collating results and interpretations. Target Gene Notebook will be released in Fall of 2017.

1359F

A graph-based pipeline to evaluate common structural variations based on haplotypes and reassembly. *S. Ji, L. Li, G. Demir, P. Komar, D. Kural.* SBGD Inc., Cambridge, MA.

Structural variations (SV) are a greater source of human genetic variation than single nucleotide variants (SNVs). However, due to the complex nature of these events, complicated downstream processing and multiple algorithms are required after initial alignment of sequenced reads. The human genome reference in its linear form cannot represent these events in their native form. Recently, multiple methods that applies a graph-based approach by representing genetic variants as distinct paths through the graph have been proposed, but mostly focused on SNVs. Here, we aim to build a graph-based SV genotyping pipeline that aligns reads originating from known SVs directly to their native positions and genotypes them. For this purpose, we first established a pipeline that evaluates common SVs and resolves SV breakpoints prior to inclusion in the graph. This is due to inconsistent breakpoints negatively affecting the graph-aligner's performance. In order to select SVs with consistent breakpoints originating from single evolutionary events, we devised multiple metrics to identify SVs common in the population (>5%) with well-defined haplotypes. An initial set of SVs in Hardy-Weinberg equilibrium with tagging SNPs (r2 > 0.9) were collected from samples of the 1000 Genomes Project phase 3. These SVs were evaluated against the Database of Genomic Variants (number of studies, platforms, algorithms, frequency in cohort, etc.), as well as SVs from the Genome of the Netherlands and the Icelandic non-repetitive, non-reference sequences (based on concordance between frequency of SV and their tagging SNPs). For each SV region, reads were sampled from individuals with tag-SNPs suggesting a homozygous SV and reassembled per individual. Scores based on consistency of breakpoints in reassembled contigs across individuals and consistent haplotypes were then collated with other metrics to provide a quantitative measure of confidence regarding the breakpoints of each SV. The pipeline can also evaluate a custom list of SVs through these steps including the 1000 Genomes SVs as one of the reference dataset. Using Illumina short read sequences from the 1000 Genomes Project and the Simons Genome Diversity Project, we show that SVs not reported in the original studies can be well genotyped through our SV-aware graph pipeline. In addition, we demonstrate that false positive variant calls that arise due to microhomology within SV regions can be rescued through this approach.

1360W

Rapid whole-genome annotation and search in the cloud: SeqAnt enables easy identification of alleles for traits of interest. *A. Kotlar[1], M. Zwick[1], D. Cutler[1], T. Wingo[2,3].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Division of Neurology, VA Medical Center, Atlanta, GA; 3) Department of Neurology, Emory University School of Medicine, Atlanta, GA.

While genome-wide association studies (GWAS) and whole-exome sequencing (WES) remain important components of human disease research, the future lies in whole-genome sequencing (WGS), as it inarguably provides more complete data. Genetic disease studies require thousands of samples to obtain adequate power, and the resulting WGS datasets are hundreds of gigabytes in size and contain tens of millions of variant alleles. Describing, prioritizing, and selecting alleles from whole-genome sequencing experiments at this scale is technically challenging, and has to date required computer science expertise. SeqAnt 2.0 (https://seqant.emory.edu/) is the first web application that make these tasks accessible to non-programmers, even for large data sets. It comprehensively describes (annotates) alleles using public data (e.g., RefSeq, dbSNP, ClinVar, CADD, PhasCons, and PhyloP), and introduces the first natural-language search engine that can locate the alleles of interest in a user's experiment in milliseconds. Instead of relying on specific terms, the SeqAnt 2.0 search engine understands English grammar, similar to modern Internet search engines like Google and Bing. To use SeqAnt 2.0, users simply upload their experimental alleles in VCF format to the SeqAnt web application, wait for the data to be processed, and then type experiment-related terms, such as gene names, diseases, or allele frequencies into the provided search bar. The alleles matching any search query can be exported for downstream analysis, and each step is automatically recorded for reproducibility. To test SeqAnt 2.0, we submitted all autosomal coding alleles from gnomAD (2.6GB). SeqAnt annotated the 4.7×10^6 variant alleles in 3.47 minutes online. To test the search engine, we then queried "breast cancer pathogenic expert-reviewed maf < .001 cadd > 20" in SeqAnt's web interface. This query returned rare predicted-deleterious gnomAD alleles associated with breast cancer that had "Pathogenic" ClinVar records reviewed by an expert panel (.02 seconds, 93 alleles). SeqAnt provides these capabilities for numerous species, including *homo sapiens*, *m. musculus*, *d. melanogaster*, *c. elegans*, and *s. cerevisiae*. As sequencing becomes more prevalent in a diversity of research settings, and as the size of individual experiments grows, SeqAnt's capabilities will become increasingly critical in lowering barriers to effective sequencing experiment data analysis by researchers and clinicians alike.

1361T

High-performance whole genome sequence variant analysis in the TOPMed project using cloud environments. *R. Kuraisa, S. Gogarten, X. Zheng, C. Laurie, D. Levine.* Department of Biostatistics, University of Washington, Seattle, WA.

The NHLBI TOPMed project is expected to sequence >120,000 whole genomes. We are developing a high-performance analysis pipeline for computing population structure, kinship and association testing of all samples combined based around GENESIS, an R/Bioconductor package for statistical analysis of samples with genetic relatedness. Since TOPMed consists of many different studies from a wide range of populations, including family studies with extended pedigrees, characterizing and accounting for these relationships is crucial in association testing. The large scale of TOPMed requires parallel processing on either a local compute cluster or the cloud. Here we present an overview of the pipeline design as well as a comparison of using it in local and cloud computing environments. We first ported our pipeline from an in-house compute cluster to Amazon Web Services (AWS) using native Amazon Machine Images (AMI) and CfnCluster, a tool used to build and manage high-performance computing (HPC) clusters using different AWS job schedulers. We then evaluated AWS Batch, a newer offering that provides cluster-like services for running batch computing jobs encapsulated in Docker containers. AWS Batch dynamically provisions the compute resources required by a job, thereby subsuming many tasks that previously needed to be managed manually. Finally, we evaluated the portability of our pipeline using the batch services model and Docker containers to Microsoft Azure and Google Cloud Platform. To enable our pipeline to run in multiple compute environments, we create cluster objects in Python that define all aspects of the environment. This allows us to separate the analysis code from the batch computing and job scheduling. Providing cost-effective access to the heterogeneous compute resources necessary to run the different steps in our pipeline in the cloud was the most challenging technical obstacle we faced. Cloud vendors tend to scale their offerings such that memory and computational performance increase proportionally, which may not map readily to the needs of different pipeline steps. The batch services model using Docker containers simplified many administrative and configuration issues and allowed us to define a virtual compute environment that better supported the needs of our pipeline for heterogeneous compute resources. Our pipeline is hosted on GitHub for collaborators and the genetics community to adapt to their needs.

1362F

Improved access to variant-centric data in ClinVar. *M.J. Landrum, J.M. Lee, M. Benson, G. Brown, C. Chen, S. Chitipiralla, L. Geer, L. Han, J. Hart, D. Hoffman, A. Irving, W. Jang, K. Katz, C. Liu, Z. Maddipatla, A. Malheiro, K. McDaniel, A. Mercado, M. Ovetsky, G. Riley, A. Sethi, G. Song, R. Villamarin-Salomon, G. Zhou, B. Holmes, B. Kattman, D.R. Maglott.* NCBI/NIH, Bethesda, MD.

ClinVar is the public archive at NCBI for interpretations of genetic variants relative to a disease or phenotype. The database includes >486,000 interpretations for >318,000 variants. Interpretations in ClinVar have been provided by more than 700 submitters from 59 countries, including clinical testing laboratories, research laboratories, OMIM and GeneReviews, and expert panels. The initial data model for ClinVar focused on the variant-disease relationship. This model was reflected in the ClinVar XML file, which is the archival form of the ClinVar dataset. However, many ClinVar users prefer to review all data for a variant, regardless of the disease. This is relevant when a gene has been associated with more than one disease, and when a disease may have several different forms, such as breast cancer. We are introducing several new representations of the ClinVar dataset that are focused on the variant, for any reported disease. First, we have developed a new XML file that is focused on the variant, rather than on variant-condition. Each variant record will be assigned an accession number with a VCV prefix. Accession numbers will be versioned, so that we can maintain a history of changes to the variant's interpretation and evidence. The variant-centric XML represents variants for which we have an explicit interpretation as well as variants that have only been reported as part of a haplotype or genotype ("included" variants). Second, a new format for the ClinVar VCF file is now available. The previous VCF file was organized around the rs number, and thus was limited to variants that were in dbSNP. In addition, for some cases data for more than one variant was reported on the same row, making parsing of the data more complicated. The new format for the VCF file is allele-specific such that each row represents a single allele. It includes any variant in ClinVar with a precise location, even if it has not yet been accessioned in dbSNP or dbVar. The new VCF format also represents both variants with explicit interpretations and "included" variants. Third, the variant-centric web display for ClinVar is being updated. The new display includes the VCV accession number so that records may be cited. It makes the aggregate and submitted interpretations for each disease more accessible. It also provides more data about individuals observed with the variant, such as their zygosity and indication for testing, and provides information about submitted experimental evidence.

1363W

Bioinformatics and Elasticsearch: The perfect combination to unify and visualize life sciences massive data. *M. Leclercq, R. Ongaro-Carcy, A. Droit.* Faculté de médecine, Université Laval, Québec, QC, Canada.

Since few years, biological datasets have grown exponentially in size and complexity, hence entering into the world of big data. DNA/RNA/Protein sequences, biomolecules structures and functional information generated by cancer and omics research (e.g. genomics, metagenomics, transcriptomics, proteomics, and metabolomics...) now represent petabytes of data. To organize and make available these data, hundreds of bioinformatics repositories are currently available, but they are disparate and biologists constantly navigate from a database to another to perform their analyses. They also have to adapt their data through various mediums (i.e. software, platforms etc.) to prepare inputs into the various available online services. Moreover, free online resources to analyze data generated by last generation technologies, such as gene expression combined with clinico-pathological and phenotype data, and link them easily with online knowledge are still lacking. To answer this need, we have developed Kibio.science, a new portal dedicated to the biology and bioinformatics scientific communities. In one view, it can provide a mean to visualize the overall content or search queries results within many repository simultaneously, such as Pubmed, Chembl, Gtex, TCGA, and many others. Kibio.science is based on Elasticsearch search engine, and Kibi, a plugin which uses Kibana visualization tool and creates the relational searches between Elasticsearch indexes (i.e. databases). This environment, once distributed on clusters of servers, makes Kibio.science very fast and flexible. Kibio.science takes care of converting any repository in Elasticsearch format and re-organize the structure of the information. Designed to be user-friendly, it offers a great user experience to play with bioinformatics knowledge, by using advanced filtering and dynamic web visualizations where everything is clickable. Simultaneous searches between many databases are possible and each selection updates automatically all graphs, heatmaps or statistical information built in every linked databases. To enhance the analytical power of kibio.science, specialized visualizations dedicated to medical or research purpose will be implemented. In the near future, Kibio.science will allow users to import their own datasets, connect them to the desired repositories and create their own visualizations. We aim to make Kibio.science a great resource for exploring, visualizing, and analyzing multidimensional Omics data.

1364T

A graph remapping framework for in silico adjudication of SNVs, indels, and structural genetic variants from genetic sequencing data. *D.H. Lee, G.T. Marth.* Human Genetics, University of Utah, Salt Lake City, UT.

Several state-of-the-art, easy to use tools are available both for short-variant detection (e.g. GATK, FREEBAYES), and structural variant (SV) detection (e.g. LUMPY, MANTRA, DELLY), but these tools often produce divergent variant calls, especially INDELs, and it is very difficult to reconcile such variants into a single, accurate set. Furthermore, while it would be highly desirable to also detect larger, structural variants (SV), existing SV detector packages are typically difficult to integrate, highly resource-intensive to run, and result in call sets that require expert manual review to reduce false positive detection rate. Our algorithm, GRAPHITE (https://github.com/dillonl/graphite) requires as input a collection of variant calls, made by one or more short-variant or SV detection tools. Typically, this starting set is high sensitivity (i.e. inclusive), but low specificity (i.e. have a high false discovery rate). We then apply a novel "variant adjudication" procedure to discard false positives, while keeping true positive calls. This is accomplished by constructing a graph from these variants (the Variant Graph) representing allelic variants as graph branches, in addition to the branches formed by the current, linear genome reference sequence. Using a graph mapping algorithm (GSSW, a graph extension of the Smith-Waterman alignment algorithm) we developed earlier, we re-map all reads from each of the samples contributing to the candidate calls. We retain candidate variants confirmed by mappings to those branches in the graph that represent the corresponding variant allele, and discard those candidates that were not confirmed by such mappings. This procedure results in a highly specific callset that also maintains the high sensitivity of the inclusive starting callset constructed by multiple primary variant calling methods. Here we present the application of this method for cross-validating structural variants calls from Pacific Biosciences data by remapping deep Illumina WGS read sets to Variant Graphs constructed using the candidate Pacific Biosciences variants, as part of the Human Genome Structural Variation Consortium (HGSVC) data analysis project. We also present GRAPHITE's application to improving the accuracy of allele frequency measurement in tumor sequencing data, which is essential for the accurate reconstruction of subclonal evolution in longitudinal tumor samples.

1365F

SAV (Sparse Allele Vectors): Efficient variant file format that scales to analysis of millions of deep genomes. *J. LeFaive, H.M. Kang, G. Abecasis.* Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.

As the number of deeply sequenced genomes grows at an unprecedented rate, efficient storage formats that scale sublinear to the number of genotypes are essential. We have designed a format for storing very large sets of genotypes and genotype probabilities that produces small file sizes and is optimized for fast association analysis. The design supports fine-level random access to variants, with no limitations on variant annotation, sample size, ploidy level or genome length. Our format capitalizes on the sparsity of genetic variation at a given locus to both compress data and reduce deserialization overhead. In addition to the I/O and computational efficiency attributed to reduced storage footprints, further computational efficiency can be achieved through sparse matrix operations. Since the proportion of rare variants continually increases with sample size, the compression ratio and efficiency of our format both improve as study sizes grow. When comparing to BCF files with sample sizes of 2,504, 9,109 and 19,879 individuals, our format reduces the storage footprints by up to 70%, 74% and 80% respectively. Accompanying our format is a C++ programming library for interfacing with it and other file formats. This library was designed for efficient association analysis and provides a mechanism to plug in linear algebra and numerical libraries, which reduces the overhead of copying data and lowers the memory footprint. Our format specification is publicly documented to allow for custom integrations. Additionally, the programming library is open source. Both can be found at https://github.com/statgen/savvy.

1366W

Pathview Web: User friendly pathway visualization and data integration. *W. Luo, G. Pant, Y.K. Bhavnasi, S.G. Blanchard, C. Brouwer.* Department of Bioinformatics and Genomics, UNC Charlotte, Kannapolis, NC.

Pathway analysis is widely used in omics studies. Pathway-based data integration and visualization is a critical component of the analysis. To address this need, we recently developed a novel R package called Pathview. Pathview maps, integrates and renders a large variety of biological data onto molecular pathway graphs. Pathview quickly became a leading tool in pathway visualization, and has been widely adopted by tens of thousands of scientists and dozens of dependent applications worldwide. Here we developed the Pathview Web server, as to make pathway visualization and data integration accessible to all scientists, including those without the special computing skills or resources. Pathview Web features an intuitive graphical web interface and a user centered design. We also provide a comprehensive online help system and multiple quick-start example analyses. Even with no special computing training and resources, users may still accomplish pathway based data visualization and integration independently. The server not only expands the core functions of Pathview, but also provides many useful features not available in the offline R package: 1) The results graphs are interactive and hyperlinked to abundant external annotation data online; 2) The server provides the latest, most complete and accurate pathway definitions and graphs by regular synchronization with KEGG source databases; 3) Users can review, replicate and share their analyses easily with free registered user accounts, which enable collaborative research and reproducible science; 4) Useful user engagement features allow users make comments and suggestions, or ask for help in designated pages. Importantly, the server presents a comprehensive workflow for both regular and integrated pathway analysis of multiple omics data. In addition, the server also provides a RESTful API for programmatic access and conveniently integration in third-party software or workflows. Pathview Web is openly and freely accessible at https://pathview.uncc.edu/.

1367T

Optimal workflow for next generation sequencing data processing using existing technology. *J.E. Martin[1], A. Larionov[1,2], G. Clark[1,3], E.R. Maher[1,3].* 1) Department of Medical Genetics, University of Cambridge CB2 0QQ, UK; 2) Statistics and Computational Biology Laboratory. Cancer Research UK Cambridge Institute; 3) Cambridge NIHR Biomedical Research Centre, Cambridge CB2 0QQ, UK.

   Introduction Next Generation Sequence (NGS) instruments are becoming the standard tool in genetics, however, this process has not been accompanied by the development of a "gold standard" pipeline for germline NGS data processing. Our goal in the present study is to determine the optimal workflow in the specific case of whole exome sequencing (WES) to maximize the cost-efficiency of the process. **Material and Methods** For our analysis, we selected a cohort of 200 individuals sequenced for WES. All the samples considered were processed using sixteen different workflows composed of the different algorithms and software available. We selected up to ten quality parameters and indices to discern the best possible workflow. **Results** Using bwa mem for the alignment compared to using bwa sampe improved the number of aligned reads by roughly a 10%. When performing and not performing BQSR and indel realignment, there were no significant differences in the number of variants called or the quality of those, but there was an increase of computing time (4:32 versus 2:24). Regarding the variant calling, there were no significant differences in the number of variants called or their quality between algorithms, while the depth of read was significantly higher in unified genotyper (73.63 versus 53.83). Also, the computational time was significantly lower for unified genotyper (2:24 versus 11:13). **Conclusions** Our results indicate an optimal workflow for the processing of NGS WES germline data which can cut the time per sample down to a tenth while maintaining the quality of the results produced.

1368F

Harnessing the power of linked data: Using semantic annotations on pediatric and adult datasets to enable cross-search. *M. Mattioni, J. Radenkovic, A. Stanojevic, N. Vukosavljevic, J. DiGiovanna, V. Mladenovic, B. Davis-Dusenbery.* Seven Bridges, Cambridge, MA.

   Advances in Next Generation Sequencing have allowed the generation of huge amounts of genomic data. At the same time, an increased amount of clinical metadata has been recorded, usually in various formats and with different methods. These metadata are becoming the compass researchers use to find and discover cohorts for further analysis. The possibility to cross-search different datasets coming from different investigations can speed data discovery and also yield to statistically significant results due to the ability to marshall a higher number of samples; moreover, researchers are empowered to ask deep queries about the data and create cohorts which intersect adult and pediatric cases at the same time. We have created an overarching data model which can represent the different metadata found in different pediatric and adult datasets, and we have harmonized the metadata using this model. We also have developed the Data Browser, a graphical interface to create deep queries in genomic data, which support cross-database search. Cross search enables users to easily explore phenotypic data in pediatric and adult datasets, to create hypothesis driven cohorts, and then to identify the connected files, which can be accessed according to the dataset's access policy they belong to.

1369W

Representing the human genome with synthetic spike-in controls. *T. Mercer[1,2], I. Deveson[1], S. Hardwick[1], J. Blackburn[1], T. Wong[1], B. Kanakamedala[1], C. Barker[1], A. Martins Reis[1].* 1) Garvan Institute of Medical Research, Sydney, NSW, Australia; 2) Altius Institute for Biomedical Science, Seattle, WA, United States.

Due to its directionality (from 5' to 3'), the human genome is a chiral object that can reversed by a mirrored plane. For example, the 5'-TGAC-3' sequence is mirrored by the chiral 5'-CAGT-3' sequence. This chiral sequence is distinct to the original human sequence, yet retains many identical properties, including the same nucleotide content, repetitiveness, sequencing and alignment properties. These chiral sequences can thereby act as synthetic DNA of RNA standards to represent almost any feature of the human genome. We terms these reference standards, *sequins* (sequencing spike-ins). We have developed *sequins* to represent instances of common and disease associated genetic variation, including germline, somatic and structural rearrangements, repeat DNA, and immune receptor recombination. We have also released RNA *sequins* that represent alternatively spliced isoforms and fusion genes. These RNA and DNA standards are mixed to form reference ladders and emulate quantitative features of genome biology, such as genotypes, mutation frequencies, clonal cell populations, gene expression and splicing. *Sequins* are typically added at a fractional concentration directly to the patient's DNA sample before library preparation, and are then sequenced together, with a proportional fraction of resultant reads deriving from the *sequins*. The alignment of the library to both a human and mirror genome partitions *sequin* reads from the accompanying sample, and allow their use as internal quantitative and qualitative controls at each step of the NGS workflow without contaminating the accompanying sample. We have demonstrated the use of sequins with tumor and normal patient samples as internal quantitative and qualitative controls during sequencing and diagnosis. We have demonstrated the use of sequins in a range of clinical applications, including whole genome sequencing in mendelian disease diagnosis, targeted panel sequencing of tumor samples, and the diagnosis of fusion genes with RNA sequencing. Sequins can be used to empirically measure diagnostic performance and quantitative accuracy of each individual NGS library. Sequins can also act as scaling factors by which to normalize between multiple samples, and thereby enable more accurate comparisons, more efficient data-sharing and standardized clinical database aggregation.

1370T

Use of a metadata-driven architecture for tools and techniques to enhance methods for data sharing and facilitate meta-analysis for genetic association. *M.C. O'Leary[1], R. Whitley[1,2], K.J. Sims[1], R.B. McNeil[3], E.R. Hauser[1,2], C.D. Williams[1,2], D. Provenzale[1,2].* 1) VA Cooperative Studies Program Epidemiology Center, Durham VA Health Care System, Durham, NC; 2) Duke University, Durham, NC; 3) Center for Clinical Research Network Coordination, RTI International, Durham, NC.

Objectives: The VA Cooperative Studies Program Epidemiology Center-Durham (CSPEC-Durham) manages a repository that contains data and specimens from multiple IRB-approved studies. We aimed to increase efficiency of repository data management by eliminating repetitive and error-prone manual steps through a metadata-driven architecture for our data management lifecycle. Specifically, we built a common set of variable descriptors (metadata) across studies and used those descriptors to generate: 1) application code for data entry and validation, 2) data dictionaries, and 3) data extracts for sharing data with approved researchers. **Methods:** We created a SQL Server database comprised of metadata tables with descriptors for each variable, including variable name, survey question, value description, minimum and maximum possible values, and other metadata. We populated the tables with variables from each study. These tables were used to generate application code for data validation, data extracts, data dictionaries, and other documentation for specific studies. We retain data maintenance history logs, snapshots of data extracts, and a history of all source code to ensure data provenance. **Results:** The metadata tables enabled the rapid generation of high-quality applications capable of automated data validations, as well as the creation of data extracts to address specific research questions or requests. Using the same tables, we have the ability to automatically extract HTML and PDF data dictionaries for whole datasets and subsets for specific research projects. Indicators flag personally identifiable information (PII) and exclude this information from extracted datasets and corresponding dictionaries as needed. **Conclusions:** We demonstrated that metadata-driven concepts related to data management and creation of study documentation can be applied to many datasets. This process will result in similarly high quality products for future researchers. We intend to monitor the use of these tools and techniques to ensure continuous quality improvement in data and documentation integrity. **Impact Statements:** We created a foundation for the automated management of data throughout its lifecycle across all studies in our repository. These metadata methods have significantly reduced the potential for errors, eliminated repetitive data management steps, and fostered a robust and secure sharing environment to facilitate future data sharing and meta-analysis for genetic association.

1371F

Improving SNP array copy-number variant calling using site-specific variance models and windowed intensity normalization. *T. Poterba[1,2,3], C. Seed[1,2,3], J. Bloom[1,2,3], D. Howrigan[1,2,3], C. Churchhouse[1,2,3], B. Neale[1,2,3].* 1) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston 02114, MA, USA.

Both common SNPs and rare, large copy-number variants (CNVs) have been shown to harbor substantial risk for heritable diseases. However, the overall contribution of CNVs to disease risk is not well understood. Recent SNP array datasets with hundreds of thousands of individuals can offer insight into the prevalence and impact of small CNVs, but existing tools to call CNVs suffer from sensitivity to noise and fail to model common error modes well. As a result, unavoidable variance in sample quality has required strict sample curation and CNV-size thresholding, reducing power to detect smaller CNVs. In order to create high-quality call sets from large cohorts, we have developed a new hidden Markov model (HMM) for CNV calling on microarray data. Spurious CNV calls usually result from either low quality SNPs or low quality samples, so we have modeled both of these error modes explicitly. The model includes per-SNP variance terms to weight emission probabilities by site quality, which reduces the false positive rate while preserving power to detect real events. We also model sample quality to dynamically adjust global priors on call probabilities, dramatically cutting the false positive rate in low quality samples and rescuing real CNVs by permitting more inclusive sample filtering. Our model is implemented using Hail, an open-source framework for statistical genetics built on Apache Spark to easily scale to hundreds of thousands of samples. Using a cohort of 1800 trios and benchmarking sensitivity and specificity against the leading method in the field, PennCNV, we demonstrate comparable CNV transmission rates for large events and improved performance for small events. This improvement will enable us to examine the role of small CNV events in contributing to the genetic burden of disease.

1372W

LUBA: A software toolbox for efficiently manipulating and analyzing NGS data. *A. Ryutov[1], M. Bootwalla[1], A. Govindarajan[1], J. Buckley[1,2], D. Maglinte[1], J. Ji[1,2], X. Gai[1,2].* 1) Center for Personalized Medicine, Department of Pathology & Laboratory Medicine, Children's Hospital Los Angeles, California, USA; 2) Keck School of Medicine, University of Southern California, California, USA.

The need for robust bioinformatics algorithms to efficiently process Next-Generation Sequencing (NGS) data continues to grow. Here we present LUBA (Lightweight Utility for Bioinformatics Analysis), a suite of command line tools for handling NGS data that include modules for phasing, coverage analysis, copy number variation (CNV) detection and calling somatic variants. **Phasing:** a computational algorithm for resolving the phase of heterozygous variant pairs in a gene from Whole-Exome Sequencing (WES) data. The algorithm assembles and connects heterozygous variants within individual reads or read pairs into "heterozygosity islands". The phase of two heterozygous variants can be directly resolved if they fall into the same "island". If not, their phase may be potentially determined from linkage disequilibrium data. The phasing algorithm was tested with our trio-WES data. For variants within the same gene, phase can be determined for 33% of all heterozygous variant pairs with 94% accuracy from the proband only exome data. **Depth:** an algorithm for generating coverage statistics from a BAM file. It is 3-7 and 3-4 times faster than SAMtools and Sambamba, respectively. The algorithm utilizes either a minimum heap or a circular array to process alignments efficiently, and has a small memory footprint. As a result, it can process more than 100 WES BAM files in two hours on a LINUX workstation. **CNV:** an algorithm to detect CNVs from WES data. The algorithm builds a baseline coverage profile for a sequencing platform by normalizing a batch of samples prepared and sequenced the same way. An input sample coverage is then normalized; and an average per-exon coverage is compared to the reference baseline to make CNV calls using a p-value cutoff. The algorithm was tested with chromosomal microarray CNV calls, and performed very well. **VAR:** a tool to make somatic variant calls from cancer WES data with high sensitivity. Comparison of LUBA VAR with several publicly available somatic variant callers such as VarDict and MuTect2 has yielded favorable results. Similarly to other LUBA tools, VAR is memory-efficient, which is especially suitable for high sequencing depth needed for oncology samples. In addition, LUBA has various general-purpose tools built with the same framework, i.e., tools for variant filtering, concordance analysis, BAM/VCF/BED processing. Our custom suite of LUBA NGS tools has demonstrated great utility in both germline and somatic diagnostic tests.

1373T

Hail: Scaling statistical genetics to tens of thousands of whole genomes.
C. Seed[1,2,3], J. Bloom[1,2,3,4], J. Goldstein[1,2,3], D. King[1,2,3], T. Poterba[1,2,3], B. Neale[1,2,3]. 1) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston 02114, MA, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 4) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

　Advances in high-throughput sequencing offer unprecedented power to discover the genetic basis of disease. However, the scale and complexity of the data are increasingly making even simple analyses cumbersome or intractable. To address this challenge, we developed Hail (https://hail.is), an open-source framework for statistical genetics that has been used to efficiently analyze tens of thousands of whole genomes and hundreds of thousands of exomes. Hail is a Python library that uses Apache Spark to provide scalability and fault-tolerance for genetic analyses. Consequently, analysts can run the same code on a laptop or large cluster without the need to manually chop up the data or manage job failures. Hail provides a flexible query language to express complex quality control and analysis pipelines with concise, readable code that reflects genetics concepts; as a result, code is less prone to error and easily shared. Hail imports data from the VCF, PLINK, and BGEN (including v1.2 of UK Biobank) formats and includes natively distributed algorithms for common variant association (linear / logistic / mixed regression, TDT, eQTL, ...), rare variant association (linear / logistic / exact burden tests, SKAT, ...), kinship, LD, principal components, and synthetic data generation. Hail has been the primary analysis platform for dozens of major studies at the Broad Institute, Oxford, Michigan, Sanger, Stanford, and UCSF. Daniel MacArthur's group used Hail to iteratively clean and analyze the gnomAD dataset of 15k whole genomes and 125k exomes in the week before ASHG 2016. Ben Neale's group used Hail to demonstrate that ultra-rare disruptive and damaging mutations influence educational attainment (Ganna et al., Nat. Neuro., 2016). Sekar Katherisan's group used Hail to analyze 16k whole genomes and 40k exomes for lipid and MI phenotypes. In 10 minutes, the initial GWAS on 8125 TOPMed genomes derived the kinship matrix, fit a linear mixed model, and computed statistics for 25M variants, compared to 2 weeks in practice with other tools. They then used Hail for programmatic, iterative fine mapping to identify independent signals at the LPA locus. Hail also enables on-the-fly analyses for online portals and will soon power a new platform for rare disease genomics. As we prepare Hail's infrastructure for a million whole genomes with deep phenotypes, we welcome the scientific community to leverage this open-source project to develop and share new methods at scale.

1374F

IMSindel: An accurate intermediate-size indel detection tool incorporating *de novo* assembly and gapped global-local alignment into split reads. *D. Shigemizu[1,2,3], F. Miya[2,3], S. Akiyama[1], S. Okuda[4], K.A. Boroevich[3], A. Fujimoto[5], H. Nakagawa[6], T. Tsunoda[2,3].* 1) Department for Medical Genome Sciences, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Aichi, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; 5) Department of Drug Discovery Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 6) Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

　Insertions and deletions (indels) have been implicated in dozens of human diseases through radically altered gene functions by short frameshift indels as well as long indels. However, the accurate detection of these indels from next-generation sequencing data is still challenging, particularly for intermediate-size indels (≥ 50bp), due to the short DNA sequencing reads. In this study, we developed a new method that predicts these intermediate-size indels using BWA soft-clipped fragments (unmatched fragments in partially mapped reads) and unmapped reads. We report a performance comparison of our method, GATK HaplotypeCaller and PINDEL, using whole exome sequencing data from the same DNA samples. False positive and false negative counts were determined through Sanger sequencing of all predicted indels across these three methods. The harmonic mean of the recall and precision, F-measure, was used to measure the performance of each method. Our method achieved the highest F-measure of 0.82 in one sample, compared to 0.53 for HaplotypeCaller and 0.59 for PINDEL. Similar results were obtained in the second sample, demonstrating that our method was superior to the other two methods for detecting intermediate-size indels. We believe that this methodology will contribute to the discovery for long indels associated with human diseases.

1375W

ENCODE Portal: A cohesive collection of genomic assays towards discovery of functional elements. *C.A. Sloan, B. Hitz, E. Chan, C. Davis, J.S. Strattan, J. Hilton, A. Narayanan, M.K. Onate, I. Gabdank, F. Tanaka, T. Dreszer, U. Baymuradov, K. Dalton, J.M. Cherry.* Genetics, Stanford University, Palo Alto, CA.

The Encyclopedia of DNA Elements (ENCODE) project is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active. As the DCC (the data coordination center) it is our task to collect, curate and disseminate the results, methods and raw data for a variety of complex assays and analyses that have been performed toward that goal. To that end, we have created the ENCODE portal (www.encodeproject.org). At this site, scientists can search, visualize, and download both raw and processed data. They can delve into the details of the experiment including the antibody and biosample validation results, the protocols, the software code, and provenance of a file. And, finally we provide insight into the methods, pipelines and standards that the consortium has developed. In addition to the ENCODE funded data, at the portal we also have data from modENCODE, mouse-ENCODE, modERN, GGR (Genomics of Gene Regulation) and Roadmap Epigenomics Mapping Consortium (ROADMAP). For the core assays like ChIP-seq, RNA-seq, and Dnase-seq, we have worked with these other projects to harmonize methods and in some cases we have provided new processing to enable integrated analysis. With over 18,000 experiments released, this is a powerful resource for the exploration of genomic assays toward the discovery of functional elements.

1376T

Cloud computing environment for hosting federated genomic databases. *W. Souza[1], B.S. Carvalho[2], C. Rocha[1], I. Lopes-Cendes[1].* 1) Department of Medical Genetics, UNICAMP, Campinas, Brazil; 2) Department of Statistics, UNICAMP, Campinas, Brazil.

Federated databases consist in using local resources for hosting data instead of submitting to centralized services. This concept facilitates sharing genomic data with restricted access to sensitive information such as individuals' data, and avoids legal issues associated to data protection. However, hosting data locally leads to difficulties in data integration from different databases because each provider may implement its own non-standard interface for sharing genomic data. Also, research groups require specialized technical assistance for configuring complex and keeping secure local databases. The Global Alliance for Genomics and Health (GA4GH) was formed to help accelerate the potential of genomic medicine to advance human health. The GA4GH Data Working Group developed data model schemas and application program interfaces (API) for standardized genomics data exchange. Software containers such as Docker, facilitates deployment of multiple services across computers. It increases security of sensitive data by isolating software solutions into standalone and lightweight images. As part of the Brazilian Initiative on Precision Medicine (BIPMed) project, we developed a containerized computing environment composed by GA4GH-based data servers and software clients. Our cloud environment provides easy deployment of one or more GA4GH reference implementation servers for hosting genomic data, GA4GHshiny application for interactive access via user-friendly graphical web interface and Beacon servers. Using the popular Docker solutions, our cloud-based computing environment allows researches groups to easily and securely host their own genomic data using local computing resources or internet services. All Docker images and documentation are freely available at https://github.com/labbcb/federated-genomic-databases.

1377F

The ENCODE annotation pipeline: Repeatable software tools for ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite experiments. *J.S. Strattan[1], T.R. Dreszer[1], B.C. Hitz[1], U.K. Baymuradov[1], E.T. Chan[1], I. Gabdank[1], J.A. Hilton[1], O.A. Jolanki[1], C.A. Sloan[1], Z. Weng[2], A. Kundaje[1], J.M. Cherry[1], ENCODE Data Coordinating Center.* 1) Stanford University School of Medicine, Department of Genetics, Stanford, CA; 2) University of Massachusetts Medical School, Program in Bioinformatics and Integrative Biology, Worcester, MA.

The ground-level annotations ENCODE applies to the epigenome and the transcriptome are produced by defined cloud-based computational pipelines that anyone can use. By standardizing the computational methodologies for analysis and quality control, results from multiple labs can be directly compared and integrated into higher-level annotations, such as ENCODE Candidate Regulatory Elements (CREs). The ENCODE Data Coordinating Center (DCC) have deployed the pipelines to a cloud-based computing environment so that the computational load can be distributed across scalable resources. Because the exact software pipelines the ENCODE DCC uses are available for anyone to run, researchers can produce analysis results and quality-control metrics from their own data that are directly comparable to ENCODE's annotations. Pipelines are available for the analysis of ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite experiments. The pipelines are open-source, easy to use, and require minimal pre-requisites. The ENCODE DCC codebase is at https://github.com/ENCODE-DCC ENCODE analyses are distributed through the ENCODE Portal at https://www.encodeproject.org/ The pipelines are also available as "ENCODE Uniform Processing Pipelines" at https://platform.dnanexus.com/projects/featured .

1378W

Quantitative assessment of the feasibility of using whole-genome sequencing data at common single nucleotide polymorphism positions to reproduce high-confidence genotype calling and copy number variation detection results from SNP microarray data. *N.S. Ten, B. Pusey, K. Chao, D. Adams, W. Gahl.* NIH Undiagnosed Diseases Program, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

The NIH Undiagnosed Diseases Program (UDP) was established in 2008 with the goal of conducting research for patients whose significant illnesses remain undiagnosed despite extensive medical workups. A specific focus of the UDP bioinformatics is the re-analysis of previously inconclusive clinical genetic studies with specialized exome, genome, and single nucleotide polymorphism (SNP) microarray sequencing analyses. SNP microarrays are utilized extensively by the UDP. Uses include recombination mapping, reconstructing missing parent genotypes from sets of siblings, and detecting copy number variations, regions of homozygosity, uniparental disomies, and mosaicisms. SNP microarrays are frequently used for quality control of next-generation sequencing (NGS) data because whole-genome sequencing should theoretically genotype the same SNPs selected in a SNP microarray. Standard commercial SNP microarrays employ redundant sampling of every SNP across multiple sites on a chip, and the fluorescence dynamic range produces high-confidence genotype calls for copy number quantitation from zero to four. For NGS data, results can have disparate coverage depths because of allele skew and high PCR amplification in library preparation for cluster generation, making copy number genotype calling difficult for exome data but within reach for low-PCR whole-genome sequencing. We tested the genotyping sensitivity of whole-genome sequencing at positions genotyped in SNP microarrays. Our comparison of SNP microarray and whole-genome sequencing-derived genotypes at over 940,000 SNPs in the Illumina OmniExpress SNP microarray suggests complete genotyping matching at 99.6% of these common positions. For discrepant or missing genotypes in SNP microarray data, we looked at differences in recombination mapping calls in nuclear family pedigrees with multiple siblings and found only minor differences. In the future, we plan to perform additional work conducting similar comparisons with other microarray-based analyses that require even more quantitation of alleles in NGS genotype calling. Concerning integer analyses of genotype calling and copy number variation, current whole-genome sequencing in the UDP appears to achieve the same quality and reliability for these specific results as that seen in SNP microarray analysis.

1379T

GA4GH Genomic Beacon Network: Security challenges and mitigation.
S. Ur-Rehman[1], J.D Spalding[1], S. Scollen[4], J. Rambla Di Argila[2], I. Lappalainen[3], ELIXIR-SE, ELIXIR-BE, ELIXIR-FR, ELIXIR-NL, GA4GH Software Security, GA4GH Beacon Project. 1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD; 2) Centre for Genomic Regulation (CRG), Dr. Aiguader 88, Barcelona 08003, Spain and Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) CSC IT Center for Science, Life Science Center Keilaniemi, Keilaranta 14, Espoo, Finland; 4) ELIXIR Hub, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD.

As the number of human genomes sequenced increases, there is a need for researchers to contextualise the genetic variation discovered against the efforts of all previously sequenced individuals to facilitate genotype to disease phenotype discoveries. The Global Alliance for Genomics and Health (GA4GH) Beacon Project in partnership with ELIXIR has created an open sharing platform for human genetic data to be discoverable. Beacons are a service that return a yes/no response when queried about the presence of a particular allele in a given chromosomal location. This allows a researcher to determine if a particular allele has been seen within a particular dataset as a precursor to applying for access to the complete dataset while respecting the privacy of the individual. Beacons can implement a tiered access model, public (anonymous user), registered (known user), and controlled. It has been shown that it is possible to re-identify an individual (determine if a particular individual's genome exists in a beacon) if the individual's genome is known a priori. In essence this involves querying multiple alleles from an individual's genome ranked in order of increasing allele frequency.Three mitigation techniques have been developed to prevent re-identification via Beacon but are only applicable to single Beacons. In order to maximise the utility of Beacons, a network of Beacons has been developed that allows search over all multiple Beacons from a single query. There are currently over 60 Beacons that include more than 200 global datasets, including 7 ELIXIR Beacons (Finland, Switzerland, Belgium, Spain, Netherlands, Sweden and EMBL-EBI - the beacon hosted at the European Genome-phenome Archive). As current mitigation techniques against re-identification only work on individual beacons, we present methods to extend the current method preventing re-identification to a Beacon network. These include the implementation of a registry, the communication of the individual(s) who are at risk of re-identification, and the mitigation techniques which are currently being deployed. We also make use of the ELIXIR AAI to track the source of queries for registered level queries. By reducing the possibility of re-identification with both individual and networked Beacons, we aim to attract more stakeholders with sensitive datasets to the Beacon Network. We are also working to improve the 'hand-off' between a positive Beacon response and other GA4GH access and data delivery services.

1380F

UDPICS, a laboratory-medical informatics system for improving translational research in the NIH Undiagnosed Diseases Program. *Z.M. Valivullah, S. Sadozai, G. Yu, D. Draper, D. Mosbrook-Davis, D.R. Adams, W.A. Gahl.* Undiagnosed Diseases Program, NHGRI, Bethesda, MD.

The Undiagnosed Diseases Program (UDP) is a translational medicine initiative based in the NIH Clinical Center which aims to diagnose patients with mysterious conditions and study unique and rare genetic disorders. The program has received over 10,000 inquiries, reviewed 3,500 applications, and seen almost 900 patients. Patient presentation is heterogeneous and our evaluation process initially treats each patient as independent. Therefore, managing patient and research data became increasingly difficult after the first years of the UDP. To allow clinicians and researchers to seamlessly collaborate on UDP cases, we built a web-based workflow management system called the Undiagnosed Diseases Program Integrated Collaboration System (UDPICS). The system is based on a commercial Ruby on Rails platform, which we expanded and customized for translational research. UDPICS integrates external biobanking, model organism, and genome analysis systems for easy access. Furthermore, we incorporated the open source tools PhenoTips for ontological patient phenotyping and Phenogrid for phenotype-genotype comparisons in humans and model organisms. UDPICS has allowed us to track patients and associated research from application to followup. Benefits of the system include rapid transit of records and tasks, reduced error in workflows, and ease of collaboration. By implementing complex process management, genomic analysis, biospecimen management, ontological phenotyping, and electronic laboratory notebooks, UDPICS has become a scalable collaborative workspace for research, clinical, and bioinformatics staff. The system is a mechanism for efficient, transparent, and scalable translational research and addresses many of the scientific and logistical problems of the NIH UDP.

1381W

Improving clinical exome interpretation using lab internal cohort data.
C. Wu[1], S. Baker[1], K. Cao[1], B. Krock[1,2], M. Sarmady[1,2]. 1) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Pathology & Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Workflow complexity and high-throughput nature of NGS is prone to producing false positive variants. Many of such artifacts are in low complexity regions of the genome, or specific to the sequencer, instruments, bench and bioinformatics processes used to yield the data. Clinical variant classification is a labor-intensive step and the presence of artifacts adds hands-on time. A common approach to reduce variants to be manually analyzed is to remove variants that are present at a high frequency in a cohort of internally sequenced and processed samples, the 'internal cohort'. This approach would remove common variants in lab-specific patient population as well as recurring technical artifacts. Here we present the essential role of using internal cohort data to remove false positive and common variants as well as challenges involved in effective removal of artifacts. In this study, VCF files from 665 unaffected and unrelated clinical exome sequencing samples at the Children's Hospital of Philadelphia were compiled to build the cohort. The individual vcfs were parental samples generated as part of the familial exome bioinformatics pipeline. The individual variant call sets were then combined into a single vcf file using *GATK* package. Subsequently, population (cohort) allele frequency (CHOP_AF) of each variant in the cohort was calculated based on the genotypes. This design enables addition of new samples without the need to recall variants on the entire cohort. To annotate patient's variants against with CHOP_AF, a python package was developed. The process compares all patient's variants to the variants in the cohort by matching chromosome, position, reference, and alternate alleles to determine whether to retain or filter out the variant using the pre-defined cut-off (i.e. 1%). On average, 41,927 raw variants across 11 clinical samples of different batches were observed in our validation samples. After internal cohort filtration of the variants, only 3,231 variants were retained on average, indicating that over 90% of the raw variants were removed from further analysis due to being too common or artifacts. We presented an efficient method to build internal cohort using unaffected samples in a clinical lab, as well as a bioinformatics strategy to filter any variant sets against the cohort. The overall approach is effective in detecting artifacts and common variants, removing over 90% of the raw variants and therefore improves the efficiency of the clinical tests.

1382T

Integrated breakpoint analysis and structural variation detection using sequencing reads from multiple sequencing technologies for an Ashkenazi trio. *C. Xiao, S. Sherry.* NIH/NLM/NCBI, Bethesda, MD.

Structural variations (SVs) play important roles in human genome organization and contribute to phenotypic diversity of human populations and diseases. Recent advancement of next generation sequencing (NGS) technologies has greatly facilitated the analysis of genomics structural variants. However, accurately detecting such structural variations with correct sizes and precise breakpoints remains a challenge. Existing structural variant detection methods mostly depend on reference-specific alignment bams or *de novo* assemblies that require intricate platform-specific workflows and significant amount of computation and storage resources. Here we introduce a novel structural variation discovery method, BreakScan, which does not rely on pre-aligned bams or pre-constructed assemblies. Instead, BreakScan models insertion and deletion events in reference sequence with breakpoint junctions observed in NGS reads, then subsequently compiles evidence for those junctions from the sequencing reads generated by multiple platforms such as Illumina, 10X Genomics, and Pacific Bioscience. The predicted structural variants are generated from event models and ranked by their supporting evidence. We have applied this method to the son of the Ashkenazi trio from Genome-in-a-Bottle (GIAB) project (ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/AshkenazimTrio) and delivered a callset to the GIAB consortium (ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/AshkenazimTrio/analysis/). Comparative analysis with other structural variant callsets has also been performed and will be discussed.

1383F

Exploiting the co-localization of trait-associated SNPs and eQTLs to identify potential biological mechanisms underlying complex diseases. *T. Xu, Z. Qin.* Emory University, Atlanta, GA.

   Genome-wide association studies (GWASs) have successfully identified a large number of disease-associated variants. However, the majority of these identified variants is non-coding, which makes functional interpretation extremely challenging. Instead of directly affecting the protein, it is likely that these variants affect the phenotype through transcription regulation. Since expression quantitative trait loci (eQTL) provide direct evidence on how genetic mutations affect gene expression, eQTL has long been recognized as an important resource for functional annotation of noncoding variants. In the present study, we proposed a novel strategy to utilize the rich eQTL information that has been accumulating in the public domain to annotate GWAS findings. To be specific, for a specific disease, we investigate whether eQTLs associated with a particular pathway are enriched around loci that harboring GWAS-identified variants. We first convert a set of GWAS variants into a set of genes through eQTL, then we conduct test similar to the gene set enrichment analysis to see if any pathway is overrepresented by the list of genes. Our method enjoys two key advantages over existing methods: first, we no longer need proximity to link a variant to a gene which has shown to be unreliable; second, eQTL allows us to identify the tissue type in which the dysregulation likely to affect the disease phenotype the most. We have implemented our strategy into an easy-to-use R package that performs functional enrichment analysis by associating non-coding genomic regions to the co-localized eQTLs. We applied our method to eQTL collections from the Genotype-Tissue Expression (GTEx) project and pathway collections from MSigDB. Using disease-related genomic regions as query, our method successfully identifies highly-relevant biological pathways for immune-related diseases and age-related neurological diseases. Tissue specificity analysis of associated eQTLs provide additional evidence of the distinct roles of different tissues in the disease mechanisms. Our tool is flexible to be applied to customized eQTL sets and pathways.

1384W

ALEC: Amplicon Long-read Error Correction for targeted long-read sequencing. *Y. Yang[1,2,3], W. Qiao[1,3], M. Delio[3], L. Edelmann[1,3], S.A. Scott[1,3].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Icahn Institute for Genomics and Multi-scale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Sema4, a Mount Sinai venture, Stamford, CT 06902.

   Long-read sequencing platforms [i.e., Pacific Biosciences (Pacbio) and Oxford Nanopore] have average read lengths of ~10-15 kb, which can be useful for interrogating structural variation, *de novo* genome assembly, variant phasing, and sequencing challenging gene regions with homologous pseudogenes. However, their single-pass error rates (~10-20%) are significantly higher than short-read error rates, particularly for insertion/deletion (indel) artifacts. Although the accuracy of PacBio single molecule real-time (SMRT) sequencing is improved by using multiple-pass circular consensus sequencing (CCS) reads, performance and output depth are limited by library length and number of passes. As such, error correction tools are typically necessary for analysis of single-pass long-read data. A common long-read sequencing correction strategy employs a hybrid assembly with high depth short-reads to identify and correct long-read errors. However, given that this strategy results in increased sequencing costs and cannot adequately interrogate non-unique regions, we developed Amplicon Long-read Error Correction (ALEC) as an informatics correction tool for targeted long-read sequencing analysis. ALEC identifies sequencing errors from SAM/BAM files generated by BWA-MEM by calculating per base substitution and indel allele frequencies, and subsequently corrects alleles below predefined or customized thresholds to their consensus nucleotide at each targeted base. Additionally, ALEC implements a linear formula to determine homopolymer allele frequency thresholds, as a linear correlation between homopolymer length and deletion error frequencies was detected in both Pacbio and Oxford Nanopore data. ALEC performance was tested on long-read amplicon data from the *CYP2D6* (~5 kb), and *RYR2* (~9.2 kb) genes derived from the PacBio RS II platform, as well as *CYP2D6* (~5 kb) amplicon data derived from the Oxford Nanopore MinION platform. Importantly, application of ALEC to PacBio data improved subread substitution errors from 1.5% to 0.05% and indel errors from 10.6% to 0.05%, and application of ALEC to Oxford Nanopore data improved substitution errors from 4.2% to 0.05% and indel errors from 20.8% to 1.4%. These data indicate that ALEC is a useful correction tool that can improve long-read sequencing accuracy, which subsequently can facilitate the incorporation of the widely used BWA-GATK pipelines for these third-generation sequencing platforms.

1385T

An alignment-based method to trim adapter and filtrate low-quality reads from pair-end FASTQ files. *R. Ye, P.C. Sham.* Department of Psychiatry, The University of Hong Kong, Hong Kong, Hong Kong.

As next generation pair-end sequencing is widely used around the world, cleaning up of raw FASTQ data is much more important than we thought before. Here we present FASTQ cleaning program (fqclean), a novel and efficient alignment-based method that trims adapter and filtrates low-quality reads from pair-end FASTQ files. Compared with previous published tools, "fqclean" does not need adapter sequence as an input parameter and has the ability to trim even 1bp-length adapter contamination. Accurate trimming and filtrating of raw data can improve sensitivity and specificity of downstream variation detection, especially single nucleotide variation, virus integration and structural variation. During trimming and filtrating, "fqclean" can also fetch barcodes which attached to 5'-end of reads and split big FASTQ files to small pieces to accelerate alignment based on Map-Reduce method.

1386F

A statistical framework for longitudinal genomic data integration. *Y. Zhang[1], Z. Ouyang[2].* 1) University of Connecticut, Storrs, CT; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Recent advances in genomic medicine resulted in accumulated longitudinal genomic data, where patients were monitored over time, their biological samples were collected at multiple time points, and the corresponding molecular profiles were measured through high-throughput assays. Longitudinal high-dimensional datasets from genomic and biomedical research may have complicated correlation structures or irregular covariance structures. Consequently, such characteristics raise challenges on dimension reduction and feature selection of longitudinal genomic data. We present a new statistical method to integrate multiple sources of information for better knowledge discovery in diverse dynamic biological processes. We demonstrate the utility of our method through simulations and applications to gene expression data of the mammalian cell cycle and longitudinal transcriptional profiling data in response to influenza viral infections.

1387W

SOAPgaea: A Hadoop/Spark based computing framework for bioinformatics in big data. *Y. Zhang, S. Li, Z. Huang, W. Hu, Y. Li, W. Chen, X. Liu, L. Fang.* BGI-Shenzhen, Shenzhen 518083, Guangdong, China.

Big data is a new normal for bioinformatics in properties of volume, variety and velocity. And three computing issues are most concerned: 1) rapidly analyze the individual sample data, 2) analyze each sample of a large dataset in a resource efficient way, 3) enable the analysis of large scale samples together in one model like population variant calling. Multiple solutions came out, including hardware accelerator and cloud computing technology. And the big data technology, built on Apache Hadoop as the core, is also gradually adopted in bioinformatics. But a sophisticated solution for all the three computing issues is never really achieved. SOAPgaea, which started at 2011 to support the big data computing at BGI, utilized the Hadoop/Spark as the core technology to solve these problems. SOAPgaea is mainly developed using Java 1.8 with C/C++ involved. New unstructured file storage (HDFS) and structured table storage (HBase, etc.) have been utilized in SOAPgaea to accelerate the data access process. Hadoop MapReduce API and Spark API were adopted for parallel processing of multiple NGS data analysis functions, including pair-end reads aggregation, window based aggregation, duplicate reads aggregation. Based on the fundamental work, SOAPgaea implements one-stop solution for NGS data analysis including variant calling and annotation and variant frequency computation. Meanwhile, SOAPgaea optimized the parallel strategy to enable the multi-sample running mode to reduce the runtime cause by the task scheduling and load unbalance. SOAPgaea completed the variant calling pipeline and annotation in 2 hours for a 30X WGS sample, from GIAB NA12878 dataset and usually takes more than 2 days, on 16 nodes (24 cores in 2*Intel Xeon E5645, 128G memory and 12*4T disk for each node) of Hadoop clusters at BGI, with sensitivities of 99.68% and 95.26% for germline SNVs and indels. Moreover, we applied SOAPgaea to 100 WES samples with 834GB gzip compressed FASTQ file from 1000 genomes project in the modes of 1/run, 10/run, 20/run, 50/run and 100/run. And it takes 50, 18, 17, 14 and 11 hours, respectively, nearly 5 times faster than single sample mode. Furthermore, we applied SOAPgaea in the joint variant calling of 1000 WES CRAM files from 1000 genomes project of phase 3, which reached a data size of 3.2 TB data in total. And SOAPgaea finished in less than 12 hours on the same cluster in one job, instead of traditional way which takes days to split and compute the data manually.

1388T

Gene set comparison analysis using both first and second moment information. *L. Zhang, W. Tang, HW. Deng.* Tulane University, New Orleans, LA.

Background: Gene set analysis (GSA) aims to determine whether a priori defined gene set, in which the genes share some common biological function, is correlated with a particular phenotype. Numerous GSA methods have been proposed to explore how the gene regulatory network differs between two or more disease status. However, most methods are developed with respect to a particular scenario, either focusing on differential mean patterns or differential co-expression patterns. Here, we present a novel method that can capture both mean and variance-covariance information regarding the association between gene sets and phenotypes of interest. **Result:** Our method tests the null hypothesis that for a pre-defined gene set, there is no difference in both first and second moment information of the genes between case and control groups. We compared the type I error and power of our method with two existing prevailing GSA methods through simulation studies. We found that our method can control the type I error rate at the pre-specified significant level, and it has an improved power to detect gene sets in various situations, especially when difference in both mean and variance-covariance features between case and control group is large. **Conclusions** In summary, we proposed a new gene set comparison test that could detect significant changes in both mean and variance-covariance patterns between two different biological conditions. Since our approach is able to capture more information regarding the association between gene sets and relevant phenotypes, it may help to find broader gene regulatory networks or biological pathways essential for disease development and progression.

1389F

A new diagnostic platform (Genomic Intelligence®) improves accuracy of whole exome and genome sequencing diagnosis in rare disease. *A. Fisher, Y. Rachmany, H. Neerman, N. Meeks.* Variantyx Inc., Framingham, MA.

Background: Whole exome (WES) and genome sequencing (WGS) is becoming the standard of care for diagnosis of rare disease. Reports show a diagnostic yield of approximately 25-40 percent in published cases. This leaves a majority of cases undiagnosed even after testing and analysis. The lack of diagnosis may be multifactorial including variants not detected due to lack coverage, synonymous, intronic or intergenic variants for which there is no known interpretive data or variants missed due to a lengthy and cumbersome curation process. Methods: A new diagnostic platform (Genomic Intelligence®) has been developed to improve accuracy and timeliness of diagnosis for those variants that may be missed due to the cumbersome curation process. Using FASTQ data, this web-based platform allows curation of each variant at the variant, gene and protein level. It collates information from common mutation databases (HGMD, ClinVar, OMIM) as well as incorporates in silico prediction programs and publicly-available allele frequencies to allow efficient and accurate curation of variants. Previously-undiagnosed cases of whole exome and whole genome sequencing were re-processed using this diagnostic platform. This included singletons as well as trio cases. Each case was reviewed and variant curation was performed by an ABMGG-certified Geneticist. The results were discussed individually by phone with the ordering clinicians to confirm a plausible diagnosis. Results: 55 previously-undiagnosed whole exome and whole genome cases were re-processed with a new possible or confirmed diagnostic rate of 55 percent. In trios the rate was 22/36 (61 Percent). In single patients, the rate was 8/18 (44 percent). In WGS, the rate was 5/7 (75 percent). In WES, the rate was 25/48 (52 percent). Conclusions: Genomic Intelligence allows for improved diagnosis of whole exome and genome sequencing as it automates a cumbersome, error-prone and time-consuming process. These results are particularly significant in that these cases were previously-undiagnosed and yet in more than half of cases, the platform discovered a diagnosis. It also demonstrates the importance of re-processing of previously un-diagnosed cases. This tool allows geneticists to focus their time and expertise on variant curation and literature review instead of information gathering. This increases the yield of genetic diagnosis allowing for better patient care.

1390W

Uncovering the genetic architecture of complex traits: A Kalman filter approach. *D. Palmer[1,2], B. Neale[1,2], N. Patterson[2,3,4].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Richard B. Simches Research Center, 185 Cambridge Street, CPZN-6818, Boston, MA 02114; 2) Stanley Center, Broad Institute, 75 Ames Street, Cambridge, MA 02142; 3) Department of Genetics, Harvard Medical School 77 Avenue Louis Pasteur, New Research Building, Boston, MA 02115; 4) Radcliffe Institute for Advanced Study, Harvard University, 10 Garden Street, Cambridge, MA 02138.

Methodological approaches have been invaluable in increasing our knowledge of the genetic contribution to complex traits. Approaches using GWAS summary statistics are particularly useful, as such data are readily available from across a raft of diseases and traits. An example of such an approach is LDscore regression [1], which has been wildly successful. The method hinges on an elegant relationship between marginal effect sizes and the smeared tagging effects of LD in that population. Extensions to the method are already highly cited in the literature (give an number here). These include examinations of the correlation between genetic architectures as well as partitioning the genome to determine the relative contribution of constituent parts to variance in phenotype [2-3]. Despite these existing extensions, in general it is not obvious how to extend LDscore to more complex models relating genetics to phenotype. We present a complementary approach, which, in the first instance, is based on Fisher's infinitesimal model. Our method has the added benefit of being readily extendable to more complex underlying models of the impact of genetics on phenotype, and naturally allows for the incorporation of priors in a well-studied statistical framework. By exploiting the approximately banded nature of LD, we can construct a Kalman filter in which true effect sizes are the hidden states. Such an approach can be used to estimate heritability, as well as the underlying true effect sizes under the model. These estimates, which account for the correlation structure imposed by LD, have a natural application in generating polygenic risk scores which may explain a greater proportion of phenotypic variation when applied across human populations. [1] Bulik-Sullivan *et al.*, 2015. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. [2] Bulik-Sullivan *et al.*, 2015. An atlas of genetic correlations across human diseases and traits. [3] Finucane *et al.*, 2015. Partitioning heritability by functional annotation using genome-wide association summary statistics.

1391T

Organize and share your bioinformatics analysis with the R package workflowr. *J.D. Blischak[1], P. Carbonetto[1,2], M. Stephens[1,3].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Research Computing Center, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL.

With rapid advances in sequencing technologies, a geneticist can quickly amass large amounts of complex data. Analyzing massive data sets requires organization and planning, which is especially challenging when juggling multiple projects at once. For example, it can be difficult to recall which version of the data or code was used to produce a particular figure. This is the motivation for our R package, workflowr: managing and organizing research projects is a complex task, and practitioners need all the help they can get. workflowr improves management of computational projects by suggesting a "workflow" designed to promote best practices for data management and code development. We chose R because it is a popular programming language for data analysis in science and particularly Human Genetics. Any researcher that is familiar with R can instantly start using workflowr. workflowr includes four key features to promote effective project management: (1) workflowr automatically creates subdirectories to separate data, code, and results files; (2) workflowr uses the revolutionary version control system Git to track the code as each individual analysis is completed without the user needing to understand the complexities of Git; (3) to support reproducibility, workflowr provides a custom R Markdown template that combines the code, figures, and explanatory text into a single HTML file per analysis and also automatically displays the version of the code that produced the results and; (4) workflowr provides instructions for transferring the HTML files online to facilitate sharing with labmates and colleagues. As a proof of concept, we previously used this framework to manage a large project investigating the technical sources of variation in single-cell RNA-sequencing data, and the results of this project can be viewed at https://jdblischak.github.io/singleCellSeq/analysis/. Our goal is that any scientist using workflowr will find it easier to write well-documented, reproducible code. Documentation and source code for workflowr are available at https://github.com/jdblischak/workflowr.

1392F

TeraPCA: A fast and scalable method to study genetic variation in tera-scale genotypes. *A. Bose[1], V. Kalantzis[2], E. Kontopoulou[1], M. Elkady[1], P. Paschou[3], P. Drineas[1].* 1) Department of Computer Science, Purdue University, West Lafayette, IN; 2) Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN; 3) Department of Biological Sciences, Purdue University, West Lafayette, IN.

One of the significant challenges in population genetics is the growing size of terabytes of data with close to one million genotyped humans. Principal Component Analysis (PCA) serves as a key tool in capturing population stratification, as it maps the genetic variation inferred from SNPs. However, a key limitation of PCA is its scalability and computational complexity when applied on tera-scale data, e.g. about one million individuals at one million SNPs. To address this issue, we propose TeraPCA, ie. we compute the top Principal Components (PCs) of tera-scale matrices using a method known as Randomized Singular Value Decomposition (RandSVD). After computing the mean-centered matrix, we invoke RandSVD to approximate, under a user-specified accuracy, the dominant PCs. Our implementation is based on multithreaded libraries such as LAPACKE, BLAS and MKL, and it can handle datasets which might exceed the amount of available system memory by performing out-of-core computations. We ran our algorithm on publicly available data sets such as the Human Genome Diversity Panel (HGDP), which we replicated to create data sets containing two thousand individuals and 1.4 million SNPs. Our implementation of TeraPCA outperforms state-of-the-art methods such as EIGENSOFT's smartpca, while producing similar qualitative results when ran on the same computing system. As a brief illustration, smartpca required 61.2 mins to compute the top 10 PCs exploiting 20 threads, whereas TeraPCA required 14.7 mins and 11.5 mins exploiting 1 and 20 threads, respectively. TeraPCA was also faster than smartpca's fastmode which converged in 15.6 mins. We also applied TeraPCA on simulated genotyped data, generated by the Pritchard-Stephens-Donelly (PSD) model for ten thousand individuals and ten million SNPs, leading to tera-size data sets and tera-sample-size scales. TeraPCA is an implementation which targets the handling of tera-scale data while combining the accurate capturing of genetic variation of the samples with a low computational overhead, and with the ability to take advantage of current multiprocessor architectures by being implemented on top of established multithreaded linear algebra routines.

1393W

Performance optimization of a genomic variant store for genotype-phenotype correlation in PhenoTips software. *P. Buczkowicz[1], S. Mastromatteo[2], T. Nguyen[1], S. Stanisor[1], D. Gross[1], M. Brudno[1,2].* 1) Gene42 Inc., Toronto, Canada; 2) Department of Computer Science, University of Toronto, Toronto, Canada.

Whole-genome and whole-exome sequencing have increased the discovery of the genetic basis for thousands of Mendelian phenotypes, thereby establishing a causal link between genotypes and phenotypes. However, even with the influx of big datasets from genome-wide sequencing efforts, approximately half of Mendelian phenotypes still have no known underlying genetic partner. This represents a large gap in the knowledge of genotype-phenotype correlation and has a significant impact on the diagnostic odysseys in medical genetics and rare disease research. A Variant Call Format (VCF) file is a standard file type used to store human genomic data and contains a list of DNA positions or segments that differ from a reference sequence. However, a VCF file for a single individual could be larger than 1 gigabyte and as a result performing queries on data from any non-trivially sized population introduces significant computational challenges. Here, we introduce a comprehensive genomic variant store compatible with open source PhenoTips software for the enhanced storage and querying of genomic data from VCF files, and capable of advanced genotype-phenotype correlation mapping. In order to interact with VCF data in real-time, it is necessary to design an effective system to store and quickly retrieve large amounts of variant data. We created a set of benchmark tests to empirically compare the performance of different data storage and processing paradigms with respect to VCF data. The storage architectures were compared based on compression ratio, query latency, and upload time in addition to the scalability across all these dimensions. We considered a variety of different open-source tools including an indexed search engine, a distributed processing system, columnar and document databases. Using the results of the benchmark tests, we chose the best performing architecture to construct a PhenoTips-integrable variant store. We show that this tool can help in providing additional insight into the study design and analytical strategies for big data storage and analysis associated with genome-wide sequencing and genotype-phenotype correlation.

1394T

From quantity to quality: A population-based approach for building reference panel imputation. *M. Cocca[1], D. Toniolo[3], P. Gasparini[1,4], M. Mezzavilla[1,2].* 1) Medical Genetics, Institute for maternal and child health IRCCS "Burlo Garofolo", Trieste, Italy; 2) Division of Experimental Genetics, Sidra Medical and Research Center, Doha, Qatar; 3) San Raffaele Scientific Institute, Division of Genetics and Cell biology, Genetics of common disorders Unit, Milan, Italy; 4) Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy.

Imputation is widely used to infer genotypes in a target population using a denser scaffold of known haplotypes (Reference Panel) and has been useful to improve the results of genome-wide association studies (GWAS). The use of a custom reference panel, containing genomes from the target population, can improve imputation quality in particular in the low and rare frequency spectrum. However, this blind approach can lead to a costly procedure in terms of time and resources, since the tendency is to build bigger reference panels to increase the variation captured. Our aim is to provide a population-based approach which rely on genetic distance (Fst) between reference and target population, to create a reference panel based on specific characteristic of the population included. We generated 4 different panels subsampling the main 1000G Phase 3 reference. Our target population is composed by 1785 individuals from Val Borbera in North Italy. The study genotypes were phased using the SHAPEIT2 software and imputation was performed using the IMPUTE2 software. We compared two IMPUTE2 statistics across the 4 panels: concordance between imputed and real genotype (r2) and info score. The panels contain from 398 to 408 samples, but have different mean Fst from the target population. One panel is composed by African populations, one only East Asian, one only European and one composed by two European populations one African and one East Asian (admixed panel). Our results show that the panel with highest r2 for variants with MAF<2% has an Fst mean of 0.005 and is the one composed by only Europeans. This panel is also the one with the highest percentage of rare variants with info score >0.4 (a common threshold to select high quality variants for GWAS studies). Moreover, the panel with lowest mean Fst to the target has 68% of rare variant that could be used in GWAS studies, in opposition to the panel with only Africans, which has Fst of 0.14 and only 44% of the imputed rare variants are informative enough for GWAS. The East Asian panel with Fst of 0.1 has the 48% of rare variants with info score >0.4 and the admixed panel with Fst of 0.05 has the 62% of the variants exploitable for GWAS analyses. Our results demonstrate also that building a population specific reference panel could be a good choice in terms of resources and time needed for the imputation process, and the inclusion of population with extreme values of Fst in a reference panel does not improve the overall imputation.

1395F

A framework for using protein structure specific features to elucidate ambiguous non-synonymous single nucleotide variants. *J.B. Jespersen[1,2,3], K. Davidsen[2,3], K. Lage[1,2].* 1) Department of Surgery, Massachusetts General Hospital, Boston, MA, USA; 2) Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 3) Department of Bioinformatics, Technical University of Denmark, Lyngby, Denmark.

When assessing genetic variants, some end up without interpretation due to ambiguity, and get marked 'Variant of Uncertain Significance' (VUS). This method particularly aims to elucidate missense mutations, which otherwise, would not be interpreted. The importance of single amino acids based on their known experimental structure, and a variety of features, both extracted from PDB as well as calculated from the protein structures. Genomic coordinates are in a robust and protein isoform specific manner, mapped to single amino acids in PDB files. Hence, the method is limited to only being able to inform on variants that has an experimentally solved protein structure, which passes a quality threshold. All data has been pre-calculated, and will be accessed when an inquiry is made on particular variant. The input format can be a VCF-file and will be returned as enriched with information on all variants that has protein structural information available.

1396W

Genoppi: A web application for interactive integration of experimental proteomics results with genetic datasets. *A. Kim[1,2], E. Malolepsza[1,2], J. Lim[1,3], K. Lage[1,2].* 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Department of Surgery, Boston, MA; 3) Massachusetts Institute of Technology, Cambridge, MA.

Integrating protein-protein interaction experiments and genetic datasets can lead to new insight into the cellular processes implicated in diseases, but this integration is technically challenging. Here, we present Genoppi, a web application that integrates quantitative interaction proteomics data and results from genome-wide association studies or exome sequencing projects, to highlight biological relationships that might otherwise be difficult to discern. Written in R, Python and Bash script, Genoppi is an open source, user-friendly framework easily deployed across Mac OS and Linux distributions and freely available at https://github.com/lagelab/Genoppi.

1397T

A unified web platform for network-based analyses of genomic data.
T. Li[1,2,3], J. Mercer[1,2], J. Rosenbluh[1,4], A. Kim[1,2], H. Horn[1,2], L. Greenfeld[1,2], D. An[1], A. Zimmer[1], J. Liberzon[1], J. Bistline[1], T. Natoli[1], Y. Li[1,5,6], A. Tsherniak[1], R. Narayan[1], A. Subramanian[1], T. Liefeld[1], B. Wong[1], D. Thompson[1], S. Calvo[1,5], S. Carr[1], J. Boehm[1], J. Jaffe[1], J. Mesirov[1,7], N. Hacohen[1,8,9], A. Regev[1,10], K. Lage[1,2,9]. 1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Department of Surgery, Massachusetts General Hospital, Boston, MA; 3) Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA; 4) Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; 5) Howard Hughes Medical Institute and Department of Molecular Biology, MGH, Boston, MA; 6) Department of Statistics, Harvard University, Cambridge, MA; 7) University of California, San Diego, CA; 8) Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Boston, MA; 9) Harvard Medical School, Boston, MA; 10) Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA.

With significant technological advances in epigenetics, proteomics and RNA sequencing, it is now possible to generate an unprecedented amount of tissue- and cell-type-specific functional genomics data that can be conveniently represented as gene networks. Integrating these networks with exome-sequencing data, or genome-wide association studies, is a cost-efficient and scalable way to identify draft cellular circuits that are enriched for genetic risk in a particular disease. However, different networks vary considerably in terms of how noisy they are and their global and local biological signals. Therefore, a major bottleneck in network-based analyses of genomic data is the quantitative comparison of biological signal in different networks to identify the optimal network dataset for a particular biological question. To address these barriers, we developed a unified web platform "Broad Institute Web Platform for Genome Networks (GeNets)". We designed a machine learning algorithm (Quack) that can learn the topological patterns of any pathway set in any network defined by users, allowing them to make detailed comparisons of the global and local biological signals of different networks. We trained Quack on 853 canonical pathways on five networks, which exhibited high area under the receiver operating charateristics curve (AUC, median 0.81) when predicting pathway membership, suggesting broad utility of Quack across different types of functional networks. To illustrate typical GeNets workflows, we analyzed 65 genes implicated in autism spectrum disorders (ASD). We chose the human protein-protein interaction network InWeb because it performed optimally in learning pathway relationship across 306 neuronal pathways compared to the other four networks. InWeb-trained Quack model identified 31 candidates topologically related to 65 ASD genes. Using GeNets visualization features, we found a significant overlap (3 genes) between the 31 candidates and genes implicated in schizophrenia and neurodevelopmental delay. Among the three genes, we discovered that *NAGA* has a brain-specific eQTL, and mutations in *NAGA* can cause Schindler disease which has overlapping symptoms with ASD, suggesting that *NAGA* could be an interesting ASD candidate gene. Overall, GeNets is a scalable and uniquely enabling computational framework for expert and non-expert users alike to perform, manage, and share analyses of genetic datasets using heterogeneous functional genomics networks.

1398F

Quantitative transcription factor occupancy across cell types, conditions, and genetic variants.
K. Luo[1,2,3,4], J. Zhong[1,2,3], A. Safi[2,5], L. Hong[2,5], A. Tewari[6], L. Song[2,5], T. Reddy[1,2,7,8,9], L. Ma[1,10], G. Crawford[1,2,5], A. Hartemink[1,2,3,11]. 1) Computational Biology & Bioinformatics Graduate Program, Duke University, Durham, NC; 2) Center for Genomic & Computational Biology, Duke University, Durham, NC; 3) Department of Computer Science, Duke University, Durham, NC; 4) Department of Human Genetics, University of Chicago, Chicago, IL; 5) Department of Pediatrics, Duke University Medical Center, Durham, NC; 6) Department of Internal Medicine, Beth Israel Deaconess Medical Center, Boston, MA; 7) Department of Biostatistics & Bioinformatics, Duke University Medical Center, Durham, NC; 8) Department of Biomedical Engineering, Duke University, Durham, NC; 9) Department of Molecular Genetics & Microbiology, Duke University Medical Center, Durham, NC; 10) Department of Statistical Science, Duke University, Durham, NC; 11) Department of Biology, Duke University, Durham, NC.

Thousands of different transcription factors (TFs) bind with varying occupancy across the human genome. Chromatin immunoprecipation can profile occupancy genome-wide, but one TF at a time, limiting our ability to comprehensively observe the TF occupancy landscape, let alone how it changes. We developed TOP, a Bayesian hierarchical regression framework to profile genome-wide quantitative occupancy of many TFs using data from a single DNase-seq experiment. TOP is supervised, but its hierarchical structure allows it to predict the occupancy of any sequence-specific TF, even those without training data. We used TOP to profile the quantitative occupancy of nearly 1500 TFs, and examined how their occupancy changed genome-wide in a number of contexts: across 178 human cell types, over 12 hours of exposure to different hormones, and across the genetic backgrounds of 70 individuals. TOP enables cost-effective exploration of quantitative changes in the landscape of TF binding.

1399W

PHESANT: A tool for performing automated phenome scans in UK Biobank. *L.A.C. Millard, N.M. Davies, T.R. Gaunt, G. Davey Smith, K. Tilling.* MRC Integrative Epidemiology Unit (IEU), School of Social and Community Medicine, University of Bristol, UK.

  Epidemiological cohorts typically contain a diverse set of phenotypes such that automation of phenome scans is non-trivial, because they require highly heterogeneous models. For this reason, phenome scans have to date tended to use a smaller homogeneous set of phenotypes that can be analysed in a consistent fashion. We present PHESANT (PHEnome Scan ANalysis Tool), a software package for performing comprehensive phenome scans in UK Biobank, including phenome-wide association studies (pheWAS) and Mendelian-randomization phenome-wide association studies (MR-pheWAS). PHESANT tests the association of a specified trait with all continuous, integer and categorical variables in UK Biobank, or a specified subset. PHESANT uses a novel rule-based algorithm to determine how to appropriately test each trait, then performs the analyses and produces plots and summary tables. We report results using PHESANT to search for the causal effects of body mass index, using the MR-pheWAS approach.

1400T

Comparative analysis of methods for discovery of germline copy-number variants from exome data. *B.D. O'Fallon[1], J. Durtschi[1], T. Lewis[2], W. Donahue[2], S. Reading[2], S. Dames[2], D. Close[2], H. Best[2,3], E. Gee[2].* 1) Biocomputing, ARUP Labs, Salt Lake City, UT; 2) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 3) Department of Pathology, University of Utah Health Sciences Center, Salt Lake City, UT.

  Purpose Sensitive and specific identification of copy number variants (CNVs) in human exomes remains a significant computational challenge, particularly for events spanning only one or two capture targets. We present a comparative analysis of six CNV detection tools, including CLAMMS, Convading, GATK (v4.1), Conifer, ExomeDepth, XHMM, and CODEX, and evaluate their performance on the detection of CNVs spanning 10 or fewer exons in germline exome data. **Methods** 44 human exomes were captured with IDT xGen Exome Research Panel v1.0 and sequenced on an Illumina HiSeq 4000 instrument. Reads were aligned with BWA MEM (v0.7.12) and duplicate marked using Picard Tools (v2.9). Nonoverlapping deletion and duplication CNVs spanning 1,2,3 and 10 exons were introduced into alignment (BAM) files by removing or duplicating reads, respectively. Caller performance was assessed by tabulating the sensitivity, specificity, and positive predictive value (PPV) at the quality threshold that maximized the F_1 statistic for each test sample. We investigated the effect on sensitivity and PPV of the number of control samples used and, by downsampling both control and test samples, the mean sample read depth. We also examined reproducibility as measured by simulated CNV call concordance from three sets of internally replicated samples. **Results** Sensitivity and PPV increased with increasing CNV size across all callers. However, we found substantial variability among methods, with differences becoming more pronounced with decreasing CNV size. For single exon CNVs, mean sensitivities ranged from 0-85% across callers, while PPV varied between 20-90%. For 3- and 10-exon deletions, sensitivity and PPV varied from 60-90% and 80-100%, respectively, for most methods. Using mean F_1 across all batches as a measure of overall accuracy, 3 of the 6 callers demonstrated increasing accuracy as the number of control samples was increased from 15 to 35. Mean F_1 for all callers improved as read depth increased from 50 to 200, although the magnitude of the effect varied considerably. Reproducibility ranged from 44-96% across methods and increased with CNV size, with nearly all callers achieving 90% or greater concordance for 10-exon CNVs. **Conclusions** These results are consistent with earlier work demonstrating limited accuracy and precision of small CNV detection from targeted NGS data, but suggest that increasing read depth and the number of control samples may increase accuracy considerably for some methods.

1401F

DNA sequence and quality value accuracy for a new Sanger sequencing instrument. *S.J. Schneider, S.R. Berosik, J. Bodeau, Y. Chu, A. Gerstner, P. Ma, D. Rodriguez, R. Schaeffer, D.C. Woo.* Thermo Fisher Scientific, South San Francisco, CA.

Sanger Sequencing with capillary electrophoresis is a cost-effective and time-efficient DNA sequencing method to generate high quality and highly accurate DNA sequence information to interrogate DNA for both Mendelian variants and now minor variants present at 5% or greater, using Minor Variant Finder software. With the introduction of the new easy-to-use Applied Biosystems SeqStudio Genetic Analyzer, with a new polymer, new spectral auto-calibration algorithm, updated KB Basecaller, and long-, medium-, or short- read sequencing options, it is important to test the accuracy of basecalls and Quality Values of SeqStudio sequence data.   Development of a new capillary electrophoresis DNA analysis platform includes the development and validation of the basecalling capability. Dye-labeled DNA fragment mobility correction as well as the capability of generating high-quality, highly-accurate sequencing content is essential for successful sequencing projects. We describe the process used to evaluate the performance of this new system, including validation of the updated KB™ Basecaller basecalling algorithm. The process starts with the creation of the mobility correction file content and includes the data generation and validation of the pure- and mixed-basecalling abilities of the KB™ Basecaller for the system using millions of bases of both genomic and FFPE-derived DNA generated on multiple SeqStudio instruments to maximize sequence variability and instrument variability.

1402W

Genotype array missing variant imputation with 78 batches comprising ~84,000 individuals. *I.B. Stanaway[1], J. Jackson[2], B. Mapes[2], M.R. Palmer[3], A. Gordon[3], F. Lin[4], Y. Joo[4], M.G. Hayes[4], A. Gharavi[5], M.D. Ritchie[6], E.A. Rosenthal[3], B. Namjou-Khales[7], M. Andrade[8], K. Kiryluk[5], G.P. Jarvik[3], D.R. Crosslin[1].* 1) Biomedical Informatics Medical Education, University of Washington, Seattle, WA; 2) eMERGE Network, Coordinating Center Nashville, TN; 3) Medical Genetics, University of Washington, Seattle, WA; 4) Northwestern, Chicago, IL; 5) Columbia University Department of Medicine, New York, NY; 6) Geisinger Health System, Danville, PA; 7) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 8) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN.

The Electronic Medical Records and Genomics (eMERGE) Network comprises 12 medical centers with electronic health records (EHR) linked to genotype data for ~84,000 participants. To create a shared resource to perform association analyses with phenotypes derived from the EHR, we imputed 78 genotype array batches against the Haplotype Reference Consortium 1.1 using the Michigan Imputation Server (MIS). A total of 5,166,562 variants were directly genotyped in any one or more array batches. We analyzed the missing genotype imputation efficiency of the various genotype arrays using the R-square (Rsq) metric reported by the MIS at 39,131,578 variants genome wide. For each array batch, we calculated the mean Rsq (range:0.089-0.68). A total of 51 array batches had a mean Rsq >0.3. We constructed a linear model using the variables, batch sample count (range:4-9,315) and the count of directly genotyped variants (range:50,911-3,410,557). Both the log of array batch sample count ($p<2\times10^{-16}$) and the log count of directly genotyped variants ($p\sim5.6\times10^{-6}$) were significantly associated with the array batch imputation performance measured using the array mean Rsq, and explained 62% and 21% of the variance, respectively. The averages of each variant's Rsq across chips had 21,924,838 variants with a mean >0.3, and 17,294,872 variants with a median >0.3. PCA ancestry analysis using imputed genotypes from chromosome 22 showed concordance with self-reported race and displayed the canonical European, African and Asian ancestry sample groupings without evidence of gross batch effects in the first two principal components. Our metrics of Rsq summaries, principal components, and identity-by-descent will assist in performing association analyses with the rich eMERGE phenotype data derived from electronic health records.

1403T

Graph embedding and visualization of genetic data using paired factor analysis. *G.T. Wang[1], KK. Dey[2], M. Stephens[1,2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Statistics, University of Chicago, Chicago, IL.

Many problems in genetics -- from characterizing gene-expression dynamics in single cell development, to inferring human population migration and admixture -- can be viewed as constructing graph embedding of high dimensional data. Currently a number of non-model based computational tools, eg Monocle, TSCAN, Slingshot, have been developed for discovering developmental trajectory in single cell data. To better understand their relative merits we have developed a Dynamic Statistical Comparison (DSC) benchmark using different types of simulated trajectories (tree, cycle and other planar graphs). We evaluated single cell specific methods as well as a number of generic approaches to dimensionality reduction and manifold learning. Our benchmark illustrates situations when certain types of noise can undermine detection power of some methods. In particular we have observed trade-off in their capability to recover global and local pattern for complex problems. In order to exploit the local linear structure in data we propose a model based approach called Paired Factor Analysis (PFA), which assigns grades of memberships to each sample with the constraint that a sample can have memberships in at most two clusters. Consequently, a graph can be embedded with factors being the nodes and samples aligning to edges. Statistical evidence for graph edges can be obtained to impose sparsity when desired, creating visualizations that automatically reveal trajectories in high dimensional space when they are present. Under the DSC benchmark, PFA is capable of recovering simulated trajectories, and is more resilient to certain noise types compared to principal component analysis. We have applied PFA to publicly available genetic data-sets including single cell RNA-seq data and genotype data of admixed individuals, providing visualizations that reflect core features in both data types. Our method is available as an R package at https://github.com/gaow/pfar.

1404F

Polaris: A collection of 220 publicly available whole genomes for sharing validated structural variants. *M.A. Bekritsky[1], P. Krusche[1], N. Johnson[2], E. Dolzhenko[2], A. Gross[2], B.R. Lajoie[2], V. Rajan[2], N. Gonzaludo[2], Z. Kingsbury[1], S.J. Humphray[1], S.S. Ajay[2], R.J. Taft[2], D.R. Bentley[1], M.A. Eberle[2].* 1) Illumina Cambridge Ltd., Chesterford Research Park, Saffron Walden, Essex, CB10 1XL, United Kingdom; 2) Illumina Inc., 5200 Illumina Way, San Diego, CA, 92122, USA.

Structural variant (SV) calling is an essential component of a clinical genomics pipeline. However, benchmarking SV caller accuracy is difficult due to the scarcity of well-defined validation methods and a dearth of sufficient whole genome sequencing (WGS) data to robustly test calling methods. Furthermore, once WGS data has been generated, researchers may be unable to share their data with others without compromising genetic privacy. We seek to address these challenges by combining publicly available WGS data with an extensive population and Mendelian genetic validation framework to facilitate sharing and benchmarking common SVs. The Polaris dataset is a collection of 220 unrelated Coriell samples, with full consent for posting of personally identifiable genetic information. The individuals within the cohort are from diverse ethnic origins and includes 70 samples from the Genetic Testing Reference Material Coordination Program (GeT-RM). Each sample has been sequenced to 35x coverage on the Illumina HiSeqX platform with 2x150 bp reads generated from a TruSeq PCR-Free library. Complete sequence data for these samples is available from the European Nucleotide Archive (PRJEB20654 and PRJEB19931). To accompany the raw WGS data, we are also releasing an initial set of >70,000 SVs that are observed in at least one Polaris sample and are consistent with population and Mendelian genetics. Candidate SVs include large deletions and insertions identified using a variety of inputs, including previously published insertions and SVs identified in the Platinum Genomes pedigree. These SVs were jointly re-called on the Polaris samples using either graph-based or depth-based structural variant calling methods. Each variant is in Hardy Weinberg Equilibrium in the Polaris dataset and an additional cohort of 3,000 samples, and satisfies pedigree consistency in the Platinum Genomes. Many of these events are also in linkage disequilibrium with nearby SNPs, providing further genetic confirmation of their validity. This data marks the beginning of a community resource to share and validate common SVs and more complex structural rearrangements that will continue to expand as more samples and further analyses are added. We hope to engage other members of the research community in our pursuit of a constantly improving and robust truth-set, and a widely accessible validation and coordination framework.

1405W

CNVs in clinical WGS: Deployment and interpretation for rare and undiagnosed disease. *A.M. Gross, S.S. Ajay, V. Rajan, C.M. Brown, K. Bluske, N. Burns, A. Chawla, A.J. Coffey, A. Malhotra, A. Scocchia, E. Thorpe, N. Johnson, E. Dolzhenko, B. Lajoie, J. Belmont, D.L. Perry, M.A. Eberle, R.T. Hagelstrom, D.R. Bentley, R.J. Taft.* Illumina, San Diego, CA.

   Here we describe the deployment of a copy-number variant (CNV) detection pipeline as a component of a clinical whole genome sequencing diagnostic test for rare and undiagnosed genetic disease (RUGD). Since late 2016 a total of 79 RUGD cases have been assessed in the context of SNVs, indels and CNVs (>10kb) in the Illumina Clinical Services Laboratory. All cases included a detailed review of medical history followed by sequencing on the HiSeq X (2 x 150 bp) and analysis using a bioinformatics pipeline that includes automated filtering, annotation, and visualization. Copy number calls that passed automated prioritization were submitted to manual quality control and variant curation before being compiled into reports. On average, we found 2.5 benign, 3 VUS-likely benign, and 4 VUS CNVs were reported per case. In 15% (11/79) cases, we reported out variants with pathogenic or likely pathogenic classifications, across a diverse set of patient phenotypes. In addition to these CNVs, our pipeline also detected a full chromosome mosaic trisomy, and a uniparental isodisomy of chromosome 15. CNVs can be the result of simple or more complex genomic rearrangements. To understand better the mechanism of the event leading the copy-number change, our pipeline integrates depth-based CNV calls with split-read based structural variants. Additionally, we assess the parent of origin of a deleted or duplicated haplotype via joint analysis of a pedigree, using inheritance patterns of small variants when a CNV arises de-novo. In practice, we used these tools to identify copy-number gains as tandem- or dispersed- duplications and link CNVs within an individual as part of chromosomal rearrangement events. We highlight the case of unbalanced translocation in an unaffected parent leading to large copy number changes in a diseased proband. Lastly, mosaic CNVs represent a unique class of variants, which are difficult to distinguish from sequencing noise. To address this we leverage data across a population of more than 3000 genomes to distinguish true copy-number changes from baseline sequencing depth variation. Among our cohort we highlight three mosaic CNVs, a 23kb deletion of a imprinted promoter, an unbalanced translocation, and a full chromosome trisomy. In conclusion, we report considerable success deploying CNV calling on top of an existing WGS assay. These results show the utility of WGS as an extendable platform for exploring the diversity of variation in RUGD.

1406T

The quantity of detected copy number variation (CNV) deletions substantially increases when coverage of whole-genome sequencing (WGS) data increases from 30x to 91x. *Y.S. Huang[1], C.Y. Tang[1], C.K. Liu[1], C.Z. Lo[2], Y.C. Lin[1], C.H. Chen[2], L.H. Li[2], A. Yao[1,2].* 1) National Center for Genome Medicine (NCGM), Academia Sinica, Taipei, Taiwan, R.O.C; 2) Institute of Biomedical Sciences (IBMS), Academia Sinica, Taipei, Taiwan, R.O.C.

   The correlation between copy number variation (CNV) and many human diseases has been established, especially in mental disorders, such as autism and schizophrenia. With advances in genome sequencing technologies, studies of CNV detection from whole-genome sequencing (WGS) data have accumulated. However, little is known about the correlation between coverage of WGS data and detected CNV events. In this study, we created a 30x and a 64x NA12878 WGS data by sampling pair-end reads from a 91x NA12878 WGS data; the 30x WGS data was a subset of the 64x counterpart. We focused on detecting CNV deletions with sizes between 10,000 and 100,000 base pairs in these NA12878 WGS data using Delly, a structural variant detector with a concise user interface, configurable genomic references, and continuous maintenance. The results indicated that the quantity of detected CNV deletions substantially increased when the coverage increased from 30x to 91x. Meanwhile, the increment of detected CNV deletions has not reached a plateau yet. To examine whether the number of detected CNV deletions in a WGS sample would approach a maximum when coverage of WGS data increases, we are currently working on creating WGS samples with various coverage by sampling pair-end reads from a 300x NA12878 WGS data and expect to present the results in the conference.

1407F

Full-spectrum copy number variation detection by high-throughput DNA sequencing. *Y. Jiang[1,2], K. Nathanson[3,4], N. Zhang[5].* 1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Division of Translational Medicine and Human Genetics, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Much effort has been devoted to elucidating the role of Copy number variations (CNVs) on human diseases and phenotypes. Such efforts have met with mixed success, with setbacks arising from the lack of reliable methods for sensitive and robust CNV profiling. In particular, in studies where it is desirable to attain deep coverage, or where the cohort is very large, whole-exome sequencing (WES) and targeted sequencing are often preferred methods, and there is a lack of reliable CNV detection tools for these platforms. For example, the Exome Aggregation Consortium has sequenced the exomes of 60,706 individuals and many cancer genomic studies use deep targeted sequencing to find low-frequency mutations within specific genomic regions. Here we propose a new method, CODEX2, to remove technical noise and improve CNV signal-to-noise ratio for all sequencing platforms including WES and targeted sequencing. A limitation shared by all current methods, highlighted by multiple independent benchmarking studies, is the lack sensitivity for common variants. CODEX2 is improved over existing methods with high sensitivity for both common and rare CNVs, thus allowing full-spectrum CNV detection. CODEX2 can be applied to two scenarios: the case-control scenario to detect CNVs that are enriched in the cases, and the scenario where control samples are not available and the goal is simply to profile all CNVs. CODEX2 is benchmarked against existing methods in three ways. First, we reanalyze the WES of the HapMap samples from the 1000 Genomes Project, where a set of experimentally validated CNVs from microarrays are used to assess performance. CODEX2 is shown to markedly improve both sensitivity and specificity over existing methods. Second, we apply CODEX2 to targeted sequencing data of melanoma cancer cell lines, where CODEX2 detects CNVs with recurrence rates that are highly concordant with those obtained from a separate cohort studied by the Cancer Genome Atlas. Finally, we perform extensive simulations benchmarking existing methods and elucidating how key variables, such as population frequency, influence sensitivity. CODEX2 is currently being applied to 12,000 WES samples in the Penn Medicine BioBank and 10,000 WES samples in the Alzheimer's Disease Sequencing Project. With increasing sequencing capacity, and increasing need to profile CNVs as a non-negligible source of genetic variation, we believe CODEX2 can be a helpful tool for the genetic and genomic community.

1408W

A clinically validated whole genome pipeline for structural variant detection and analysis. *A. Kaplun.* Variantyx Inc., Framingham, MA.

With the cost of whole genome sequencing (WGS) continuing to decrease we are fast approaching the point of inflection where WGS testing will become more economically feasible, facilitating broader access to the benefits that are helping to define WGS as the new diagnostic standard. Many benefits are inherent in the sequencing technology itself. PCR-free DNA preparation methods eliminate amplification-related issues producing better coverage of amino acid coding regions. In addition comprehensive coverage of intronic and intergenic regions ensures the identification of potentially relevant transcription factor binding site, enhancer and other regulatory variants. Importantly, WGS provides unique opportunities for detection of structural variants, but such analyses have not previously made their way into clinical practice. We have developed a clinically validated WGS pipeline for highly specific and sensitive detection of structural variants. Using a combination of breakpoint analysis, read depth analysis and *de novo* assembly of tandem nucleotide repeats and tri-nucleotide tandem repeats, the pipeline identifies structural variants down to single base pair resolution. False positives are minimized using calculations for loss of heterozygosity and bi-modal heterozygous variant allele frequencies to enhance heterozygous deletion and duplication detection respectively. To facilitate clinical interpretation, identified variants are annotated with phenotype information derived from HGMD® Professional and population allele frequencies derived from DGV. Single base pair resolution enables easy visual inspection of potentially causal variants using the IGV genome browser as well as easy biochemical validation via PCR. Patient cases demonstrating clinical utility of the pipeline will be presented.

1409T

Robust identification of deletions in next generation trio sequence data based on clustering of Mendelian errors. *K.B. Manheimer[1], N. Patel[1], F. Richter[1], J. Gorham[2], A. Tai[2], J.G. Seidman[2], C.E. Seidman[2], B.D. Gelb[1,3,4], A.J. Sharp[3].* 1) Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics, Harvard Medical School, Boston, MA; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY.

Identification of structural variation is important as they contribute to many types of disease including autism, schizophrenia and congenital heart disease. Next generation sequencing (NGS) data can be used to identify CNVs, often relying on changes in read depth as implemented by various tools such as XHMM for whole exome sequencing (WES) and CNVnator for whole genome sequencing (WGS). These tools have limited sensitivity, however, primarily because coverage is inherently variable across the genome.   Here, we present an orthogonal method to identify deletions from NGS data based on the clustering of Mendelian errors (MEs). Hemizygous loci (due to an underlying deletion) are assigned homozygous genotypes by variant callers. With the availability of trios, these loci can be detected as MEs. We developed the ME method (MEM) to identify deletion CNVs in 2600 WES trios and 350 WGS trios from the Pediatric Cardiac Genomics Consortium (PCGC). We compared MEM to XHMM and CNVnator for *de novo* deletion detection in WES and WGS, respectively.   From 2600 WES trios, MEM identified 171 putative deletions. Digital droplet PCR confirmed 91/104 for a positive predictive value of 87.5%, better than the 80% rate achievable with XHMM. Using MEM and XHMM, 74 *de novo* deletion CNVs were called: 37 by both tools, 4 MEM only, and 33 XHMM only. MEM false negatives were attributable to insufficient numbers of MEs, principally related to small deletion size (<30 kb). An additional strength of MEM was its ability to detect other sources of MEs: DNA contamination (n=6), broken trios (n=20), and uniparental disomy (n=8).   From 350 WGS trios, MEM identified 6676 putative deletions (19.1 deletions/proband), of which 11% were exonic. Using MEM and CNVnator, 33 *de novo* deletions were called: 9 by both tools, 8 MEM only, and 15 CNVnator only. MEM WGS false negatives were primarily for deletions smaller than 10 kb. Robustness of MEM for calling deletion CNVs from WGS was assessed using the trio from Genome in a Bottle for which PacBio long-read sequencing was available; MEM called 26/29 deletions (90%), although the boundaries for 2 true positive calls were inaccurate.   In conclusion, MEM is a robust method for calling inherited and *de novo* deletion CNVs from WES and WGS trio data. As it is orthogonal to read depth-based methods, it can obviate the need for PCR-based confirmations. MEM also detects UPD, which is missed by other NGS-based callers.

1410F

Automated parameter tuning for more accurate CNV calling in WES/WGS data. *M. Wiewiórka[1], W. Kuśmirek[1], M. Okoniewski[2], T. Gambin[1].* 1) Institute of Computer Science, Warsaw University of Technology, Warsaw, Mazowieckie, Poland; 2) Scientific IT Services, ETH Zurich, Zurich, Switzerland.

Background: Accurate detection of CNVs from sequencing data is required to provide a comprehensive molecular diagnosis. However, there is still a lack of best practices developed for germline CNV calling pipelines and therefore both implementation and tuning of such a pipeline in a diagnostic lab is a challenge for a number of reasons. First of all, a great selection of algorithms and contradictory conclusions from CNV calling benchmarks makes it difficult to choose the most suitable method for a given dataset. Additionally, different target designs and coverage profiles may greatly impact the accuracy of a particular tool. Furthermore, there are no clear recommendations for the optimal selection of control samples that are required for joint-calling algorithms. Finally, a fine-tuning of the specific CNV caller parameters is problematic without access to a golden standard call set. **Methods:** To address these issues we developed a method capable of automated parameter tuning for the most popular CNV calling algorithms, including XHMM and CODEX. The main idea behind our approach is to modify read coverage profiles of an user dataset to imitate existence of additional rare CNVs that we further use for accuracy assessment. Their coverage characteristics (i.e. relative changes in depth of coverage), allele frequencies, lengths and genomic locations are derived from the set of validated CNVs in 1000 Genomes samples and corresponding sequencing data. Since the processing of BAM files, required to obtain high resolution coverage profiles, is a time consuming problem, we used Apache Spark, i.e. a distributed computing framework. In addition, we proposed a novel columnar data format for storing precomputed read depth to further speed up calculations. **Results:** We demonstrated the utility of our framework on both publicly available subset of 1000 Genomes data and locally available dataset. Results indicated that our solution can substantially improve the CNV calling performance of state-of-the-art methods when compared to their default settings. Concluding, proposed approach simplifying CNV calling pipelines implementation may give rise to higher adoption of CNV analysis from sequencing data in clinical laboratories.

1411W

Integrative DNA copy number detection and genotyping from sequencing and array-based platforms. *Z. Zhou[1], W. Wang[2], L. Wang[2], N. Zhang[3].* 1) Graduate Group in Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Copy number variations (CNVs) are gains and losses of DNA segments that have been shown to be associated with disease. Previously, many large disease cohorts perform CNV analysis with data from SNP-array. However, with the cost of Next Generation Sequencing (NGS) dropped dramatically, those large cohorts redo their CNV analysis with technologies such as whole exome sequencing (WES) and whole genome sequencing (WGS), where both SNP-array and NGS data are available. An integrated analysis is expected to improve resolution and accuracy, yet there is a lack of tools that efficiently combines data from both platforms, nor has there been a systematic study of the expected accuracy improvement achievable by integration. Here we propose a statistical framework, integrated Copy Number Variation Analysis (iCNV), which can be applied to multiplatform study designs: WES only, WGS only, SNP array only, or any combination of SNP and sequencing data. iCNV allows platform specific normalization to remove platform specific biases, uses allele specific reads from sequencing and B-allele frequency from array data to improve detection sensitivity and genotyping accuracy, and achieves fast computation through a Hidden Markov Model (HMM). Using iCNV, we compare CNV calling integrating multiple platforms versus simply performing intersection or union of detections made by the two platforms and show that properly integration across platforms increases sensitivity and specificity. We further compare CNV detection sensitivity between single platform versus joint platforms by an in silico spike-in study and find that array data greatly improves detection power when added to WES, while not too much when added to WGS. Finally, we compare iCNV to other commonly used CNV calling methods when analyzing solely WGS data on a cohort containing related individuals, where Mendelian concordance-based quality metrics show that iCNV improves upon existing methods owing to its improved normalization pipeline and utilization of allele-specific reads.

1412T

CNVs from targeted NGS data: Building a cohort for validation and semiautomatic regression testing in a diagnostic setting. *M. Ziegler, D. Becker, Y. Dinçer, S. Eck, K. Pindl, K. Mayer, K. Gemmeke, HG. Klein.* Center for Human Genetics, Martinsried, Germany.

The validation of custom-made bioinformatics analysis systems in a GxP regulated setting is a major challenge due mostly to lack of reference material as well as to system complexity. Reference material for CNVs with exon level resolution is not available at all. Therefore, we set up a patient cohort, which underwent both targeted NGS analysis (Agilent QXT, NextSeq500) and further specification via MLPA (multiplex ligation-dependent probe amplification). The cohort consists of 277 patients with various phenotypes (e.g. Long QT syndrome, Marfan syndrome, cystic fibrosis, famililal hypercholesterolemia) who underwent more than 420 MLPA analyses. Probes from 19 MLPA kits were lifted from hg18 to GRCh37p13 and GRCh38p10 coordinates and linked to patient specific probe signals. The matrix was transferred to patient specific files in browser extensible data format (BED) in order to generate a dataset that can be evaluated semi-automatically. The dataset consists of about 20,000 probe positions encompassing 1.2 Mb of analyzed region. Genomic regions of interest were enriched and sequenced for all cohort individuals. For enrichment, we used seven different targeted panels with an average region size of about 300 kb per panel. FASTQ files were demultiplexed and mapped to the reference genome (GRCh38p10) via Picard tools. The resulting binary alignment map files (BAM) were sorted for enrichment version and gender, and stored for subsequent CNV analysis. Since the resulting file format of our custom-made CNV pipeline is BED, MLPA and NGS results are semi-automatically comparable. Therefore, we set up a comparison script to determine key figures for analytical quality (sensitivity, specificity, accuracy and precision). The MLPA dataset also allows fast and precise change management of the CNV pipeline.

1413F

Towards "gold standard" sequence-resolved structural variants in benchmark human trio reference samples. *J.M. Zook[1], L. Chapman[1], N.F. Hansen[4], F.J. Sedlazeck[5], M. Salit[2,3], Genome in a Bottle Consortium.* 1) Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD; 2) Material Measurement Laboratory, National Institute of Standards and Technology, Palo Alto, CA; 3) Department of Pathology, Stanford University, Palo Alto, CA; 4) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, Rockville, MD; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

The Genome in a Bottle Consortium (GIAB) has characterized an Ashkenazi trio from the Personal Genome Project (NIST Reference Material 8392) with 12 short, long, and linked read sequencing and mapping methods. Datasets are public without embargo for analysis and methods development by the community. We have characterized ~3.7 million small variants as well as reference calls for ~90% of the genome with estimated errors of ~2 FPs and 2 FNs per million variants. To extend this characterization to larger indels and structural variants, we collected analyses of variants >=20 bp from 33 bioinformatics methods and five technologies. Nineteen discovery and refinement methods produced sequence-resolved calls using local or global assembly or split reads, giving a precise prediction of deletion breakpoints, inserted sequences, and complex changes. These methods produced at least one sequence-resolved call in 82,386 of 88,348 candidate SV regions where we merged all candidate SV calls within 1000bp. We designed three different integration approaches to address challenges in comparing large variants, which are frequently in tandem repeats (>50% of all calls) and not precisely characterized. (1) We compared the sizes of deletions predicted in each candidate region and found 20,218 deletions in the trio where >=2 technologies had size predictions within 20%. (2) We used SURVIVOR to compare predicted breakpoints and found 27,304 deletions, 14,045 insertions, and 117 inversions where >=2 technologies had variants of the same type and breakpoints within 1000bp. (3) We used SVcomp to compare sequence-resolved calls, accounting for differing representations in repetitive regions. SVcomp found 30,493 (matched exactly), 33,533 (<=2% different using three distance metrics), and 45,667 (<=10% different) variants of all SV types supported by >=2 technologies. Current work includes evaluating these potential benchmark calls using genotyping/evaluation methods, machine learning, experimental confirmation, and crowd-sourced manual curation of visualizations. These results represent a significant step in GIAB work towards improved benchmarking of large variants in research and clinical settings.

1414W

A simple, flexible and EHR-agnostic platform for rapid cohort identification and characterization in a large and diverse biobank environment. *F. Mentch[1], H. Qiu[1], M. March[1], J.J. Connolly[1], L. Hermannsson[1], P.M.A. Sleiman[1,2], H. Hakonarson[1,2].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

Data to support many as yet undiscovered gene-disease and gene-disease-environment associations is very likely already captured in large biobanks set up for that purpose. One limiting factor in realizing these discoveries is the effort required to identify hundreds of clinically consistent patient cohorts using complex EHR records. Within our biobank environment with tens of thousands of genotyped subjects with full patient EHR records, we have developed a framework that standardizes and automates a significant percentage of the cohort selection effort while not requiring any particular EHR data schema. Schema independence facilitates augmenting our data sources with deep phenotype data not captured in the institution-wide EHR when it is available directly from clinicians. We have currently used this platform for over 300 distinct cohort searches from 160 data sources encompassing more than 7,000 unique diagnosis, medication, procedure and other search terms. Search metadata is permanently stored and full search results are stored in an associated data warehouse for as long as needed. As a benchmark, we find that this platform automates from 60-90% of the effort required to implement phenotype algorithms we have run that were developed by the eMERGE consortium. Other benefits of this platform include ready reuse or refinement of all or selected components of an existing search, the ability to apply extensions developed for one search to all past and future searches, and the ability to query across search metadata and results.

1415T

ClinGen Allele Registry: Linking information about human genetic variation across the web. *P. Pawliczek[1], R. Patel[1], A. Jackson[1], N. Shah[1], C. Bizon[4], T. Nelson[5], S. Dwight[7], M. Wright[7], S. Harrison[2], T. Strande[6], B. Powell[6], R. Freimuth[4], L. Babb[3], P. McGarvey[8], H. Rehm[2], S. Plon[1], A. Milosavljevic[1] on behalf of the ClinGen Resource.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Laboratory of Molecular Medicine, Partners HealthCare Personalized Medicine; 3) GeneInsight a Sunquest Company, Boston, MA; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota; 5) Geisinger autism and developmental medicine, Lewisburg, PA; 6) University of North Carolina, Department of Genetics, Chapel Hill, North Carolina; 7) Stanford University, School of Medicine, Stanford, CA; 8) Georgetown University Medical Center, Washington, DC.

The Clinical Genome Resource (ClinGen) project developed the ClinGen Allele Registry (CAR) to enable layering of information about nucleotide and amino acid variants across the Web. More than 650 million currently registered variants are linked to >1.2 billion variant records in public databases including ClinVar, ExAC, gnomAD, dbSNP, MyVariant.info, and UniProtKB. A relatively small (~1,500) but growing number of variants is registered individually by users via the web interface (reg.clinicalgenome.org). The interface includes a number of web apps including both query tools and tools supporting contribution of new layers of variant information. Variants may be registered by batch upload (up to 50M variants per request) using HGVS identifiers or VCF files, registration of ~4M variants from a typical single whole-genome sequencing run taking a few minutes. In contrast to the Data Warehousing approach where variants are de-duplicated and assigned locally unique identifiers as part of the Extract-Transform-Load process, CAR provides canonical identifiers ("CA IDs") and globally unique dereferencable URIs serving variant information via Linked Data standards (including RDF). Robust, reliable and well-documented JSON-LD APIs (reg.clinicalgenome.org under button "API specification") enable linking of variant information across the Web and use of CAR services by applications such as the ClinGen Pathogenicity Calculator (Genome Med. 2017;9(1):3) and ClinGen variant curation interface. Using the Allele Data Model defined by the ClinGen Data Model Working Group, CAR tracks variants across >500K reference sequences (genomic, transcripts, and amino-acid), including NCBI36, GRCh37, and GRCh38 genome builds as well as most RefSeq, ENSEMBL, and LRG transcript reference sequences. A highly optimized alignment-based index maps variants of size <10Kb across genomic, transcript, or amino acid reference sequences, producing corresponding HGVS identifiers and variant URIs. The CAR resolves the identity of a wide range of variants such as indels, retroposon insertions, duplications, and haplotypes. The CAR tracks relations between nucleotide and amino acid variants, thus helping pool evidence for or against pathogenicity across nucleotide variants with similar effects on protein function that would otherwise have been fragmented and poorly accessible. This work was funded by the NIH National Human Genome Research Institute (NHGRI) grants U01 HG007436-01 and U41HG006834.

1416F

Finding possible missed variant candidates through coverage comparison among varying sequencing technologies. *H. Fauni, C. Adams, K. Chao, T. Ten, C. Lau, B. Pusey, D. Adams, W. Gahl.* NIH Undiagnosed Diseases Program, NHGRI, National Institutes of Health, Bethesda, MD.

The NIH Undiagnosed Diseases Program (UDP) was started in 2008 with the goals of making diagnoses and conducting research for individuals with serious medical illnesses remaining undiagnosed despite extensive evaluation. Next-generation sequencing, both exome and genome, have been important tools for a subset of these families. Since the start of the UDP, more than 700 families have been studied with these techniques. Less than 10% of cases yield a clear diagnosis, a success rate that is expected given the extensive prior evaluations (often including clinical exome sequencing) of study participants. Exome sequencing has evolved substantially from the time it was first employed for clinical diagnostics. Adequate coverage rates for genes known to be associated with human disease have improved from 80-85% to over 95%. Some improvements have been due to techniques that specifically target genes associated with known diseases. Other techniques have the potential to improve coverage of the remaining genes in the exome. We posited that exome coverage has improved during the last 5 years in the UDP. We sought to quantify the number of exonic base positions that remained unevaluated due to poor coverage in older UDP cases compared with the current cases. Such positions represent a potential pool of sources for new diagnoses in older UDP cases. We calculated the exome coverage of individuals that were sequenced using different exome capture kits. We evaluated coverage in 30,522 hg19 UCSC transcripts and stratified the number of transcripts into three categories: no coverage, 0<x<20 mean coverage, and ≥20 mean coverage. Here we will provide summary data that describe and quantify the missed bases in our older cases and discuss the relative merits of rerunning those older cases to obtain improved coverage and aid new diagnoses.

1417W

Diagnostic variant prioritization using a statistical framework for patient genome interpretation. *N. Stong[1], T.J. Hayeck[1], B. Copeland[1], V. Aggarwal[1], K. Wang[1], D. Goldstein[1], A. Allen[1,2].* 1) Columbia University Medical Center, New York, NY; 2) Duke University, Durham, NC.

A central challenge of diagnostic sequencing is judging the "exceptionalism" of individual variants, and therefore, their potential to severely affect a patient's phenotype. The current variant prioritization paradigm relies on the application of a series of filters designed to enrich for pathogenic variants. Such filters include variant frequency, impact of the variant on the ultimate gene product (e.g., polyphen, sift, etc.), intolerance and/or conservation of the region in which the variant is found, etc. Each of these filtering steps require a threshold, which, though informed by experience and other considerations, is still, fundamentally ad hoc. Here we present a statistically principled framework for quantifying the exceptionalism of individual variants. Our scoring system (the **surprise index**) characterizes the probability of seeing a variant as extreme or more extreme than the test variant in terms of a series of annotations (frequency, deleteriousness, intolerance, conservation, etc.) if the variant were randomly sampled from a large population control dataset, e.g. gnomAD. The population controls constitute a null distribution, and, as such, the probability we compute is analogous to a p-value. To illustrate the framework's utility, the surprise index was calculated for SNVs in individuals with previously identified pathogenic *de novo* mutations. In a ranking of functional SNVs (7939-9237 qualifying variants per trio) the surprise index ranks the pathogenic variant as the top variant in 1 trio, in the top 10 in 4 trios, and in the top 100 in all 7. These results are especially notable, given: (1) the framework does not explicitly use trio information, i.e. it is naïve to which variants are actually *de novo*; and (2) phenotype has not been incorporated in the prioritization. A subsequent refinement, in which phenolyzer scores are used to identify genes of interest, further improves performance: the surprise index ranks the pathogenic variant as the top variant in 2 trios, in the top 10 in 5 trios, and in the top 20 in all 7 trios. These results suggest that the proposed framework effectively prioritizes pathogenic variants in the context of individual patient genomes.

1418T

Insights into the performance of whole-exome sequencing technologies. *Y. Yu, H. Hu, J. Fowler, Y. Ye, M. Hildebrandt, H. Zhao, P. Scheet, X. Wu, C. Huff.* Epidemiology, UT MD Anderson Cancer Center, Houston, TX.1Department of Epidemiology, The University of Texas MD Anderson Cancer Center.

A number of new technologies have been developed to support whole-exome capture and sequencing, presenting researchers with many configuration options when designing whole-exome sequencing (WES) experiments. To evaluate the relative performance of available technologies, we conducted a series of WES experiments varying the following design options: genome fragmentation (sonication vs. enzymatic), average DNA fragment length, library preparation vendor (Kapa vs. Agilent), exome capture product (Agilent Clinical Research Exome, or CRE, vs. Nimblegen MedExome, or ME), and whether to PCR amplify libraries prior to target capture. All sequencing was performed on an Illumina HiSeq 4000 using 2x150bp paired-end (PE) reads with eight samples pooled per lane, which on average generated ~45 million read pairs per sample. Our results demonstrate that under optimal conditions, the CRE and ME products provided comparable exome coverage, on average covering of 94.5% and 92.6% of coding DNA sequence (CDS) based on RefSeq annotation at ≥20X for CRE and ME, respectively. Further, we found that sonication fragmentation using Covaris-sheared DNA generated ~20% more unique fragments compared to enzymatic fragmentation (KAPA HyperPlus), and smaller DNA fragment sizes of 200bp average length produced ~10% greater coverage than larger fragment lengths due to increased on-target coverage. To investigate the relationship between coverage depth and variant identification accuracy, we evaluated the proportion of detectable variants that could be accurately called as a function of sequencing coverage. Here, we define detectable variants as those can be identified when we pool all data from different libraries together (~50 million read pairs per sample). Pooling eight samples onto one HiSeq 4000 lane, which generated a depth of ~250X per sample, we were able to identify 99.97% and 99.85% of all detectable variants for CRE and ME, respectively. We also found only 0.62% of all detectable variants were discordant between CRE and ME, nearly all of which could be filtered by applying additional variant QC. Through permutation analysis, we found 99.7% and 99.01% of all detectable variants could be identified for CRE and ME, respectively, if 16 samples were pooled into one lane. In our study, we evaluated most recent WES technologies and proposed an optional design for a WES protocol, which balances the sensitivity and specificity of genotyping and experiment cost.

1419F

Hepatic eQTL meta-analysis leads to identification of novel *cis*-eQTL in pharmacogenes. *A.S. Etheridge[1], P.J. Gallins[2], D. Jima[2], C. Molony[3], F. Wright[2], Y.H. Zhou[2], F. Innocenti[1,4].* 1) Eshelman School of Pharmacy, Univ of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Bioinformatics Research Center, North Carolina State University, Raleigh, NC; 3) Computational Biomedicine, Pfizer, Inc., Cambridge, MA; 4) Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Given the importance of the liver in drug absorption, disposition, metabolism, and excretion (ADME), as well as in drug response and toxicity, hepatic expression quantitative trait loci (eQTL) studies are crucial for elucidating how genetic variation influences inter-individual variability in therapeutic outcomes and may help guide strategies to obtain maximal efficacy and safety for patients. We performed a meta-analysis, combining microarray data and imputed genotype data from four human liver eQTL studies (n=1180), in order to increase statistical power to identify novel hepatic *cis*-eQTL. Within each study we tested for associations between genotype and gene expression, with sex, age, genetic ancestry, and experimental factors as covariates. A meta-t-statistic was calculated from the SNP/gene pair t-statistics for each liver study and p-values were generated using the normal distribution. After adjusting for multiple testing, the 1405 genes with significant *cis*-eQTL (adjusted p-value < 7.7 E-9) were filtered for relevance to ADME processes and drug response using a gene set compiled from the PharmGKB knowledge base. At a Bonferroni level of 7.7 E-9, 130,055 significant *cis*-eQTL were identified. We have identified genetic variants, not previously reported in the literature, that associate with the expression of genes important in the disposition of drugs that demonstrate a narrow therapeutic index and/or wide variability in pharmacokinetics or pharmacodynamics. Among these novel associations is rs4239798 in carbonyl reductase 1 (*CBR1*; adjusted p-value 7.8 E-25), an enzyme involved in the metabolism of the widely used chemotherapeutic doxorubicin. *CBR1* is the predominant hepatic enzyme catalyzing the reduction of doxorubicin to the less antineoplastic, but more cardiotoxic, metabolite doxorubicinol. rs4239798 was associated with increased *CBR1* expression (magnitude of effect = 11.3), indicating the potential for increased metabolism of doxorubicin and offering a possible mechanism for the up to 10-fold interpatient variability in doxorubicin plasma concentrations. Additionally, increased production of doxorubicinol may help explain acute cardiotoxicity that occurs in approximately 11% of doxorubicin treated cancer patients. Additional studies are needed to verify the role of rs4239798 in patient response to doxorubicin.

1420W

Detecting sex specific mRNA and miRNA – eQTLs: Insight into sex biased gene regulation. *J.J. Shen, W.L. Yang.* University of Hong Kong, Hong Kong, Hong Kong.

Despite sex being an important epidemiological and physiological factor, not much is known about how sex works to interact with genotypes to result in different phenotypes. Both messenger RNA (mRNA) and microRNA (miRNA) may be differentially expressed between the sexes in different physiological conditions. Using data from the geuvadis consortium, we use whole transcriptome lymphoblastoid cell line (LCL) data from 338 European samples to look for sex-specific expression quantitative trait loci (eQTLs). We use both mRNA and miRNA expression data to search for genotypes that are differentially associated with mRNA and miRNA between sexes, which we call sex specific eQTLs (ss-eQTLs). Using linear regression, we uncover 72 mRNA ss-eQTL and 4 miRNA ss-eQTLs at genotype minor allele frequency (MAF) cutoff of 0.05 and a further set of 12 mRNA ss-eQTLs of higher confidence at genotype MAF cutoff of 0.2. We present the data in an easy to search eQTL browser at http://147.8.193.83/ss_eQTL. We believe the ss-eQTLs presented will assist researchers in uncovering the basis of sex-biased gene expression regulation and ultimately help us understand the genetic basis of differences in phenotypes between sexes.

1421T

Factorbook: ENCODE ChIP-seq and DNase-seq data visualizer. *M. Purcaro, H. Pratt, J. Moore, Z. Weng.* Program in Bioinformatics and Integrative Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

The goal of the Encyclopedia of DNA Elements (ENCODE) project has been to characterize all the functional elements of the human genome. These elements include expressed transcripts and genomic regions bound by transcription factors (TFs), occupied by nucleosomes, occupied by nucleosomes with modified histones, or hypersensitive to DNase I cleavage, etc. Chromatin Immunoprecipitation (ChIP-seq) is an experimental technique for detecting TF binding in living cells, and the genomic regions bound by TFs are called ChIP-seq peaks. ENCODE has performed and compiled results from tens of thousands of experiments, including ChIP-seq, DNase, RNA-seq, Hi-C, ChIA-PET, and eCLIP-seq. Factorbook is a web-based analysis tool that integrates all public ENCODE ChIP-seq TF, histone, and DNase-seq data. Factorbook is peak-centric, with peak data being summarized and displayed in several different manners. For TFs, aggregation plots display average histone mark signals in a +/-2kb regions centered on TF peak summits. Nucleosome profiles show the effect of TF binding on the location of nucleosomes using MNase-seq data, and peaks are similarly split into proximal and distal by TSS distance. For all ChIP-seq experiments, we performed de novo motif discovery using the MEME-ChIP software suite. Discovery is performed on 100-basepair regions centered on the summits of the top 500 ChIP-seq peaks as determined by signal rank; up to five motifs are discovered for each set of 500 regions. Motifs are assessed for quality using the FIMO tool from the MEME software suite; motifs not meeting both an FDR threshold and an occurrence threshold are considered to have failed quality assessment and are grayed out on final display. Using kmedoid clustering, we have also grouped motifs together to produce averaged motifs. These averaged motifs are indexed to allow a full-text search for motifs of interest, and clicking on an individual motif provides a list of the most similar motifs discovered in other experiments.

1422F

Prioritizing phenome-wide associations using epigenome fine-mapping. *M.D. Ritchie[1,2], A. Verma[1,2], Y. Zhang[3].* 1) Biomedical and Translational Informatics, Geisinger Health System, State College, PA; 2) The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA; 3) Department of Statistics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA.

Phenome-wide association studies enable us to investigate genetic associations across many complex human diseases. However, it is still challenging to prioritize associations with genetic variants in high linkage disequilibrium (LD) and ultimately identify causal variants from association studies. Genetic impact of the variants has been addressed by using the predicted effect of variants from tools like Variant Effect Predictor(VEP), SNPeff, PolyPhen or functional annotation of non-coding regions from Roadmap Epigenome. But, such approaches are limited to answer the functional implication of the variants and tissue-specific effect on gene expression. Here we propose a method to prioritize genetic associations within LD haplotype blocks by using epigenetic functional annotations and gene expression data simultaneously. We used a 20-chromatin state model generated by the IDEAS method in 127 Epigenomes within 200 base pair window of a genomic region. We overlaid the SNP coordinates with Epigenome regions and annotated SNPs with the predicted chromatin state. We also utilized RNA-seq data on 56 Epigenomes provided by Roadmap Epigenome Consortium. We calculated the correlation between the chromatin state in 200 base pair windows and RPKM measures from 56 Epigenomes for each gene. We then used the correlation metric (r2) to prioritize association results within LD haplotype blocks for each phenotype/trait. The correlation measure suggests the effect of chromatin state on gene expression and hence the impact of variants within that 200 base pair window. Using chromatin state, we can infer functional consequence of the variants mapping to those states, and the correlation metric takes us one step further to narrow down the variants potentially impacting gene regulation. Our approach is broadly applicable to many genome and phenome-wide association studies to prioritize genetic associations in a high-throughput manner.

1423W

New statistical tools to simulate, analyze, and assess the performance of CRISPR regulatory screens. *G. McVicker, P. Fiaux, H. Chen, I. Luthra, C. O'Connor.* Salk Institute for Biological Studies, La Jolla, CA.

CRISPR regulatory screens are a powerful new approach for the systematic identification of functional regulatory sequences in the human genome. In a regulatory screen, single guide RNAs (sgRNAs) are designed to target potential regulatory sequences of target genes for mutation, silencing, or activation. Cells that express sgRNAs are sorted into low- and high-expression pools, such that cells containing sgRNAs that reduce the target gene's expression are more likely to end up in the low-expression pool. After sorting, the sgRNAs in each pool are sequenced and the differences in sgRNA frequencies between pools are analyzed to identify putative regulatory regions. Several analysis methods have been used to analyze screening data, however the performance of these methods has not been rigorously assessed, mainly because there is no test dataset with known regulatory and non-regulatory sequences. To address these limitations, we have developed a generative model that can be used to simulate data from CRISPR regulatory screens under a variety of realistic experimental conditions. We have used this simulation framework to generate screening datasets and to assess the performance of existing methods. In general, while we find that ranked statistics from existing methods perform well (the top hits are usually true positives), the p-values are poorly calibrated and it is unclear what threshold should be used to consider a putative regulatory sequence 'significant'. We are using the results from this simulation framework to guide experimental design and to develop new analysis methods with informative p-values. In addition, we have conducted our own regulatory screen for sequences that regulate the gene *GATA3* in the Jurkat T cell line using a paired-guide design that tiles deletions over a 2MB region surrounding the target gene.

1424T

CRISPinatoR: A web-based sgRNA design tool that accounts for post-transcriptional influences on protein translation. *Y. Yeu[1], R. Tuladhar[2], J.R. Clemenceau[1], J.T. Piazza[2], L. Lum[2], T. Hwang[1].* 1) Dept. of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH; 2) Dept. of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX.

The CRISPR/CAS9 genome editing tool has transformed our approach to gene interrogation and genomic medicine. Targeting of Cas9 to specific genomic sequences is achieved by a single-guide RNA (sgRNA) that pairs with a complementary DNA sequence preceding a protospacer adjacent motif. Whereas much effort has been devoted to the development of computational approaches for minimizing the off-targeting activities of sgRNAs, a systematic approach to predicting perturbations in DNA-encoded epigenetic regulatory mechanism that contribute to CRISPR/Cas9 induced changes in protein translation is lacking. Here we have developed a web-based tool called CRISPinatoR that not only provides off-target predictions for sgRNAs but also anticipates their propensity for generating foreign proteins either as a consequence of compromising mRNA splicing elements found within exonic sequences (Exon Splicing Enhancers) or altering the coding frame of pseudo mRNAs. We discuss the power and limitations of CRISPinatoR for gene deletion and rescue campaigns.

1425F

Pioneering an efficient migration of 10,000 whole genomes: Catching up with the latest human genome assembly. *S. Graf[1,2,3], M. Haimel[1,2,3], B. Tolhuis[4], T. Karten[4], O. Shamardina[1,3], C. Penkett[1,3], K. Stirrups[1,3], F.L. Raymond[1,5], W. Ouwehand[1,3], J. Karten[4], NIHR BioResource - Rare Diseases (NIHRBR-RD) consortium.* 1) NIHR BioResource - Rare Diseases, Cambridge University Hospitals, Cambridge Biomedical Campus, Cambridge, United Kingdom; 2) Department of Medicine, University of Cambridge, Cambridge, United Kingdom; 3) Department of Haematology, University of Cambridge, Cambridge, United Kingdom; 4) GENALICE B.V., Harderwijk, The Netherlands; 5) Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom.

It has been three years since the Genome Reference Consortium (GRC) released the latest version of the Human genome assembly GRCh38. Many large scale genome sequencing projects, however, still work with the even older version from 2009 as reference to base their analysis on, due to computationally intensive short read alignment and variant calling approaches. The NIHRBR-RD study has sequenced over 10,000 subjects diagnosed with rare diseases since 2013 using Illumina's 100bp, 125bp and 150bp read length PCR free whole genome sequencing (WGS) biochemistry at 30x on average. In order to evaluate potential gains by moving onto GRCh38 we selected 1036 samples for a pilot study to assess factors including read length bias, samples sequenced in duplicate or triplicate, parent-offspring trios, carriers of known disease causing variants, and unexplained cases. To assess differences between GRCh37 and GRCh38 we deployed the Genalice monolithic secondary analysis suite to align WGS data from 1036 samples and call aggregated variants for each reference. We determined the robustness of the migration process by looking at compute and storage footprints, quality of alignments and variant calls, and identification of known causative variants. The short reads from binary alignment/map (BAM) files are aligned to the respective reference genome and stored in Genalice Aligned Read (GAR) format. On average, alignment takes 17 minutes and variant calling 3 minutes per sample. The size of a GAR file is 5% of the original BAM. Compared to GRCh37 0.25% more reads are mapped. The Population Calling module aggregates variants into a random access Genalice Variant Map (GVM), from which one compressed 57 GB multi-sample VCF file is exported. The number of unique autosomal variants increases from GRCh37 to GRCh38 by 8.8% for SNPs and by 1.5% for INDELs. All 426 disease causing variants previously identified in rare disease patients were recovered with both GRCh37 and GRCh38. We assessed the quality of variants using common (Ts/Tv ratio, INDEL length distribution) and population specific (Hardy-Weinberg equilibrium) metrics, which are in keeping with other large scale sequencing projects like GnomAD for GRCh37. When extrapolating to 10,000 samples, as the process scales linearly, the realignment to GRCh38 and variant calling is estimated to take only 2 weeks while reducing the storage requirements and maintaining alignment and variant calling quality.

1426W

Finding associated variants in genome-wide associations studies on multiple traits. *L. Gai[1], E. Eskin[1,2].* 1) Computer Science, UCLA, Los Angeles, CA; 2) Human Genetics, UCLA, Los Angeles, CA.

Many variants identified by genome-wide association studies (GWAS) have been found to affect multiple traits, either directly or through shared pathways. There is currently a wealth of GWAS data collected in numerous phenotypes, and analyzing multiple traits at once can increase power to detect shared variant effects. However, the vast majority of studies consider one trait at a time. Studies that do analyze multiple traits are typically limited to sets of traits already believed to share a genetic basis. Traditional meta-analysis methods for combining studies are designed for use on studies in the same trait. When applied to dissimilar studies, meta-analysis methods can be underpowered compared to univariate analysis. This is major limitation as the degree to which a pair of traits share effects is often not known. Here we present a flexible method for finding associated variants from GWAS summary statistics for multiple traits. Our method estimates the degree of shared variant effects between traits from the data. Using simulations, we show that our method properly controls the false positive rate and increases power at varying degrees of relatedness between traits. We then apply our method to real datasets in a wide variety of disease and medically relevant traits.

1427T

A map of highly constrained coding regions in the human genome. *J. Havrilla, B. Pedersen, R. Layer, A. Quinlan.* Department of Human Genetics, University of Utah, Salt Lake City, UT.

Interspecies sequence conservation summarizes the degree of genetic constraint over vast evolutionary periods. In contrast, catalogs of genetic variation from thousands of exomes and genomes enable the inference of more recent constraint from extreme paucities of genetic variation. While existing techniques such as pLI and RVIS summarize constraint for entire genes, it is clear that single metrics do not capture the variability in constraint that exists with each protein coding gene. Therefore, to address this limitation, we have devised a linear model to identify significantly constrained regions across the entire protein coding exome by leveraging the deep resource of genetic variation observed among 60,706 exomes in the Exome Aggregation Consortium (ExAC). Constrained coding regions (CCRs) arise when the observed distance between missense variants in ExAC — a proxy for constraint — is much greater than expected by chance.   We demonstrate that our most constrained coding regions are significantly enriched for variants that are pathogenic for autosomal dominant disease, de novo mutations underlying developmental delay, and genes that are predicted to be essential in humans. As a proof of principle, we have also validated on a benign set constructed from the 123,136 exomes in ExAC v2 that are not present in ExAC v1, since ClinVar benign variants are mostly found in ExAC v1. CCRs are also found in many genes known to be associated with severe phenotypes and important cellular function. Moreover, in our studies of 15 infants afflicted with early infantile epileptic encephalopathy, the causal de novo mutation lies within CCRs ranking above the 95th percentile. Therefore, CCRs facilitate the prioritization of genetic variants underlying autosomal dominant disease and, perhaps more importantly, enable the identification of previously unknown coding regions that are likely to have fitness consequences if mutated.

1428F

VIVA: A collaboration tool for processing, storing, exploring, and sharing of next generation sequencing (NGS) data. *M. Kohram, P.J. Dexheimer, N.A. Felicelli, C.D. Van Horn, P. Komaravolu, S.D. Huitger, J.M. Hendershot, P. Maleki, G. Labilloy, J. Chen, A.M. Rupert, P.S. White.* Cincinnati Children's Hospital, Cincinnati, OH.

Whole exome and genome sequencing projects generate large quantities of data that require organized storage, versioning, filtering and annotating. Genomic research groups often design custom workflows to accommodate particular NGS data analysis needs that lack interoperability standards for data re-use and sharing. We have developed an application suite (VIVA) that has been deployed across a complex academic healthcare environment to provide an institutional variant data commons for NGS data. VIVA seamlessly integrates file storage, metadata, and a variant query tool with a web-based front-end that allows for easy information retrieval. VIVA's backend is designed using a microservice architecture with services communicating through RESTful APIs. A central API integrates all services, exposing a public RESTful API to the outside world. VIVA has been containerized using Docker and a custom deployment utility allows for fast and simple deployment of the application. The web interface of VIVA enables biologists and clinicians without any computer science background to upload FASTQ files and launch an automated workflow for uniform processing of sequence projects. Variant calling in VIVA uses the standard GATK HaplotypeCaller pipeline, and variants are annotated with the Ensembl Variant Effect Predictor tool and dbNSFP annotations. Workflows are automatically launched directly from NGS data outputs at local sequence providers. Results are managed in a variant knowledgebase that provides a facile user interface for variant search and exploration. VIVA allows grouping of samples together by project, phenotype, or any other user specified metadata. A robust permission structure supports assignment of user level permissions to make samples available to a subset of researchers or to a public share so that all users may have access. A primary goal of VIVA is to encourage collaboration among researchers who may benefit from discovering shared interest in a particular gene or variant. To accomplish this, VIVA allows researchers to form collaboration groups to share subsets of their project data or specific queries with others. Each member of the collaboration may contribute all or a portion of their data for the entire collaboration group to see. Future work will enable VIVA to recommend collaborations to users based on query results and variants of interest.

1429W

How low can you go? Recommendations for ultra-low input RNA-sequencing. *E.A. Tsai[1], P. Cullen[1], M. Mingueneau[2], H. Li[1], W. Chen[1], D. Sangurdekar[1], N. Allaire[1].* 1) Computational Biology and Genomics, Biogen, Cambridge, MA; 2) Neuroimmunology, Acute Neurology and Pain, Biogen, Cambridge, MA.

RNA-seq is a common tool for identifying differentially expressed genes (DEGs) across biological conditions. Due to quantity requirements, RNA is often extracted from whole tissue rather than sorted primary cells. Cell type heterogeneity in tissue-derived RNA could mask underlying gene expression changes, so we turn to low input protocols to enable profiling of specific cell populations. However, amplification bias from increased PCR cycles in these protocols may impair the ability to detect low or moderately expressed DEGs. We designed a standardization experiment to investigate the properties of low input RNA-seq and used the results to set guidelines for library preparation, quality control thresholds, and data analysis. The transcriptomic signature of low input RNA-seq has been reported,[1] but its effects when sequencing primary tissue and on the DEGs identified has yet to be explored. We sequenced 4 replicates of RNA stock standards, Stratagene Universal Human Reference RNA and Ambion Human Brain Reference Total RNA. As a reference, standard input RNA sequencing experiment was prepared using Illumina TruSeq v2 Kit with 500ng of RNA. The ultra-low RNA input experiments were prepared using Clonetech SMART-Seq v4 Kit with 10ng, 1ng, 100pg, and 10pg of RNA, spanning the recommended range of detection. We generated ~50 million 150bp PE reads per sample on Illumina HiSeq 2500. We used STAR, RSEM, and DESeq2 to align, quantify, and identify DEGs. We compared the calculated fold changes for each input quantity against those identified by qPCR (n=821).[2] The Spearman correlation coefficient ρ was 0.82, 0.85, 0.84, 0.79, and 0.65 for 500ng, 10ng, 1ng, 100pg, and 10pg, respectively. With the exception for 10pg, the fold changes correlated well with qPCR validated genes ($p<10^{-15}$). Applying standard thresholds for DEGs, sensitivity was similar across kits (50-60%), with 10ng having least sensitivity. False positive rates for 10ng (11%), 1ng (11%), and 100pg (14%) were lower than the rate for TruSeq (19%) and at 25% for 10pg. The number of unique splice junctions detected is significantly less ($p<0.05$) at 100pg and 10pg. While DEGs are detected at similar sensitivity between normal input and low input quantities greater than 100pg, more work is needed to understand the effects of low input sequencing on other properties of the transcriptome including isoform detection. [1]Bhargava et al., *S. Sci. Rep.* 4, 3678 (2014). [2]MAQC-III Consortium, *Nat. Biotech.* 24, 1151 (2006).

1430T

Impact of polygenic risk on changes in biomarkers over time due to lifestyle intervention and aging. *M. Conomos, J.C. Lovejoy, A.T. Magis.* Arivale Inc., Seattle, WA.

It is well known that polygenic scores created by aggregating the effects of hundreds or thousands of genetic variants across the entire genome are often correlated with measurements of quantitative traits or disease risk in a cross-sectional population. However, associations of polygenic scores with changes in phenotypes in response to lifestyle intervention or with aging have been less well studied. In order to investigate such longitudinal associations, we utilized a cohort of over 2000 individuals enrolled in a lifestyle intervention program (Arivale, Seattle, WA), for whom WGS data was collected. In addition, clinical biomarkers, including laboratory tests and protein levels, were measured from blood multiple times over the course of at least one year. We identified associations of changes in body mass index (BMI), HDL cholesterol, LDL cholesterol, and triglycerides with corresponding polygenic scores constructed using association data available from previously published meta-analyses of GWAS. The differential changes in these biomarkers implies that individuals with higher polygenic risk may not only be more likely to have unhealthy levels at a specific time point, but may also have more difficulty improving those levels following lifestyle intervention. Additionally, we identified interaction effects indicating different relationships between clinical biomarkers and age at different levels of polygenic disease risk. For example, we identified multiple proteins for which the relationship with age is perturbed among individuals at higher polygenic risk for Alzheimer's disease. These results suggest that increased polygenic disease risk may also disrupt the changes in certain phenotypes that correspond with healthy aging.

1431F

Small RNA-Seq workflow in Strand NGS: Detection of novel small RNAs. *H.K. Adil, S. Ramachandrula.* Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India.

Single cell sequencing techniques are being commonly applied for studying DNA and RNA. A research group designed Small-seq, a method for sequencing small RNAs from single cells and employed it for studying hESCs (naive and prime human embryonic stem cells) and HEK (human embryonic kidney cells). We reanalyzed this data set (GSE81287) using the Small RNA Seq analysis workflow in Strand NGS software. Raw data from GSE81287 was imported into Strand NGS and aligned to human hg19 reference with readily compiled small RNA gene annotations comprising of miRNA annotations from miRBase, tRNA from tRNAscan-SE, and snoRNA, snRNA and piRNA from Ensembl gene model. Adaptors were trimmed and unique small RNA molecules were found by collapsing UMI (Unique Molecular Identifiers) families. Upon quantification, expression profiles of known miRNAs matched with those reported by the authors. Sample correlation indicated that the cells within a cell type group showed consistent expression patterns. We could also distinguish between naive and primed cell types based on expression profiles, indicating the potential of many miRNAs and some snoRNAs as cell type biomarkers. Target genes of active miRNAs were identified. Gene Ontology and Pathway analysis of the target genes indicated a role for miRNAs of naive hESCs in maintenance of pluripotency. In addition, using the novel small RNA detection algorithm in Strand NGS, we identified a large number of novel small RNA regions in hESCs. This algorithm clusters reads and these clusters are categorized as novel and known regions using annotations. These known regions are used to build a decision tree classification model. Each novel region is ascribed a gene type using this classification model. We observed that miRNAs were the most abundant in novel category followed by snoRNAs and tRNAs. In some of the cells, novel small RNAs constituted as high as 70% of the sequenced molecules. This helped us in understanding the complexity of the small RNA transcriptome in hESCs and highlighted how most of this remains uncharacterized. We intend to showcase the bioinformatics utility of novel small RNA detection capability of Strand NGS apart from its ability to handle end-to-end analysis of large scale single cell small RNA sequencing data.

1432W

Estimating the impact of allele specific expression on detecting genetic associations. *J. Dannemiller.* Psychology Dept., Rice University, Houston, TX.

In allele specific expression (ASE), the two coding alleles of a gene are expressed at significantly different rates. ASE is thought to result from a regulatory variant in *cis* with the coding variant that alters the transcription rates of the two coding alleles resulting in a fold factor that can deviate from the expected value of 1. Valle et al. (2008) showed that the fold factors between two differentially expressed coding alleles in the gene *TGFBR1* varied substantially across a sample of 138 patients ranging from approximately 1:2 to 4:1. This variation in the transcription rates of the two *coding* alleles has the potential to alter the apparent strength of the association between the coding genotype and a downstream phenotype by adding unexplained phenotypic variance. This will happen *if the regulatory variant is unknown or is not in strong LD with the coding locus*. To estimate the impact of ASE on the percentage of downstream phenotypic variance explained (PVE) by a simple bi-allelic coding SNP (A/G), we used Gillespie's (1976) method to simulate the transcription/translation of this coding SNP combined with bootstrap sampling of fold factors from the distribution in Valle et al. (2008). A simple downstream phenotype was constructed that depended both on the relative biological activities (G:A, 1.10:1.0) and on the amounts of the two proteins at steady state produced the two alleles. Additional Gaussian *phenotypic* noise was injected to further modulate the strength of the simulated association. 1200 mono-allelelic steady-state protein levels (each averaged across 200 isogenic cells) were simulated for each of the four haplotypes: CA, CG, TA, TG. From these 4800 simulated protein levels, the required diplotypes were obtained again using bootstrap sampling, and 1200 phenotypes were computed per experiment. The major results with low LD were: 1) undetected ASE significantly reduced the PVE by the coding genotype, 2) the PVE with ASE seldom rose above approximately 3%-4% even when the coding SNP explained as much as 60% of the phenotypic variance in the absence of ASE. Paradoxically, ASE with strong LD enhanced the coding genotype signal in the phenotypic variance probably because ASE was assumed to operate by *multiplicatively* altering the transcription rates of the two alleles. These results point to the many ways in which *cis*-regulated differential expression of two alleles can either obscure or enhance the genotypic signal in the phenotypic variance.

1433T

Unsupervised pattern discovery in noncoding variants enables identify their potential functional consequences. *H. Yang[1,2], R. Chen[1,2], Q. Wang[1,2], Q. Wei[1,2], Y. Ji[1,2], X. Zhong[2,3], B. Li[1,2].* 1) Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee, United States of America, Center for Quantitative Sciences, Vanderbilt University, Nashville, Tennessee, United States of America; 3) Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America.

   Comprehensive catalogue of genetics variants which predispose disease risk is key to advancing the understanding disease pathogenesis and drug treatment. Given the complexity of genome non-coding regions, the catalogue of disease causal variants is far from complete yet. Increasing evidence shows that the regulatory function of causal variants is correlated with diverse functional annotations. In this study, we aim to leverage various biochemical and evolutionary evidence to measure the functional impact of non-coding variants through an integrative framework. Specifically, we integrated a total of 887 functional annotations including histone modifications, CTCF marks, DNase, CAGE peaks, super-enhancer regions, transcription start sites (TSS) distance, CpG island marks, chromatin interactions regions and conservation scores and developed a Bayesian framework based on graph model to identify functional variants in an unsupervised fashion. uVAR captures the inherent functional patterns of non-coding variants and constructs the joint distribution to assess the predictive power of functional variants through complex and correlated annotation data. With a purely data-driven perspective, uVAR automatically reveals five distinct patterns which are drastically different from the known functional elements. Leveraging the new insight of the non-coding region revealed in our framework, we developed a new scoring method to assess the functional impact of non-coding variants. We show that uVAR outperforms other state-of-the-art methods on classification of clinically significant SNPs, fine-mapped disease SNPs, eQTLs and experimentally verified SNPs in reporter assays. Note that the joint distribution was learned without distinguish between SNPs and indels, and we also illustrate that uVAR displayed a remarkable ability to handle non-coding indels. In addition, we further integrated other clustering information learned from known functional elements into our approach and achieved further improved performance. Finally, we applied uVAR to all variants in the 1000 Genomes Project data and made the functional prediction publicly available through our website: https://medschool.vanderbilt.edu/cgg/.

1434F

Ultra-sensitive profiling of eukaryotic and viral communities residing inside and outside the human body across 300 individuals and 1736 built environments. *N. LaPierre[1], S. Mangul[1], I. Mandric[3], N. Wu[4], D. Koslicki[5], E. Eskin[1,2].* 1) Computer Science Department, University of California, Los Angeles, Los Angeles, CA; 2) Human Genetics Department, University of California, Los Angeles, Los Angeles, CA; 3) Computer Science Department, Georgia State University, Atlanta, Georgia; 4) The Scripps Research Institute, La Jolla, California; 5) Mathematics Department, Oregon State University, Corvallis, Oregon.

   High throughput sequencing has spurred the development of metagenomics, which involves the direct analysis of microbial communities in various environments such as soil, ocean water, and the human body. However, most existing methods use marker genes or k-mers, which do not make use of the maximum possible amount of reads and thus limit sensitivity. Additionally, most work in metagenomics has focused on bacteria and archaea, neglecting to identify other key microbes such as viruses and eukaryotes. Here we present a method, MiCoP, that uses fast-mapping of reads to a comprehensive reference database of full genomes from all types of microbes to achieve maximum read usage and enable the analysis of the virome and eukaryome in each sample. We show that our method is accurate on simulated data and uses a much higher portion of the reads and identifies many more species than previously-reported results on real datasets from the Human Microbiome Project and the MetaSUB consortium.

1435W

Investigations of unmapped reads from human exome sequencing. *R. Sood[1], S. Ennis[2], J. Gibson[1].* 1) Faculty of Natural and Envornmental Sciences, University of Southampton, Southampton, Hampshire, SO17 1BJ, United Kingdom; 2) Human Genetics & Genomic Medicine, Faculty of Medicine, University of Southampton, Duthie Building (MP 808), Tremona Road, Southampton, SO16 6YD, United Kingdom.

Up to 10% of reads, per sample, in sequencing studies do not map to the intended reference genome. 'Unmapped reads' are usually discarded and not analysed. In this study we analysed unmapped paired-end reads from human exome sequencing of 245 Inflammatory Bowel Disease (IBD) patients and 113 non-IBD 'control' samples. We aimed to determine if these reads originate from regions of the genome which are under-mapped, and if so if they contain clinically important genes. Secondly we aimed to reveal if these reads represent other organisms, which show a different profile of organisms in IBD versus non-IBD samples. Exome data (Agilent Sureselect 51/60Mb V4/5/6) was mapped using BWA (v0.7.12) to the hg38 human reference. Unmapped reads were extracted and filtered by a PHRED score >20 where both reads of the pair were unmapped. Reads were then mapped with BLAST (v2.6.0) against NCBI databases for viruses, plants, fungi, bacteria, mammals and humans, taking the best match per read. Unmapped reads were mapped to the human database only, taking the best match, and annotated with GENCODE v24 genes. Matches were processed using MEGAN6. Classification of reads at kingdom and species levels do not show differences in clustering between case and control samples using PCA analysis. Few fungal or viral reads were found, though sequencing control phiX174 was found in 196 samples, comprising ≥50% unmapped reads. Bacterial reads were detected in 323 samples, top species were commensal, associated with the oral and upper respiratory tract. The greatest percentage of plant matches were in the case group, where plant species were ≥50% of unmapped reads in 42 samples. Most common matches were to *G. hirsutum* and *A. thaliana*. For 61 samples ≥ 50% of unmapped reads matched metazoan, in particular primates such as *P. troglodytes*, likely due to high sequence similarity with *H. sapiens*. Mapping only to the human database matched on average 9% of reads, reflecting 136 genes. 306 samples (total 69,448 reads) matched to the gene *CABIN1*, involved in T-cell receptor signal transduction. Two genes linked with IBD, *TJP1* and *AKT3* had 157 and 96 samples with matches respectively, but few reads per sample. No reads mapped to any gene identified by ACMG as important for clinical reporting. Results show few differences when comparing kingdoms between case and controls, though more plant matches were observed in cases. Matches to human may highlight regions of the genome which are problematic for mapping.

1436T

High-speed mosaicism detection for agnostic genome-wide detection, using the DGX-1 array processor on an NHGRI website for the genetic community. *T. Markello[1], M.G. Gordon[1], F. Gu[1], V.L. Arthur[2], M.J. Ombrello[2], J. MacDowall[1], W.A. Gahl[1], B. Pusey[1].* 1) Med Gen Branch, NIHOffice of the Clinical Director, NHGRI, and NIH Undiagnosed Diseases Network and Program, NIH, Bethesda MD /NHGRI, Bethesda, MD; 2) Translational Genetics and Genomics Unit Institution: National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Mosaicism, in DNA extracted from a sample of multiple cells taken from a single individual, is known to be associated with a variety of human diseases and is expected to contribute to undiagnosed disease complexity. Kolmogorov-Smirnov goodness of fit regression of cumulative distribution functions, and Circulatory Binary Segmentation methods, have both been published for detection and quantification of mosaicism. Here we describe how the first method was re-coded into a vector-graphic method that runs efficiently even at the limit of low quality signal/noise seen in some samples that could be used for genome wide agnostic detection of mosaicism. The code has been optimized to run on 8 high performance GPU processors (NVIDIA's Tesla V100) integrated in a single NVIDIA DGX-1 server. It is capable of analyzing SNP chip data at the rate of ≈500-900 samples per day. Four large sets of SNP data (2 public GWAS and two private) comprising over 5000 samples total, were run in preliminary testing. Those data were manually cross checked for false positive and false negative detections. For three of the data sets, the system detected mosaicism in 1/500 samples. For the UDN data set of 950 samples, it detected a rate of 1/200, suggesting an enrichment for mosaicism in this cohort. To share this resource, an outward facing website was constructed to allow investigators to submit data in the form of b-allele and log R ratio formats and receive a report for the positions, size and degree of mosaicism for their samples. We hope it will aide in the ability of smaller groups to automatically scan for mosaicism in SNP data, and we are working to extend this to sequencing data with sufficient read depths.

1437F

A survey of genetic variant frequency in 220,000 Han Chinese individuals. *Z. Huang[1], N. Rustagi[1], F. Tian[2], J. Li[1], X. Ge[1], F. Xia[1], Y. Zhang[2], K. Wang[2], H. Lin[3], Y. Gao[2], D. Zhou[2], F. Yu[1,2].* 1) Molecular and Human Genetics Department, Baylor College of Medicine, Houston, TX 77054; 2) Berry Genomics, Beijing, 100015, China; 3) Aliyun Computing, Hangzhou, 310099, China.

Large scale sequencing projects have been aggregated (such as in ExAC and gnomAD) to provide the frequency of genetic variants in multiple populations, which plays a vital role in the prediction of variant pathogenicity and selection of mutation panel for carrier screening. We explored an alternative approach to survey the frequency of genetic variants in millions of humans interrogated already using Non-Invasive Prenatal Testing (NIPT). As a proof of concept, we accrued 214,277 NIPT samples of Han Chinese population sequenced with 36 base pair single-end short reads. The average sequencing coverage is 0.05x and 97% of covered region is covered by only one read. The ultra-low sequencing coverage and short read length impose serious challenges for the estimation of frequency of high fidelity variants. We addressed the challenge of high fidelity calling with a read-level machine learning based variant caller, significantly reducing sample level false discovery rate from 74% in the raw sequencing data to 3% while maintaining the sample level sensitivity of 67%. Overall we called 25.8 million SNPs (transition-to-transversion ratio of 2.72). In coding region, 4.05% of SNPs discovered in our NIPT data set are novel to ExAC, and 3.89% are novel to gnomAD (all populations). We developed a statistical estimator of allele frequency with error estimated from sequencing coverage variation. The allele frequency estimated using our method is in concordance with the allele frequency of East Asian populations (EAS) in ExAC and gnomAD, and the allele frequency error linearly correlates with the allele frequency difference between our data set and ExAC. Our method unveiled the power of ultra-low coverage sequencing with unprecedented large sample size and high fidelity results. The unique summary data generated using our method provides an independent survey of SNP allele frequencies in Han Chinese population down to f~0.01%, offering an important reference resource for population specific carrier allele frequencies.

1438W

PheWeb: Do-it-yourself PheWAS. *P. VandeHaar[1], S.A. Gagliano[1], C. Clark[1], E. Schmidt[1], M. Flickinger[1], L.G. Fritsche[2], G.R. Abecasis[1].* 1) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway.

To allow intuitive browsing, comparison and interpretation of continuously growing publicly-available genetic association results across a wide range of complex traits we developed a free online tool, PheWeb. We aggregated GWAS summary statistics across over 100 traits into the first public release of PheWeb. We implemented a wide range of interactive visualizations, including Manhattan, QQ, PheWAS and regional LocusZoom plots. All variants tested for association with a GWAS are visualized in an interactive Manhattan plot, which is accompanied by a QQ plot displaying genomic control inflation factors stratified by allele frequency. Users can dive into particular regions of interested with LocusZoom and explore nearby genes with various genomic annotations. For each variant, we provide a PheWAS plot that displays its association summary statistics across all available traits. For each gene, we provide the traits that show significant associations with variants within the gene, and the region is visualized in a LocusZoom-like view. To confirm known associations or to discover novel signals, an analyst can search PheWeb by gene, variant identifier, or trait. PheWeb is available at pheweb.sph.umich.edu. We are continuously expanding the available collection of summary statistics to aid in the discovery of novel findings. Additionally, we provide an easy and well-documented procedure to set up a fully functional PheWeb browser for private in-house datasets.

1439T

A novel method of phenotype similarities of rare-disease patients using gene-pathway-phenotype relationships. *S. Ogishima, S. Nagaie.* Tohoku Medical Megabank Organization, Tohoku University, Sendai-shi, Miyagi, Japan.

Precision medicine is precisely stratified medicine. Precision medicine aims at provision of the best available medicine for each patient based on stratification into disease subclasses with common genetic and phenotypic factors of disease (Robinson, 2012). The precise stratification of patients into disease subclasses requires precise measures of genetic and phenotypic factors. As for genetic factors, many measurements of sequence similarities have been proposed before, which are supported by the theory of molecular evolution. On the other hand, as for phenotypic factors, Resnik, Jiang and Pesquita have proposed phenotypic similarities with considering semantic relationships in an ontology [Resnik 1995; Jiang and Conrath 1997, Pesquita et al. 2007], which can be applied for calculation of patient similarity. In this study, we proposed a novel method of phenotype similarities of rare-disease patients using gene-pathway-phenotype relationships which take into account the structure of the ontology and the specificity of each term. We then evaluated our methods on simulated phenotype data of rare-disease patients with phenotypic noise and imprecision, and compared our method with 13 measures including the PhenoDigm measure [Smedley et al. 2013] and the simGIC measure [Pesquita et al. 2007].

1440F

PQC: A phenotype checking and tracking tool. *N.W. Rayner[1,2,3], M.I. McCarthy[1,2,4].* 1) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 3) Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, United Kingdom.

Genetic studies have traditionally focussed on a single disease and collected only the directly relevant phenotypes, yielding small easily verifiable phenotype data sets. With the rapid growth in the size of genome wide association studies and the use of Electronic Health Records (EHRs) the number of phenotype data points has increased dramatically. The requirement for these to be validated still remains; however the number of data points can now run into the billions, orders of magnitude too large for an individual to check. To address this issue we have developed a Perl program to automatically summarise phenotype files of any size. The program requires only a tab delimited text file. From this it automatically attempts to determine whether the phenotype contains numeric, text, mixed or coded values and a summary is produced based on this determination. If there are <20 unique entries or <0.1% of the total (whichever is the greater) then the field is treated as a coded set of variables and the numbers of each are summarised; otherwise the total number of missing and non-missing entries is reported. For fields identified as a continuous numeric phenotype, e.g. height or weight, the program reports the mean, median, standard deviation (S.D.) and lists the samples greater than 3 S.D.s from the mean. These simple text summaries provide an overview of the data and have allowed us to detect otherwise hard to find errors, such as swapped characters, e.g. O with 0, erroneous values and multiple missing values. The program also produces histograms for all numeric phenotypes, both coded and continuous, providing an additional, visual, verification of the distribution of the data. In addition, the program can track changes between phenotype releases by generating a per-sample and per-phenotype "fingerprint", using the number of missing, numeric, text and coded entries and the total number of entries. These fingerprints are then compared to ascertain whether a sample or a phenotype entry has been changed. The program has been developed to process very large files without requiring extensive compute resources; to date we have used the program on a UK biobank data set consisting of 12,000 phenotype entries and 500,000 individuals. The program is able to summarise and plot all 6 billion entries in <22 hours using a single processor in a workstation. The program is freely available to download and will work on any Unix based system.

1441W

Recombination rate estimation in large-scale genetic data. *S. Choi, P.F. O'Reilly.* MRC Centre for SGDP, King's College London, London, United Kingdom.

Knowledge of recombination rates across the genome is critical in the study of human genetics, given the impact of recombination in local linkage disequilibrium (LD) structure, its association with genome instability and the potential role of recombination in selection and speciation. Recombination rates can be estimated through sperm typing, pedigree analysis, and can also be inferred based on patterns of LD. LD-based statistical methods, such as LDhat, have the advantage over sperm typing and pedigree analyses that a fine-scale map of recombination rates can be estimated across the entire human genome. With the recent availability of very large-scale genetic data, an improved map of recombination rates may be achieved by applying LD-based methods. However, most of the current methods, including LDhat, are computationally intensive, and thus may be infeasible for estimating recombination rates in biobank scale genetic data, eg. hundreds of thousands of samples. Thus, there is a need for the development of novel methods tailored to the estimation of recombination rates in large-scale genetic data. We investigate the performance of a new algorithm to estimate fine-scale recombination rates via simulations of genetic data with more than 100k individuals. When applied to large sample sizes, the method has sufficient power to provide an accurate fine-scale map of recombination rate of the human genome. An improved map of recombination rate may lead to a better understanding of local LD structure, more accurate inference of structural variants such as inversion polymorphisms, and enable powerful investigation of potential associations between recombination rates and disease susceptibility loci. .

1442T

Automated abstraction of phenotypes from electronic health records facilitates genetic diagnosis. *K. Wang[1], J. Son[1], C. Yuan[1], N. Stong[1], L.E. Bier[1], L. Wang[2], F. Shen[2], E.E. Groopman[1], K. Kiryluk[1], A.G. Gharavi[1], H. Liu[2], D.B. Goldstein[1], W.K. Chung[1], C. Weng[1].* 1) Columbia University, New York, NY; 2) Mayo Clinic, Rochester, MN.

Electronic health records (EHRs) capture rich fine-grained longitudinal phenotypic manifestations of various disorders. Integration of such phenotypic data and knowledge on genotype-phenotype relationships promises to enable high-throughput phenotype-driven diagnosis of genetic diseases. However, how to best extract phenotypes from noisy and heterogeneous EHR data remains a major challenge. Here we developed EHR-Phenolyzer, a novel high-throughput EHR phenotype extraction and analysis pipeline. EHR-Phenolyzer first applies a biomedical concept extraction system called MetaMap followed by heuristic concept filtering to extract relevant phenotype terms from clinical genetic notes and normalize them using Human Phenotype Ontology (HPO) terms. It then feeds these HPO terms into Phenolyzer, a published tool to prioritize candidate genes based on phenotypic manifestations. In the context of clinical exome sequencing, this list of top candidate genes can potentially increase diagnostic yield and decrease analysis time. We assessed EHR-Phenolyzer on 28 pediatric rare disease cases with confirmed genetic diagnoses. First, an experienced physician performed manual chart review on clinicians' notes (written before any genetic test) and summarized relevant phenotype terms. On average, 8 terms were used per patient. Analysis of these expert-compiled terms using Phenolyzer ranked causal genes among the top 50 genes for 12/28 cases, far more than expected by chance (p<2.2e-16), supporting the effectiveness of Phenolyzer for gene prioritization in diagnostic analysis. Next, we applied EHR-Phenolyzer to automate the above process and generated ~20 terms per patient on average. In 16/28 cases, the causal gene was ranked among top 50 genes, supporting that this automated approach is likewise effective. To assess the generalizability of our method, we replicated this finding on several independent EHR datasets from other studies or other institutions. Finally, we propose a disease prediction framework that infers patients with potential genetic syndromes directly from EHR data. This EHR-based method will help clinical geneticists in making key decisions, such as whether to order a genetic test and if so, which modality to use (such as clinical microarrays, clinical exomes, or repeat expansion testing). In summary, EHR-Phenolyzer enables comprehensive utilization of the phenotypic information within EHRs, facilitating the implementation of genomic medicine.

1443F

Dynamic aggregation of external evidence and standardized manual curation of genetic and experimental evidence in ClinGen's interfaces promote curation quality and consistency. *M. Wright[1], S. Dwight[1], M. Choi[1], K. Dalton[1], C. Thomas[1], J. Zhen[1], D. Azzariti[2], M. DiStefano[2], S. Harrison[2], R. Patel[3], P. Pawliczek[3], E. Rooney Riggs[4], B. Seifert[5], N. Strande[5], R. Ghosh[3], J. Berg[5], C. Martin[4], S. Plon[3], H. Rehm[2], C. Bustamante[1], J.M. Cherry[1], Clinical Genome Resource (ClinGen).* 1) Stanford University, School of Medicine, Stanford, CA; 2) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA; 3) Baylor College of Medicine, Houston, TX; 4) Geisinger Health System, Danville, PA; 5) University of North Carolina at Chapel Hill, NC.

At the core of ClinGen's (clinicalgenome.org) mission is the accurate, consistent curation of the clinical relevance of genes and variants to inform precision medicine and research. Additionally, ClinGen seeks to share its Clinical Validity Classification framework for gene curation and the tools built to support both this and the ACMG/AMP Guidelines for variant interpretation with the broader community, as well as to promote access to shared evidence. Guided by the above objectives, ClinGen's Gene and Variant Curation Interfaces (GCI and VCI) provide access to evidence from a wide variety of genomics resources as well as evidence curated manually from multiple sources. These sources include literature, research labs, clinics, and clinicians. All evidence is accessible to any registered user, allowing all users to view and evaluate the same set of information. In keeping with the principle of shared evidence, the interfaces promote curation by all curators in a guided, standardized manner according to the ClinGen curation frameworks. The VCI programmatically aggregates external data representing multiple evidence types from many existing genomic resources, thereby providing users with efficient, one-stop-shop access to specified, up-to-date data. It achieves this by querying multiple APIs (Application Program Interfaces) as soon as a user enters a unique variant ID and dynamically displaying all returned data for that variant. These external service APIs include those from dedicated data sources such as ClinVar, the ClinGen Allele Registry (CAR), and OMIM, and from aggregating data services such as MyVariant.info, MyGene.info, the Variant Effect Predictor, and the Ontology Lookup Service. The requirement of a unique variant identifier (ClinVar VariationID or CAR ID) representing a canonical allele insures the accurate collection of transcript, population, and *in silico* predictive data for the desired allele. This use of APIs for the aggregation of evidence in the VCI represents a modern approach that eliminates the need for storing and continually updating externally derived evidence in the curation database itself. Combined with shared manually curated evidence, it allows the ClinGen interfaces to provide a vehicle for the consistent, standardized curation of genes and variants with respect to disease for all users. Both interfaces are currently in production (curation.clinicalgenome.org) with continued development focused on new features and data types.

1444W

Effects of filtration on imputation in clusterised variants. *C.M. Charon[1], R.S. Allodji[2], JF. Deleuze[1].* 1) CNRGH/CNG, CEA Paris-Saclay Institut Francois Jacob, Evry, Essonne, France; 2) Institut Gustave Roussy, Biostastics and epidemiology, Unit 1018 INSERM, France.

The standards based on the same principles of filtration are commonly used for genome-wide association studies and imputations. To investigate the effect of imputation with and without pre-filtration of SNP followed by post-filtration on the variants imputed at a conservative and less stringent threshold, we categorised them in different classes. We used 1,031 individuals from diverse ethnicities and compared also their allele frequencies with the 1,089 NCBI recorded individuals, within a 2Mb of chr20 dbSNPB37.p13, after curation of the variants database. Only the markers removed during quality control (QC) and absent from the reference genome were not imputed, with shapeit 2 and impute 2. Hence, to maintain genotyped variants the pre-filtration could be less stringent, but no significant differences in that number was observed between the imputation prior and after pre-filtration of SNP. High correlation between frequencies minor alleles generated after imputation were found between both conditions. We didn't find any significative differences between the frequencies, except within the range of very rare and rare variants. However, the magnitude of those differences were small and became non significative under the hypothesis testing their true mean differences m = 1.85E-05. There was a slight loss of information after QC. The presence of low quality genotyped variants prior to imputation did not impair neither their imputation quality that showed maximum, nor the imputation of the remaining good quality SNP. When an indel and polymorphism were present at the same locus, there was dual imputation. Variants with no repository names, with position for sole identifier produced unreliable maf. Null allele were detected in NCBI database, by comparison with the imputed variants, and vice et versa, NCBI showed records while many null alleles would be imputed. We also considered as unreliable, SNP with null allele frequencies in both the imputed results and the NCBI dbSNP B37.p13, having very poor quality scores. An addition of 0.5 in the post-filtration imputation score stringency from 0.3 to 0.8 lead to a decrease in the number of SNV < 0.01 maf by 1.8 fold, in both conditions with and without prior QC filtration, and by a 2.5 fold in the number of SNV < 0.001 maf and lowered by half the number of very rare variants (< 5E-04). With an average maf > 0.01, a standard threshold, showed a mean score > 0.8 whether QC was performed or not.

1445T

Modeling and analysis of RNA structuromes. *Z. Ouyang.* The Jackson Laboratory for Genomic Medicine, Farmington, CT.

Determining the structures of RNAs is key to understanding their functions. RNA structure regulates almost all aspects of RNA metabolism. Moreover, structural dysregulation of RNAs contributes to the etiology of various diseases. Ultra high-throughput sequencing technologies are revolutionizing the study of RNA structures of the whole transcriptome (named RNA structurome). However, the modeling and analysis of RNA structuromes remain challenging due to the complexity of such data sets. We have developed a statistical framework for studying RNA structuromes. It models high-throughput enzymatic and chemical probing data and generates RNA structural information at single-nucleotide resolution. Genome-wide RNA structural information facilitates the understanding of the structure and function of the genome.

1446F

Development and validation of computational approaches for GWAS-informed target gene identification. *S.K. McFarland[1,2], J.C. Ulirsch[1,2], S.K. Nandakumar[1,2], L.M. Mateyka[1,2], V.G. Sankaran[1,2].* 1) Boston Children's Hospital, Boston, MA; 2) Broad Institute, Cambridge, MA.

In recent years, functional screens using both short hairpin RNAs (shRNAs) and CRISPR/Cas9 genome editing have become increasingly common for gaining insight into a variety of biological phenomena. In particular, such functional perturbation approaches hold considerable promise for uncovering the underlying biology in genome-wide association studies (GWAS), as well as other human genetic studies. Indeed, a major challenge in GWAS follow up has been in the identification and validation of target genes underlying diseases or traits of interest. However, analytic approaches for such functional screens have not been comprehensively evaluated and the results obtained from these approaches can vary considerably. Using a high-throughput shRNA screen in primary human hematopoietic cells focused on the interrogation of loci associated with red blood cell traits as a test set, we benchmark a number of established approaches, including MAGeCK, MAGeCK MLE, and PheLiM. Each of these methods has unique strengths, but also notable limitations, to account for in the analysis of GWAS-based functional screens. Particular to this study, MAGeCK MLE's employment of a design matrix robustly leverages the longitudinal time course information of the screen, and PheLiM's sequence-based approach extends our ability to dissect phenotypic contributions beyond the annotations denoted for each shRNA. However, a number of shortcomings also become apparent, including difficulty accounting for off-target effects, dealing with non-random panel design, and accommodating the integration of a priori controls. Building upon these observations, we have developed and utilized a linear mixed model approach that seeks to incorporate these advantages while minimizing the noted shortcomings, providing a robust framework to optimize the selection of candidate genes that appear enriched for orthogonal metrics of causality. We demonstrate how utilization of this linear mixed model approach allows us to confidently identify candidate gene targets underlying the majority of tested loci in our screen. Our results also demonstrate the broader value of this computational approach for analysis of a variety of large functional genomic datasets.

1447W

OASIS: Omics Analysis, Search and Information System for biological discovery in whole-genome sequence and trans-omics datasets. *J.A. Perry, K.A. Ryan, B.D. Mitchell, J.R. O'Connell.* Program for Personalized and Genomic Medicine and Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, 21201, USA.

   Background: The volume of omics data is virtually exploding. New technologies are providing thousands of whole genomes as well as transcriptomes, methylomes, proteomes and metabolomes. Computers are running nonstop to find associations with hundreds of phenotypes in our quest to understand biology and conquer disease. But how are we to sort through this overwhelming volume of data? How will we quickly translate all this information into knowledge? **Objective:** The key to enabling and accelerating discovery is to effectively deliver new information to biological researchers in a way that directly connects with their thought processes. Effective delivery requires data visualization, a broad spectrum of annotation, and real-time, supplemental analysis for fine-mapping, conditional analysis, and exploring the LD structure. An integrated analysis and search system will enable researchers to quickly follow targeted research questions, generate new hypotheses, and accelerate the discovery process. **Method:** To demonstrate this concept, an Omics Analysis, Search and Information System (OASIS) was populated with annotated GWAS results generated with the MMAP software using phenotypes from 11 studies of the Old Order Amish and 3 genotype platforms including TOPMed WGS, Illumina Exome BeadChip, and 1000 Genomes imputed dosages using the Affymetrix GeneChip 6.0. OASIS was constructed using HTML/Javascript for form-based user input, a MySQL database, and Perl/CGI scripts all running on a Linux/Apache webserver. Real-time analysis is achieved with system calls to MMAP for conditional, multi-covariate, and SKAT analysis. Visualizations include boxplots & histograms plus LocusZoom & Haploview for LD exploration. Links by variant/gene are provided to NCBI (dbSNP, Gene, Variation Viewer), Broad (HaploReg, GTEx) and WashU (Roadmap) along with annotation from Annovar and WGSA. **Result:** 1.4 billion association results with p < 0.05 from 27 million genotypes are searchable in OASIS. Phenotypes include 1200 traits from different Amish research studies (e.g. Calcification, Osteoporosis, Diabetes, Longevity, Wellness), 200 traits from metabolomics datasets and expression levels for 15,000 RNA-Seq transcripts. Researchers can easily search for associations by any combination of trait, gene(s), genomic region(s), rsid(s), p-value, effect size, allele frequency. Boxplots quickly show effect sizes. Real-time supplemental analysis options address follow up questions and new hypotheses.

1448T

Leveraging genomic data for Bayesian analysis of high-throughput regulatory functionalization assays. *A. Ghazi, C. Shaw.* Baylor College of Medicine, Houston, TX.

   The 1000 Genomes project estimates that a typical genome has around 500,000 genetic variants over-lapping known regulatory regions. Moreover, genome-wide association studies continue to identify large numbers of non-coding variants associated with a wide variety of phenotypes and disease states. Because of the prevalence of this type of human variation, functional analysis of variation in the non-coding and regulatory genome is a central challenge. The Massively Parallel Reporter Assay (MPRA) provides a high-throughput experimental method for measuring the transcriptional effects of genetic variants through the use of next generation sequencing. Despite its promise, MPRA exhibits high noise while simultaneously suffering a heavy statistical burden from multiple testing corrections, thus yielding an estimated sensitivity of roughly 20% in recent publications. In order to provide a higher sensitivity to resolve the variant-modulated transcriptional landscape, this research aims to leverage genomic data from projects like ENCODE to formulate a Bayesian statistical framework for MPRA data. Under this approach, the external data sources are used as prior knowledge that define variant similarity metrics that are used to formulate an informative empirical Bayesian prior on variant effects. Preliminary analysis of publicly available MPRA data show substantial agreement with traditional frequentist methods on variants with large effect sizes and an increased detection rate among variants with lower magnitude effects. Large numbers of new candidates are also identified by this new approach. These results are being validated by testing a subset of the putatively functional variants across a spectrum of effect sizes using gold-standard luciferase assay functionalization. Overall the method presents an important contribution to data integration and to improve the power and methodology employed for MPRA studies.

1449F

Integrating functional genomics knowledge to construct comprehensive models for complex traits and translational studies. *B. Li[1], A. Frase[2], M.D. Ritchie[2].* 1) The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA; 2) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA.

   Rapidly advancing technologies provide the opportunity of using a plethora of omics data to study complex traits, which are likely the outcome of the interplay among biological variations at different regulation levels. However, due to the lack of integrative annotation, mechanistic studies of complex traits remain limited to one single type of omics data at a time. Thus, to help construct a thorough and comprehensive genetic model behind complex traits, we propose a more comprehensive functional unit that more precisely reveals interactions between different levels of biological features. To address this issue, we have developed "omic units" and "omic modules". The omic units are the collection of identifiers to define a single gene. We use two major databases, Ensembl and Entrez Gene, as the core for gene identifiers and expand this to other types of relating identifiers, including exon, transcript, gene symbol, protein, and so on. On the other hand, omic modules describe how each omic unit is regulated by functional elements and how it interacts with other omic units. This is achieved by integrating functional regions and regulatory interactions into the LOKI database developed by our lab. Along the way of building omic units and modules, there are several obstacles to be overcome, such as resolving discrepancy and redundancy between databases of the same and different types of data. For omic units, the discrepancy was found in cross-references provided by Ensembl and Entrez where Ensembl genes link to multiple Entrez genes and vice versa. This is caused by the differences in the way each database records gene information. Currently, by improving the omic unit algorithm, we were able to successfully reduce the number of annotation discrepancies to a few hundred Entrez/Ensembl genes. As for omic modules, we compared annotation coverage on GWAS Catalog variants given by different databases or methods, including VISTA, FANTOM5, TargetFinder, and IDEAS. We found that these databases of similar types of data had little redundancy and annotation achieved the greatest coverage on GWAS Catalog variants when all databases and methods were used. Though in its early stage, we expect omic modules will be of great value to help construct thorough and comprehensive models behind complex disease.

1450W

Custom targeted design workflow for next generation sequencing. *B. Marosy, D. Mohr, A. Scott, K. Doheny.* Johns Hopkins Genomics, Center for Inherited Disease Research, IGM, Johns Hopkins University - School of Medicine, Baltimore, MD.

   Genome Wide Association (GWA) and linkage studies that identify a chromosomal region(s) may require additional follow-up sequencing in order to find a causal variant(s) that contributes to the genetic trait. Targeted sequencing is a powerful tool for querying regions of interest in large sample cohorts at high depth for relatively low cost compared to whole exome or whole genome sequencing. However inefficient design of custom probes across regions of homology and repetitive elements can cause a decrease in the selection metrics which lead to increased sequencing costs, rendering custom capture as an expensive option. By utilizing a rigorous design workflow, these challenges can be addressed prior to manufacturing of a custom product, aiding in reducing unforeseen sequencing costs. Here we have developed a workflow for designing custom targeted panels and present cases where improved sequencing conditions were met. Using the UCSC genome browser, genes and genomic elements of interest are identified for inclusion in the design. GALAXY provides tools to extract exons, add flanks, merge or subtract regions and calculate base coverage. Probes are designed across regions of interest in a tiered fashion, reducing stringency parameters for each subsequent pass of uncovered target regions not covered by a probe in a previous pass. Sequences of the probes from each pass are analyzed using BLAT to identify any potentially problematic probes that will affect selection due to homology. Threshold settings within BLAT can be varied according to the project needs, scaling the stringency of the analysis to either increase coverage or balance sequencing costs. Application of BLAT analysis can improve sequencing selection metrics by >20%. Depending on the capture size, this could prevent as much as 50% of added sequencing costs due to inefficient design.

1451T

RNASeqFPro, a full processing pipeline for RNA-Seq differential gene expression analysis. *M. Pjanic[1,2], C. Miller[1,2], T. Quertermous[1,2].* 1) Division of Cardiovascular Medicine, Stanford University, Stanford, California, United States of America; 2) Cardiovascular Institute, Stanford University, Stanford, California, United States of America.

RNA-Seq gene expression analysis comprises multiple steps, including quality control, mapping, read counting and differential expression (DE) analysis. Here we present a pipeline for automated RNA-Seq DE analysis based on top isoform selection. RNASeqFPro is a combined bash/R pipeline that automates the process of RNA-Seq DE analysis. It is designed so that fastq files placed in a working folder are the only prerequisite for running, in addition to installing dependencies and providing a meta information file. RNASeqFPro can perform analysis on paired-end or single-read RNA-Seq data for mouse mm10 and human hg19 genomes. RNASeqFPro sorts and pairs fastq files, then, performs FastQC quality control, and maps fastq files to the reference genomes using STAR second pass mapping. It uses feature-Counts to perform read summarization on the GENCODE mouse or human transcript collection. It calls fileMulti2TableMod1, an awk script we developed to generate the mastertable and writes an R script to select the top transcript and convert GENCODE to RefSeq identifiers using biomaRt. Finally, it writes an R script and performs DESeq analysis of differential gene expression, creates output tables and generates graph outputs using GO term annotations and goseq package. For the human genome it also calls RGSEPD script that will generate multiple graphical outputs using GO annotation. We also created two pipeline versions that use Kallisto for fast pseudo-alignment of the reads to the GENCODE transcripts that dramatically reduced processing time. Both STAR and Kallisto versions of RNASeqFPro are based on the transcription noise model and selection of the top expressed transcript for differential gene expression (DGE) analysis, therefore considering other transcripts rising from the same locus as transcriptional noise not contributing to the effective gene expression level. RNASeqFPro is a DGE pipeline, based on the selection of main transcript isoforms to minimize statistical influences of the transcriptional noise. RNASeqFPro performs all steps in RNA-Seq gene expression analyses in a sequential manner, and thus significantly saving time in script preparation for each individual run. RNASeqFPro should be useful for researchers with less computational experience as it requires a single executing command.

1452F

Drug side effects and adverse events are predicted by genetics of their intended targets. *P.A. Nguyen, A. Deaton, P. Nioi, L.D. Ward.* Amgen Inc, Cambridge, MA.

Human genetics is a powerful tool to mitigate drug discovery risks. To increase the likelihood of efficacy, it has been proposed that genes associated with the clinical indication should be nominated as drug targets. Historical precedence supports this concept: target-indication pairs with genetic evidence have been shown to be more likely to progress to later phases of development and to get approved (Nelson et al., Nat Genet, 2015). Target safety is another important risk driver, as clinical and non-clinical toxicity are the major reasons for drug failure (Waring et al., Nat Rev Drug Disc, 2015). We propose that genes with causal variants associated with adverse event phenotypes not related to the indication should not be pursued as drug targets, in order to avoid potential toxicities in clinical trials. To test this concept, here we performed a retrospective analysis on historical drug development data to evaluate the power of human genetics of intended targets in predicting drug toxicities. We applied a machine learning approach, using multivariate logistic regression with LASSO regularization, to predict drug side effects and adverse events from the genetics of their intended targets, while controlling for drug indication, modality, and delivery route. For marketed or withdrawn drugs, side effects were obtained from drug labels (SIDER) and physicians' reports (OFFSIDES/FAERS). For drugs that have been in clinical development, adverse events were obtained from clinical trials (Cortellis Clinical and Drug Design APIs). Drug-target and target-genetic phenotype associations were obtained from public and commercial sources. We built two sets of models: (1) predicting side effect profiles of marketed drugs (N=934) and (2) predicting the adverse events seen during clinical development of drugs (N=4639). Both sets of models showed that genetic associations (GWAS and Mendelian) with an organ system predicted side effects and adverse events within that organ system for 9/21 and 13/21 organ systems tested, respectively. Furthermore, absence of genetic phenotypes was predictive of the lack of side effect/adverse event phenotypes for 1/21 and 5/21 organ systems, respectively. The models were most predictive for endocrine system, blood and lymphatic system, and heart disorders. To our knowledge, our study provides the first unbiased and quantitative evidence suggesting that genetics can be used to predict not only efficacy but also safety of drug targets.

1453W

PALMER: A novel pre-masking method for detecting mobile element insertions using long-read sequencing technology. *W. Zhou[1], S.B. Emery[2], D.A. Flasch[2], Y. Wang[2], K.Y. Kwan[2,3], J.V. Moran[2,4], J.M. Kidd[1,2], R.E. Mills[1,2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, 100 Washtenaw Avenue, Ann Arbor, MI 48109, USA; 2) Department of Human Genetics, University of Michigan Medical School, 1241 East Catherine Street, Ann Arbor, MI 48109, USA; 3) Molecular and Behavioral Neuroscience Institute, University of Michigan Medical School, 109 Zina Pitcher Place, Ann Arbor, MI 48109, USA; 4) Department of Internal Medicine, University of Michigan, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA.

Mobile genetic elements, including Long INterspersed Element-1 (LINE-1), *Alu,* and SVA retrotransposons, occupy approximately 46% of the human genome and have been shown to play an important role in disease. Several algorithms have been developed to detect mobile element insertions (MEIs) using short-read, next-generation sequencing technologies and have been quite successful at identifying candidate MEIs with a low false discovery rate. However, due to the limitations of aligning short reads to a reference sequence, there is likely an underestimate of MEIs that lie in more complex or repetitive regions of the genome as these sequence reads are typically filtered out using existing computational pipelines. The advent of single-molecule real-time sequencing technology presents an improved long-read sequencing platform to better resolve such difficult regions of the genome and allow for a deeper interrogation for the presence of previously overlooked MEIs. Here, we present a new method naming Pre-mAsking Long reads for Mobile Element inseRtion (PALMER), which first pre-masks long-read sequences containing known reference MEIs and then searches against a library of MEI sequences to detect non-reference MEIs within the remaining unmasked sequences. Using high coverage (~70x) PacBio whole genome sequencing data of a well-established CEPH pedigree sample, NA12878, we show that PALMER identifies many LINE-1 insertion events, including a number of full-length copies, which have not been reported previously. Comparisons with fosmid paired-end sequencing libraries and other orthogonal data support that the majority of candidates represent *bona fide* MEIs. PALMER is further able to identify characteristic features of MEIs (*e.g.* short target-site duplications, 3' poly-A homopolymer runs, and 3'-transduction sequences), and can be extended to work with other types of MEIs found in non-human model organisms. This approach should enable a more holistic view of retrotransposon insertions in the human genome and provide additional utility for studies using long-read sequencing technology.

1454T

General validation framework using semi-supervised learning on complex cfDNA clinical assays. *K.R. Haas, K.A. Beauchamp, J. Tratner, K. D'Auria, C. Oyolu, H. Haverty, D. Muzzey.* Counsyl Inc., South San Francisco, CA.

Background With advancements that enable detection of a greater range and complexity of genetic variation, there are increasing challenges in effectively validating clinical genomic tests and proving their analytical performances. This difficulty is especially conspicuous in the rapidly growing fields of cell-free DNA sequencing, such as noninvasive prenatal screening (NIPS) and circulating-tumor DNA analyses, where one aims to detect miniscule signal from placental or tumor DNA, respectively. In a typical bioinformatics pipeline, a variant is called positive if the confidence score (e.g., z-score, likelihood) exceeds established thresholds, which can vary depending on the type of variant, assay chemistry, and calling algorithm. Tuning this threshold and then evaluating sensitivity and specificity can be difficult (1) when reference samples with consensus genotypes are scarce and (2) when the underlying biology of the assay necessarily causes the signal of positive samples to approach the limits of detection. **Methods** To overcome this challenge, we have developed a Bayesian graphical modeling approach capable of deconvoluting and parameterizing the negative and positive distributions from empirical data of unlabeled samples. One begins by postulating the interrelation between the latent distributions for variant incidence, allele fraction, and sample classification with the observed confidence score. Inference of the posterior predictive distribution—which assigns probabilistic outcome labels to the previously unlabeled data—is performed using Markov Chain Monte Carlo (MCMC). We then parametrically scan a range of classification thresholds and calculate sensitivity and specificity to construct a ROC curve with confidence intervals given by bootstrap sampling. **Results and Conclusions** We present an illustrative example of how this statistical approach has been leveraged to evaluate major algorithmic enhancements to Counsyl's NIPS using 34,000 random clinical samples, including >500 positives. From the ROC curve, we determined that sensitivity increased significantly, with a ~2x reduction in the false-negative rate, while maintaining 99.9% specificity. Furthermore, we discovered the relationship between sequencing depth, minor allele fraction and confidence score to decipher the limits of detection. This strategy effectively determines analytical performance from a real world collection of test samples for cutting-edge genetic assays.

1455F

Winston: Optimizing parallelized variant calling from large-scale whole genome sequencing data. *B. Jew, J.H. Sul.* University of California–Los Angeles, Los Angeles, CA.

Recent advancements in sequencing technologies allow the generation of an abundance of genetic information. This data can yield crucial insight into the genetic basis for human diseases and other complex traits through the identification of associated variation. This task is computationally intensive, requiring the use of a multitude of bioinformatics tools that align and process sequencing data before variant calling. Each step can require a significant amount of processing time and resources, introducing issues with scalability and reliability for analyzing large-scale genomic data. Winston, a parallelized variant calling pipeline, utilizes the scalable and efficient approach utilized by Churchill [Kelly et al. Genome Biology (2015) 16:6] for fast analysis of sequencing data adhering to GATK best practices. Winston improves upon this regional method by optimizing resource usage, data processing, and robustness for greater efficiency and reliability. While Churchill generates thousands of non-informative and intermediate files and on the order of gigabytes per sample, Winston generates no additional files beyond processed data. The pipeline significantly reduces disk usage and speeds up analysis by avoiding the excessive script and intermediate file generation carried out by Churchill, saving several gigabytes of storage and reducing computational time by roughly five percent per sample. This brings Churchill's analysis of the 1000 genomes dataset on 400 AWS EC2 instances in a week down by nearly half a day while minimizing storage requirements and preventing stalled analysis due to hardware failure. The responsive framework employs a checkpoint and monitor system that detects errors in analysis and availability of resources, rerunning and holding analysis as necessary. This optimized method will allow efficient and accurate discovery of genetic variation in the rapidly growing pool of sequencing data by appropriate utilization of parallel compute resources. The pipeline is currently being used for several large-scale whole genome sequencing projects at UCLA.

1456W

Pipeline for DNA-seq analysis: Streamlining of data processing, quick and accurate variant calling, and annotating variants. *A. Pal, R. Bettadapura, M.W. Saifi, P. Agarwal, S. Ramachandrula, M. Nagarajan, P. Chitipothu, H.K. Adil, S.C. Cherukuri.* Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India.

An ideal DNA-Seq pipeline must (A) identify somatic mutations with high accuracy and reproducibility, eliminating those that are clinically non-relevant and (B) annotate each mutation with its functional consequences. Strand NGS, our flagship NGS tool, contains a best-practices pipeline for somatic whole exome samples. The pipeline comprises alignment, deduplication, local realignment around indels, variant calling and annotation, and structural variant and copy number detection. Each stage in the pipeline has been specifically tuned to detect low frequency point mutations, copy number variations (CNVs), and structural variants (SVs) that occur in a somatic context. Gain of somatic mutations is often the cause for developing drug resistance. Identifying and annotating these mutations are essential in a clinical context, where biological understanding and therapy design are the two major goals. Using our somatic best-practices pipeline, we reanalyzed the whole exome sequencing samples of a drug-resistant basal cell carcinoma (BCC) dataset, GSE58374. As a part of variant calling, low frequency SNPs were detected and annotated with dbSNP. Further, we identified damaging somatic variants, annotating each of them with the COSMIC database and HGVS notation. We also analyzed a 66 sample cohort (GSE58376) of targeted re-sequencing data from untreated and drug-resistant BCC samples. As reported by the authors, we found that *SMO* and/or genes downstream of it in HedgeHog signaling pathway were mutated in all tumor samples. Deletions in *PTCH1* and *TP53* were also found in most tumors. As reported in the re-sequencing study, we found that the resistant tumors had a much higher occurrence of mutations in *PTCH1* and *SMO*. In each of the high coverage samples, the variant support view helped visualize the supporting information at and in the neighborhood of these mutations; the elastic genome browser helped inspect CNVs and SVs spanning thousands of base pairs. Taken together, our visualization tools helped validate the findings in the studies above. The pipelines in Strand NGS while primarily streamlining sequencing data analysis, also aid in easy retrieval, and compact storage. In addition, sequencing data downstream analysis can be performed from multiple omic technologies aided with statistical algorithms, dynamic visualizations, and interpretations, thereby allowing to comprehend the experimental data in a real biological context through pathway analysis.

1457T

A reference haplotype panel for genome-wide imputation of short tandem repeat variants. *S. Saini[1], M. Gymrek[1,2].* 1) Computer Science and Engineering, University of California, San Diego, San Diego, CA; 2) Department of Medicine, University of California, San Diego, San Diego, CA.

Despite conducting Genome-Wide Association Studies (GWAS) using tens of thousands of individuals, traditional GWAS based on single nucleotide polymorphisms (SNPs) still fail to explain the majority of heritability for many complex traits. One hypothesis for this is that many causal variants are not tagged well by common SNPs. Short Tandem Repeats (STRs) are one of the largest contributors of de novo mutations in humans, are implicated in dozens of Mendelian disorders and are likely to play an important role in complex traits. Next-generation sequencing technologies allow genotyping of complex variant types, including STRs. However, availability of such data is currently limited to at most several thousand samples for a given trait. On the other hand, SNP genotype array data is available for hundreds of thousands of samples. A framework to impute STR genotypes into SNP data would allow testing for STR associations with hundreds of traits across massive sample sizes. However, STR imputation is challenging for several reasons: high STR mutation rates cause lower linkage disequilibrium between STRs and nearby SNPs; and relationships between STR alleles and SNP haplotypes can be complex. Thus, constructing phased reference panels for imputation from unrelated samples has so far remained elusive. Here, we leverage a large sequencing dataset of related individuals along with existing bioinformatics methods to phase STRs onto SNP haplotypes for accurate imputation of STRs into existing SNP datasets. We apply a family based approach to extract SNP-STR haplotypes in 2,000 samples from 500 families sequenced as part of the Simons Simplex Collection. SNPs are jointly genotyped using GATK and phased using ShapeIt. STRs are genotyped using HipSTR with phased SNPs as additional input. We tested imputation using our panel at a set of known pathogenic STRs and achieved 70% concordance between true and imputed STRs. These loci are highly polymorphic and thus represent a lower performance bound. Most errors result from rare SNP haplotypes or STR alleles not well represented in the reference panel. Finally, we evaluate genome-wide imputation accuracy using thousands of samples where both SNP haplotypes and STR are available. Overall, this work provides a high quality reference panel of phased STR-SNP haplotypes that will enable imputing STRs into hundreds of thousands of existing SNP datasets.

3021W

Computational pipeline for single nucleus RNA sequencing and its application to human skeletal myotubes. *Q. Wang[1,2,3], J.L. Marshall[2,3], N. Habib[2,4,5], I.A. Davidi[2], L. Caplan[2], F. Barthelemy[6,7], Y. Li[2,4,5], I.M. Armean[2,3], N. Khanlou[8], D.A. O'Connell[2,3], E.D. Douine[7,9], B.B. Cummings[2,3,10], F. Zhang[2,4,5,11,12], M.C. Miceli[6,7], J. Wanagat[13], S.F. Nelson[7,8,9,14], A. Regev[2,15], D.G. MacArthur[2,3,16].* 1) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) The Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) McGovern Institute of Brain Research, MIT, Cambridge, MA, USA; 6) Microbiology Immunology and Molecular Genetics, University of California, Los Angeles, CA, USA; 7) Center for Duchenne Muscular Dystrophy, University of California, Los Angeles, CA, USA; 8) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA, USA; 9) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 10) Harvard Medical School, Program in Biomedical and Biological Sciences, Boston, MA, USA; 11) Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA; 12) Department of Biological Engineering, MIT, Cambridge, MA, USA; 13) Division of Geriatrics, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 14) Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, CA, USA; 15) Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, MIT, Cambridge, MA, USA; 16) Center for Mendelian Genomics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Rapid development of single cell RNA sequencing techniques have made it easier for researchers to characterize the heterogeneity of cell population and discover new biomarkers for different cell types. Even for those tissues that single cell dissection is not possible, single nucleus RNA sequencing (sNuc RNA-seq) can be applied in a similar way. On the other hand, data analysis of those single cell or nucleus data is extremely challenging, for both technical and biological noisiness, combined with often low amount of RNA expression level. Here, we have developed our computational pipeline that is tailored for single nucleus RNA-seq data. The pipeline will include - Spectral clustering based method for classification of cell population into different sub population, which takes into account of extreme sparseness of expression data, expression biases of nuclei compared to whole cell - Quantitative comparison with publicly available tissue specific gene expression data (Genotype-Tissue Expression = GTEx), to map and compare the identified cell types into known categories - Pseudo temporal path estimation based on tree construction from clusters and investigation of previously known developmental stage-specific marker genes together with standard methods such as Principle Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE) and differential gene expression analysis. The pipeline will be applied to different in vitro human myotube models, to examine the differentiation efficiency or maturity of each model by characterizing the potentially heterogeneous population of in vitro myotube cell culture. Further, we intend to perform single nucleus RNA-seq and apply this pipeline to human cells with muscle disease (genetically engineered/patients derived) to characterize the transcriptional changes in muscle disease patients in comparison with the cultured myotube models. Combining single nucleus RNA-seq with clustering, differential expression, and pseudo-time reconstruction methods, we aim to deepen our understanding of progression muscle diseases, and ultimately identify novel therapeutic targets.

1458F

Hypertriglyceridemia as a presenting sign of HMG-CoA synthase deficiency. *E. Conboy[1,2], F. Vairo[2], K. Agre[1,2], M. Schultz[3], R. Ridsdale[3], D. Deyle[1,2], D. Oglesbee[2], E. Klee[2], B. Lanpher[1,2].* 1) Clinical Genomics, Mayo Clinic, Rochester, MN, USA; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA; 3) 3)Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA.

We report an 8-month-old patient with decreased consciousness after a febrile episode and reduced oral intake. He was profoundly acidotic but his lactate was normal. Serum triglycerides were markedly elevated and HDL cholesterol was very low. The urine organic acid analysis during the acute episode revealed a complex pattern of relative hypoketotic dicarboxylic aciduria, which was suggestive of a potential fatty acid oxidation disorder, but plasma amino acids and acylcarnitine species were undiagnostic. An MRI showed extensive brain abnormalities concerning for a primary energy deficiency. Whole exome sequencing revealed heterozygotic *HMGCS2* variants, a gene encoding for mitochondrial 3-hydroxy-3-methylglutaryl-CoA (3-hydroxy-3-methylglutaryl-CoA synthase-2 deficiency 3-hydroxy-3-methylglutaryl-CoA 3-hydroxy-3-methylglutaryl-CoA 3-hydroxy-3-methylglutaryl-CoA 3-hydroxy-3-methylglutaryl-CoA HMG-CoA) synthase-2 (HMGCS2), which catalyzes the irreversible and rate-limiting reaction of ketogenesis in the mitochondrial matrix. Autosome recessive HMGCS2 deficiency is characterized by hypoketotic hypoglycemia, vomiting, lethargy and hepatomegaly after periods of prolonged fasting or illness. HMGCS2 synthesizes HMG-CoA, a precursor to ketone bodies as well as mevalonate and cholesterol. The biochemical and molecular findings were unexpected as 3-hydroxybutyrate and acetoacetate were present in urine during crisis. A retrospective analysis of the urine organic acid analysis identified putative 4-hydrox-6-methyl-2-pyrone, a recently reported biomarker of HMGCS2 deficiency. There was also a relative elevation of plasma acetylcarnitine as previously reported in one case. Our patient highlights a unique presentation of HMGCS2 deficiency caused by novel variants in *HMGCS2*. This is the first report of HMGCS2 deficiency with a significantly elevated triglyceride level and decreased HDL cholesterol level at presentation. Hence, HMGCS2 deficiency should be included in the differential diagnosis of individuals with coma induced by fasting or illness, and who present with hypertriglyceridemia, or low HDL cholesterol levels in childhood. Although HMGCS2 deficiency is a rare disorder with unspecific symptoms, with the advent of next-generation sequencing, and the recognition of novel biochemical biomarkers, the incidence of this condition may become better understood.

1459W

Characterizing NAFLD/NASH drug targets using metabolomics and genetics. *D.M. Waterworth[1], A. Sahlqvist[2], J. Freudenberg[1], D. Rajpal[1], R.A. Scott[2], L. Lotta[3], C. Langenberg[3], N. Wareham[3].* 1) GlaxoSmithKline, King of Prussia, PA; 2) GlaxoSmithKline, Stevenage, UK; 3) MRC Epidemiology Unit, University of Cambridge, UK.

Non-alcoholic fatty liver disease (NAFLD) is the liver manifestation of the metabolic syndrome and can progress to non-alcoholic steatohepatitis (NASH), a common cause of liver cirrhosis and hepatocellular carcinoma. In order to identify drug targets for NASH we reviewed the literature for genes that have been reported to be associated with NAFLD or NASH. Five genes have been identified by GWAS and an additional six by candidate gene studies where other evidence of disease involvement was available. To date these genes have been characterized mainly using clinical biomarkers (e.g. lipids, glucose, LFTs) and outcomes (diabetes and CVD), but access to metabolomic profiles allows for the opportunity to more deeply understand the role of these proteins in metabolism. We selected 16 NAFLD-associated SNPs from these 11 genes and looked for associated metabolites in up to 7500 subjects from the EPIC-Norfolk study where Metabolon profiles were available, using an FDR threshold of ≤0.1 to delineate significant associations. Six genes demonstrated significant associations (*TM6SF2*, *MBOAT7*, *PPP1R3B*, *GCKR*, *SOD2* and *TRIB1*). Hierarchical cluster analysis was applied to look for similarities and also variants from 3 genes (*PPP1R3B*, *MBOAT7* and *TM6SF2*) were mapped onto metabolic pathways using MetaCore. *GCKR* variants were highly pleiotropic, TRIB1 variants were enriched for lipid pathways and the *SOD2* variant was limited to branch chain amino acid pathways. The E167K in *TM6SF2* NAFLD risk allele was associated with higher retinol and lower lysophosphotidylcholine (lysoPC), which has been shown to differentiate between NAFLD with and without insulin resistance. The *MBOAT7* SNPs were associated with lower choline phosphate and phosphoethanolamine (PE). LysoPC and PE are key metabolites in predicting the degree of NASH in NAFLD patients. The NAFLD risk allele for rs4240624 in *PPP1R3B* was associated with higher vitamin A (retinol) and lower sarcosine and glycine. Glycine is known to be correlated with liver fat, and is a rate limiting step in glutathione metabolism, a key detoxifying agent in the liver. It is predicted that higher expression of *PPP1R3B* would result in reduced liver fat, lower circulating lipids and improvements in insulin sensitivity and therefore is a promising target for NAFLD. Thus a targeted analysis of circulating metabolites can provide a window into liver metabolism and help prioritize targets for intervention.

1460T

Elucidation of the complex metabolic profile of cerebrospinal fluid using an untargeted biochemical profiling assay. *L. Hubert[1], A.D. Kennedy[2], K.L. Pappan[2], T.R. Donti[1], A.M. Evans[2], J.E. Wulff[2], L.A.D. Miller[2], V.R. Sutton[1], Q. Sun[1], M.J. Miller[1], S.H. Elsea[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Metabolon Inc., Durham, NC, USA.

We sought to determine the molecular composition of human cerebrospinal fluid (CSF) and to identify the biochemical pathways represented in CSF to understand the potential for untargeted screening of inborn errors of metabolism (IEMs). Biochemical profiles for each sample were obtained using an integrated metabolomics workflow comprised of four chromatographic techniques followed by mass spectrometry. Secondarily, we wanted to compare the biochemical profile of CSF with those of plasma and urine within the integrated mass spectrometric-based metabolomic workflow. Three sample types, CSF (N = 30), urine (N = 40) and EDTA plasma (N = 31), were analyzed from retrospectively collected pediatric cohorts of equivalent age and gender characteristics. We identified 435 biochemicals in CSF representing numerous biological and chemical/structural families. Sixty-three percent (273 of 435) of the biochemicals detected in CSF also were detected in urine and plasma, another 32% (140 of 435) were detected in either plasma or urine, and 5% (22 of 435) were detected only in CSF. Analyses of several metabolites showed agreement between clinically useful assays and the metabolomics approach. An additional set of CSF and plasma samples collected from the same patient revealed correlation between several biochemicals detected in paired samples. Finally, analysis of CSF from several different pediatric cases including gamma-aminobutyrate transaminase, adenylosuccinate lyase (ADSL), dihydropteridine reductase (DHPR) deficiency, and pyridoxine-dependent epilepsy demonstrated the utility of untargeted global metabolic phenotyping as a broad assessment to screen samples from patients with undifferentiated phenotypes. The results indicate a single CSF sample processed with an integrated metabolomics workflow can be used to identify a large breadth of biochemicals that could be useful for identifying disrupted metabolic patterns associated with IEMs.

1461F

Infant weight gain trajectories linked to oral microbiome composition. *K. Makova[1,2], D. Blankenberg[3], A.C.L. Parodi[4], I.M. Paul[1,5], L.L. Birch[6], J.S. Savage[7,8], M.E. Marini[7], J.L. Stokes[5], A. Nekrutenko[3], M. Reimherr[9], F. Chiaromonte[1,9,10], S.J.C. Craig[1,2].* 1) Center for medical genomics, Penn State Univ, University Park, PA, 16802, USA; 2) Department of Biology, Penn State University, University Park, PA, 16802, USA; 3) Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA, 16802, USA; 4) Department of Mathematics, Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano, 20133, Italy; 5) Department of Pediatrics, Penn State College of Medicine, 500 University Drive, Hershey, PA, 17033, USA; 6) Department of Foods and Nutrition, 176 Dawson Hall, University of Georgia, Athens, GA, 30602, USA; 7) Center for Childhood Obesity Research, Penn State University, University Park, PA, 16802, USA; 8) Department of Nutritional Sciences, Penn State University, University Park, PA, 16802, USA; 9) Department of Statistics, Penn State University, University Park, PA, 16802, USA; 10) Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, 33, Pisa, 56127, Italy.

Gut microbiome composition has been linked to obesity in adults and adolescents, but less is known about its association with rapid weight gain in early childhood – a strong risk factor for obesity later in life. Moreover, connections between oral microbiome and infant weight gain have not, to our knowledge, been explored. Here we studied the relationships among weight gain in early childhood, children's gut and oral microbiomes, and mothers' oral microbiomes, in 226 mother-child dyads. As part of a large study of behavioral interventions on feeding habits, children were followed during the first two years of life -- with weight and length measured at seven time points. We used these data to identify children with rapid weight gain, and to derive growth curves with Functional Data Analysis (FDA) techniques. We also collected oral and gut microbiome samples at the end of the two-year period, and surveyed them with 16S sequencing. Information on diet and other factors was also available for our analysis. We found significant links between weight gain during the first two years of life and the oral microbiome as established by the end of this period. Significantly lower diversity and elevated Firmicutes-to-Bacteroidetes ratio were observed in the oral microbiomes of children with rapid (vs. non-rapid) weight gain, and the abundances of several bacterial taxa showed significant associations with children's growth curves. Also interestingly, the diversities of children's and mothers' oral microbiomes were significantly correlated, suggesting a familial influence. Finally, we found significant associations between diet and children's gut microbiomes but, somewhat surprisingly, not between the latter and weight gain. Our results suggest that, by the age of two years, the oral microbiome may have already established patterns that reflect an association with accelerated weight gain -- in contrast to the gut microbiome, which does not yet harbor the obesity signatures many researchers identified in later life stages.

1462W

Integrative analysis of genome, epigenome and transcriptome data from adipose tissue of obesity in Koreans. *B.-J. Kim, S. Moon, I-U. Koh, N.H. Choi, J.H. Yun, Y.J Kim.* Divisioin of Genome Research, Korea National Institute of Health, Cheongju-si, Chungcheongbuk-do, 28159, South Korea.

Obesity is defined as abnormal or excessive fat accumulation, resulting negative effect on health. Interplay of multiple factors such as environmental, behavioral and genetic factors can cause obesity. Nowadays, it is emerging as a major problem in public health. With technological advances of genome-wide association study (GWAS), many risk factors associated with obesity have been identified. However, the vast majority of discovered variants lie in non-coding regions, it is not easy to find the functional relevance between genetic loci and diseases. Recently, owing to recent progress of omics technologies, various analytical approaches to link genome data using epigenome and transcriptome has been studied. Moreover, several integrative studies suggested functional roles of genomic and epigenomic loci. Here, we aimed to find functional relevance between variation and phenotype. To achieve this, we conducted integrative analysis of genome, epigenome and transcriptome data obtained from adipose tissue. Three kinds of omics data were extracted from 65 same individuals (22 case and 43 controls by BMI cutoff point of 25 kgm^2). We used an Axiome-based customized array named Korea Biobank Array, an Illumina Infinium EPIC 850K, and an Illumina HT-12 V4 for genome, epigenome and transcriptome data, respectively. Regression analysis of case-control samples in genome, epigenome and transcriptome data and integration analysis between these data were performed. From the regression analysis between case-control samples, we confirmed known genetic loci. In addition, we identified 5 differentially methylated regions including on *IL10RA* gene reported in GWAS catalog and 33 differentially expressed gene including obesity related genes such as *LEP, AGPAT9, APOE* and *CALCRL* genes. Moreover, integrative analysis results showed that gene expressions which are highly correlated with methylated regions were mainly enriched in signal transduction, metabolism, and hemostasis. Our result provides researchers with the comprehensive information to understand genotype and phenotype interactions in obesity.

1463T

RNA sequencing analysis identifies differentially expressed genes in Lymphoblastoid Cell Lines (LCLs) generated from diabetic retinopathy patients. *Y. Cho[1], H. Choi[1], K. Lee[2,3], K. Kang[4], C. Nho[2,3].* 1) Biomedical Science, Hallym University, Chuncheon, Gangwon-do, South Korea; 2) Convergence Research Center for Smart Farm Solution, Korea Institute of Science and Technology, Gangneung, South Korea; 3) Natural Products Research Center, Korea Institute of Science and Technology, Gangneung, South Korea; 4) Department of Ophthalmology, The Catholic University of Korea Incheon St. Mary's Hospital, Incheon, South Korea.

Diabetic retinopathy (DR) is an ophthalmic complication of diabetes and has a prevalence rate of 14.2% in diabetic patients. Diabetic retinopathy can cause blurred vision and, in severe cases, blindness. The treatment of DR is mainly focused on surgical operations such as laser treatment. However, the development of DR drugs for not only therapeutic but also preventive purposes is necessary in order to reduce medical expenses. In the hope to identify drug targets for DR, we tried to detect differentially expressed genes (DEGs) in DR patients by applying RNA sequencing analysis. In this study, RNA sequencing experiments were performed using 6 samples of lymphoblastoid cell lines (LCLs) (generated from 3 DR patients and 3 normal subjects). After adaptor trimming and removing sequence reads with poor quality (Q < 30) by Trim Galore program (version 0.4.2), TopHat and Bowtie programs were employed to map RNA reads to human reference genome hg38. The number of genes expressed in the cells were counted by HTseq algorithm and DEGs were estimated by DESeq2 program. A total of 43 genes were identified as DEGs that fulfill our selection criteria (adjusted P-value < 0.01 & fold change > 4). To understand the functional role of up- or down-regulated genes in DR cells, we carried out the KEGG pathway analysis and Gene Ontology analysis through DAVID database. In addition, up-regulated or down-regulated genes in the significant KEGG pathways were adopted to generate protein-protein interaction networks using the GeneMANIA module in Cytoscape program. In this presentation, we will discuss and interpret the potential biological roles of these DEGs in the DR development based on the results from these analyses.

1464F

Optimizing simultaneous isolation of biomolecules and cells from cryopreserved adipose tissue for omics applications. *E.C. Lawrence[1], S.L. Rosenthal[1], M.A. Schusterman II[2], L.E. Kokai[2], J.P. Rubin[2], E.E. Kershaw[3], Z. Urban[1].* 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Plastic Surgery, University of Pittsburgh, Pittsburgh, PA; 3) Medicine, Division of Endocrinology and Metabolism, University of Pittsburgh. PA.

Adipose tissue is notoriously problematic for biomolecule isolation given its high lipid and low protein content. Obesity is a global epidemic and affects an array of comorbidities, thus defining its molecular basis is necessary. Parallel sets of genome, transcriptome and proteome data can provide insight into these mechanisms in an inclusive and targeted fashion. The goal of this study was to compare 11 current simultaneous extraction methods for such comprehensive downstream omics analyses of adipose tissue. We used adipose tissue collected from panniculectomy (n=6) in triplicates with 20-100 mg of input tissue according to the manufacturer's protocol (kits) or established published protocols (Trizol). DNA and RNA concentrations and quality was assessed using UV spectrophotometry. Protein concentration was measured using the Bradford assay. Following these experiments, we determined that the Trizol-based method performed the best, particularly with respect to RNA quality ($OD_{260/280}$=1.6-2.0), although all methods tended to yield variable amounts of DNA, RNA, and protein for extractions from immediate and stored samples. We then sought to identify ideal storage conditions for biomolecules extracted using the Trizol method: whole tissue stored in Qiagen's RNAlater solution or homogenized tissue stored in Trizol reagent. We also hoped to improve quality with column purification for DNA and RNA. Preliminary results (n=1) suggest on-column purification improves quality for RNA (RNALater, mean $OD_{260/280}$=2.00, Trizol, mean $OD_{260/280}$=2.22), but not for DNA (RNALater, $OD_{260/280}$=4.24, Trizol, mean $OD_{260/280}$=-14.01). Additionally, we are assessing the viability of cells isolated from crypreserved lipoaspirates. To this end, we have collected 4 g of tissue in 0.2 M trehalose with 0.5 M DMSO and stored samples at either -80°C or in liquid nitrogen. At present, we have successfully isolated and cultured cells stored in trehalose < 5 days from both storage conditions. Our results corroborate previous studies indicating variability among quality and quantity of biomolecules extracted using existing simultaneous methods. Considerable within-subject variability exists, likely due to difficulty in obtaining precise amounts of tissue and heterogeneity in the tissue composition. While this study is ongoing, we have demonstrated that existing protocols require modification to optimize the consistency, purity and yield of biomolecules and cells from adipose tissue.

1465W

Loss of function variant in *NFKB1* causes autoimmune lymphoproliferative syndrome-like disease. *F. Vairo[1,5], M.M. Patnaik[2], N.K. Shenoy[2], P.N. Pichurin[3], J.L. Kemppainen[3], M.J. Smith[4], W. Ranatunga[4], R.S. Abraham[4], E.W. Klee[1,3,5].* 1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 2) Department of Hematology, Mayo Clinic, Rochester, MN; 3) Department of Clinical Genomics, Mayo Clinic, Rochester, MN; 4) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 5) Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Autoimmune lymphoproliferative syndrome (ALPS) is a primary immunodeficiency (PID), characterized by immune dysregulation. Mutations in genes such as *FAS, FASLG, FADD, CASP8, CASP10, NRAS*, and *KRAS* are usually associated with an ALPS or ALPS-like phenotype. Approximately 30% of patients with an ALPS-like phenotype (lymphadenopathy (LAD) +/- splenomegaly and autoimmune (AI) cytopenias) do not have a confirmed genetic defect. The presence of LAD and AI cytopenias does not necessarily support an underlying diagnosis of ALPS, as other PID conditions could also be associated with this phenotype, including CTLA4 haploinsufficiency, or *STAT3* gain-of-function variants. Whole exome sequencing (WES) is an approach utilized in diagnostic odyssey cases to identify the underlying genetic cause for disease. Here, we report on a male patient who first presented at 20 years of age with idiopathic thrombocytopenic purpura (ITP). Family history was notable for several maternal relatives, including his mother, who had evidence of splenomegaly and ITP. He later developed neutropenia, axillary cellulitis, recurrent sore throat, cervical and mediastinal lymphadenopathy and Coombs positive hemolytic anemia. Flow cytometry for double-negative T cells (DNT; CD3+CD4-CD8-, TCRab+) revealed an expansion in the DNT clone, with these cells expressing B220+. This result suggested a diagnosis of ALPS, and targeted NGS was performed for *CASP8, CASP10, FADD, FAS, FASLG, ITK, KRAS, MAGT1*, and *NRAS,* which were negative. Evaluation for other relevant genes associated with AI and LAD, such as *CTLA4, PIK3R1* and *LRBA* were also negative. Given the ALPS-like clinical phenotype and negative targeted gene panels, WES was performed on the proband and his parents. An inherited heterozygous loss-of-function novel variant in *NFKB1* (p.Ser237*) was identified as one of the top candidate variants. *NFKB1* mutations have been reported with an autosomal dominant form of common variable immunodeficiency (CVID) with variable clinical phenotypes. The B cell phenotyping of this patient revealed normal isotype class-switching and memory B cells, but there was a significant increase in transitional (T1) B cells. Protein modeling indicates that heterodimer formation of p50 and p65 subunits of NFkB is affected, which suggests that there may not be normal nuclear translocation or downstream signaling. Additional functional and protein simulation studies are underway to characterize this novel *NFKB1* variant.

1466T

Novel human T cell receptor variable gene alleles revealed by long-amplicon TCRβ repertoire sequencing with Ion Torrent. *T. Looney[1], A. Glavin[1], S. Pabla[3], S. Glenn[3], L. Miller[2], D. Topacio-Hall[2], E. Linch[2], A. Pankov[1], A. Zheng[1], J. Conroy[3], C. Morrison[3], G. Lowman[2], M. Andersen[2], F. Hyland[1].* 1) Thermo Fisher Scientific, South San Francisco, CA; 2) Thermo Fisher Scientific, Carlsbad, CA; 3) OmniSeq, Inc., Buffalo, NY.

Background Human T cell antigen receptors play a critical role in protective adaptive immune responses to infectious disease and cancer but are also implicated in autoimmune disease. The antigen specificity of the T cell receptor is determined in part by the sequence of the CDR and Framework regions encoded by the TCRβ variable gene. Previous studies of population sequencing data indicate that current antigen receptor allele databases such as IMGT fail to capture a significant portion of human variation, though interpretation of this data is challenging. Here we sought to use long-amplicon multiplex sequencing of rearranged TCRβ receptors to validate putative novel human variable gene alleles previously recovered from of 1000 genomes data. **Methods** TCRβ rearrangements were amplified from cDNA template from 85 donors using AmpliSeq-based multiplex Framework 1 and Constant gene primers to produce ~330bp amplicons. Samples were sequenced in multiplex using the Ion Torrent S5 530 chip to produce ~1.5M raw reads per sample. Raw data was uploaded to Ion Reporter for clonotyping and identification of rearrangements containing variable gene sequences absent from the IMGT database. Putatively novel sequences were compared with those reported in the Lym1k database of alleles recovered from 1000 genomes sequence data. **Results** We identified 15 variable gene alleles that are absent from the IMGT database but present in the Lym1k database, the majority of which result in amino acid changes to the CDR or Framework regions of the T cell receptor. We also find evidence for novel variable gene alleles that are absent from the Lym1k database, potentially due to challenges in inferring receptor alleles from short-read population sequencing studies. **Conclusion** We find evidence for significant human diversity in TCRβ variable gene alleles beyond what is currently represented in the IMGT database. TCRβ sequencing using multiplex Framework 1 and Constant gene targeting primers is ideally suited for the study of the role of T cell antigen receptor diversity in health and disease.

1467F

Transcriptome altered by lytic human cytomegalovirus infection on human foreskin fibroblast cells (HFF) using RNA-seq. *H. Li, S. Yang, L. Feng, Y. Ran, X. Yang, T. Zhou, Z. Deng.* Dept. Biotechnology, Jinan University, Guangzhou, Guangdong, China.

Human cytomegalovirus (HCMV) is a ubiquitous, highly host-specific herpesvirus that causes severe, sometimes life-threatening disease in congenitally infected newborns as well as in immunocompromised individuals such as bone marrow allograft transplant recipients and AIDS patients . To date, no effective vaccine or antiviral drugs is available to prevent or treat HCMV infection. The most important and effective factor to prevent the development of serious complications after HCMV infection dwells in cellular immunity in humans, which may involved transcriptional alterations. Whereas a virus can potentially modulate the level of cellular mRNA by various mechanisms, identification of the cellular transcription alterations that are repressed or activated after HCMV latent infection could unveil the dynamic virus-host interaction and facilitate the development of antiviral therapy. Here, we profiled the expression of mRNAs in HFF cells using the emerging RNA-seq to investigate the transcriptional changes during HCMV lytic infection. At 18 hours post HCMV infection(MOI:1), a total of 12,124,539 sequence reads was obtained. 12,306 protein-coding genes were observed in Refseq database. Differential gene expression analysis identified 97 differentially expressed genes (DEGs) between HCMV-infected and mock-infected HFF cells, including 91 up-regulated genes and 6 down-regulated genes. These regulated genes were involved in immune response and defense response, all of which may be implicated in viral pathogenesis. These findings have provided a dynamic scenario of differentially expressed candidate genes at the virus-host interface and clearly warrant further experimental investigation associated with HCMV infection.

1468W

Simultaneous digital measurement of protein and mRNA content by massively parallel single cell sequencing to better identify T cell subsets. *C. Chang[1], E.Y. Shum[1], N. Bansal[1], J.C. Martin[2], J. Ghadiali[2], J.C. Tsai[1], D. Rosenfeld[1], H.C. Fan[1].* 1) BD Genomics, Menlo Park, CA; 2) BD Biosciences, San Diego, CA.

High throughput single cell RNA sequencing recently emerged as a powerful tool to profile complex and heterogeneous cell populations and dynamics. However, the lack of information on protein expression can make identifying cell types that have conventionally been defined by cell surface markers challenging, as mRNA and protein expression are often not tightly correlated. T cells in particular, contain relatively low abundance of transcripts, and different T cell subsets often exhibit highly similar transcriptional profiles. Here we demonstrate a novel approach of using oligo-conjugated antibodies to measure protein expression by sequencing, which enables simultaneous detection of protein and mRNA expression in a single cell. The antibody specific oligos are captured, amplified and sequenced alongside mRNA in a single workflow using BD™ Resolve, a massively parallel single cell analysis system. To demonstrate the power of the application, we created an oligo-conjugated antibody panel that includes many common T cell markers and applied the assay to human PBMCs. The detection of protein expression by oligo-conjugated antibodies was highly sensitive and specific. And the addition of protein marker measurement provided more distinct and robust clustering of single cell expression profile, especially in the T cell compartment, such as the separation of naïve CD4 vs CD8 T cells, and naïve vs memory T cells. In particular, we were able to identify a rare and recently described T cell subset, stem memory T cells, which is a long-lasting memory T cell population with stem cell-like properties. The method presented here can transform both single cell transcriptional profiling and high parameter proteomics to further efforts in elucidating complex biological systems, understanding disease states and enabling more effective biomarker discovery.

1469T

Modulation of Notch pathway in adult mouse airways by antisense oligonucleotides. *T.R. Grossman, M. Carrer, J. Crosby, B.P. Monia, C.H. Hart.* Antisense Drug Discoverey, Ionis Pharmaceuticals, Carlsbad, CA.

Multiple respiratory disorders, including asthma, cystic fibrosis and chronic obstructive pulmonary disease (COPD), are characterized by aberrant goblet cells metaplasia, which leads to excessive mucus production and insufficient mucus clearance. Previous studies have suggested that Notch signaling is critical for controlling cell fates of the lung epithelium in both healthy and diseased conditions. We designed, screened, and tested chemically modified antisense oligonucleotides (ASOs) that reduce the expression of RNAs produced from genes encoding various Notch-signaling components to further dissect this signaling system. Our goal is to better understand the therapeutic potential of promoting the trans-differentiation of mucus-secreting cells (club and goblet cells) into mucus clearing cells (ciliated cells). Ultimately, we hope to restore the normal mucociliary clearance in diseased airways through therapeutically controlled remodeling of the lung. In wild type mice and in a mouse model for severe asthma, we found that systemic and local administration of ASOs targeting Notch components can result in the significant reduction (ranging from 50 to 80%) of the target mRNAs in the murine lung. Remarkably, we were able to reduce secretory-cell prevalence and induce ciliated-cells within the airway epithelium and in the severe asthma model this robust trans-differentiation led to a significant improvement in lung function. Taken together, These observations extend a foundation for therapeutic modulation of Notch signaling to control cell fate decisions in the adult lung and suggest that ASO targeting of the different Notch components could lead to promising new therapeutic options for the treatment of respiratory disorders including as asthma, cystic fibrosis and COPD.

1470F

The translational utility of metabolomics in the integrative omics era: A case study in asthma. *J. Lasky-Su[1], R. Kelly[1], B. Chawes[1,2], H. Wolsk[2], K. Bønnelykke[2], A. Litonjua[1], H. Bisgaard[2], S. Weiss[1].* 1) Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark.

Once environmental and genetic variants for complex diseases are identified, it is often difficult to determine the causal mechanisms through which they operate to affect disease. Metabolomic profiling has the distinct advantage of being a marker that reflects the cumulative sum of past and current environmental and genetic exposures leading to the disease. We demonstrate the translational utility of metabolomics through the integration of metabolites with the variants at the well-replicated *ORMDL3* asthma locus and the protective effect of vitamin D supplementation. Using mediation analysis and cox regression in the context of 2 prebirth cohort clinical trials (VDAART & COPSAC) that randomized pregnant mothers to high and low doses of vitamin D, we identified a causal mechanism by which the *ORMDL3* variant (rs12936231) and prenatal vitamin D exposure operate on asthma risk via metabolites in the sphingolipid pathway. Specifically we used cox regression to demonstrate that the effect of vitamin D supplementation is modified by rs12936231 in both VDAART(N=806) and COPSAC(N=581). A meta-analysis of the trial results showed an increasing protective effect of vitamin D by decreasing number of *ORMDL3* C risk alleles: HRs for GG,GC,CC=1.07,0.79,0.50 (individual $p < 0.05$, interaction $p = 0.062$). In order to identify metabolites that may be mediating the observed relationship between vitamin D and asthma incidence, we performed logistic regression and mediation analysis and subsequently identified 11 metabolites associated with both vitamin D and asthma ($p < 0.05$), where sphingomyelin was identified as a significant mediator of the vitamin D induced reduced risk of asthma ($p = 0.04$). Finally, we determined that sphingomyelin levels were inversely correlated to the number of *ORMDL3* risk alleles. Mediation analysis determined that 10% of the effect of the vitamin D intervention was mediated through 4 metabolites, including sphingomyelin and glyco-syl-N-palmitoyl-sphingosine, 2 key components of the sphingolipid metabolism pathway. These results suggest a causal pathway, in which prenatal vitamin D supplementation leads to an upregulation of the sphingolipid metabolism pathway in the offspring, which in turn has a protective affect against asthma. Furthermore, we demonstrate that *ORMDL3* modifies this relationship. This approach can be generalized to identify causal mechanisms for genetic and environmental risk factors through the metabolome to enable translational understanding.

1471W

Global gene expression patterns in X-autosome balanced translocation patients. *A. Di-Battista[1], M. Moysés-Oliveira[1], A. Reymond[2], M.I. Melaragno[1].* 1) Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, São Paulo SP, Brazil; 2) University of Lausanne, Lausanne, Switzerland.

Balanced structural rearrangements may result in position effects by modifying specific *cis*- and *trans*-acting DNA regulatory element, thus changing the overall chromatin topology and DNA accessibility. These alterations could lead to differential expression patterns, altering the phenotype. Patients with balanced X-autosome translocations with breakpoint in Xq usually show premature ovarian failure, even with no gene disruption related to gonadal function at breakpoints. Expression profile studies in these patients may help in elucidating the repositioning effect of chromosome segments on gene regulation. We performed RNA-seq experiments in three patients with balanced X-autosome and premature ovarian failure and their respective controls in order to identify differential expression patterns between groups. Libraries for RNA-seq were prepared using IlluminaTruSeq RNA sample preparation kit. cDNA libraries were amplified by PCR and samples were sequenced as a pool on an Illumina HiSeq machine (read length 100 bp, single-end). For analysis we used the following pipeline: Tophat2, for reads alignment, and Cufflinks, to assemble transcripts. As reference files, we used iGenomes data. Cuffmerge and Cuffcompare steps were used in order to merge our transcripts assemblies and compare them to annotation files. For transcripts quantification and to perform differential expression analysis we used Kallisto-Sleuth package. As preliminary results, we found at least 23 differentially expressed transcripts, 16 of which were downregulated and 7 upregulated in patients. The different expression patterns observed between groups might be due to derivative chromosomes repositioning within interphasic nucleus, and further analyses may reveal a possible correlation of these transcripts with gonadal function. Financial support: FAPESP 2014/11572-8.

1472T

Tau phosphorylation is impacted by rare AD-associated AKAP9 mutations specific to African Americans. *T. Ikezu[1], C-D. Chen[1], E. Zeldich[1], A. DeLeo[1], K.L. Lunetta[3], D. Fallin[4], C. Abraham[1], M.W. Logue[1,2], L.A. Farrer[1].* 1) Boston University School of Medicine, Boston, MA; 2) National Center for PTSD at VA Boston Healthcare System, Boston, MA; 3) Boston University School of Public Health, Boston, MA; 4) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

Background: We previously identified by whole exome sequencing two rare variants of *AKAP9* (rs144662445 and rs149979685) in African American (AA) cohort. Compared to non-carriers, the odds of late-onset Alzheimer disease (AD) is 2.75-fold greater among individuals having one or both of these variants (AKAP9+). This study examined the effect of *AKAP9* mutations on AD-related processes. Methods: We examined lymphoblastoid cell lines (LCLs) from AD cases and controls with (AKAP9+, n=11) and without (AKAP9-, n=17) *AKAP9* variants. The LCLs were infected with the viral vectors expressing human amyloid precursor protein (APP) with Swedish mutation or human 2N4R Tau. Cell lysates were analyzed for total APP and total/phosphor-Tau (pTau) by ELISA, and media was analyzed for amyloid-beta peptide 1-40 (Aβ40) by ELISA. For measurements of Tau and pTau, cells were treated with or without rolipram, a phosphodiesterase type 4 inhibitor. Tau was also immunoprecipitated for proteomic analysis of its posttranslational modifications and interactomes. Aβ40/APP and pTau/Tau ratios were analyzed in a random effects model comparing AKAP9+ and AKAP9- LCLs with covariates for AD status, APOE genotype and a random effect for multiple measures from the same cell line. Results: AD was not associated with a difference in the mean Aβ40/APP ratio, the rolipram-treated pTau/Tau ratio, or the rolipram-untreated pTau/Tau ratio (all p>.3). Similarly, *APOE* genotype was not associated with a difference in mean Aβ40/APP or treated or untreated pTau/Tau raito (all p>.1). AKAP9+ was not significantly associated with Aβ40/APP (p=0.22) or with pTau/Tau in the untreated LCLs (p=0.098), but was significantly associated with pTau/Tau in LCLs treated with rolipram, with higher pTau in AKAP9+ lines (p=2.5x10-6). Tandem mass spectrometry of Tau-immunoprecipitates revealed different set of posttranslational modifications of Tau and its interactomes between AKAP9+ and AKAP9- LCLs, suggesting a biological contribution of the *AKAP9* mutation in metabolism and molecular interaction of Tau. Conclusions: The association of these *AKAP9* risk variants with tau phosphorylation and interactomes in LCLs suggests that *AKAP9*—and these AD risk variants in particular—impact the posttranslational modification of tau, a central mechanism of AD pathogenesis. However, negative results (e.g. the lack of association with AD status and *APOE*) are inconclusive until further experiments are performed on neuronal tissue.

1473F

Somatic mutations are abundant in focal cortical dysplasia. *V.S. Almeida[1], S.H. Avansini[1], M. Borges[1], F.R. Torres[1], F. Rogerio[2], B.S. Carvalho[3], F. Cendes[4], I. Lopes-Cendes[1].* 1) Department of Medical Genetics, School of Medical Sciences; 2) Department of Anatomical Pathology, School of Medical Sciences; 3) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, IMECC; 4) Department of Neurology, School of Medical Sciences; University of Campinas.

Introduction: Malformations of cortical development (MCD), including focal cortical dysplasia (FCD), can cause epilepsy and are often associated with the occurrence of refractory seizures. FCD is characterized by alterations in the cortical cytoarchitecture also observed in other MCDs, such as in Tuberous Sclerosis (TS) and Hemimegaencephaly (HME). Recently, mosaic mutations were detected in TS, HME and in rare cases of FCD; however, it is still unclear if only a single mutation event may lead to FCD or whether somatic mosaicisms is a more widespread genetic mechanism. **Materials and Methods:** Deep whole exome sequencing was performed on genomic DNA extracted from brain tissue resected by surgery (BTRS) and blood samples of 12 patients with FCD. We performed capturing and enrichment with Nextera® Expanded Kit (Illumina®). Samples were sequenced following a 200bp paired-end protocol in a Hiseq2500 (Illumina®) to achieve at least 200x of average coverage. We aligned sequences using BWA-MEM and performed realignment around SNPs and indels, quality recalibration and variant calling using the Genome Analysis Toolkit (GATK). We evaluated mosaicisms using Mutect2 and Strelka. Variants were classified as mosaic mutations when less than 10% of reads are not aligned to the human genome reference and are present only in BTRS. Variants were filtered prioritizing frameshift, missense, nonsense and splicing site mutations. In addition, we also focused in variants not described previously in the normal population (reference individuals) or variants whose minor allele frequency (MAF) is ≤ 0.01. Effect of variants was evaluated using Variant Effect Predictor (VEP). **Results:** We identified potentially deleterious mosaic mutations in BTRS in **all** 12 patients analyzed. Mosaic mutations were identified in 28 genes belonging to the mTOR pathway and in 40 genes belonging to the TAU pathway. More interestingly, **all** 12 patients have somatic mutations in BTRS in at least two genes. **Discussion/Conclusion:** Somatic mutations were identified in genes with functional roles and expression in the central nervous system. Somatic mutations were abundant in FCD tissue, affecting several genes from the mTOR and Tau pathways, and were detected in all samples examined. Therefore, somatic mosaicism seems to be a much more abundant and widespread genetic phenomenon than previously described in FCD. .

1474W

Proteomics analysis of the dentate gyrus isolated from two different animal models of mesial temporal lobe epilepsy. *A. Morato do Canto[1,2], A.H.B. Matos[1,2], B.A. Bertelli[1,2], A.S. Vieria[1,2,3], R. Glioli[4], B. Henning[2], B. Carvalho[2], I. Lopes Cendes[1,2].* 1) Medical Genetics, School of Medical Science - UNICAMP, Campinas, SP, Brazil; 2) Brazilian Institute for Neuroscience and Neurotechnology - BRAINN, Campinas, SP, Brazil; 3) Biology Institute, UNICAMP, Campinas, SP, Brazil; 4) Multidisciplinary Center for Biological Investigation of Laboratory Animals (CEMIB); University of Campinas (UNICAMP), Campinas, SP, BRAZIL.

Mesial temporal lobe epilepsy (MTLE) is the most common form of epilepsy in adults. In most patients with MTLE there is a distinct lesion in the mesial temporal structures, including the hippocampal formation. The dentate gyrus (DG) is a cortical region that is an integral portion of the larger functional structure named hippocampal formation. There are numerous features of the DG that makes it unique in a neuroanatomical and functional way. We studied the DG in its dorsal (dDG) and ventral (vDG) anatomical regions in two MTLE animal models, the perforant pathway stimulation (PPEM - n=5) and the pilocarpine model (PiM - n=5). We performed proteomics analysis using Label-free quantification (LTQ-Orbitrap) and the software MaxQuant, Perseus and R. In the dDG from the PPEM we found that the protein CYS-C was up-regulated suggesting the activation of defense mechanisms against the injury caused in the model. We also identified an extensive process of astrogliosis in both regions (dorsal and ventral) resulting in higher expression of the proteins GFAP and Vimentin. In the vDG of the PPEM we found altered proteins responsible for the structure of Gap junctions suggesting changes in the electrical synapses and in ion homeostasis. By contrast, in the dDG of the PiM we found changes in energy metabolism such as in the Glycolisis and gluconeogenis pathways as well as in Ca^{2+} signaling. Moreover, for the vDG from the PiM we identified changes in signal transduction involving mTORC2, which is related to many processes such as cellular metabolism and cytoskeleton properties. In conclusion, we show that different models of MTLE present remarkable differences in proteomics profiles, indicating that epileptogenesis was induced by different biological pathways in these two models.

1475T

RNA sequencing and proteomics approaches reveal novel multi-cellular deficits in the cortex of Rett syndrome mice. *N.L. Pacheco[1], M.R. Heaven[2], L.M. Holt[1,3], D.K. Crossman[4], K.J. Boggio[5], S.A. Shaffer[5], D.L. Flint[2], M.L. Olsen[1,3].* 1) Cellular, Developmental, and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL; 2) Vulcan Analytical, LLC, Birmingham, AL; 3) School of Neuroscience, Virginia Polytechnic and State University, Blacksburg, VA; 4) UAB Heflin Center for Genomic Science, Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; 5) Proteomics and Mass Spectrometry Facility, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Shrewsbury, MA.

Rett syndrome (RTT) is an X-linked neurodevelopmental disorder caused by mutations in the transcriptional regulator MeCP2. RTT is characterized by having apparently normal development until 6-18 months, when a progressive decline in motor and language functions begins and breathing abnormalities and seizures present. Despite intense research, the molecular targets of MeCP2 and their contribution to the disease are unknown. Here we present the first comprehensive transcriptomic and proteomic analysis in a RTT mouse model. Examining whole cortex tissue in symptomatic males (*Mecp2*-null) with wildtype littermates, we have identified 391 genes (FDR < 0.05), and 465 proteins (p < 0.1) considered to be significantly, differentially expressed. These data indicate RNA metabolism, proteostasis, monoamine metabolism, and cholesterol synthesis are disrupted in the RTT proteome. Hits common to both data sets indicate disrupted cellular metabolism, calcium signaling, protein stability, DNA binding, and cytoskeletal cell structure. Finally, in addition to confirming disrupted pathways and identifying novel hits in neuronal function, our data indicate aberrant myelination, inflammation and vascular disruption. Intriguingly, in opposition to what is typically observed in most neurological diseases, there is no evidence of reactive gliosis. Instead, gene, protein, and pathway analyses suggest astrocytic maturation and morphological deficits. To further investigate this finding, we performed transcriptomic analyses on acutely isolated cortical astrocytes throughout postnatal development of *Mecp2*-null and wildtype littermates. Preliminary analysis of transcriptomic data from symptomatic *Mecp2*-null male cortical astrocytes (postnatal day 60+) also indicates disrupted genes and pathways associated with astrocytic maturation. Transcriptomic analyses of the remaining cortical astrocytic developmental time points are ongoing. Collectively, these analyses support previous works indicating widespread CNS dysfunction and may serve as a valuable resource for those interested in cellular dysfunction in RTT.

1476F

Metabolomic studies of a brain-specific mouse model of tuberous sclerosis complex: Major changes in the methylation pathway. *M.J. Gambello[1], J.O. McKenna[1], D.J. Kapfhamer[1], J. Kinchen[2], P.L. Hall[1], B. Wasek[3], T. Bottiglieri[3].* 1) Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Metabolon, Inc Durham, NC; 3) Baylor Institute of Metabolic Disease Dallas, TX.

The tuberous sclerosis complex (TSC) is an autosomal dominant disorder causing substantial neuropathology. TSC is the quintessential mTORopathy as loss of either causative gene, *TSC1* or *TSC2*, leads to unopposed mTORC1 kinase activation and results in cell growth and division. This unrestrained anabolic state requires a major reprogramming of cellular metabolism to support macromolecular biosynthesis including increased NADPH production and de novo purine and pyrimidine generation. Many studies elucidating relevant pathways have focused on cell lines, but little is known about the metabolic reprogramming that occurs in TSC-affected organs in vivo. Using a brain-specific mouse model of TSC, we performed an unbiased metabolic screen of hippocampal lysates to identify *Tsc2*-dependent metabolic changes. We annotated a total of 434 compounds, of which 103 metabolites were upregulated and 68 downregulated in mutant hippocampus compared to control ($p < 0.05$). Rapamycin, an mTORC1 inhibitor, had an effect on 42/103 upregulated metabolites and 29/68 downregulated metabolites. We found that Tsc2 deficiency caused changes in redox homeostasis and in glucose, purine and neurotransmitter metabolism. Intriguingly, we also observed significant changes in methylation metabolism that were sensitive to rapamycin. Targeted metabolomic analysis of cortical lysates revealed the same changes: decreased methionine, cystathionine, and SAM (S-adenosylmethionine) - the main methyl donor, as well as decreased SAM/SAH ratio. There was also a marked increase in betaine, an alternative methyl donor. We hypothesize that increased methionine utilization is needed to support protein synthesis, which in turn alters the methylation potential in neurons and/or glia. In support of this hypothesis we also found alterations of SAM dependent reactants and products in the mutant brains such as guanidinoacetate (precursor of the energy metabolite creatine), 3-methoxy-tyramine and O-methyl dopa (neurotransmitter metabolites), and putrescine (polyamine). Alterations in SAM-dependent methylation reactions may be an important aspect of TSC neuropathology, and be amenable to nutritional or pharmacologic therapy.

1477W

Kinome profiling of neural stem cells (NSC): Kinome profiling of NSC derived from induced pluripotent stem cells (iPSC) of Huntington's disease patient. *A. Baharani[1,2], E. Scruten[1,2], S. Napper[1,2].* 1) Dept of Biochemistry, Univ of Saskatchewan, 107 Wiggins Road, Saskatoon, Saskatchewan SK-S7N5E5, Canada; 2) Vaccine and Infectious Disease Organization (VIDO), 120 Veterinary Road, Saskatoon, Saskatchewan SK-S7N5E3, Canada.

Authors: Akanksha Baharani, Erin Scruten, Scott Napper   Primary cell lines derived from patients with Huntington Disease (HD) were used to investigate physiologically relevant molecular aberrations in a cellular context. We used commercially available primary HD neuronal stem cells (NSCs), that were reprogrammed from induced pluripotent stem cells (iPSCs) isolated from an HD patient; as a model; to investigate the signaling aberrations associated with HD pathogenesis. The primary HD cell lines we used were sampled, karyotyped and characterized as having 45 polyglutamine (CAG) repeats. We utilized a custom-designed peptide array comprising 300 unique peptides corresponding to functionally annotated phospho-sites of known signaling intermediates. Lysates from primary HD cells and control human cells were applied on our peptide arrays and phosphorylation intensities corresponding to individual peptides quantified. The intra-array peptide replicate phospho-intensities were normalized and data presented as fold-change difference in peptide-specific phosphorylation between the HD cells and control cells. Our analyses led to the identification of various signaling intermediates that were differentially phosphorylated in HD cells compared to control cells. Using quantitative immunoblotting analyses, we validated a subset of the identified targets that included AKT1 Ser 473, LIMK1 Thr 508 and PTEN Thr 382 in HD cells. We confirmed that, in HD cells, the hypophosphorylation of AKT1 at Ser 473 led to reduced inhibition of the AKT substrate, GSK3b leading to increase phosphorylation at Ser 9. Further, we also validated the hyperphosphorylation of PTEN at Thr 382, a phospho-site associated with promoting PTEN stability. Given that PTEN is a negative-regulator of AKT, our data suggest that the AKT signaling axis is negatively regulated in HD cells, leading to reduced neuronal cell survival. Likewise, we also validated LIMK1 Thr 508 as a hypophosphorylated phospho-site in HD cells. We further found that the hypophosphorylation of LIMK1 at Thr 508 along with reduced phosphatase activity of Slingshot (SSHL1) at S987, led to increased phosphorylation of their substrate, Cofilin at Ser 3. The deregulation of LIMK1, Slingshot and Cofilin represents HD-specific pathway which is linked to disordered cytoskeletal dynamics. Taken together, our data identified key signaling proteins deregulated in HD cells which may be suggestive of a prognostic significance in HD pathogenesis.

1478T

Bridging the therapy gap for rare genetic disorders: Comprehensive high-throughput drug repurposing screening to identify potential new treatment opportunities. *C.M. Maher[1], B.C. Gay[1], J. Rader[1], C.J. Schulz[1], J.M. Andresen[1], A.C. Gerlach[2], B.M. Antonio[2], S. Santos[2], K.M. Padilla[2], O. Devinsky[3], M. Might[4], S. Petrou[5], D.B. Goldstein[6], G.R. Stewart[1].* 1) Pairnomix LLC, Maple Grove, MN; 2) Icagen Inc, Durham, NC; 3) Neurology Dept, New York University Medical Center, New York, NY; 4) Hugh Kaul Personalized Medicine Institute, University of Alabama at Birmingham, Birmingham, AL; 5) Florey Institute of Neuroscience and Mental Health, University of Melbourne, Parkville, VIC, Australia; 6) Institute for Genomic Medicine, Columbia University, New York, NY.

The power of genetic testing has led to rapid advancements in identifying a growing number of genes directly linked to rare diseases. While genetic testing has provided great insight into the pathoetiology of such disorders, diagnostic discoveries rarely translate into improved therapies. Pairnomix was created to bridge this gap between genetic diagnosis and precision medicine. Starting with a patient's genetic sequencing report, Pairnomix constructs a personalized *in vitro* cellular model that incorporates their specific mutation through site-directed mutagenesis. The functional consequences of the mutated model are characterized in relation to a wild-type model and an assay specific to those consequences is developed and optimized for use in high-throughput drug screening (HTDS). That assay is then used to screen a library of ~1,300 clinically available compounds, mostly taken from the Prestwick library. A panel of targeted inhibitors or potentiators (if available) are also routinely screened. A compound is considered a "lead" if it has activity that is greater than two standard deviations from the mean activity of all compounds. Leads from the HTDS are then re-screened using compound from a second source supplier across a 10-point concentration curve to characterize maximal activity in the assay and to identify an EC_{50}. Finally, Pairnomix relays the information from these experiments back to the patient's physician so together they can make better informed healthcare decisions. Here, we describe the use of the Pairnomix approach for a patient with uncontrolled epilepsy who carries an R1872Q mutation in the *SCN8A* gene that encodes the Nav1.6 sodium channel. In a HEK293 cellular model of the mutation, mutated Nav1.6 channels were found to display an increased sensitivity to activation and a delayed inactivation profile that contributed to a gain-of-function phenotype. In an HTDS seeking to identify inhibitors of sodium influx, 90 compounds were identified that provided significant inhibition in the assay. Of these 90 compounds, the majority have not been implicated in ion channel modulation. This approach demonstrates the utility of HTDS to identify and repurpose existing drugs for new therapeutic indications.

1479F

Reproducibility in iPSC omics: An international multicentre study. *V. Volpato[1], J. Smith[2], C. Sandor[1], J. Ried[3], A. Baud[4], A. Handel[1], S. Newey[1], F. Wessely[1], M. Attar[1], E. Whiteley[1], S. Chintawar[1], A. Verheyen[5], A. Artati[6], E. Sharma[1], W. Heywood[4], R. De Filippis[3], K. Janssen[3], P. Reinhardt[7], I. Royaux[5], P. Peeters[5], G. Terstappen[3], M. Graf[8], R. Livesey[2], C. Akerman[1], K. Mills[4], R. Bowden[1], G. Nicholson[1], Z. Cader[1], V. Lakics[3], C. Webber[1], IMI StemBANCC Consortium.* 1) University of Oxford, Oxford, United Kingdom; 2) University of Cambridge, Cambridge, United Kingdom; 3) AbbVie Deutschland GmbH & Co. KG, Ludwigshafen, Germany; 4) UCL, London, United Kingdom; 5) Janssen, Antwerp, Belgium; 6) Helmholtz Zentrum München GmbH, Neuherberg, Germany; 7) Max Planck Institute for Molecular Biomedicine, Munster, Germany; 8) Hoffmann-La Roche Ltd, Basel, Switzerland.

Reproducibility is crucial to scientific discovery. Human induced pluripotent stem cells (iPSC) are providing valuable in vitro disease models, especially for hitherto inaccessible human tissues. Using agreed protocols across five laboratories within the StemBANCC consortium (three academic and two pharma), we assessed the cross-site reproducibility of a cortical iPSC differentiation protocol in terms of identifying robust differences between the same two iPSC lines, one control and one derived from an Alzheimer's patient carrying a PSEN1 variant. These two cell lines were cultured in triplicate across the five different laboratories with transcriptomic and proteomic data generated at two time points during differentiation, including measurement of the β-amyloid 40/42 ratio. We found that a decreased β-amyloid 40/42 ratio in the PSEN1 line was reliably detected across all five laboratories. However, while the gene expression profiles clustered by cell line within each laboratory's experiment, between laboratories high variation in both omics readouts between the two lines significantly confounded replicating their comparison. Factor analyses (RUVs) demonstrated that the biological variation from the two covariates of interest (cell line and time point) could be revealed after estimating and removing factors of unwanted variation, with "Laboratory" identified as the single largest source of variation. Examining the technical biases, we identified significant confounders in the iPSC-based cell line comparisons such as passage number, feeding patterns, use of frozen neural progenitors among others, that we caution should be explicitly controlled in iPSC study designs. Single-cell analyses further revealed that cell type compositional heterogeneity varied by laboratory as well as cell line, and contributed significantly to the masking of genotypic effects in multi-site analysis, most notably with pathways altered in different directions in different cell populations. Despite significant efforts to replicate the same iPSC experiment across multiple laboratories, omics comparisons were confounded by significant variation. Our study advises caution when comparing different experiments, and identifies many experimental variables that need to be controlled and reported.

1480W

Targeted enrichment without amplification and SMRT Sequencing of repeat-expansion disease causative genomic regions. *T. Clark, Y. Tsai, D. Greenberg, J. Ekholm.* PacBio, Menlo Park, CA.

Targeted sequencing has proven to be an economical means of obtaining sequence information for one or more defined regions of a larger genome. However, most target enrichment methods are reliant upon some form of amplification. Amplification removes the epigenetic marks present in native DNA, and some genomic regions, such as those with extreme GC content and repetitive sequences, are recalcitrant to faithful amplification. Yet, a large number of genetic disorders are caused by expansions of repeat sequences. Furthermore, for some disorders, methylation status has been shown to be a key factor in the mechanism of disease. We have developed a novel, amplification-free enrichment technique that employs the CRISPR/Cas9 system for specific targeting of individual human genes. This method, in conjunction with SMRT Sequencing's long reads, high consensus accuracy, and uniform coverage, allows the sequencing of complex genomic regions that cannot be investigated with other technologies. Using human genomic DNA samples and this strategy, we have successfully targeted the loci of a number of repeat expansion disorders (*HTT*, *FMR1*, *SCA10*, *C9orf72*) and disease-associated homonucleotide stretches (*TOMM40*). With this data, we demonstrate the ability to isolate hundreds of individual on-target molecules and accurately sequence through long repeat stretches, regardless of the extreme GC-content, followed by accurate sequencing on a single PacBio RS II SMRT Cell or Sequel SMRT Cell 1M. The method is compatible with multiplexing of multiple targets and multiple samples in a single reaction. Furthermore, this technique also preserves native DNA molecules for sequencing, allowing for the possibility of direct detection and characterization of epigenetic signatures. We demonstrate detection of 5-mC in human promoter sequences and CpG islands.

1481T

Novel treatment approaches for autism spectrum disorder: An in vitro model. *L. Boccuto[1,2], L. Cascio[1], K. Champaigne[2], K. Jones[1], R. Pauly[1], S. Srikanth[1], C.F. Chen[1], S. Dunn[1], C. Skinner[1], C.E. Schwartz[1,2].* 1) Research, Greenwood Genetic Center, Greenwood, SC; 2) Clemson University, Clemson, SC.

Utilizing the Biolog Phenotype Microarray (PM) panel, our group demonstrated a reduced utilization of the amino acid tryptophan (Trp) as an energy source by lymphoblastoid cell lines (LCLs) from individuals with autism spectrum disorder (ASD), as compared to typically developing (TD) controls. This finding suggested potential targets such as the kynurenine pathway and the mitochondrial respiratory chain. In order to identify potential candidate drugs for novel treatments, 10 ASD and 10 TD LCLs were tested on a customized Trp plate, before and after exposure to several compounds added to the medium: supplemental energy sources (Trp, valine, glucose), intermediates of the kynurenine pathway (kynurenate, 3-OH-kynurenine, quinolinate), antipurinergic (suramin) and anti-oxidant (sulforaphane) drugs. The results confirmed significantly reduced utilization of Trp in untreated ASD LCLs and showed a normal metabolic response when the same ASD LCLs were exposed to an increased amount of Trp (3 mM) and sulforaphane (50 µM). Functional tests on mitochondrial metabolism exposing the same cell lines to respiratory chain inhibitors showed that rotenone (1 µM) reproduces in TD cells the same metabolic profile observed in untreated ASD LCLs. Additional experiments were conducted on a second, independent cohort of 10 ASD and 10 TD LCLs, utilizing both the customized Trp plate and the PM-Tox1 plate, which focuses on mitochondrial metabolism. The cells were exposed to different concentrations of Nutrifriend, a supplement of omega-3 fatty acids (ω-3) and vitamin D3, and a placebo (supplement without ω-3). The results replicated the significant differences in Trp utilization between ASD and TD samples and showed reduced utilization by ASD cells of glucose, inosine, galactose, and α-keto-glutaric acid, confirming a disruption in mitochondrial aerobic metabolism. The supplement proved to be significantly more efficient than the placebo (*P* values <0.05) in eliminating the metabolic differences between ASD and TD LCLs in both Trp and PM-Tox1 plates. Overall, our studies show how the PM platform can be used to successfully investigate the effect of potential treatments on the function of presumed target pathways involved in the pathogenesis of ASD. The results also suggest novel approaches to normalize metabolic abnormalities detected in ASD cells and indicate a potential beneficial effect of anti-oxidants and supplementation of Trp and ω-3 in individuals with this disorder.

1482F

LCM-Seq: Single cell-type whole genome bisulfite sequencing and transcriptomic profiling in post-mortem brain. *D.M. Almeida[1,2], G. Chen[2], M.A. Davoli[2], N. Mechawar[2,3], G. Turecki[2,3].* 1) Integrated Program in Neuroscience, McGill University, Montreal , Quebec, Canada; 2) McGill Group for Suicide Studies, Douglas Hospital Research Center, Verdun, Quebec, Canada; 3) Psychiatry, McGill University, Montreal , Quebec, Canada.

The epigenome and transcriptome of a cell constitutes an essential piece of cellular identity and accounts for the multifaceted complexity and heterogeneity of cell types within the mammalian brain. During neurodevelopment, spatiotemporal control over gene expression through epigenetic regulation of promoters and enhancers leads to precisely defined cellular fates. Each discrete cellular population is also differentially influenced by extrinsic signals from their local environments and neighbouring cells. Thus, while a wealth of studies have investigated epigenomic and transcriptomic alterations underlying the neurobiology of psychiatric or neurological illnesses, the use of bulk-tissue homogenates have masked their ability to determine cell-type specific molecular dysfunctions. Here we describe our progress on a pipeline that employs laser capture microdissection (LCM) of prefrontal layer V pyramidal cells, based on their well defined morphology, followed by downstream whole-genome bisulfite sequencing (WGBS) and transcriptomic profiling. Using this method, we achieve a 59% mapping rate efficiency, bisulfite conversion rates within the expected rage and optimal coverage of CpG sites for WGBS. RNA sequencing resulted in a mapping efficiency rate of 98% and captured a wide distribution of transcripts that map back to 13,141 genes. The major utility of this pipeline is its capacity to allow for the investigation of DNA methylation and expression patterns from the exact same dissected population of cells derived from post-mortem human brain tissue. LCM is compatible with immunohistochemistry as well as in-situ hybridization for the visualization and capture of specific cell types. Our goal is to use this pipeline to investigate cell-type specific molecular dysfunction underlying depression and suicide.

1483W

Single-nuclei transcriptomics in the brains of individuals with depression who died by suicide. *C. Nagy[1,2], M. Maitra[1,2], J.F. Theroux[1,2], H. Djambazian[3], G. Turecki[1,2].* 1) Douglas Mental Health University Institute; 2) McGill Group for Suicide Studies; 3) McGill University and Genome Quebec Innovation Centre.

Purpose: Molecular changes are typically measured in tissue homogenates or cellular fractions containing millions of cells pooled together. However, most tissue types, particularly the brain, have heterogeneous cellular composition. Multiple neuronal and glial subtypes with specific gene expression patterns are likely to be distinctly modified in a diseased state. As such, the normal variation of gene expression between cell types can mask specific changes when obtaining molecular information from cellular homogenates. The detection of subtle transcriptomic alterations such as those expected in psychiatric, would benefit greatly from single cell resolution. **Methods:** Using post-mortem brain tissue obtained from the Douglas Bell Canada Brain bank we explored the frontal cortex of 16 suicide completers and 16 matched healthy controls. Bulk nuclei were isolated from BA8/9 using an Optiprep™ gradient and single nuclei were captured using 10x Genomics' Chromium technology, a droplet based protocol. Each droplet encapsulates a nucleus, lysis buffer, reverse transcriptome reagents and a hydrogel bead with a DNA oligonucleotide. After encapsulation, the nuclei are lysed and the mRNA in each nucleus is reverse transcribed and tagged with a cell barcode and unique molecular indices (UMIs). The RNA undergoes a linear in vivo transcription base on a T7 priming site to produce amplified RNA (RNAa). The amplified RNA is then fragmented and reversed transcribed, adding the Illumina sequencing primers. A bulk library is made from the uniquely barcoded cells providing adequate material and reducing library prep costs. The tagged cDNA was sequenced using Illumina's HiSeq 4000 platform to approximately 50x coverage per cell. We used a custom Cell Ranger pipeline to de-multiplex the RNA sequencing reads. Cells were separated into clusters based on their transcription profile, and each cluster was further sub-divided into clusters for differential analysis between groups. **Results:** Using this single-nuclei approach, we can identify unique transcriptomic subtleties between groups. The features that cause these differences, need further study. These data are the stepping stone for identifying specific genes and networks of genes whose expression is dysregulated in depression.

1484T

Personalized dosing of dichloroacetate by clinical genotyping assay. *T. Langaee[1,3], R. Wagner[3], L. Horne[2], P. Starostik[2], L.A. Lawson[2], P. Stacpoole[2].* 1) College of Pharmacy, University of Florida, Gainesville, FL; 2) College of Medicine, University of Florida, Gainesville, FL; 3) Medosome Biotec LLC, Alachua, Florida.

Abstract Dichloroacetate (DCA) represents the first targeted therapy for pyruvate dehydrogenase complex deficiency (PDCD) by stimulating residual pyruvate dehydrogenase complex (PDC) activity. A federally-funded, multi-center Phase 3 Trial of DCA in PDCD is underway. DCA is metabolized to glyoxylate by glutathione transferase zeta1 (GSTZ1). Current data clearly establish variation in *GSTZ1* haplotype as the principal variable influencing DCA kinetics and dynamics in humans and occurs independently of subject age. The FDA has stipulated that were DCA to be approved for PDCD, labeling must include the need for genetics-based dosing. We aimed to develop and validate a genetic testing kit at Medosome Biotec clinical laboratory for *GSTZ1* genotyping to investigate the plasma pharmacokinetics (PK) of DCA as a function of *GSTZ1* haplotype and also determine *GSTZ1* haplotype status in individuals who would be treated with DCA. Forty-five healthy volunteer study participants signed consent and completed *GSTZ1* genotyping by providing buccal cells and whole blood. DNA samples were genotyped by TaqMan® for 3 functionally important *GSTZ1* SNPs (rs7975, rs7972 and rs1046428). There was 100% genotype concordance between DNA samples from blood and buccal swab. Prior studies dichotomized subjects into two classes on the presence of at least one EGT allele, such that EGT carriers metabolized DCA faster than EGT non-carriers. Eight participants were selected to complete a standard pharmacokinetic study based on *GSTZ1* genotype: 4 fast metabolizer (EGT carrier) and 4 slow metabolizer participants (EGT non-carrier). Each participant received a singled oral dose of 25 mg/kg of DCA (IND 028625) as a liquid formulation (25 mg/mL) each morning for 5 days, with formal PK evaluation conducted on day 5. Data from 8 subjects shown in the table are consistent with the premise that the EGT haplotype confers relating faster metabolism of DCA.

ID	Haplotype status	Plasma t½ (hr)	AUC (min*mg/mL)	Plasma CL (mL/min/kg)
	EGT carrier			
18	EGT/EGT	1.39	4171.42	6.05
22	EGT/KGT	3.32	6128.56	4.08
34	EGT/EGT	2.14	3469.33	7.27
41	EGT/EGT	1.33	3525.55	7.1
	EGT non-carrier			
11	EGM/KRT	28.35	83136.05	0.3
26	KGT/KGT	5.68	11341.26	2.21
33	KGT/KGT	2.68	5422.64	4.61
40	KGT/KRT	12.54	19173.44	1.32

We conclude that these preliminary data establish the validity and practicability of our rapid genotyping/haplotyping procedure for genetic-based DCA dosing.

1485F

Sequencing a baby for an optimal outcome: A clinical genomic application of newborn hearing screening. *A.B.S. Giersch[1,2,8], Y.I. Chekaluk[1], M.S. Cohen[2,4], K. Gregory[1], J. Hochschild[5], M.A. Kenna[2,3], P. Levesque[4], J. Manganella[3], J. Shen[1,2,8], M.R. Toro[1], C.C. Morton[1,2,6,7,8].* 1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Boston Children's Hospital, Boston, MA; 4) Massachusetts Eye and Ear, Boston, MA; 5) Harvard University, Cambridge, MA; 6) The Broad Institute, Cambridge, MA; 7) The University of Manchester, Manchester, UK; 8) Harvard Medical School Center for Hereditary Deafness, Boston, MA.

Purpose: SEQuencing a Baby for an Optimal Outcome (SEQaBOO) is a research project which seeks to integrate high-throughput genomic approaches into routine newborn screening for hearing loss. In this project, we will test the hypothesis that rapid discovery of the exact cause of a newborn's hearing loss will benefit clinical management. Background: Today, congenital hearing loss and other subtle birth defects are recognized in newborns through newborn screening, allowing early interventions that limit life-long disabilities. Genetic factors contribute to the majority of congenital hearing loss in the US. The etiologies are highly heterogeneous, and pathogenic variants in hundreds of genes have already been identified. Although not life threatening, hearing loss requires a number of adjustments by the family and patient to optimize quality of life. Cochlear implants can provide hearing in the near normal range, but they do not correct the underlying molecular defect. Development of whole genome sequencing technologies provides a new opportunity as well as challenge for researchers and clinicians dedicated to improving the lives of newborns with birth defects to investigate whether genetic etiologies of congenital defects can be more accurately and efficiently defined and whether improved genetic diagnosis translates into superior clinical care. Methods: The project will enroll approximately 500 newborns who do not pass their initial newborn hearing screen along with their biological parent(s) (trios) for whole genome sequence analysis. If parents choose not to enroll in the WGS study arm, they may choose to participate in an annual survey study arm. Genetic information pertinent to the infant's hearing loss will be returned to the child's physicians to guide management. Both study arms includes annual surveys to ascertain general health, including speech and language development, in addition to hearing status, and parental attitudes toward genomic sequencing. Summary: SEQaBOO aims to analyze and assemble genomic datasets, perform clinical genomic research of hearing loss identifiable through newborn screening, and explore implications of integration of genomic sequencing into newborn screening. All of this will inform the care and management of newborns with congenital hearing loss and allow us to investigate factors associated with society's acceptance of this new technology for "optimal outcome" of a newborn baby.

1486W

Building and scaling the world's fastest clinical-grade whole genome based pipeline. *Y. Ding, S. Nahas, S. Chowdhury, L. Van Der Kraan, L. Puckett, C. Yamada, S. Batalov, N. Veeraraghavan, R. Hovey, J. Cakici, S. Caylor, L. Salz, D. Oh, P. Mulrooney, W. Benson, P. Laird, D. Dimmock, S. Kingsmore.* Clinical Genome Center, Rady Children's Institute for Genomic Medicine, San Diego, CA.

Undiagnosed genetic diseases have a profound impact on the development of infants and young children. They are also the leading cause of death in infants, especially in Neonatal Intensive Care Units (NICU) and Pediatric Intensive Care Units (PICU). Efficacious medical interventions depend critically upon early diagnosis. Next generation sequencing such as the whole exome sequencing (WES) and the whole genome sequencing (WGS) have dramatically increased diagnosis rates. However, the turnaround time (TAT), test cost, data processing and data interpretation/reporting remain major obstacles for clinical facilities trying to perform WES or WGS. In October 2016 Rady Children's Institute for Genomic Medicine (RCIGM) established a clinical genome center at Rady Children's Hospital (RCH) San Diego CA. The laboratory is equipped with Illumina Hiseq2500 and HiSeq4000 instruments which allow performance of 120 rapid 40X WGS (rWGS) per month with an optimized rWGS workflow. From sample acquisition to a provisional diagnosis provided, our average TAT is 6 days and fastest TAT is 37 hours. In order to ensure the accuracy of the rWGS sample identity, Short Tandem Repeat (STR) allele profiling is employed. The laboratory has also optimized protocols to confirm potentially causative variants identified by rWGS using Applied Biosystems Genetic Analyzer. To date, over 300 individuals from more than 100 families have been sequenced and analyzed. To further shorten our current TAT and reduce sequencing cost, we are evaluating a novel library construction method and adopting a novel, rapid Illumina sequencing platform (NovaSeq6000 S2). Early diagnosis in NICU and PICU is further assured by utilizing a hybrid cloud-based ultra-rapid bioinformatics analysis and reporting pipelines, which includes single nucleotide variant (SNV), short insertion or deletion (INDEL) and structural variant detection, as well as the concordance and contamination checking for sample integrity. These which are specifically optimized for the short TAT requirements, yet remain uncompromising of performance. We present protocols and results of validation studies for clinical rWGS with return of results in less than 37 hours.

1487T

Disease relevant structural variation analysis by next-generation mapping. *A. Hastie, A.W. Pang, J. Lee, E.T. Lam, T. Anatharaman, W. Andrews, X.Y. Zhang, M. Saghbini, H. Cao.* Bionano Genomics, San Diego, CA.

Next-generation mapping (NGM) using the Saphyr™ System developed by Bionano, offers high-throughput, genome-wide visualization of extremely long DNA molecules in their native form. It allows researchers to interrogate genomic structural variations (SVs) in the range of one kilobase pairs and above. It uses extremely long range information to span interspersed and long tandem repeats making it suitable for elucidating the structure and copy number of complex regions of the human genome, such as loci with complex pseudogene and paralogous gene families. Since NGM is a *de novo* process and with molecules analyzed being longer than almost all genomic repeats, NGM is able to detect a wide range of SVs including insertions of novel sequence, tandem duplications, interspersed duplications, deletions, inversions and translocations. No other single technology can discover the breath and quantity of large SVs. The high throughput, cost-effective, genome wide nature and comprehensiveness of the SV types detected is driving NGM's application to the analysis of human genomes for the detection of SVs potentially involved in disease pathogenesis, including in diagnosing rare disease and cancer. We present several *in silico* and biological validation experiments that demonstrate the sensitivity and specificity of NGM for the detection of insertions, deletions, inversions and translocations compared to benchmark studies using short read and long read sequencing. We also show the application of NGM to study somatic variation in a breast cancer cell line, finding hundreds of somatic structural variations otherwise undetected. Finally, we applied NGM to several leukemia patient samples to find more than 50 cancer related SVs in each patient. NGM is a fast and cost effective method for the detection of a broad range of traditionally refractory SVs across the human genome. .

1488F

A systems biology approach to the understanding of asthma severity through the integration of metabolomic, transcriptomic and epigenetic networks. *R.S. Kelly[1], W. Qui[1], B.L. Chawes[2], Y. Virkud[1], K. Blighe[1], J. Celedon[3], S.T. Weiss[1], D. De Meo[1], J. Lasky-Su[1].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark; 3) Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh of the University of Pittsburgh Medical Center, University of Pittsburgh, Pittsburgh, PA 15260, USA.

   Background: Asthma emerges from a complex interplay between genetics and environmental exposures. Single-omic analyses have provided some insight into asthmatic lung function phenotypes, but the underlying biological pathways are still poorly understood. Integrative omicsanalysis may provide important insights into the mechanistic basis of this common chronic condition. Methods: 328 children with asthma from the 'Genetic Epidemiology of Asthma in Costa Rica' study underwent metabolomic (8,185 metabolite features), transcriptomic (25,060 gene probes) and epigenetic (118,721 CpG sites) profiling in blood extracted at recruitment (mean: 9.1 years). Weighted gene co-expression network analysis (WGCNA) in each omic dataset was used to independently identify modules of co-regulated metabolites, gene-probes, and gene-associated CpG-sites. Modules significantly associated with metrics of asthma severity; airway obstruction, reversibility, and hyper-responsiveness, were identified and their biology explored. Significantly correlated metabolite-gene and gene-gene module pairs were identified and the constituent variables submitted for integrated pathway analysis using IMPaLA. Results: WGCNA identified eight metabolomic modules, eight transcriptomic modules, and ten epigenetic modules. Of these, six metabolomic, four transcriptomic and three epigenetic modules associated (p<0.05) with asthma severity metrics based on their eigenvalue. The modules were enriched for asthma relevant processes including immune, mitotic and metabolic processes. Between-omic module associations were identified; in particular a 'sphingolipid metabolism' metabolite module associated with a transcriptomic (p=0.05) and epigenetic (p=0.01) module both of which were enriched for immune processes and included *ORMDL3*; a sphingolipid biosynthesis regulator and validated asthma gene. Conclusions: This study demonstrates that integrating multiple omic technologies using a systems biology approach provides a more informative, interpretable and biologically meaningful picture of asthma severity relative to single-omics analyses. Integration of correlated omic modules expanded and refined the single omic findings, linking dysregulated immunity to asthma severity via *ORMDL3* and sphingolipid metabolism. Therefore metabolomics can provide a mechanistic basis for the role of a number of asthma genes.

1489W

An integrative view of genetic and transcriptional regulatory interactions of the human placenta. *F. Delahaye[1], C. Do[2], Y. Kong[1], B. Tycko[2], F. Hughes Einstein[3], R. Wapner[4].* 1) Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) Institute for Cancer Genetics, Columbia university, New York, NY; 3) Department of Obstetrics and Gynecology, NYU, New York, NY; 4) Department of Obstetrics and Gynecology, Columbia, New York, NY.

   Functional genomics relies on the integration of genomics, epigenomics and transcriptomic information toward the discovery of early biomarkers and insights into disease etiology. The constant improvement in sequencing technologies has resulted in a surge in the number and complexity of multi-omics datasets, providing unprecedented opportunities to map the regulatory landscape, while imposing substantial analytical challenges. From present genomics association studies and quantitative trait loci analysis it became evident that significant efforts are still necessary to accurately depict the regulatory landscape. Yet, they allow the identification of inherent limitations to the current assays and analytical approaches such as the impact of cell subpopulation, genetic background and the need for broader data integration. These limitations have motivated the present study where we provide insight into the regulatory landscape of the human placenta and offer integrative analytical approaches to overcome the limitations of current platform. We intensively characterized the effects of genetic variation on gene expression and DNA methylation, and developed a comprehensive approach to identify methylation-sensitive transcription. We distinguish two profiles of interaction between expression and DNA methylation, revealing linear or bimodal interactions, reflecting differences in genomic context, transcription factor recruitment and cell subpopulation effects. We present here the extensive placenta-specific genome-wide analyses from the largest cohort studied to date. By combining multiple datasets, we overcome the limitations inherent to each individual dataset, improving our understanding of the intricacy of the regulatory mechanism in placenta. We show the complex interactions of genetic and transcriptional regulatory mechanisms, defining expression and methylation quantitative trait loci. We note the limitations of the transcriptional and DNA methylation studies performed, and how they could lead to over-interpretation of findings unless accompanied by stringent analysis. By publicly releasing such a reference dataset, we will considerably improve the interpretability of future (epi)genome-wide association studies.

1490T

Sequence read length effects on differential gene expression analyses using RNA-sequencing technologies. *A.C. Shetty[1], A. Mahurkar[1], C.M. Fraser[1,2], D.A. Rasko[1,3], V. Bruno[1,3], J.D. Hotopp[1,3,4].* 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland; 2) Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland; 3) Department of Microbiology & Immunology, University of Maryland School of Medicine, Baltimore, Maryland; 4) Greenebaum Cancer Center, University of Maryland, Baltimore, Maryland.

   Background: Since the onset of sequencing, there has been an exponential increase in sequencing throughput with an inconsequential decrease in costs. Sequencing of RNA molecules (RNA-Seq) to measure gene expression and assess differential expression between conditions has become increasingly powerful and an effective analytical approach. The Illumina HiSeq platform is commonly used for RNA-Seq generate 50 – 200 base pair (bp) paired-end reads. **Hypothesis:** Sequencing read lengths, library type and library depths have a significant effect on RNA-sequencing analytical results and need to be adjusted based on factors that are influenced by the variations in genome complexity and study designs. **Methods:** We analyzed 6 different transcriptomic datasets involving 5 reference organisms with varying genome size, gene number, gene length, exons per gene, and intron length. Raw 101 bp sequencing reads for each dataset were aligned to their reference genomes and read counts were computed for their genes as defined by their respective reference annotations. The raw counts were normalized for library size and gene dispersion and compared between conditions to determine differential expression of genes (DEGs). The analytical steps were repeated for read lengths of 72 bp, 54 bp and 36 bp as well as for individual single-end reads or paired-end reads. **Results:** We observed that alignment specificity decreases with increasing read length, except for less complex genomes that have no or a limited number of introns. Principle component analysis (PCA) and hierarchical clustering of samples for each dataset illustrated the factors influential in differentiating between samples from different conditions. We observed read length playing an influential role in multiple datasets further amplified when compared between paired-end and single-end sequencing techniques. When comparing the detected DEGs across the different read lengths, we observed highest correlations between analysis using the 54 bp and 72 bp reads. **Conclusions:** We observed differences in alignment statistics and detected DEGs in varying datasets influenced by the complexity of the genome, read length and sequence read pairing. This implies that longer reads do not necessarily mean better alignment and analytical results when analyzing complex transcriptomic datasets using existing methods.

1491F

Optimization of CRISPR/Cas9-mediated gene editing in monkey embryos. *X. Luo[1,2], B. Su[1].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China; 2) Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China.

   The RNA-guided Cas9 nuclease from the microbial clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system have been used to facilitate efficient genome editing in many model and non-model animals. However, its application in non-human primates is still at early stage, and the efficiency of gene modification in primate embryos remains rather low. Here, using zygotic injection of Cas9 mRNA and multiple single-guide RNAs with an optimized protocol, we achieved a high knockout efficiency of 86.67% for *MCPH1*- a human autosomal recessive primary microcephaly gene. Furthermore, through homology directed repair (HDR), we obtained monkey embryos with humanized *MCPH1*, and the efficiency is 21%. The improved gene editing efficiency in monkeys may serve as a useful tool in human disease and human evolution studies.

1492W

Development of a dual platform strategy for targeted DNA sequencing in genetic screening. *C. Schumacher, A. Wood, S. Sandhu, J. Lenhart, L. Kurihara, V. Makarov, T. Harkins.* Swift Biosciences, Ann Arbor, MI.

The use of next-generation sequencing (NGS) to assess genetic variation in genes involved in both inherited diseases, as well as cancer, is critical. Traditional NGS assays utilize short-read (SR) chemistry to enable sequencing of biomarkers in accessible parts of the genome, devoid of difficult to sequence motifs and high homology to pseudogenes. Newer long-read (LR) assays offer additional insight into the genetic architecture of known genes including haplotype structure, detection of large insertions/deletions, variants present in regulatory elements, and copy number variations. Both massively parallel screening techniques are economical, and a turn-around time of less than 48 hours makes them ideal for high-throughput use. To assess the utility of both SR and LR platforms, Swift Biosciences has developed single tube, multiplexed amplicon assays for both *CFTR* and *BRCA1/BRCA2* (*BRCA*) genes. DNA from MUTCF-2 (Coriell Institute) with known *CFTR* variants was used in both the SR and LR assays. In the SR assay, an 87-amplicon panel comprehensively covered all exons (≈10Kb) in *CFTR*. 10ng of DNA was used as input for multiplexed PCR, and the products were adapted and sequenced on an Illumina MiniSeq. The LR assay covered all the same exons, and 6 complete intronic regions, using 21 amplicons (≈54Kb). 25ng of DNA was used as input for the multiplexed PCR, barcoded for sample multiplexing, and adapted for Pacific Biosciences (PacBio) sequencing on an RS II. To assess the *BRCA* genes, DNA from BC01 (Coriell Institute) was used as input into the 246-amplicon (≈23Kb) SR panel which comprehensively covers all *BRCA* exons. A 35-amplicon panel (≈86Kb), was used to cover these same exons and 20 complete intronic regions with the LR assay. Both assays were adapted and sequenced as described for *CFTR*. Libraries from both SR assays were sequenced with >95% on-target and coverage uniformity. Based on the quality of the consensus sequence and the low number of multiplexed amplicons, the LR assays require up to 10x less sequencing reads per amplicon to call the same variants as the SR assays. The same number of variants were detected in both assays, despite the difference in sequencing depth. The LR assay identified indels, such as a 40bp deletion in *BRCA*, that are challenging to detect with short reads, indicating that large insertions/deletions are more robustly detectable with LR technology.

1493T

Enrichment of long reads for Mendelian disease using the Oxford Nanopore MinION. *E. Farrow[1,2,4], M. Gibson[1], J. Johnston[1], B. Yoo[1], G. Twist[1], S. Soden[1,2,4], C. Saunders[1,3,4], N. Miller[1], I. Thiffault[1,3,4].* 1) Center for Pediatric Genomic Medicine, Children's Mercy-Hospitals and Clinics, Kansas City, Missouri 64108; 2) Department of Pediatrics, Children's Mercy-Hospitals and Clinics, Kansas City, Missouri 64108; 3) Department of Pathology and Laboratory Medicine, Children's Mercy-Hospitals and Clinics, Kansas City, Missouri 64108; 4) University of Missouri-Kansas City School of Medicine, Kansas City, MO 64108.

The exponential decreases in sequencing costs, combined with the high quality score and accuracy of short read sequencing technology has led to an explosion of genetic testing. This approach has been particularly powerful for point mutations and small insertions and deletions. However, despite widespread utilization, diagnostic rates for Mendelian disorders range from 25-50% in part, due to the limited ability of short read sequencing to detect variation beyond point mutations and small insertions and deletions. Improvement in diagnostic rates will require overcoming challenges in the accurate detection of copy number variants, determination of repetitive sequence lengths, and resolution of pseudogenes. Long read sequencing is a potential tool for improving genetic testing as it addresses many of the current limitations of short read sequencing. However, the cost of long read sequencing combined with lower quality and overall accuracy has limited its utilization. The relatively low cost of the Oxford Nanopore MinION sequencer, both in capital expense and in sequencing costs, offers an attractive option for obtaining long reads. To explore the use of long read sequencing technologies for genetic testing, we have developed a novel enrichment protocol for long reads (average size of ~5Kb) that are sequenced using the MinION. The protocol is amenable to both small panel and whole exome enrichment using kits from Integrated DNA Technologies. Preliminary studies using exome enrichment resulted in 93% of bases on target. Overall output was initially low utilizing a single Mk1 R9 flow cell. Enrichment of a 170 gene panel resulted in approximately 1.19 Gb of raw data, which would enable multiplexing of up to 3 samples. The technology is rapidly improving, and using the newest SpotON Mk1 R9.4 flow cell we obtained more than 2Gb of high-quality reads from a single flow cell. To assist with error correction, samples were also sequenced using Illumina instrumentation (MiSeq or HiSeq 2500/4000). The total time required for sample preparation and sequencing on the MinION is one week. Current studies are ongoing evaluating the resolution of pseudogenes and triplet repeat expansions (*CYP2D6* and *FMR1,* respectively). In sum, we present a robust enrichment strategy for sequencing long reads, which has broad applications.

1494F

NGS pretesting and QC using Illumina Infinium Arrays. *J. Romm, H. Ling, P. Zhang, I. McMullen, P. Witmer, K. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. To reduce both reagent and sample waste, CIDR routinely performs sample pretesting to evaluate DNA quality, detect contamination/sample mixtures, confirm biological gender assignments, and check for Mendelian inconsistencies and inbreeding prior to initiating laboratory processes. For NGS pretesting, CIDR uses a dense Illumina Infinium array, which permits extended analyses for 1st and 2nd degree relationships, unexpected relationships, ethnicity and large chromosomal anomalies In addition, we calculate concordance and sensitivity to call heterozygotes, providing a critical quality control measure for NGS data.

1495W

900 exomes for rare disease research: Outcomes of the 2016 BBMRI-LPC WES call in collaboration with EuroBioBank and RD-Connect. *S. Beltran[1,2], M. Bayes[1,2], J. Dawson[3], B. Fusté[1,2], M. Gut[1,2], S. Laurie[1,2], H. Lochmüller[3], E. López-Martín[4,5], L. Matalonga[1,2], L. Monaco[6], G.B. van Ommen[7], D. Piscia[1,2], M. Posada[4,5], S. Sims[8], R. Thompson[3], C.M. Wang[6], E. Zeggini[8], I. Gut[1,2], BBMRI-LPC Consortium.* 1) Bioinformatics Analysis Group, Centro Nacional de Análisis Genómico (CNAG), Barcelona, Catalonia, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK; 4) Institute of Rare Diseases Research, IIER-ISCIII, Madrid, Spain; 5) Centre for Biomedical Network Research on Rare Diseases, CIBERER, Madrid, Spain; 6) Fondazione Telethon, Milan, Italy; 7) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 8) Wellcome Trust Sanger Institute, Hinxton, UK.

The 2016 BBMRI-LPC WES Call offered a unique free-of-charge opportunity to genetically diagnose rare disease patients with biological samples deposited within the EuroBioBank network. The program sequenced and analyzed 900 Whole Exomes from 17 projects, each with 2-3 principal investigators (PIs) from different countries. These projects included rare genetic eye diseases, congenital myasthenic syndromes, ataxia, neuromuscular diseases, CMT, inborn errors of metabolism, mitochondrial diseases, familial gastric neuroendocrine tumors, gastroschisis, disorders of glycosylation and 3-methylglutaconic aciduria, albinism, rett-like syndrome, Pyruvate Kinase Deficiency and Sudden Cardiac Death. The Informed Consent had to allow data sharing for research through controlled access repositories and databases such as the EGA (https://ega.crg.eu/) and RD-Connect (https://platform.rd-connect.eu). Sequencing was conducted at the Centro Nacional de Análisis Genómico (CNAG-CRG, Spain) and at the Wellcome Trust Sanger Institute (WTSI, UK) and was finished in early 2017. Raw data will be submitted to the EGA for long-term storage. Clinical and phenotypic information was collected with RD-Connect's customized PhenoTips instance, using standards such as the Human Phenotype Ontology (HPO), OMIM and Orpha codes. Exomes were processed through the RD-Connect validated analysis pipeline and were made available through the platform once the PIs had fulfilled all the requirements. The RD-Connect platform allows researchers to analyze and interpret their genotype:phenotype data privately for up to 6 months before it is shared with other authorized users. The user-friendly online system includes numerous annotations and filters as well as tools such as Exomiser and Genomiser. It also enables anonymized data sharing through APIs from initiatives such as the GA4GH/IRDiRC MatchMaker Exchange and GA4GH Beacon. The BBMRI-LPC WES Call represents an important move towards developing a best-practice paradigm for secure data and sample sharing for diagnostics and gene discovery in rare disease. We will report on the challenges and lessons learned from conducting such a complex transnational collaborative initiative. We will also present the up-to-date diagnostic yield of the project as a whole while focusing on some success stories.

1496T

Performance comparison of two exome enrichment systems for enhanced coverage of disease-associated regions. *L. Tian[1], R. Pellegrino Silva[1], J. Garifallou[1], M. Hegde[2], H. Hakonarson[1,3,4], A. Santani[1,5].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) PerkinElmer, Inc; 3) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania.

   Exome sequencing is increasingly used as the primary diagnostic test in patients with rare genetic disorders. To ensure the quality of the clinical exome sequencing tests, it is imperative for clinical and research laboratories to be aware of the technical complexities with the commercially available products that can have a critical impact on the sensitivity of mutation detection. We assessed two recently available commercial DNA target enrichment methods, namely SureSelect Clinical Research Exome V2 (Agilent) and SeqCap EZ MedExome (Roche NimbleGen). Both target enrichment systems include enhanced coverage of select disease associated genes. The CRE v2 design was curated by the consortium of the Emory Genetics Laboratory and the Children's Hospital of Philadelphia. Target enrichment libraries from one individual (NA12878) were prepared following each manufacture's protocol, and pooled libraries were run on an Illumina HiSeq2500 with 2x101 paired end reads. The sequencing data were processed on Edico genome's Dragen BioIT processor. Target enrichment metrics were calculated using Picard and GATK DepthOfCoverage tool.   We compared the performance of two enrichment systems based on a comprehensive metrics, including coverage of CCDS, RefSeq, Ensembl, and GENCODE databases, depth of coverage, evenness of coverage, enrichment efficiency, influence of GC content, cumulative distribution of on-target sequence coverage in 1) target bases, 2) coding exons, 3) 5' UTR and 3'UTR, 4) splice junctions, 5) promoters in the following regions 1) ~4600 disease-associated genes, 2) 59 ACMG secondary finding v2.0 genes, 3) 70988 HGMD/ClinVar variants in disease-associated genes.   We evaluated the sensitivity and specificity of SNV and small INDEL calls generated from the two enrichment systems by comparing the calls with the high-confidence calls published by the Genome in a Bottle consortium. We concluded that both enrichment systems 1) show high target coverage efficiency (>70%), 2) yield highly sensitive and accurate variants calls in the ~4600 disease associated genes, 3) have a negative correlation between sequencing coverage and extreme GC content. Each system presents its own strength and weakness. CREV2 consistently demonstrated improved performance in coverage over disease associated genes and higher coverage of known pathogenic mutations.

1497F

Improvement in automated NGS sample preparation workflows: Combining the Agilent Bravo Liquid Handling Platform with Covaris Focused-ultrasonicators. *C. Whitman[1], M. Visitacion[1], J. Terry[2], G. Durin[2].* 1) Agilent Technologies, Santa Clara, CA; 2) Covaris Inc, Woburn, MA.

   As the field of Next-Generation Sequencing (NGS) continues to expand, more and more genomic researchers are seeking automation as way to increase their sample throughput while maintaining data quality and sample traceability. In addition to have a robust liquid handling platform there is a need to provide a simplified user interface as most new users are not automation experts. Covaris Focused-ultrasonicators employ Adaptive Focused Acoustics™ (AFA) technology for the controlled mechanical shearing of nucleic acids. AFA hydrodynamic shear force-based fragmentation is purely mechanical and designed to be isothermal, ensuring both unbiased fragmentation and high recovery of DNA. It is such recognized as the industry standard for DNA shearing for NGS sample preparation. The Covaris microTUBE is specifically designed and engineered for compatibility AFA technology and small sample volumes. The 96 microTUBE Plate is an automation friendly consumable that is directly compatible with standard automation deck positions. For smaller throughput experiments the 8 microTUBE Strip, with newly developed H Slit septa, provides a versatile solution with a SBS-format plate adaptor that can handle up to 12 tube strips. While most of the available automation workflows have primarily been centered on NGS Library Preparation and Target Enrichment protocols, there have been only a few commercial solutions presented around the initial DNA sample transfer from the Covaris microTUBE consumables. Here we introduce an automated workflow for extracting sheared DNA samples from a 96 microTUBE Plate and 8 microTUBE Strip using the Agilent Bravo NGS liquid handling platform. We show the successful transfer of samples from each microTUBE product and demonstrate that sample and data qualities are comparable between manual and automated handling. Automation is a mandatory requirement for any high-throughput project that reduces variability, improves workflow efficiency and increases walk-away time. Utilizing the Agilent Bravo Liquid Handling Platform for automating extraction of sheared DNA from the Covaris 96 microTUBE Plate and 8 microTUBE Strip offers a reliable and reproducible solution for high-throughput studies. For Research Use Only. Not for use in diagnostic procedures.

1498W

Spike-in controls designed for detecting sample bleeding and misidentification in sequencing workflows. *J. Kinman, S. Piehl, A. Morris, D. Fox, C. Firmani, M. Toloue.* PerkinElmer, Austin, TX.

Cross contamination in sequencing workflows is a serious issue that has led to inaccurate data interpretation. Verification and elimination of sample ambiguity is critical for the widespread use of NGS. Understanding the extent and source of miscalled reads, even at low levels, is important for accurate variant calling and can affect statistical power in the detection of genetic associations. Traditional library preparation methods lack the ability to effectively screen for procedural, instrumental, and human error in sequencing workflows. A fully validated set of spike-in controls can identify these errors and ensure samples are correctly labeled and identified with minimal impact on sequencing data. These spike-ins offer sequencing service providers and scientists a tool to assess sample purity, cross-contamination, and index carryover; giving them confidence in the integrity of their data and a confirmatory checkpoint for when errors do occur.

1499T

NEBNext Ultra II FS DNA: An enzyme-based, single tube fragmentation system for library construction. *V. Panchapakesa, L. Apone, K. Duggan, B. Langhorst, T. Shtatland, C. Song, P. Liu, C. Sumner, C. Rozzi, M. Arn, D. Munafo, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, 240 County Road, Ipswich, MA 01938.

Success on NGS platforms begins with optimal sample preparation, which has become a bottleneck in NGS workflows. We have developed a robust library preparation method that integrates enzyme-based DNA fragmentation with end repair and dA-tailing in a single tube, resulting in the elimination of expensive equipment to fragment DNA. This streamlined method is compatible with a wide range of DNA samples and input amounts, resulting in reproducible and reliable fragment sizes. Genomic DNA (100pg to 500ng) from various sources was fragmented, end repaired and dA-tailed in a single step, followed by adaptor ligation in the same tube. The sample was then cleaned up and sequenced on an Illumina platform. For low input samples, PCR amplification was performed prior to sequencing. Libraries constructed using intact genomic DNA and this novel library preparation method produced substantially higher yields than those generated using standard methods and mechanically fragmented DNA. The greatest differences were observed with the lowest DNA inputs and most challenging samples, including FFPE where a 2- 4 fold increase in yield was observed with the Ultra II FS DNA kit. Sequencing quality of libraries generated with inputs ranging from 100pg to 500ng DNA show similar coverage uniformity, GC bias and fragment size distribution. Ultra II FS generates high-quality libraries from a broad range of DNA sample amounts and qualities. Sequencing metrics of FS libraries are similar or superior, to those produced with mechanically sheared DNA. This easily automatable, single tube, workflow has a wide range of NGS applications including whole genome, whole exome or targeted sequencing, and single cell RNA-seq.

1500F

The Personal Genome Project Canada: Whole genome sequences and comprehensive medical annotation of 56 Canadians. *M.S. Reuter[1], S. Walker[1], B. Thiruvahindrapuram[1], J. Whitney[1], I. Cohn[2], R.K.C. Yuen[1,3,4], B. Trost[1], N. Sondheimer[5], M.R. Hildebrandt[6], J. Ellis[4,6], M. Fiume[1,7], T.A. Paton[1], S. Pereira[1], D. Merico[1,8], J.R. MacDonald[1], J. Buchanan[1], T. Nalpathamkalam[1], W.W.L. Sung[1], R. Patel[1], G. Pellecchia[1], J. Wei[1], A. Hazell[9], J. Davies[9], N. Monfared[5,10], S.M. Hosseini[1,3,11], F. Keeley[12], H.C. Lee[13], M. Szego[1,14], C.R. Marshall[1,10,15], S.W. Scherer[1,3,4,13].* 1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Clinical Pharmacology and Toxicology, The Hospital for Sick Children, Toronto, ON, Canada; 3) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 6) Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto, ON, Canada; 7) DNA stack, Toronto, ON, Canada; 8) Deep Genomics Inc., Toronto, ON, Canada; 9) Medcan Health Management Inc., Toronto, ON, Canada; 10) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 11) Ted Rogers Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, ON, Canada; 12) Departments of Biochemistry & Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; 13) McLaughlin Centre, University of Toronto, ON, Canada; 14) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 15) Department of Molecular Genetics, Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada.

Rapid advances in sequencing technology promise that genetic information will be increasingly used in mainstream medicine. Established a decade ago, the Personal Genome Project Canada is a foundational research project for ethics protocol development, sample collection, data generation, medical annotation and delivery of findings, all designed for impact in the context of the Canadian health care system. A unique feature of this project is that data are returned to participants by a genetic counsellor. A prime objective was to determine the clinical utility of whole genome sequencing (WGS) when offered to population (healthy) controls. We followed first-come-first-serve enrolment at our website (http://www.personalgenomes.ca) and have registered 1107 participants, who undergo an eligibility screening and provide informed consent to publish their data. Using blood DNA, we performed WGS (Illumina HiSeq X) with an average of 38-fold coverage. High-resolution microarray (Affymetrix Cytoscan-HD) and methylation arrays (Illumina EPIC) were also performed, and induced pluripotent stem cell lines are being generated. Here, we report on the analysis of the genomes of the first 56 participants. We assessed all clinically relevant genetic variants. Rare variants (<5%) were emphasized, and those either predicted to result in null alleles or reported to be disease causing, were further assessed according to American College of Medical Genetics recommendations. Of 1587 analyzed variants, 1409 were (likely) benign or of uncertain significance; 169 sequence-level variants and 9 CNVs were classified as (likely) pathogenic, mostly associated with autosomal recessive (162) or semidominant (12) inheritance patterns. Four (likely) pathogenic (semi)dominant sequence variants (in *BRCA1*, *LMNA*, *LZTR1* and *SLC7A9*) and a mosaic (70%) loss of one X chromosome were considered to be of clinical relevance. Several variants with evidence from functional studies remained of uncertain significance but clinically suspicious (in *ANK2*, *CDH1*, *CHMP2B*, *ELN*, *KCNE2*, *TRPM4*). A pathogenic mitochondrial variant in *MT-TV* was found with low level heteroplasmy (7%). We identified an average of 3.2 disease alleles per individual. However, substantial burden of clinically ambiguous alleles and pathogenic variants with insufficient data on disease penetrance hampered individual risk predictions. In 12/56 participants (21%), our research findings may have health planning/management implications.

1501W

Comprehensive discovery of genomic variation from the integration of multiple sequencing and discovering technologies. *X. Zhao[1], M. Chaisson[2], Human Genome Structural Variation Consortium.* 1) Massachusetts General Hospital, Boston, MA; 2) University of Washington, Department of Genome Sciences, Seattle, WA.

Genomic variants range in size from single nucleotide variants to large structural variants spanning megabases of DNA. Such variation plays an important role in evolutionary selection and can also be a critical driver of human disease spanning the spectrum from neurological disorders to cancers. Comprehensive discovery and sequence resolution of structural genomic variation is a difficult problem that has limited our ability to make disease associations. Here, we applied 7 different genomic technologies and 23 independent algorithms to parent-child trios from diverse ancestries to discover haplotype-resolved SVs at nucleotide resolution. We independently integrated SVs discovered by different platforms, followed by systematic comparison and further integration to achieve a comprehensive SV callset. On average, we identified and validated more than 500,000 indels (1-49 bp) and 30,000 copy number variants (≥ 50bp) within each child, representing a 10-fold increase over previous 1000 Genomes Project reports derived from standard short read libraries. We also discovered ~235 high confident large inversions and complex SVs per genome, which were often partially misrepresented or omitted by other studies. We further present an analysis of these SVs for their prevalence across the genome as well as their potential impact on gene expression. We show that near complete physical haplotype resolution with 10X Genomics and strand-seq data allows long-read sequences to be accurately partitioned and assembled by haplotype, significantly increasing our yield of structural variation. Application of strand-seq based approaches coupled with optical mapping provides access to large-scale inversions in the human population that were previously underrepresented and where the current human reference genome is likely in error. Our results suggest that haploid resolution of diploid genomes and combinations of orthogonal approaches depending on variant size and class are required to accurately resolve the majority of human genome structural variation. We anticipate that these data will enable a more complete understanding of the evolutionary and functional consequences of these variants in human populations, and serve as the new gold standard for human genetic variation.

1502T

Detection of small exonic CNVs, SNPs and INDELs in a single assay. *S. Melo, K. Jeong, A. Vadapalli, A. Ashutosh, J. Ghosh, G. Amparo, M. Ramaiah, D. Roberts.* Agilent Technologies, Santa Clara, CA.

Deletions and duplications in genomic DNA larger than 50 base pairs, collectively known as Copy Number Variations (CNVs), have been recognized as a major source of genetic variation. While many CNVs have no obvious phenotypic effect, many others have been implicated in numerous genetic conditions and cancer, underlying the importance of CNV detection. Next generation sequencing, particularly targeted sequencing of exons, is becoming the predominant approach to identify and characterize genomic variants and mutations. A wide variety of methods and computational solutions that use this type of data to infer CNVs have previously been described. However, measurements along only exonic regions can pose significant limitations to resolution of smaller CNVs spanning one or two exons. To enable higher resolution of CNV calls in and around exons of interest, we have designed custom target enrichment probes flanking exons, described as custom OneSeq probes. These probes were designed to augment Agilent SureSelect exon panels, which enables detection of single nucleotide polymorphisms (SNPs), deletions and insertions smaller than 50 bp (INDELs), and exon level CNVs spanning hundreds to thousands of bp in one single assay. Here, we demonstrate improved detection of small exonic CNVs in several clinically important regions, including DMD, PARK2, and MECP2. Importantly, the addition of custom OneSeq probes refines predicted breakpoints, and reduces the rate of false positives, without sacrificing detection of exonic SNPs/INDELs. Thus, Custom OneSeq probes together with exonic probes allow detection of small exonic CNVs, SNPs and INDELs in a single assay.

1503F

An optimized single tube workflow for robust, low input NGS library preparation. *G. Durin[1], E. Weiss[1], J. Terry[1], J. Donaldson[1], U. Thomann[1], E. Daviso[1], J. Sampson[1], J. Lenhart[2], S. Sandhu[2], L. Kurihara[2], J. Laugharn[1].* 1) Covaris Inc, Woburn, MA, USA; 2) Swift Biosciences, Ann Arbor, MI, USA.

As Next Generation Sequencing (NGS) moves into the clinical diagnostics space, requiring higher throughput and sensitivity, existing processes need to be improved to ensure accurate, precise data outputs. Mechanical DNA shearing using Covaris Adaptive Focused Acoustic (AFA) technology is recognized as the gold-standard for DNA fragmentation in NGS library preparation. AFA hydrodynamic shear force-based DNA fragmentation is strictly mechanical and isothermal, and thus guarantees a highly controlled process while yielding random, unbiased DNA fragment distributions. We present a new DNA shearing vessel, the Covaris oneTUBE that is optimized for low DNA input and allows single tube DNA shearing and NGS library preparation. This new engineered polymer vessel incorporates features that effectively control acoustic cavitation to enable reproducible and precisely tuned hydrodynamic shear forces. Its low binding surface guarantees high DNA recovery as well as ease of volume handling down to 5 µl and picogram DNA inputs. Additionally, this vessel is fully compatible with magnetic separation procedures, thermal cyclers and liquid handling platforms. oneTUBE design is adaptable for scaled-up SBS formats, such as 96, 384, and higher well densities. We present data that demonstrates precise and accurate DNA shearing in the Covaris oneTUBE with optimized truSHEAR buffer, and validation using a wide range of DNA input quantities and %GC. Additionally, processing times are up to two times faster than in traditional vessels. For NGS library construction, DNA shearing in the oneTUBE was paired with Swift Biosciences Accel-NGS 2S Kit. The 2S kit preserves complexity from low input samples due to enhanced end repair chemistry for highly efficient adapter ligation in a single tube workflow. This combination enables for the first time DNA shearing, end repair, and adapter ligation in the same vessel, thereby eliminating cumbersome sample transfers. We present sequencing data from libraries constructed with this workflow and compare them to libraries constructed according to standard "with transfer" protocol. Using genomic DNA from species with varying %GC, no adverse biases in number of PCR duplicates or coverage uniformity were observed at inputs ranging from 10 pg to 100 ng. PCR-free human whole genome sequencing on an Illumina HiSeq X using 200 ng Coriell NA12878 will also be presented, where the oneTUBE workflow produced higher library yields than the standard workflow.

1504W

Microbiome technology comparison: Axiom microbiome array, 16S rRNA sequencing and metagenomics. *J. Foster[1], R. Almond[2], L. Cobb[2], S. Welsh[2], G. Tilston[2], M. Shah[3], A. Pirani[3], C. Bloor[1], G. Scopes[1].* 1) Thermo Fisher Scientific, Paisley, United Kingdom; 2) UK Biobank, Stockport UK; 3) Thermo Fisher Scientific, Santa Clara, United States.

Historically microbiome research has hinged upon the targeted sequencing of rRNA and whole genome (metagenomic) sequencing for organisms that could not be cultured in the lab. With the release of the Axiom™ Microbiome Array an orthogonal approach to the elucidation of diverse microbioata is possible. As part of a UK Biobank potential sample collection protocol, a pilot study was conducted to determine optimal sample collection, storage and treatment methods. Results for metabolomics, 16S rRNA sequencing, Metagenomics and the Axiom Microbiome Array were generated, allowing the three genomics technologies to be compared. Comparison involved 77 DNA extractions from 7 samples aliquoted in triplicate; one aliquot for each technology. Participants were age-matched samples outside of the UK Biobank cohort. Samples were subjected to one of four treatments (Native, DNA/RNAase-free water, Omnigene, RNA Later) and underwent one of three storage conditions: immediate storage at -80°C and storage at 4°C for 24 or 48 hours then transferred to -80°C prior to extraction. Microarrays were processed in the Thermo Fisher UK Service Lab and sequencing carried out at two independent academic collaborator sites. Microbiota identification was also conducted independently at the microarray/sequencing source site. Cross technology comparison was performed at Thermo Fisher with no re-processing of the reported microbe identifications. Comparison was primarily done at the genera level, that being the lowest common denominator between the three technologies with the 16s rRNA data unable to produce higher resolution species-level identifications. The Axiom Microbiome array compared favourably to 16s rRNA sequencing in respects to within-participant reproducibility of identifications, QC pass rate of samples and resolution of identifications. Metagenomics provided a wealth of high resolution identifications to corroborate those of the Axiom Microbiome array with the trade-off of laborious complex data analysis and high per-sample cost. In addition to bacterial identifications, we also present Axiom Microbiome Array results from the four other super kingdoms of life: Archaea, Fungi, Protozoa and Viruses.

1505T

A spin column and magnetic bead-based approach for the isolation of host and bacterial DNA from human feces. *D. Wieczorek, S. Lewis, B. Hook, T. Schagat.* Promega Corporation, Madison, WI.

Fecal samples contain an abundance of genetic material and can be collected non-invasively, making them an ideal sample type for human disease and microbiome research. However, the presence of humic acid and other amplification inhibitors in human fecal samples can present significant challenges, as the incomplete removal of these contaminants can affect downstream applications. To effectively purify DNA from these samples, lengthy and hazardous protocols are often necessary. As an alternative, we present two approaches for purification of DNA from feces, one compatible with manual spin columns and another magnetic particle based method adaptable to small, automated instrumentation as well as high throughput platforms. The performance of these methods was compared to another commonly used manual spin column-based technique and DNA quantity, quality, amplifiability, and performance in 16s microbial sequencing were evaluated. In this testing, the methods we describe demonstrated higher DNA yields and equivalent amplification and sequencing performance when compared to the established method. These flexible techniques yield pure, concentrated DNA from fecal samples without the use of organic solvents or cumbersome spin steps for use in rigorous downstream applications.

1506F

Preparing small RNA libraries from low input and single cell total RNA.
S. Shore[1], JM. Henderson[1], EWT. Tam[2], JF. Quintana[2], FWN. Chow[2], K. Chhatbar[2], D. Keys[1], AP. McCaffrey[1], G. Zon[1], A. Buck[2], RI. Hogrefe[1]. 1) Research and Development, TriLink BioTechnologies,San Diego, CA; 2) Centre for Immunity, Infection and Evolution, University of Edinburgh, Edinburgh, UK.

 Next-generation sequencing (NGS) has revolutionized nucleic acid analysis, and is revealing new details of diverse RNA functions. Many biological samples contain limited amounts of RNA, thus it is important to use a library preparation method yielding highly specific products from low input amounts. To achieve high sensitivity detection, it is vital to minimize side reactions and contaminants such as adapter dimers that waste flow cell space resulting in unproductive NGS reads. A challenging sample type to prepare for sequencing is small RNA (sRNA), non-coding RNA sequences, such as miRNA. Due to their short sequence lengths, a sRNA library is time-consuming to separate away from adapter dimer prior to NGS. Therefore, we developed a technology using modified adapter chemistries to reduce adapter dimer formation otherwise occurring during sRNA library preparation. By significantly reducing adapter dimer formation in the NGS library, small RNA libraries can be purified by rapid and automatable magnetic bead-based RNA size selection eliminating the need for gel purification. In addition, this reduction of adapter dimer allows for successful NGS library prep when using significantly lower total RNA inputs. For example, using modified CleanTag™ adapters, gel purified libraries generated from the inputting of 10 and 1 nanogram (ng) total human brain RNA resulted in 86% of total reads mapped to miRNA and less than 1% of total reads attributed to adapter dimer. In contrast at 10 ng input RNA the gel purified libraries from a commercially available kit resulted in only 40% of total reads mapped to miRNA while 50% of total reads attributed to adapter dimer. While commercial kits do not produce usable data at 1 ng, by replacing gel purification with magnetic bead purification at 1 ng input, CleanTag™ libraries resulted in sufficient miRNA reads and less than 10% adapter dimer. We have demonstrated broad applicability by successful sequencing low RNA input biological samples such as extracellular vesicles, plasma, urine, FACS sorted and immunoprecipitated samples that have previously proven difficult to sequence. CleanTag™ library sequencing results at 10 picogram total human brain RNA input show the potential to sequence small RNA at the single cell level. These results demonstrate the value of adapter dimer suppression and the utility of CleanTag™ technology for both low input and high throughput library preparation.   .

1507W

TaqMan Advanced miRNA assays to simultaneously study expression of miRNA and mRNA from serum samples. *H. Veereshlingam[1], M. Buchholz[2], K. Varma[1], S. Brown[1].* 1) ThermoFisher Scientific, South San Francisco, CA; 2) Philipps-Universitaet Marburg, 35043 Marburg, Germany.

 MicroRNAs (miRNA) are a class of small non-coding RNAs (approximately 21 nt long) that bind complementary sequences in target mRNAs to specifically regulate gene expression. miRNAs play important roles in regulation of gene expression during cell development, differentiation, proliferation and apoptosis. Misregulation of miRNAs and their targets has been associated with several diseases including cancer. The interplay between miRNA and mRNA has been found to be important in cancer development and progression. Simultaneous expression studies of miRNA and mRNA can be valuable in understanding molecular mechanisms that may potentially have an underlying role in various diseases. We demonstrate the technical verification of a novel method to reverse transcribe and pre-amplify miRNA and mRNA from sample limiting serum research samples using TaqMan® Advanced miRNA cDNA synthesis Kit. Based on results from previous studies, Dr. Buchholz's research group identified a signature of 55 mRNA and 41 miRNA targets that may help to distinguish between benign and malignant pancreatic tissues in the future. In this study these previously identified miRNA and mRNA targets have been used to study expression pattern differences in serum from Normal and Test samples. These miRNA and mRNA targets were spotted on Custom TaqMan array cards to facilitate investigation of 4 samples on each card. Initial results indicate that miRNA and mRNAs can be reliably quantified from a single reverse transcription reaction.

1508T

Complete, streamlined, reduced-bias workflow for RNA extraction and small RNA library preparation from serum and plasma samples. *R. Wong[1], S. Eun[2], A. Lukasiewicz[2], K. Allen[2], D. Simpson[1], A. Morris[2], M. Toloue[2].* 1) Queen's University Belfast, Belfast, Northern Ireland; 2) PerkinElmer, Austin, TX.

   The profiling of small RNAs, miRNAs in particular, in human plasma and serum samples is a rapidly expanding field, largely due to the great promise of cell-free RNAs as biomarkers. Here we describe a workflow using a phenolic extraction reagent along with magnetic bead-based cleanup for isolation of total RNA from human serum and plasma, followed by reduced bias small RNA NGS library preparation. We have developed a protocol using a phenolic extraction reagent combined with a magnetic bead purification for isolation of total RNA, including small RNA, from up to 600 µL of plasma or serum in a microcentrifuge tube format. The use of magnetic bead-based purification greatly increases throughput, time to complete, and user-friendliness by avoiding time consuming and tedious alcohol precipitation or column-based cleanups. Isolated RNA is then used as input into a reduced-bias library preparation protocol that uses the proven technology of adapters with randomized ends to reduce ligase bias. We also describe a protocol modification that greatly reduces the number of reads mapping to tRNA and YRNA fragments, which can often be a substantial number of reads in libraries created from serum or plasma RNA. This modification results in a greater number of reads mapping to miRNAs which, in combination with bias reduction, allows for detection/discovery of more miRNAs at equal sequencing depth or the ability to sequence samples less deeply than other popular protocols while obtaining comparable data. Here we describe a streamlined, user-friendly workflow for isolation of total RNA and generation of reduced-bias small RNA sequencing libraries from plasma and serum samples. This tool will be valuable to the growing number of researchers exploring extracellular small RNA for various reasons, including biomarker discovery.

1509F

Study of touch DNA in simulated situations for forensic purposes. *F.T. Goncalves, D.O. Francisco, J.H. Rios, C. Fridman.* Legal Medicine, Ethics and Occupational Health, Medical School, University of São Paulo, Sao Paulo, Sao Paulo, Brazil.

   In a crime scene, many types of biological traces can be found, among them, fingerprints are the most commonly. Currently, partial fingerprints can be used to obtaining genetic profiles based on genomic polymorphisms, assisting the elucidation of criminal investigations. However, many variables may interfere with the obtaining of these profiles, for example, the type of the surface on which the biological material was deposited. In Brazil, crimes with cold weapons are very common and happen daily. The most significant are those committed with knives, a tool easily found in any residence. Therefore, trying to assist the forensic work, this study aimed to analyze the feasibility of obtaining STR profiles from three different kind of knife handle: wood, steel and polypropylene. Forty simulated experiments were conducted with three women. Each woman manipulated one of the knives during one hour. The genetic material provided from the fingerprints left on the handle knives were collected with swab and the DNA was extracted by two methodologies in different days. In part of the experiments Chelex 5% was used to extraction and the other part of experiments this step was conducted with Swab Solution and the samples were amplified by multiplex PCR using the kit for Human Identification *PowerPlex® Fusion* System. Results showed that the wood handle knife was the best to recover the biological material deposited by hands, because 20 of the 24 markers were found in the STR profiles. On the other hand, profiles obtained from polypropylene were partials. However, when Chelex was used in the extraction, the quantity of allelic drop-out were higher than when Swab Solution was used. No DNA profile were obtained in the steel knife handle with any methodology used. These results showed that it was possible to obtain DNA profile from fingerprint from two of three types of knife grip materials that may aid in criminal investigations. These data are part of a larger study that evaluate the feasibility of obtaining STR profiles on door handles and steering wheels of cars and computer keyboards, as well as the evaluation of other variables such as gender of participants and different climatic conditions.

1510W

Design of Axiom Asia Precision Medicine Genotyping Array optimized for East and South Asian populations using improved SNP selection algorithms. *A. Mittal, J. Gollub, S. Kaushikkar, T. Webster, J. Schmidt.* Thermo Fisher Scientific, Santa Clara, CA.

Most existing pan-population genotyping arrays offer less than ideal high coverage of East and South Asian populations. We have designed Applied Biosystems™ Axiom™ Asia Precision Medicine Array that includes markers ideally suited for imputation in East and South Asian populations. Marker selection was accomplished by extending our previously-described imputation-aware array design algorithm [Hoffmann et al., 2012] using an approach inspired by the multiple-population tag SNP selection algorithm [Carstensen et al., 2014]. The set of markers only occupies 40 % of a typical Axiom array, leaving 60 % for additional content that can be customized, while achieving coverage of up to 93% for variants with minor allele frequency > 0.05 and up to 82% for variants with minor allele frequency > 0.01 in all East and South Asian populations. This compares favorably to the coverage achieved by Axiom™ Precision Medicine Research Array (PMRA). Marker selection was initially achieved in 168 CPU-hours on 180-core Linux cluster with 8GB of RAM/core. We have further demonstrated the ability to execute marker selection using a cloud-based cluster that can reduce development time by up to 6x, permitting flexible customization of the imputation panel if needed. We have extended this population specific approach to coding, LoF, and potentially pathogenic variants, which are also typically biased toward European populations on other arrays. We have identified such variants, taking their alternate allele frequencies for East and South Asian populations into account, from public databases such as ClinVar (May 2017), NHGRI GWAS catalog (May 2017) and PharmGKB, as well as additional recent publications. In summary, we describe a custom genotyping array that efficiently captures genomic variation in East and South Asian populations and offers a powerful approach for precision medicine research initiatives in Asian populations. The approach presented can readily be applied to one or more populations. *For Research Use Only. Not for use in diagnostic procedures*.

1511T

Pathogenic variants that alter protein code often disrupt splicing. *R. Soemedi[1,2], K. Cygan[1,2], C. Rhine[2], J. Wang[2], C. Bulacan[3], J. Yang[4], P. Bayrak-Toydemir[5], J. McDonald[5], W. Fairbrother[1,2,6].* 1) Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI; 3) Department of Computer Engineering, Brown University, Providence, RI; 4) Department of Computer Science, Brown University, Providence, RI; 5) Department of Pathology, University of Utah, School of Medicine, Salt Lake City, UT; 6) 6Hassenfeld Child Health Innovation Institute of Brown University, Providence, RI.

The evolution of genome sequencing technology has led to an unprecedented speed and accuracy for the detection human genetic variation. However, the lack of tools to identify causative variants from genome sequencing data has greatly limit the promise of precision medicine. Previous studies suggest that one-third of disease-associated alleles alter splicing. We analyzed 4,964 published disease-causing exonic mutations reported in Human Gene Mutation Database (HGMD) using a novel massively parallel splicing assay (MaPSy) to identify exonic variants that alter pre-mRNA splicing. Our results showed an 81% concordance rate with splicing in patient tissue (Soemedi et al., 2017). Approximately 10% of exonic mutations altered splicing, mostly by disrupting multiple stages of spliceosome assembly. We discovered that alleles causing splicing defects cluster in disease-associated genes (for example, haploinsufficient genes). Interestingly, diseases that are more frequently caused by intronic splice-site mutations also have more exonic splicing mutations (ESM). This likely reflects disease processes that occur through loss-of-function mechanisms. Furthermore, we used the large dataset of ESM generated with MaPSy to build a Random Forest prediction model for exonic splicing mutation. Many exonic properties were found to sensitize an exon to splicing mutations. Finally, we have created a public web-server that enables the visualization of MaPSy results (http://fairbrother.biomed.brown.edu/ESM_browser/). In conclusion, MaPSy facilitates the first large-scale identification and characterization of ESM. The system effectively translates to 5K implementations of basic mutational approaches and can be further adapted to other mutation panels, thus accelerating the efforts to characterize all sequence variation.

1512F

Low-input transcript profiling with enhanced sensitivity using a highly efficient, low-bias and strand-specific RNA-seq library preparation method. *M. Arn, E. Yigit, K. Krishnan, D. Rodriguez, B. Langhorst, T. Shtatland, C. Sumner, D. Munafo, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, 240 County Road, Ipswich, MA.

RNA-seq (RNA sequencing) has undoubtedly become the most popular method for transcriptome analysis. It is widely used, both for gene expression analysis and for other applications, including detection of mutations, gene fusions, alternative splicing, and post-transcriptional modifications. Recent improvements in next generation sequencing technologies (NGS) and sample barcoding strategies have resulted in significant cost reduction. As RNA-seq is increasingly adopted for molecular diagnostics, the quality and reproducibility of library preparation methods become more important. In addition, demand for library preparation methods that produce successful NGS libraries from ultra low input RNA or precious clinical samples is increasing. To overcome these challenges, we have developed a streamlined RNA-seq library preparation method that retains information about which strand of DNA is transcribed, which is now available as NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina®. Strand specificity is important for the correct annotation of genes, identification of antisense transcripts with potential regulatory roles, and correct determination of gene expression levels in the presence of antisense transcripts. Our method can be used across a wide range of input RNA (5-1000 ng total RNA) without any modifications to the protocol, making it a convenient method for RNA-seq library preparation. More importantly, GC content analysis, gene body coverage and gene expression correlation show that these important parameters remain consistent across varying inputs, even though input amounts vary by two orders of magnitude. As a result, our method has increased sensitivity and specificity for low-abundance transcripts, and reduced PCR duplicates and sequence bias, delivering high quality strand-specific data. Our method is compatible with poly A-tail enriched and ribosomal RNA depleted samples as well as RNA extracted from FFPE (formaldehyde fixed paraffin embedded) tissue samples, and is amenable to large-scale library construction and automation.

1513W

High-throughput single-cell genomic profiling with droplet microfluidics. *D. Eastburn, M. Pellegrino, A. Sciambi, S. Treusch, K. Gokhale, J. Jacob, T. Chen, K. Jones, W. Oldham.* Mission Bio, Inc., South San Francisco, CA.

Single cell analysis tools are crucial for studying the role that rare or heterogeneous cells play in disease progression. While there have been recent advances in single-cell RNA-Seq methods, current approaches for genome profiling with single-cell DNA-Seq continue to be limited in throughput and content. The Mission Bio Mosaic Single Cell Platform enables the characterization of genetic variation within cancer cell populations by utilizing a novel approach that barcodes amplified genomic DNA of individual cells confined to microfluidic droplets. The barcodes are used to reassemble the genetic profiles of single-cells from next-generation sequencing data. A key feature of the approach is the "two-step" microfluidic workflow. The workflow first encapsulates individual cells in droplets, then lyses the cells and prepares the genomic DNA for amplification with proteases. Following this lysate preparation step, the proteases are inactivated and droplets containing the genomes of individual cells are paired with molecular barcodes and PCR reagents. We demonstrate that the two-step microfluidic approach is superior to workflows without the two-step process for efficient DNA amplification on up to 5,000 individual cells per run with high coverage uniformity and low allelic dropout of targeted genomic loci. To apply our single-cell sequencing technology to human tumor samples, we developed a targeted panel to simultaneously sequence regions of interest across 19 genes frequently mutated in acute myeloid leukemia (AML) including *TP53*, *DNMT3A*, *FLT3*, *NPM1*, *NRAS*, *IDH1* and *IDH2*. Using this panel, we were able to sensitively assay for changes in clonal architecture within AML samples collected longitudinally at the time of diagnosis, remission and relapse. Our data correlated well with variant allele frequencies (VAFs) obtained from bulk sequencing data; moreover, the single-cell nature of our approach enabled the unambiguous correlation of multiple mutations within subclones not possible with bulk measurements. Collectively, our results show a greater degree of heterogeneity in AML tumor samples than is commonly appreciated with traditional sequencing paradigms. The Mosaic platform enables cost-effective targeted sequencing of SNPs and indels at the single-cell level at unprecedented cell throughput and can be applied to additional disease states with customization of targeted regions of interest.

1514T

Highly efficient transcriptome profiling method for single-cell or low input RNA. *K. Krishnan, Y. Bei, S. Guan, J. Borgaro, T. Shtatland, B. langhorst, T. Evans, E. Dimalanta, N. Nichols.* New England Biolabs, 240 County Road, Ipswich, MA.

RNA sequencing has been widely used to determine gene expression profiles of diverse tissues, cell types, developmental stages and diseases. Most of these studies have been based on population analysis using thousands of cells. Such ensemble studies, however, disguise the potentially significant variations between individual cells. To overcome this limitation, single-cell RNA-seq is emerging as a powerful approach to characterize gene expression heterogeneity within phenotypically identical or complex cell populations and in rare cell types. We have developed a simple and robust single-cell or low input RNA-seq workflow to generate full-length cDNAs that can easily be converted into sequence-ready Illumina libraries with the newly developed NEBNext Ultra II FS DNA Library Preparation Kit (which has an enzyme-based DNA fragmentation method incorporated into the end repair and dA-tailing step). The cDNA can also be used in other workflows to generate libraries for alternate platforms. A variety of single-cell types or low input (2.5 pg-200 ng) Universal Human Reference RNA (UHR) with an RNA spike-in from External RNA Controls Consortium (ERCC) was used to synthesize and amplify full-length cDNAs, in a single tube using a template switching mechanism. The resulting cDNA libraries are converted into Illumina libraries and sequenced on the Illumina Nextseq 500. Paired-end reads are first downsampled, adaptor-trimmed and mapped to the corresponding reference genomes with Hisat 2, and transcriptomes with Salmon. Libraries are assessed across a spectrum of metrics including mapping rate, 5'-3' coverage, %rRNA, transcript level correlation between input amounts, and sensitivity to low abundance transcripts. Using this method, high quality RNA sequencing data was obtained from a variety of single-cell types, and low input RNAs. The data shows strong gene expression correlation (Pearson r>0.9) between inputs that span over five orders of magnitude. In addition, we observe excellent gene body coverage, and high sensitivity as demonstrated by detection of single copy ERCC RNA molecules. We have developed a highly specific and sensitive single-cell/low input RNA-seq method that consistently generates high quality data. It is streamlined and amenable to large-scale library construction as well as high throughput automation. We envision this method facilitating novel discoveries in the area of low-input transcriptome applications.

1515F

Tissue-of-origin identification of forensic material by targeted sequencing of stable regions of RNA using RNase-H2. *L. Parker-Katiraee[1], R. Fleming[2], N. Dunn[2], M. McNeill[1], Y. Wang[1], Y. Bao[1].* 1) Product Development, Integrated DNA Technologies, Redwood City, CA; 2) Forensic Research and Development Programme, Institute of Environmental Science and Research Ltd (ESR), Auckland, New Zealand.

Tissue-of-origin identification of forensic material can provide important information regarding the nature of a crime. However, identification is seldom performed due to the amount of material required, the degraded nature of RNA found at crime scenes, and the high cost. Here, we used a highly multiplexed PCR targeting tissue-specific RNAs that are predicted to be more resistant to degradation (StaRs – Stable regions of RNA), followed by next-generation sequencing. RNase-H2 dependent PCR (rhPCR) chemistry was used to circumvent the formation of primer dimers that occur when primers are multiplexed in low inputs of DNA/cDNA. rhPCR uses 3′ blocked primers containing one RNA base that is used for primer activation by RNase H2. The RNase H2 cleaves at RNA-DNA heteroduplexes when there is a perfect match, which also removes the 3′ blocking modification. Initial workflow and assay performance were evaluated on synthetic templates and universal RNA samples using a multiplexed pool of 71 primer pairs designed to identify the expression of tissue-specific transcripts in circulatory blood, menstrual blood, semen, saliva, vaginal material, or skin. Additional primer pairs were included to identify reference transcripts and spike-in controls. We observed on-target rates as high as 85% in universal RNA samples. Using titrated amounts of cDNA from body fluids, including blood, menstrual fluid, and semen, we tested the sensitivity of the assay and the exclusive expression of the stable markers. Our data suggest that StaRs amplified by rhPCR can successfully predict tissue of origin in representative forensic samples.

1516W

Quality deep-sequencing miRNA data from matched fresh and FFPE cells for expression analysis profiling. *K. Tokarz, T. Vilboux, R. Iyer, D. Thach, M. Ahmad.* Inova Translational Medicine Institute, Falls Church, VA.

Micro-RNAs (miRNA) are small noncoding RNAs that posttranscriptionally regulate gene expression and are currently used as biomarkers in many diseases, including cancer. Formalin is the most widely used fixative and is the standard preservation method for diagnostic pathology. Many molecular analyses including Next Generation Sequencing can show discrepancies in the data depending on the sample processing and storage. In this study, we investigated the variability of the miRNA repertoire from 10 pairs of formalin-fixed paraffin-embedded (FFPE), and their freshly collected counterparts. The samples were prepared on a liquid handler using the HTG EdgeSeq miRNA Whole Transcription Assay. Processed samples underwent library preparation and were sequenced on a MiSeq instrument. The libraries gave quality deep sequencing data for both fresh and FFPE embedded cells. Our laboratory implemented a filtering procedure, removing any probes where the sum of counts for the given probe was less than ten across the samples. This gave us 1498 unique miRNA for the samples after filtering. In order to evaluate the expression of each miRNA probe in proportion to the total counts per million of the run, a $\log_2$(CPM) was applied. Statistical analysis gave Pearson correlation data between the sample types across all cell lines. This study shows that our process produces a high quality data from FFPE samples, which are typically a challenging sample type for next generation sequencing. Because of the volume of collections of surgical specimens and archived tissues from retrospective studies of human diseases, the ability to isolate and sequence miRNA from FFPE samples with high confidence will provide a powerful tool to identify new biomarkers and understand the mechanism of many diseases. We will now begin to look for specific miRNA expression profiles that can be correlated to the clinical information for these known Coriell cell lines in order to explore various genetic disorders.

1517T

Evaluation of NXType™ NGS high-resolution HLA typing kit. *S. Khor[1], Y. Hitomi[1], Y. Okudaira[2], A. Masuya[2], Y. Ozaki[3], M. Ueta[4], K. Nakatani[1], M. Nagato[5], T. Ogawa[5], C. Sotozono[4], S. Kinoshita[4], H. Inoko[2], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Tokyo, Japan; 2) GenoDive Pharma Inc., Honatsugi, Kanagawa, Japan; 3) The Center of Medical Innovation and Translational Research, Graduate School of Medicine, Osaka University, Osaka, Japan; 4) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 5) Research and Development Section, Wakunaga Pharmaceutical Co., Takata, Hiroshima, Japan.

The human MHC region has been shown to be associated with a wide range of diseases.The recent advancement in the next generation sequencing (NGS) technology has definitely helped in increasing the resolution (up to 4-field) of HLA genotyping. However, it remains a challenge in interpreting the results from NGS platform due to biases including systematic sequencing error by sequencing platform, difference in the sensitivity and specificity of HLA calling algorithm, ability to resolve HLA alleles ambiguity by incorporating phasing information, and more importantly incompleteness of the current IMGT database. In order to access the sensitivity and specificity of the NXType™ NGS HLA typing kit (One Lambda), a total of 105 Japanese samples were sequenced for HLA class I genes using the Ion Torrent Personal Genome Machine (PGM) following the NXType protocol. HLA calling was performed using the HLATypeStream v1.1.0.11 with IMGT/HLA databses of 3.27.0.3. Concordance rates were evaluated by comparing HLATypeStream results with Luminex-based HLA typing result for up to 2-field result. With the default setting of HLATypeStream and without manual interpretation of the results, the allelic concordance is more than 99% for HLA-A, HLA-B and HLA-C. Other commercially available HLA calling software and open-access software were also evaluated. Careful inspection of the final results and occasionally manual interpretations of results are necessary especially for rare or novel HLA alleles.

1518F

European Genome-phenome Archive: Finding, applying for, and accessing controlled-access data. *J. Spalding[1], J. Almeida-King[1], A. Carreño-Torres[2], M. Ding[1], P.A Garcia[1], A. Gil[2], J. Kandasamy[1], G. Kerry[1], O.M. Llobet[2], A. Lloret-Villas[2], M. Moldes[2], A. Semf[1], N. Spataro[2], S. de la Torre[2], S. Ur-Rehman[1], A. Navarro[2,3,4], R. Guigó[2], P. Flicek[1], H. Parkinson[1], J. Rambla[2], T. Keane[1].* 1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 2) Centre for Genomic Regulation (CRG), Dr. Aiguader 88, Barcelona 08003, Spain and Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Institute of Evolutionary Biology, Barcelona, Spain; 4) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

The European Genome-phenome Archive (EGA) is a controlled-access archive, which provides secure long-term storage and distribution of human genetic and phenotypic data for biomedical research. Access to datasets is made by application to data access committees in accordance with donor consent agreements. As of May 2017, the EGA stores over 4.2 petabytes of data available for download from over 965,000 unique samples in over 1100 studies. EGA hosts reference data collections and supports data re-use, such as data derived from the UK10K project, WTCCC, Human Induced Pluripotent Stem Cells Initiative (HipSci), RD-Connect and the International Cancer Genome Consortium (ICGC). The EGA is working closely with the Global Alliance for Genomics and Health (GA4GH) to improve the way controlled access data can be found, applied for, and accessed. In collaboration with GA4GH we have been mapping consent codes to the Data Use Ontology (DUO) that classifies data use restrictions to 5 main classes, which can be further refined by referencing secondary classes and other ontologies. The use of DUO aids data discoverability by allowing researchers to determine which data can be used for specific research projects. Ontologies are machine-readable, allowing fully automated data access and providing a basis for supporting interoperability with other relevant GA4GH access projects. To access data at EGA, the EGA downloader enables any file in EGA to be delivered to a destination in an encrypted format and accessed via a FUSE layer. This allows applications to access these files without decryption irrespective of file format, which leads to the possibility of data being hosted in cloud environments securely. EGA is working with the GA4GH to develop a standardised encryption container format allowing random access to common NGS formats while the data remains encrypted at rest. Combining these GA4GH projects with a Beacon allows a researcher to find a dataset by variant, immediately determine the DUO code for the dataset, electronically apply for and be granted access, and visualise the region of interest with the bam file remotely as if the data was open access and hosted on their own computer.

1519W

Machine learning applied to single-molecule electronic DNA mapping for structural variant verification in human genomes. *B. Bready, J.R. Davis, B.S. Grinberg, M.D. Kaiser, J.S. Oliver, J.M. Sage, L. Seward.* Nabsys, Providence, RI.

The importance of structural variation in human disease and the difficulty of detecting structural variants larger than 50 base pairs has led to the development of several long-read sequencing technologies and optical mapping platforms. Frequently, multiple technologies and *ad hoc* methods are required to obtain a consensus regarding the location, size and nature of a structural variant, with no single approach able to reliably bridge the gap of variant sizes between those readily detected using NGS technologies and the largest rearrangements observed with optical mapping. Often, structural variants larger than 10 kilobases are not detected. To address this unmet need, we have developed a new software package, *SV-Verify™*, which utilizes data collected with the Nabsys High Definition Mapping (*HD-Mapping™*) system, to perform hypothesis-based verification of putative deletions. We demonstrate that whole genome maps, constructed from data generated by electronic detection of tagged DNA, hundreds of kilobases in length, can be used effectively to facilitate calling of structural variants ranging in size from 300 base pairs to hundreds of kilobase pairs. *SV-Verify* implements hypothesis-based verification of putative structural variants using supervised machine learning. Machine learning is realized using a set of support vector machines, capable of concurrently testing several thousand independent hypotheses. We describe support vector machine training, utilizing 1089 deletions and 4637 negative controls from a well-characterized human genome. Plots delineating the specificity versus sensitivity of each of the support vector machines will be presented. We subsequently applied the trained classifiers to another human genome, evaluating > 5000 putative deletions, demonstrating high sensitivity and specificity for deletions from 300 base pairs to hundreds of kilobases. Over 78% of deletions called by three or more technologies were confirmed by *SV-Verify*.

1520T

Automation of ultra-high molecular weight DNA isolation and labeling for genome mapping. *P. Lynch, A. Lee, W. Stedman, C. Gadd, D. Bollish, M. Saghbini, M. Borodkin.* Bionano Genomics, San Diego, CA.

Structural variations (SVs) are known to play an important role in human genetic disease. Those SVs include deletions, insertions, and other complex rearrangements of genomic material. They are frequently disrupting or altering genes and regulatory elements. SVs are used as markers in the clinical diagnosis of diseases such as childhood syndromic disorders (Ex.: DiGeorge syndrome) as well as in cancers. Next Generation Mapping (NGM) using the Saphyr™ system from Bionano Genomics provides high-speed, high-through-put genome mapping with unparalleled structural variation detection. These variants are able to be detected by utilizing ultra-high molecular weight DNA. Molecules that range from hundreds of kilobases to over a megabase are linearized and imaged using NanoChannel array technology. These images are then used as input to the Bionano Solve *de novo* genome map assembly and SV calling pipeline.   High quality, ultra-high molecular weight DNA is essential to enable the Saphyr system to identify structural variants. The current Bionano Prep™ protocol is a three day process of harvesting cells, embedding in agarose gel plugs, isolation and purification of DNA, and sequence-specific labeling. As the size of studies focusing on SVs increases, there is a need to develop automated ultra-high molecular weight protocols to enable processing large numbers of samples. Here we present a novel, automated protocol for preparation of ultra-high molecular weight DNA from cell lines and blood. This protocol consists of using a standard liquid handling instrument with custom hardware developed specifically for ultra-high molecular weight DNA isolation and labeling. This method significantly decreases the hands-on time needed while improving genome map contiguity and SV sensitivity.   We present data from multiple replicates of DNA isolation from the same donor's cells prepared using this automated method. Each of the replicates was run on a Saphyr chip to sufficient coverage for haplotype-aware SV detection using the hands-off, adaptive loading protocol that significantly reduces instrument hands-on time and improves throughput. SVs were detected using the latest automated Bionano Solve assembly and SV detection pipeline and were compared between the replicates. We observed very high rates of concordance between replicates, as well as a greater number of SVs detected compared to the current commercially available protocol.

1521F

Application of electronic mapping to metagenomic characterization of a complex microbial sample. *J.M. Sage, B. Bready, A.P. Catalano, J.R. Davis, M.D. Kaiser, J.S. Oliver, L. Seward.* Nabsys, Providence, RI.

The microbiome is a complex and dynamic mixture of microbes inhabiting many surfaces on and within the human body. While increasing lines of evidence point to the crucial role that these microbes play in maintaining health, disruptions in population compositions also correlate with a variety of diseases. The characterization of microbe populations has historically been limited to known genetic markers of strains such as 16S rRNA genes. More recently, shotgun sequencing has been employed in an attempt to provide a more accurate measure of biodiversity. While whole metagenome sequencing is more comprehensive than single marker typing, short read lengths prevent accurate quantitation of related strains within a mixture as well as consistent characterization of large-scale structural variation that can significantly impact pathogenicity. To address these issues, Nabsys has applied the *HD-Mapping*™ platform to the identification and quantitation of microbial strains in the context of a single defined complex sample. *HD-Mapping* employs fully electronic detection of tagged single DNA molecules, hundreds of kilobases in length, at a resolution superior to existing optical mapping approaches. The preservation of long-range information coupled with superior resolution allows for highly sensitive and specific genome detection and structural characterization. Nabsys single-molecule reads derived from this complex mixture were mapped to a database of microbial references resulting in identification of strains within the sample. Titrating the relative abundance of a single bacterial genome in a complex sample resulted in quantitation that was linear over 3 orders of magnitude. In addition, assembly of the mixed reads into independent, map-level contigs allowed for the identification and structural comparison of distinct but related strains contained within the same sample.

1522W

A comprehensive workflow for copy number variation identification from whole-genome sequencing data. *B. Trost[1], Z. Wang[1], S. Walker[1], B. Thiruvahindrapuram[1], J.R. MacDonald[1], W.W.L. Sung[1], S. Pereira[1], J. Whitney[1], A.J.S. Chan[1,2], G. Pellecchia[1], M.S. Reuter[1], S. Lok[1], R.K.C. Yuen[1], C.R. Marshall[1,3], D. Merico[1,4], S.W. Scherer[1,2,5].* 1) The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 3) Genome Diagnostics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Canada; 4) Deep Genomics Inc., Toronto, Canada; 5) McLaughlin Centre, University of Toronto, Toronto, Canada.

Whole-genome sequencing (WGS) is poised to become a first-tier genetic test, replacing less comprehensive technologies such as karyotyping, microarrays, and exome sequencing. A remaining hurdle to broad uptake has been to demonstrate accurate detection of copy number variations (CNVs) from WGS data. Here, we describe a comprehensive workflow for identifying germline CNVs from short-read WGS data using read depth-based computational algorithms. Using multiple datasets to evaluate sensitivity, false discovery rate (FDR), and reproducibility, we empirically determined best practices for several stages of the CNV-detection workflow, including DNA library preparation, sequencing, quality control and assurance, reference mapping, computational CNV identification, and filtering based on technically-difficult genomic regions. Based on our findings, we make several recommendations, including the use of PCR-free DNA library preparation methods, sequencing to an average depth of ~30x, maintaining minimum standards for uniformity of sequencing depth, and using consensus predictions from the CNV detection algorithms ERDS and CNVnator. We compared CNVs detected using our workflow with a comprehensive, multi-technology benchmark of CNVs in the HuRef genome. For the highest-confidence CNVs in this benchmark (those detected by two or more different technologies), deletions >10 kb (the size range typically detectable by high-resolution microarrays) could be identified with 100% sensitivity and 0% FDR (71% and 25%, respectively, for all deletions >1 kb). Unsurprisingly, detecting duplications was less accurate than detecting deletions. We used our workflow to detect rare, genic CNVs in individuals with autism spectrum disorder (ASD), and both deletions and duplications in this subset of CNVs could be detected highly accurately, with 100% of CNVs chosen for validation using orthogonal methods (ranging in size from 3 kb to 33 Mb) successfully confirmed. In the subset of individuals with ASD in which both clinical microarray experiments and WGS were performed, WGS detected all clinically-relevant CNVs that were detected by microarray, as well as additional pathogenic CNVs smaller than 10 kb. Thus, CNVs of clinical relevance can be found from WGS with a detection rate exceeding microarrays, positioning WGS as a single assay for variant detection in clinical genomics.

1523T

Implementation of an automated sample quality control tool in a whole exome sequencing workflow. *A. Regala[1], M. Liu[2], E. Viering[3], J. Petersen[4], M. Beckhaus[4], S. Wolf[4].* 1) Agilent Technologies, Santa Clara, CA; 2) Agilent Technologies, La Jolla, CA; 3) Agilent Technologies, Waldbronn, Germany; 4) DKFZ Genomics and Proteomics Core Facility, High Throughput Sequencing Unit, Heidelberg, Germany.

Sample quality control is of major importance in the Next Generation Sequencing (NGS) workflow as it avoids time- and cost-intensive repetition of experiments. Here the German Cancer Research Center (DKFZ) demonstrates how an automated electrophoresis system can be used to control an exome sequencing workflow to assure the generation of high quality sequencing data. 88 genomic DNA samples from formalin-fixed paraffin-embedded (FFPE) tumor tissues were analyzed. Samples that passed the initial quality assessment, with respect to sample quantity and integrity, were submitted to the library preparation workflow. Depending on the DNA integrity, a suitable DNA fragmentation protocol was used to enable library preparation from degraded DNA. Implementation of quality control steps at various steps throughout the protocol allowed the monitoring of the workflow by checking for library preparation artifacts, as well as controlling sample concentrations and average fragment sizes to increase read number during sequencing. The stringent quality control criteria used by the DKFZ allow for an increase in efficiency during library preparation and ensure reliable sequencing results in high throughput sample analysis.

1524F

Capillary electrophoresis using the SeqStudio™ Genetic Analyzer as part of genome editing workflows. *S. Jackson[1], S. Kumar[2], N. Ravinder[2], K. Varma[1].* 1) Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA; 2) Synthetic Biology, Thermo Fisher Scientific, Carlsbad, CA.

Genome editing is poised to revolutionize our understanding and control of diverse biological systems. Critical to any genome editing workflow is monitoring the efficiency and accuracy of the steps leading to a designed edit. Thermo Fisher Scientific offers complete toolkits for designing, verifying and analyzing the effects of a genome editing experiment. To illustrate this workflow, we used the Thermo Fisher Scientific Cloud-based CRISPR Search and Design tool to identify guide RNAs against human HPRT and RELA loci. Guide RNAs and mRNA encoding Cas9 enzyme (Invitrogen) were introduced into HEK293 cells and incubated for 78 hours. Primers were designed that flanked the targeted sequence. Amplified products from primary pools of edited cells were analyzed for efficiency of editing on SeqStudio Genetic Analyzer by three methods: 1) by fragment analysis using GeneArt™ Genomic Cleavage Detection Kit, 2) by Sanger sequencing of the mixed pool of transformants, and 3) by subcloning into a plasmid (Invitrogen) and Sanger sequencing individual clones. The efficiency measured by these three methods correlated extremely well. Cells from the primary pool were diluted and plated to single cell density; these cultures were then sequenced to confirm that they were monoclonal and to establish sequence of the edit that was introduced. Together, these results demonstrate that the portfolio of tools available from Thermo Fisher Scientific can provide an integrated solution that facilitates genome editing experimentation.

1525W

Chromosomal integration of libraries of full-length mutant genes with associated barcode tags. *X. Jia, V. Chen, M. Maksutova, S. Jayakody, R. Lemons, J. Kitzman.* Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Genome editing provides a powerful toolset to interrogate sequence variants' functional impacts via targeted genomic integration in living cells. Recently, these techniques have been multiplexed to tile short genomic targets, such as single exons, for replacement with pools of mutations. However, current approaches have been limited by the inefficiency of homology-directed repair as well as the necessity of individually targeting disperse sequences such as the multiple exons comprising typical mammalian genes. To overcome this technical bottleneck, we have devised Blunt Library Integration of Tagged Sequences (BLITS), which leverages Cas9-mediated homology-independent targeted integration (HITI) to knock-in complex libraries of full-length sequences into a genomic "safe harbor" locus. A key aspect of BLITS is that each integrated gene variant is pre-tagged with a unique molecular identifier (UMI), with the variant-to-UMI pairing established by sequencing prior to genomic integration. Consequently, during downstream pooled functional selections, it is only necessary to track the identity and abundances of these short UMIs rather than the full sequence of each mutant gene. As a proof of principle, we integrated an allelic series of human cDNAs into the AAVS1 genomic safe-harbor locus. We demonstrate that BLITS can achieve site-specific knock-in of ~3 kbp fragments at an efficiency of ~10% in cultured human cells, compared to ~1% for homology dependent repair of typically much shorter regions. In pilot functional selection experiments, we show that tracking integrated UMI counts reduces sequencing costs, and we model the corresponding improvements in accuracy and reproducibility in functional effect measurements. We believe BLITS will be a useful tool for high-throughput structure-function studies of complex libraries of gene variants.

1526T

Monitoring guide RNA synthesis for CRISPR/Cas9 genome editing workflow. *M. Liu[1], A. Padmanaban[2].* 1) Agilent Technologies, Inc., La Jolla, CA; 2) Agilent Technologies, Bangalore India.

Bacterial clustered regularly interspaced short palindromic repeats (CRISPR) - associated protein 9 (Cas9) system has increased in popularity as a genome editing tool for targeted mutations, insertions, deletions and gene knock-out studies. CRISPR genome editing has also proved superior to Zinc-finger nucleases (ZFN) and transcription activator- like effector nucleases (TALENs) due to its simplicity and easy programmability. In CRISPR, a guide RNA (gRNA) is used to recognize and introduce a double standard break (DSB) in a target DNA. The DNA repair mechanism triggered after the break is then exploited to introduce an insertion/deletion (indel) in the case of non-homologous end joining (NHEJ), or precise genetic modification if a homology-directed repair (HDR) pathway is triggered. A critical part of the CRISPR-Cas9 tool is the design and synthesis of the gRNA that comprises T7 promoter sequence, target sequence, and protospacer adjacent motifs (PAM). Monitoring the transcription of the gRNA is critical to the workflow to ensure successful gene editing. Here we present an automated electrophoresis approach for monitoring the synthesis, integrity, and functional activity of gRNAs created for a CRISPR/Cas workflow.

1527F

Copy number variants can be detected using next generation sequencing. *R. Drouin[1], J. Morin-Leisk[2], B. Lace[1], C. Bouffard[3], H. Gao[2].* 1) Div Med Genetics, Dept Pediatrics, Laval University, Quebec City, Quebec, Canada; 2) Fulgent Genetics, Temple City, CA; 3) Div Genetics, Dept Pediatrics, University of Sherbrooke, Sherbrooke, Quebec, Canada.

Background. Next generation sequencing (NGS) is frequently used to quickly make a diagnosis. NGS data is primarily used for detection of SNPs and small indels. But diagnostic findings commonly include copy number variants (CNVs), which can also be detected by NGS, which broadens the scope of NGS as a method for identifying diagnostic genetic changes. Objectives. The case studies described here aim to demonstrate the clinical utility of CNV analysis using NGS data. Specifically, the presentation of cases with a multi-exon deletion or duplication, detected by NGS, that appear to explain the clinical phenotype of the individuals tested. Methods. Genomic DNA, extracted from peripheral leukocytes, barcoded and enriched for coding exons, was sequenced using an Illumina HiSeq. Sequencing results were aligned to the GRCh37.p13 (hg19) human reference sequence. Single nucleotide variants and short insertions and deletions were detected by VarScan. CNV analysis was performed using Fulgent-developed computational tools and verified by qPCR. Cases. Here we describe six children in whom a CNV, detected by NGS, was identified and is likely diagnostic. We found: (1) a paternally inherited duplication of all coding exons of the *UBE3A* gene; (2) a *de novo* deletion of all coding exons of the *PRRT2* gene; (3) a maternally inherited 2-exon duplication of the *GRIA3* gene, located on the X-chromosome; (4) a *de novo* whole-gene deletion of the *RBCK1* gene; (5) a *de novo* whole-gene duplication of the *TUSC3* gene; and (6) a paternally inherited deletion of all coding exons of the *TANGO2* gene with a maternally inherited c.218G>C (p.Arg73Pro) variant on the other allele. These CNVs were confirmed by microarrays to be part microdeletion/microduplication that included several genes. Conclusions. These findings support the clinical utility of using NGS data to detect CNVs, since they provide faster and optimal diagnosis and healthcare management. From a clinical, ethical and social perspective, it is beneficial for the treating physicians, their patients, and their patients' families.

1528W

HiSeq X and NovaSeq optimization for clinical applications. *K. Walker[1], Q. Wang[1], D. Kalra[1], H. Doddapaneni[1], J. Hu[1], Z. Momin[1], J. Santibanez[1], N. Osuji[1], C. Nessner[1], G. Metcalf[1], Z. Khan[1], J. Farek[1], W. Salerno[1], Y. Han[1], H. Dinh[1], E. Boerwinkle[1,2], R. Gibbs[1], D.M. Muzny[1].* 1) Baylor College of Medicine HGSC, Houston, TX; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

High-throughput parallel nucleotide sequencing has revolutionized genomic research and reshaped applications in clinical diagnostics. The HiSeq X Ten and NovaSeq platforms further expand these opportunities by providing unprecedented capacity and the opportunity for routine clinical whole genome sequencing (WGS). The Human Genome Sequencing Center (HGSC) at Baylor College of Medicine has more than three years' experience with the HiSeq X Ten platform and recently acquired a NovaSeq 6000. These platforms have now been benchmarked for WGS, relative to our extensive experience with whole exome sequencing (WES), and are now optimized for validation in the HGSC's clinical diagnostic laboratory environment (HGSC-CL). To date, we have analyzed more than 37,000 30X human WGS on the HiSeq X platform in the research environment. These studies have included cohorts for studying common diseases, samples of inherited cancers and families with Mendelian disease. PCR-Free library methods were evaluated and implemented for optimized coverage in GC-rich regions. Metrics for coverage, sample integrity and variant representation were established to ensure consistently high quality WGS. Genome coverage metrics include 90% of genome covered at 20x and 95% at 10x with a minimum of 86×10^9 mapped and aligned bases with ≥Q20. Platform sensitivity and precision at 30X coverage was determined to be 97.8% and 99.6% respectively. Our primary finding is that effective clinical WGS will require ≥40X coverage to provide high consistent exonic coverage of clinically relevant genes similar to that generated in clinical WES. The NovaSeq platform offers to both improve WGS and lower the cost of WES. We acquired a NovaSeq 6000 in March 2017 and preliminary data shows similar performance to HiSeq X in read alignment rates and %coverage at 10X and 20X. Additionally, %read error rates and %duplicate reads are lower for the NovaSeq and suggest the instrument could ultimately produce a higher yield at greater quality. A direct comparison of SNP calls from these two platforms were found to be highly concordant (98.11%) and both have >96% overlap to NIST standards.

1529T

An optimal long-read workflow for human genome sequencing. *J. Lenhart, S. Sandhu, L. Kurihara, V. Makarov, T. Harkins.* Swift Biosciences, Ann Arbor, MI.

The goal of precision medicine is to treat disease with targeted therapies predicated on highly accurate diagnostic sequencing. Short-read sequencing technology offers low error-rates, high throughput capabilities, and low sequencing costs; however, short-read sequencing has a limited ability to map reads in regions beset by low complexity (i.e. repetitive elements) or high polymorphic density while having minimal ability in identifying large insertions or deletions. An alternate long-read sequencing technology, Single Molecule Real-Time (SMRT®) offered by Pacific Biosciences® (PacBio®) provide continuous, low bias DNA reads upwards of 60 kb in length. SMRT technology is ideally suited to produce readily mapped reads in regions that would typically be undetected with short-read technologies. These capabilities ideally position the long-range sequencing PacBio platform to provide enhanced coverage and assembly metrics for whole-genome sequencing and potentially *de novo* assembly of human genomes, bringing the goal of precision medicine within reach. However, standard workflows are laborious and time consuming. Here, we provide a complete workflow with the Swift Accel-NGS® XL Library Kit for the acquisition of ultra-long (> 30 kb) subreads for the PacBio sequencing platforms. The Accel-NGS XL Library Kit for PacBio combines Swift's proprietary repair and ligation chemistry to offer a significant improvement in library conversion efficiency. Furthermore, we highlight several strategies to further optimize average human library subread read length (> 10 kb), from producing a high quality initial shear using the Diagenode Megaruptor2 and selection of ultra-long library molecules with the Sage BluePippin™, to improving integrity of the final library with the provided Swift terminal repair module. These long-read libraries can result in increased coverage of hard-to-sequence regions of the human genome, facilitating detection of both small and large structural variants, improved phasing of distant SNPs, along with precise elucidation of repetitive repeat number. Ultimately, by enabling an improved, complete genomic landscape for precision medicine, better understanding of variants can be realized.

1530F

Digital gene expression of up to 96 targets in 96 samples for cell line screening with nCounter® PlexSet ™. *G.T. Ong, C. Merritt, P. Webster.* Nanostring Technologies, Seattle, WA.

Cell line screening studies require highly efficient protocols for studying many samples in parallel. We have developed a lyse-and-go protocol for digital gene expression profiling of 96-samples by 96-genes in parallel. This protocol is based on nCounter PlexSet Reagents which enable multiplexing of samples and gene targets. The NanoString nCounter Analysis platform uses a novel molecular barcoding technology to measure multiplexed gene expression. The assay counts fluorescent barcodes hybridized to targets to provide precise digital data. This platform is used in a wide variety of research applications. The standard nCounter gene expression assay can be used to process 12 samples per run. We recognize an additional need for a higher sample-throughput assay which would enable researchers to quickly evaluate up to 96 multiplexed targets in samples processed in a 96-well plate-based format. The PlexSet protocol is simple and fast and can be divided into 3 steps: 1) lyse cells, 2) set-up hybridization with crude lysate; 3) initiate automated post-hybridization processing and digital counting. We will present data showing that up to 96 high-quality, digital data points for gene expression in each of 96 samples can be generated from crude cell lysates in an efficient, high-throughput protocol.

1531W

NCBI resources for visualization and analysis of genome assemblies. *V.A. Schneider, P. Meric, N. Bouk, E. Weitz, V. Ananiev, W.R. Anderson, A. Kuznetsov, P.A. Kitts, F. Thibaud-Nissen, T. Murphy, K.D. Pruitt.* IEB, NIH/NCBI, Bethesda, MD.

Ongoing technological developments in sequencing, coupled with a decrease in cost, are contributing to growth in the number of human genome assemblies in public databases. Meanwhile, improvements in assembly algorithms are making the production of de novo diploid assemblies a reality. The availability of these data present exciting new opportunities for analysis and discovery in many areas of basic and clinical genomic research. However, effective and efficient interpretation of these data requires compatible tools and resources that can be used as part of analysis workflows. Genome Data Viewer (GDV), NCBI's new genome browser, provides graphical displays for all RefSeq annotated eukaryotic genome assemblies and is replacing Map Viewer. Built on NCBI's existing browser technology used for the Variation Viewer and 1000 Genomes Browser, GDV makes use of the wealth of sequence-associated data stored at NCBI, including gene annotation, variation, transcript expression, alignments, and GWAS results, to offer users a rich and diverse set of tracks for display and analysis. We will present several recent feature additions to GDV that support improved navigation and analysis of supported genomic assemblies and their annotation, as well as its integration with other NCBI resources such as Assembly, Gene, dbGaP, GEO and BLAST. We will also discuss recent updates that improve the management and display of user-supplied data in GDV. In addition, we will present new features and datasets available in the NCBI Remapping Service, which allows users to move feature annotations between different versions of the same assembly or different assemblies. These include a new stand-alone software package and improvements specifically targeted for remapping of data to diploid assemblies. We will also highlight other NCBI resources, such as the Genome Decoration Page, Bulk Conversion Service, and Clone DB that further support the analysis and communication of genome scale datasets. Together, this collection of public resources will help users take full advantage of the growing collection of genome assemblies.

1532T

Genomic DNA analysis using automated pulsed-field capillary electro-phoresis. *J. Uthe, K. Wong, S. Siembieda.* Advanced Analytical Technologies, Inc. 2450 SE Oak Tree Ct, Ankeny, IA, USA, 50021.

Advances in the third-generation sequencing platforms from small genomes analysis to de-novo assembly of complex genomes impose high quality standards for large genomic DNA samples. Preparation and qualification of genomic DNA (gDNA) samples remains a bottleneck in the workflow involving long read sequencing platforms. Pulsed field gel electrophoresis (PFGE), a classical method for assessment of size and integrity of gDNA preparations, takes hours to days to complete. Thereby an automated pulsed field capillary electrophoresis system (FEMTO *Pulse*™) that can assess the size and integrity of high molecular weight gDNA samples in a fraction of a time compared to PFGE gels was developed. In the FEMTO *Pulse* system, superior resolution, qualitative and quantitative gDNA analysis and automated gDNA quality scoring can be obtained in 70 minutes to 3 hours as compared to the 14 - 20 hours PFGE analysis. In addition, the FEMTO *Pulse* system requires only picograms to femtograms of gDNA for analysis thus conserving precious gDNA samples for downstream applications. With an enhanced sensitivity protocol, as little as 5 pg of gDNA (comparable to gDNA yield from a single cell) can be analyzed on FEMTO *Pulse*. Key features of the FEMTO *Pulse* system will be presented including sizing, resolution, sensitivity and reproducibility of gDNA analysis along with comparison of the system to classical PFGE gels. The FEMTO *Pulse* system streamlines the quality assessment of gDNA samples and offers innovation to the third generation sequencing workflow by eliminating the long wait times associated with the agarose-based PFGE gels, reducing labor and decreasing costs by quickly excluding unqualified samples.

1533F

Targeted sequencing using a long-read sequencing technology. *I. Mc-Laughlin, P. Baybayan, R. Hall, J. Harting.* Pacific Biosciences of California, Inc., Menlo Park, CA.

Target enrichment by PCR in massively parallel sequencing is a common approach to studying variation in the human genome and its relation to genetic diseases whether somatic or germline in origin. Complications in genotype calling in clinically relevant genes with nearby homologous pseudogenes, such as *CYP2D6* and *CYP2D7*, or typing assignments due to incomplete haplotype phasing, such as in HLA typing, can arise in methods using short amplicon sequencing. Pacific Biosciences' ability to generate highly accurate, long sequencing reads makes it well suited for studying genomic variation in a wide range of amplicon sizes from 250 bp to > 10 kb. With the recent introduction of the Sequel System, high-throughput analysis of multiplexed samples opens the door to cost-effective sample processing in the diagnostic setting. Circular Consensus Sequencing (CCS) for amplicons ranging in length from 250 bp to 3 kb and Long Amplicon Analysis (LAA) for amplicons greater than 3 kb in length represent two workflows that can generate the high-quality data (> QV50) required to study a range of genetic variation from minor variants and insertion/deletion polymorphisms to SNP phasing on the Sequel System. CCS generates high intra-molecule consensus accuracy by enabling the polymerase to sequence the same molecule multiple times. This approach allows for reliable identification of minor variants with allelic frequencies as low as 1% with 6,000 CCS reads. Unlike single molecule CCS, LAA generates a highly accurate consensus from multiple molecules using clustering. LAA can be used to generate phased, full-length consensus sequences for multiple genes in a single sequencing run. A benefit of LAA in HLA typing, as an example, is imputation-free analysis due to long haplotype phasing of 3-5 kb amplicons. Barcoding options to enable sample multiplexing are available for both CCS and LAA workflows. A PCR primer-based approach is available with the flexibility to redesign primers with up to 384 barcodes. A second approach using up to 96 barcoded adapters is available for customers not able to update existing PCR primer designs. Data analysis for either approach is straightforward and simple through SMRT Analysis software consisting of demultiplexing, reference alignment, and consensus sequence generation. Here we present workflows for single-molecule sequencing of amplicons ranging from 250 bp to > 10 kb in length for human genetic analysis.

1534W

Irreversible inactivation of ribonuclease A on a surface by UV LED. *T.L. Thompson, J. Pasquantonio, G. Eliason.* Phoseon Technology, Hillsboro, OR.

RNases are pervasive and when contaminating critical surfaces can present problems for laboratories engaged in RNA sequencing and analysis. RNases, specifically RNase A, are difficult to irreversibly inactivate in the absence of long-term heat treatment (1) or harsh chemicals. Such methods may be incompatible with common laboratory materials or complicate subsequent biochemical reactions. UV light at 254 nm is known to act on RNase A via the effect of the aromatic amino acids proximal to disulfide bonds (2). While UV methods for inactivation of RNase A are well known for mercury arc lamp sources, fast, complete, irreversible inactivation has been difficult to achieve. We report here the first use of high irradiance UV LED sources for enzyme inactivation. In this study, an RNase A solution (0.02 U/ml) was used to contaminate glass slides and allowed to dry. The samples were then exposed to 275 nm UV LED light at irradiances of 316 mW/cm^2 or 635.4 mW/cm^2. UV LED exposure times of the dried samples ranged from 0 to 30 minutes. Samples were resuspended from the slides in nuclease free water. RNase A activity of the resuspended samples was assayed using RNaseAlert Kit (IDT) for each illumination condition. Reactions were monitored for 2 hours at 1 minute intervals. Exposure to 316 mW/cm^2 275 nm source decreased V_{max} from 163 RFU/min at a 3 minute exposure and 9 RFU/min for a 15 minute exposure to reach RNase A negative control levels at a 30 minute exposure. Illumination with the 635.4 mW/cm^2 275 nm source reduced V_{max} to match the RNase A negative control at an exposure time of less than 5 minutes. Total radiant exposure to reach RNase A activity equivalent to the negative controls was 190 J/cm^2 for the 635.4 mW/cm^2 source and 568.8 J/cm^2 for the 316 mW/cm^2 source. These results show that both radiant flux (power) and radiant fluence (dose) contribute to rapid inactivation of the enzyme. We conclude that high power 275 nm UV LED irradiation represents a novel, fast and convenient irreversible inactivation method for RNases on surfaces. 1) Zale and Klibanov, Why does Ribonuclease Irreversibly Inactivate at High Temperatures. *Biochem* **25**:5432-5444 (1986). 2) T.K. Rathinasamy, L.G. Aurgenstein, Photochemical yields in Ribonuclease and Oxidized Glutathione Irradiated at Different Wavelengths in the Ultraviolet. *Biophys. J.* **8**:1275-1287 (1968).

1535T

iMETHYL: An integrative database of whole genome DNA methylation analysis combined with whole genome and whole transcriptome analyses of over 100 Japanese individuals. *S. Komaki[1], Y. Shiwa[1], R. Furukawa[1], T. Hachiya[1], H. Ohmomo[1], R. Otomo[1], M. Satoh[1], J. Hitomi[2], K. Sobue[2], M. Sasaki[2], A. Shimizu[1].* 1) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan; 2) Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan.

Purpose DNA methylation (DNAm), a key regulator of transcription, is being explored in epigenetic studies. Particularly, epigenome-wide association studies have identified numerous associations of DNAm signatures with phenotypes in humans. However, public databases of DNAm in humans that can be used as references for epigenetic studies are limited. Here, we introduce a DNAm database, iMETHYL, which integrates whole DNA methylation analysis, whole genome analysis, and whole transcriptome analysis. Methods Blood samples were collected from 209 apparently healthy participants (mean age ± SD, 56.9 ± 13.4; %Female, 51.7) of the Tohoku Medical Megabank Project, Japan. From the whole blood samples, CD4+ T-lymphocytes (CD4T, N = 102), CD14+/CD16− monocytes (N = 102), and CD16+ neutrophils (N = 106) were isolated. DNA and RNA extracted from each cell type were subjected to whole genome bisulfite sequencing and whole transcriptome sequencing, respectively. Simultaneously, whole genome sequencing was performed using DNA extracted from the whole blood samples. Based on the DNAm status, gene expression levels, and genotypes obtained in this study, expression Quantitative Trait Locus (eQTL), DNA methylation Quantitative Trait Locus (mQTL), and expression Quantitative Trait Methylation (eQTM) analyses were performed. Results We obtained a comprehensive dataset of DNAm (~24,000,000 autosomal CpG sites) and gene expression (~16,000 genes) profiles of CD4T, monocytes, and neutrophils and single nucleotide variants (~4,000,000 SNVs). We summarized these genetic data and developed a public database named iMETHYL (available at: http://imethyl.iwate-megabank.org). On the iMETHYL browser, regions of interest can be specified using gene symbols, dbSNP rsIDs, and genomic positions; the mean, variation, and distribution of DNAm and gene expression levels and frequency of genotypes of the regions can be referred to with respect to cell-type specificity. Association analyses of eQTL, mQTL, and eQTM are still ongoing. Conclusion We established a public DNAm database, iMETHYL, based on specific cell types with wide genome coverage. Although improvement of iMETHYL is still under progress, this integrative database should be a useful reference for multi-omics analyses. Associations between DNAm signatures, gene expression profiles, and genotypes will appear on iMETHYL when the analyses are finished, which will make iMETHYL a more powerful database.

1536F

CNVs from whole exome sequencing and array: First comparative study in a Brazilian clinical cohort. *E.A. Zanardo[1], G.M. Novo-Filho[1], M.M. Montenegro[1], A.S. Brasil[1], S.N. Chehimi[1], F.A.R. Madia[1], F. Montenegro[2], A.M. Nascimento[1], L.S.A. Costa[2], F.B. Piazzon[2], Y.G. Oliveira[1], A.T. Dias[1], T.V.M.M. Costa[1], J.G. Damasceno[1], F. Kok[2], C.A. Kim[3], L.D. Kulikowski[1].* 1) Laboratório de Citogenõmica, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil; 2) Mendelics Análises Genômica, São Paulo, Brazil; 3) Unidade de Genética, Departamento de Pediatria, Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

   Over the last years, the identification of CNVs from arrays and whole-exome sequencing (WES) data has become a common practice for research. Consequently, the demand for more efficient clinical methods has increased. In this study, we evaluated 32 Brazilian patients with multiple congenital malformations and developmental delay by WES technique for CNVs assessment using the software CoNIFER (Copy Number Inference From Exome Reads). We used the array for comparison and confirmation of the results and verification of possibility of the other CNVs. We identified several different genomic alterations in all patients by the array and/or WES analysis. The array identified 466 CNVs, among then 190 alterations were deletions, 244 duplications and 32 loss of heterozygosity. The exome identified 1322 CNVs, wherein 890 were deletions and 432 were duplications. We were able to verify that this technique identified a greater amount (839/1322) of small CNVs with a maximum size of 50 kb. To measure the consistency of exome with array, we determined the overlap of CNVs identified between them. Among all alterations we found 60 CNVs consistent in both techniques evaluated, corresponding to 35 deletions and 25 duplications. So the rates of proportion of consistent changes regarding to inconsistent changes in the same technique were 12.9% for array and 6.5% for exome. The CNV identification through WES data has some shortcomings and still needs better standardization. However, a high amount of changes detected by exome were small changes and often impossible to be identified in array coverage. So there is value in identifying CNVs using WES data but extra caution needs to be taken into consideration due to the high false positive rate and additional confirmation is highly recommended in this analysis.

1537W

Hitting the target: An analysis of noncoding alterations as captured by panels and diagnostic exome sequencing at a commercial lab. *B. Schoenfeld, M. Towne, H. Laduca, P. Reineke, H. Vuong, S. Tang.* Ambry Genetics, Aliso Viejo, CA.

   Multigene panels offer comprehensive targeted analysis of genes of interest and can be useful for diagnosis in clinically-variable or genetically-heterogeneous disorders. Overtime, diagnostic exome sequencing (DES) is more frequently ordered as a 1st or 2nd tier test. We assessed the technical sensitivity of DES for variants in noncoding regions to test the belief that panels offer far better intronic and promoter coverage than DES can. Using *in silico* analysis, we explored the coverage and sensitivity of DES for detecting VUS, VLP or MUT located in promoter or intronic regions that were reported on cancer and cardiac panels offered by Ambry Genetics. Corresponding nucleotide positions were interrogated in data from 100 randomly-selected DES samples to determine the mean sequence coverage at each position. In total, 1675 noncoding alterations in 108 genes were reported. Most alterations were classified as VUS (1172; 70.0%), and the remaining were split between MUT (217; 13.0%) and VLP (286; 17.0%). In sum, 1207 (72.1%) had mean coverage $\geq$20X on DES, with only 4.0% (20/503) of VLP/MUT alterations with mean coverage <20X. Most noncoding variants with insufficient coverage on DES (428/468; 91.5%) were located in the promoter regions of 3 genes. The remaining 40 alterations without sufficient coverage on DES were located in intronic regions and had a median coverage of 16.6X. Reported alterations in canonical splice sites (+/-1 or +/-2) were more often reported as VLP or MUT than those located 3, 4 or 5 nucleotides from the intron-exon junction; however the percent of loci with $\geq$20X coverage was high (96.9%) for all positions +/-5 of the intron-exon junction. Among 14 variants (13 MUT/ VLP and 1 VUS) reported beyond the 5th intronic nucleotide in panels, 10 had coverage $\geq$20X on DES (71.4%). In summary, DES has adequate coverage of variants within +/-5, but less sufficient coverage of deep-intronic and promoter region alterations. The vast majority (96.0%) of noncoding MUT/VLP by panels are technically detectable by DES. The number of reported noncoding VUSs on panels greatly outweighs those reported on DES, likely causing the belief that DES has inferior coverage. Intronic alterations beyond +/-2 are mainly uninformative. DES reporting filters to focus on variants at canonical splice sites and established VLP/MUT in deep-intronic regions provide the most informative results while reducing VUSs and lessening the burden of uncertain results for clinicians.

1538T

Evaluation of PCR followed by high-resolution melting analysis and synthetic constructs on four commercial thermocyclers. *K. Dawkins[1], L. Chambers[1], Z. Kabir[1], V. Mukherjee[1], K. Nunnally-Rivera[1], J. Moll[1], E. Sagini[1], L. Papazisi[1], L. Jiang[1], J. Huuskonen[1], I. Knight[1,2].* 1) Canon U.S. Life Sciences, Inc., Rockville, MD; 2) Canon BioMedical, Rockville, MD.

   IntroductionPCR followed by high-resolution melting (HRM) analysis is an accurate and effective method for genotyping, and many manufacturers have incorporated HRM technology into their thermocyclers. The Novallele™ genotyping assays were designed and optimized using one thermocycler and the accompanying software. The purpose of this study was to demonstrate that the Novallele genotyping assays are robust and are easily transferred between different thermocyclers without compromising assay reliability.MethodsFour HRM-capable thermocyclers were used to compare the performance of six unlabeled-probe assays and six small-amplicon assays. PCR robustness and assay reliability were tested using the software provided with each thermocycler. Synthetic double-stranded DNA constructs were designed for each assay to provide materials to genotype each mutation, and this approach facilitated genotyping target and non-target mutations when genomic DNA was not available.ResultsAll twelve assays accurately identified their target genetic variation on each of the four thermocyclers. PCR amplification and melting temperature difference between the wild-type and variant samples was consistent from instrument to instrument. In addition, reliable genotyping was achieved for these assays using a wide concentration range of approximately 6–60ng. The synthetic genotyping controls showed similar wild-type to homozygous variant melting temperature difference, validating the concept of using these as a relevant control. For two thermocyclers, four assays required minor optimization of either a change to the annealing temperature by ≤1.5°C or an extra five amplification cycles. On another instrument, five assays required minor optimization of increasing the annealing, denaturing, and extension hold times in order to maintain an equivalent degree of performance.ConclusionOur study indicates that carefully designed and optimized genotyping assays can perform robustly on different HRM-enabled thermocyclers with modest, if any, adjustment to the cycling conditions. In addition, the use of controls resulted in comparable genotyping results to genomic DNA, validating their use if genomic DNA is not available. The Novallele genotyping assays are for Research Use Only. Not for use in diagnostic procedures.

1539F

Merck-custom Affymetrix Axiom™ genotyping array performance evaluation. *J. Xu[1], B. Scherer[2], W. Hussain[2], P. Shaw[3].* 1) Department of Clinical Biostatistics, Merck & Co., Inc., North Wales, PA; 2) Nelson Biological Laboratories, Piscataway, NJ, USA; 3) Department of Clinical Pharmacogenomics, Merck & Co., Inc. West Point, PA.

   The Merck-custom Affymetrix Axiom™ Genotyping Array was designed based on the back bone of the Axiom™ UK Biobank Genotyping Array and modified to improve imputation of common variants to over 10 million genetic variants in broader ethnic populations often found in global clinical (clin. PGx) studies. There were four experiments conducted to evaluate the array performance: 1) 96 genetic samples genotyped in two facilities to evaluate the within plate and between plate reproducibility; 2) 270 genetic samples called by different genotype calling algorithms (SNP specific prior against generic prior) to evaluate algorithm performance, the concordance rate against 1000 Genome Project were higher in at least 90K assayed SNPs using SNP specific prior compared to generic prior; 3) 7,090 samples called in single batch and multiple batches to evaluate batch effect, the average per SNP concordance rate is 99.98% between single and multiple batches, which translates into only 1.2 / 7,090 discordant genotype calls per SNP on average; and 4) 2 positive control samples were compared with 1000 Genome Project in each plate to assess the plate by plate variability over time. Call rate and concordance rate within the same sample data were compared. The concordance rates against 1000 Genome Project are on average 98.7% and 98.1% for the two samples, respectively, and consistent across two year period.

1540W

Streamlined, efficient, and uniform molecular inversion probe capture for targeted sequencing. *E. Boyden, E. Tsung, D. Maganzini, G. Porreca, M. Umbarger.* Good Start Genetics Inc, Cambridge, MA.

Molecular inversion probe (MIP) capture represents an attractive targeted sequencing sample preparation technology as it enables simultaneous library construction and highly specific target enrichment. However, the adoption of MIP technology for targeted sequencing applications has been relatively limited in part due to probe design challenges and the perceived complexity of the MIP capture workflow. We have optimized MIP sequence design and the associated experimental workflow to minimize these barriers to adoption. More specifically, we implemented a new algorithm to design probes that tile along both strands of the genome in a strand-alternating fashion. The resulting design mitigates the risk of allele dropout, a common pitfall of other amplicon-based targeting strategies, while simultaneously improving capture/sequencing depth uniformity and per-locus capture efficiency, and reducing sequencing costs and sample input requirements. Across 5 MIP panels targeting a collection of polymorphic loci and coding regions of genes throughout the human genome, we observed that greater than 98% of the target bases were within 20% of the mean coverage, and the fold-80 base penalties (the amount of additional sequencing required to raise the 20th percentile coverage to the mean) ranged from 1.3 to 2. Achieving this level of sequence coverage uniformity generally did not require empirical probe concentration optimization or 'rebalancing', a strategy often used with other target capture technologies to improve uniformity. We also employed unique molecular identifiers (UMIs) to enable PCR duplicate removal and low frequency somatic variant detection. Using UMI-barcoded MIPs, we derived median per-locus capture efficiency estimates of ~12% and observed that allele fractions of 5% can be detected reliably with 10 ng of DNA input. Finally, we have developed a streamlined MIP capture protocol that can be completed in less than 12 hours. This single-tube protocol is easily automated, utilizes dual indexed barcode primers to minimize the risk of index swapping, and can be scaled to run hundreds to thousands of samples simultaneously. Collectively, our highly optimized target capture technology represents an attractive option for highly efficient and uniform, streamlined, and scalable targeted sequencing.

1541T

Use of a molecular inversion probe (MIP) system for the detection of copy number variants. *K. Jefferson, R. Bannen, M. Brockman, S. Suresh, T. Richmond, D. Burgess.* Roche Sequencing Solutions, Madison, WI.

Copy number variants (CNVs) are large (>50bp) indels which often have been associated with inherited disease, disease susceptibility, neurological disorders, and cancer. We evaluated the MIP-based HEAT-Seq target enrichment method for use in rapid and sensitive SNV detection. Here, we report the validation and application of the HEAT-Seq target enrichment assay for the efficient detection of CNVs at known disease loci. The HEAT-Seq target enrichment system is a targeted NGS method based on optimized, multiplexed molecular inversion probes (MIPs). It utilizes an advanced version of MIPs, with unique molecular identifiers (UID or UMI) incorporated into probes in order to identify PCR duplicates, improve sensitivity, and reduce false positive calls. The probes in the HEAT-Seq Choice panel used in this study were selected from an empirically tested database containing the entire protein-coding exome plus non-coding medically relevant regions of the human genome. This panel targets CNVs in Charcot-Marie-Tooth type 1A, Smith-Magenis Syndrome, and Burkitt's Lymphoma, including the protein-coding regions of the disease genes and nearby non-disease genes. The targets were chosen to overlap affected regions in which CNVs previously have been characterized in known samples. The HEAT-Seq target enrichment system was originally developed as a rapid single-nucleotide variant detection discovery and validation platform with a simple workflow. Using CNVkit software for genome wide copy number detection and visualization, we were able to detect CNVs with multiple duplications and deletions. The results of this study show the HEAT-Seq target enrichment system can also be used for CNV detection, broadening the applications for this technology in the study of inherited disease, neurology, and cancer.

1542F

ELOVL5, an epigenetic biomarker, predisposes for the risk of type 2 diabetes mellitus with inflammation. *H. Lee, J. Hwang, M. Go, H. Jang, B. Kim.* NIH Korea, Cheongju, South Korea.

Genome-wide DNA methylation has been implicated in human complex diseases. Here, we aimed to unravel epigenetic biomarkers for type 2 diabetes (T2D) underlying obesogenic environments. In a blood-based DNA methylation analysis of 11 monozygotic twins (MZTW) discordant for T2D, we discovered genetically independent candidate methylation sites. We replicate the T2D-association at a novel CpG-signal in the ELOVL5 gene specific to T2D-discordant MZTW. ELOVL5 gene is significantly associated with differential expression in adipose tissues from unrelated T2D patients and in human pancreatic islets. Further functional studies in cellular models demonstrated that ELOVL5 is increased in compliance with the gene expression of increased endoplasmic reticulum (ER) stress and hepatic gluconeogenesis. Also, ELOVL5 expression is increased with IL6, ELOVL6, SCD, and pIkBa in arachidonic acid treated SK-Hep I human liver cell line. These data suggest that increase of ELOVL5 methylation is associated with increased ER stress mediated diabetic risk. An integrative multi-target tissue analysis might reveal new functional and clinical implications related to epigenetic control in human T2D.

1543W

Star strand miR-192 (miR-192*) as an overlooked metabolic regulator in pre-diabetic liver. *K.K. Miu, X. Zhang, TY. Ha, HH. Cheung, B. Feng, WY. Chan.* School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong.

Circulating levels of miR-192 in serum was reported to rise in both the plasma of glucose-intolerant mice and in patients under pre-diabetic state. Nonetheless, its presence in the serum had been suggested as a sensitive marker of liver damage. Interestingly, the star strand miR-192 (miR-192*) was predicted to target multiple genes related to fatty acid (FA) metabolism and inflammation. In fact, a recent study demonstrated that miR-192* expression in human visceral adipose tissue negatively-correlates with serum fat content. We discovered that miR-192* rather than miR-192 was lower in the liver of *db/db* mice when compared to heterozygous lean control. In addition, we confirmed expression changes in some of the predicted targets of miR-192* in their liver. Furthermore, key genes related to hepatic anabolism of FA or for its transport were upregulated correspondingly, potentially resulted from loss of the miRNA, hence aggravating diabetic conditions. Indeed, we noted expression changes of those genes upon transfection of miR-192* mimics to isolated primary mouse hepatocytes. Likewise, miRNA mimics reduced pro-inflammatory interleukin-6 (IL6) signalling in the hepatic cell line HepG2, with IL6R, IL6ST, STAT3 predicted as common target genes for both mature miRNAs. To demonstrate the unique activity of miR-192* in regulating liver metabolism, we constructed an artificial shRNA stem-loop construct for its overexpression. Active uptake of saturated FA palmitic acid-BSA complexes was prevented in LO2 human liver cell clones stably transfected with the construct. Its physiological action was checked by replenishment via hydrodynamic injection of the construct *in vivo*. Finally, liver cell lines with altered expression of miR-192* were constructed for subsequent transcriptomic analysis to confirm its target gene ontology. In retrospect, our investigation suggested that the regulatory role of miR-192* in liver metabolism might be overlooked. Therefore, we forecast that the loss of miR-192* in liver aggravated diabetic conditions with dysregulated FA metabolism.

1544T

Physical interaction in human beta-cells between islet eQTL sites and target gene promoters at loci associated with type 2 diabetes risk and glycaemic traits. *J. Torres[1,2], V. Nylander[2], D. Downes[3], M. Thurner[1,2], C. Burrows[2], J. Hughes[3], M. McCarthy[1,2,4], A. Gloyn[1,2,4], WIGWAM Consortium.* 1) Wellcome Trust Centre for Human Genetics, The University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes Endocrinology & Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 3) Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom; 4) NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

   Determining the effector transcripts at loci associated with type 2 diabetes (T2D) risk and glycemic traits remains a challenge as the majority of GWAS signals map to non-coding regions. Given that pancreatic islet enhancers are strongly enriched for GWAS variants, we previously mapped expression quantitative trait loci (eQTLs) in primary human islets and identified eQTLs co-localizing with T2D-associated variants. To determine if islet eQTLs physically interact with the promoters of their target genes, we used next-generation (NG) Capture-C to generate high-resolution interaction maps in EndoC-βH1 cells - a glucose-responsive, human β-cell line. Capture probes were designed to target promoters of effector genes corresponding to 9 cis eQTLs that co-localize with variants associated with either T2D or glycemic traits (linkage disequilibrium r2 > 0.8). Significantly interacting restriction fragments were determined based on a rank-product test implemented in peakC. We found that all 9 islet cis eQTLs co-localizing with T2D and glycemic trait-associated SNPs map to the interaction domains of their target gene promoters in EndoC-βH1 cells. These include variants at the *GCK*, *ADCY5*, *AP3S2*, and *WARS* loci and two independently mapped eQTLs at the *DGKB* locus. At the *CAMK1D* and *STARD10* loci, eQTLs map to significantly interacting contact sites that indicate stable enhancer-promoter looping events. Moreover, these contact sites map to open chromatin regions in both islets and EndoC-βH1 cells and correspond to signatures of active gene regulation in human islets. These maps support gene regulatory mechanisms involving T2D and glycemic trait-associated variants facilitated by enhancer-promoter interactions and show how integrating eQTL and epigenomic annotations with 3D interaction maps can advance resolution of causal genes at loci associated with human disease.

1545F

The molecular basis of increased diabetes susceptibility in carriers of the PGC1α (Ser482) risk allele. *R. Vandnebeek[1,2], N. Perwaiz-Khan[1,2], A. Besse-Patin[1,3], M. Leveille[1,3], JL. Estall[1,2,3].* 1) Experimental Medicine, McGill University, Montreal, Quebec, Canada; 2) Institut de recherches cliniques de Montreal, Montreal, Quebec, Canada; 3) Cellular and Molecular Medicine, University of Montreal, Montreal, Quebec, Canada.

   A Gly482Ser SNP within the coding region of human peroxisome proliferator activated receptor (PPAR)-γ coactivator 1 (PGC-1α) is associated with type 2 diabetes susceptibility (T2D) and other metabolic diseases. It has been shown that the Serine expressing variant increases the relative risk of T2D; however, the functional significance of this single nucleotide polymorphism (SNP) has not yet been characterized. By cycloheximide chase, we determined that a serine residue at amino acid position 482 decreases protein stability, compared to the 482Gly counterpart. Replacing this residue with either an alanine residue (phosphorylation-null) or an aspartate residue (phosphomimetic) led to protein stabilization and destabilization, respectively, when expressed in INS-1 pancreatic β-cells. Upon addition of a proteosomal/lysosomal inhibitor (MG132), we observed equal PGC-1α protein levels, suggesting that this modification leads to proteosomal degradation. These data provide evidence that there is a phosphorylation event at the 482Ser site and that this post-translational modification regulates stability of the protein. This regulation was also observed when constructs were expressed in HepG2 human hepatocytes, suggesting this mechanism is not beta cell specific. Furthermore, the 482Gly variant had enhanced expression of target genes involved in metabolic pathways compared to the Serine containing variant. Using mice that are homozygous for the glycine variant (generated using CRISPR/Cas9) offers us many benefits: it allows us to study the effect of endogenous Gly and Ser variants on gene/protein stability and, consequently, on co-activator activity. Taken together, our data suggests that links between the Gly482Ser SNP variants and metabolic disease may be due to reduced protein stability, and thus, coactivator function. Understanding how the variant affects PGC-1α gene and protein stability, as well as function, may help to design novel therapeutics to treat those who possess this SNP and are at heightened risk of diabetes complications.

1546W

Cytosine methylation predicts renal function decline in American Indians. *R. Hanson[1], C. Qiu[2], G. Fufaa[1], S. Kobes[1], C. Gluck[2], J. Huang[3], Y. Chen[3], R. Nelson[1], W. Knowler[1], K. Susztak[2].* 1) Diabetes Epidemiology Clin Res, NIDDK, Phoenix, AZ; 2) Dept of Medicine, Dept Genetics, University of Pennsylvania Perelman School of Med, Philadelphia, PA; 3) Dept of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Med, Philadelphia, PA.

Diabetic nephropathy accounts for most of the excess mortality experienced by individuals with diabetes, but understanding of the molecular mechanisms by which nephropathy develops is limited. In the present study, we analyzed the association of cytosine methylation levels with decline in estimated glomerular filtration rate (eGFR) in 181 diabetic Pima Indians who participated in a longitudinal study. Methylation levels at 397,063 CpG sites were measured with the Infinium HumanMethylation450 Beadchip in peripheral blood specimens. Rate of decline in eGFR (over an average of 6 years) was assessed using a mixed model and analyzed for association with methylation levels. Methylation levels at 77 sites showed significant association with eGFR decline after correction for multiple comparisons (false discovery rate < 0.05). Using a split-sample validation procedure we found that a model including methylation level at two probes (cg25799291 and cg22253401) improved prediction of eGFR decline in addition to currently used predictors (baseline eGFR, albumin:creatinine ratio [ACR]) with percent of variance explained improving from 23.1% to 35.9% ($P=3.6\times10^{-6}$). A risk score based on these two sites significantly predicted development of end-stage renal disease (80 cases, hazard ratio=1.46 per SD, P=0.01). The 3490 probes at which methylation significantly associated with eGFR decline (P<0.05 and effect size in the top 1%) were significantly more likely than non-significant probes to localize to gene promoter regions (31% versus 23%) and to localize to validated transcription binding sites in ENCODE kidney cell lines (81% versus 73%). This indicates that many of the associations are likely functional in the human kidney, as the associated sites localize to regulatory regions of the genome. Weighted gene co-network regulatory analysis showed that these sites were enriched for genes with metabolic functions and apoptosis (r=-0.15, P=0.04). Four of the probes that associated with eGFR decline in blood samples also showed significant association with fibrosis in microdissected human kidney tissue in a separate cohort (n=94). These analyses suggest that cytosine methylation levels may be useful biomarkers of disease progression in diabetic nephropathy, and they strongly suggest that epigenetic factors contribute to development of diabetic kidney disease.

1547T

Dissecting features of epigenetic variants underlying cardiometabolic risk using full-resolution epigenome maps. *F. Allum[1,2], X. Shao[1,2], A. Hedman[3], J. Vijay[1,2], T. Kwan[1,2], F. Guénard[4], M.M. Simon[1,2], M. Lathrop[1,2], A. Tchernof[5], M. McCarthy[6,7,8], P. Deloukas[9,10], T. Pastinen[1,2], M.C. Vohl[4], E. Grundberg[1,2].* 1) Department of Human Genetics, McGill University , Montreal, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 3) Cardiovascular Medicine Unit, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden; 4) Institute of Nutrition and Functional Foods (INAF), Université Laval, Québec, Canada; 5) Québec Heart and Lung Institute, Université Laval, Québec, Canada; 6) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 7) Oxford Centre for Diabetes, University of Oxford, Churchill Hospital, Oxford, United Kingdom; 8) Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom; 9) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 10) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom.

Complex diseases such as obesity are caused by joint action of predisposing genetic and environmental factors. Genome-wide studies (GWAS) have mined multiple metabolic disease-linked variants enriched in non-coding regions, making translation into biological knowledge challenging. Linking variants to function by cellular traits such as CpG methylation in disease-relevant tissues can aid interpretation. Most large-scale epigenome-wide studies (EWAS) use targeted arrays (e.g. 450K array) biased to static promoters, preventing full-resolution mapping of functional CpG methylation. As such, we implemented MCC-Seq to simultaneously profile methylomes and genotypes in active regulatory regions at single-base resolution. Using an adipose MCC-Seq panel, we linked adipose methylation (1.3M CpGs; N=199; ULaval) to cardiometabolic risk through an EWAS of circulating lipid levels, identifying 12,987 TG-CpGs and 6,015 HDL-CpGs (5% FDR) enriched in adipose enhancers but depleted in promoters, validating the targeting of dynamic regions. Further specific mapping of lipid-CpGs revealed key positional differences at these elements with variant "hotspots" enriched for adipogenesis transcription factor binding motifs. Using heritability values (h^2) from the MuTHER cohort (N~650; 450K array), genetic vs. environmental contribution dissection of our lipid-CpGs showed enrichment among heritable sites (h^2>0.7; >1.6-fold; p<2.2e-16). The genomic density of nearby lipid-CpGs (+/-250bp) found heritable sites clustering in regions (>49%) and environmentally driven sites occurring more as singletons (>72%). Of note, clustered heritable lipid-CpGs showed marked enrichment within adipose-unique promoters that themselves are enriched near genes with adipose-specific expression (log2-fold>2; p<0.05). By linking genetic variation to adipose methylation (metQTL) we highlight key examples of heritable genetically driven lipid-CpGs (metQTL p<1.0e-8) where driving SNPs overlap Global Lipids Consortium GWAS findings. Finally, we validated lipid-CpGs in the MuTHER cohort (450K array), replicating 1,283/4,192 (31%) TG-CpGs and 99/254 (39%) HDL-CpGs within +/-250bp. We show the fine-mapping potential of MCC-Seq by noting that >80% of replicated MuTHER lipid-CpGs exhibit stronger lipid associations at nearby non-overlapping sites. We demonstrate the value of full-resolution epigenetic profiling in functional regions to dissect and map variants to better understand the etiology of complex traits.

1548F

An epigenome-wide association study in whole blood of measures of adiposity among Ghanaians: The RODAM Study. *K.A.C. Meeks[1], P. Henneman[2], A. Venema[2], T. Burr[3], C. Galbete[4], I. Danquah[4,5], M.B. Schulze[4], F.P. Mockenhaupt[6], E. Owusu-Dabo[7], C.N. Rotimi[8], J. Addo[9], L. Smeeth[9], S. Bahendeka[10], J. Spranger[11,12,13], M.M.A.M. Mannens[2], M.H. Zafarmand[1], C. Agyemang[1], A. Adeyemo[8].* 1) Department of Public Health, Academic Medical Center - University of Amsterdam, Amsterdam, The Netherlands; 2) Department of Clinical Genetics, Academic Medical Center - University of Amsterdam, Amsterdam, The Netherlands; 3) Source BioScience, Nottingham, United Kingdom; 4) Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Germany; 5) Institute for Social Medicine, Epidemiology and Health Economics, Charité – Universitaetsmedizin Berlin, Berlin, Germany; 6) Institute of Tropical Medicine and International Health, Charité – University Medicine Berlin, Berlin, Germany; 7) Kumasi Centre for collaborative Research, College of Health Sciences, KNUST, Kumasi, Ghana; 8) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, United States; 9) Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom; 10) MKPGMS - Uganda Martyrs University, Kampala, Uganda; 11) Department of Endocrinology and Metabolism, Charité–University Medicine Berlin, Berlin, Germany; 12) German Centre for Cardiovascular Research (DZHK), Berlin, Germany; 13) Center for Cardiovascular Research (CCR), Charité–University Medicine Berlin, Berlin, Germany.

Background: Epigenome-wide association studies (EWAS) have identified DNA methylation loci involved in adiposity. However, EWAS on adiposity in sub-Saharan Africans are lacking despite the high burden of adiposity among African populations. We undertook an EWAS for anthropometric measures of adiposity among Ghanaians aiming to identify DNA methylation loci that are significantly associated. Methods: The Illumina 450k DNA methylation array was used to profile DNA methylation in whole blood samples of 547 Ghanaians from the Research on Obesity and Diabetes among African Migrants (RODAM) study. Differentially Methylated Positions (DMPs) and Differentially Methylation Regions (DMRs) were identified for BMI and obesity (BMI ≥30 kg/m[2]), as well as for waist circumference (WC) and abdominal obesity (WC ≥102cm in men, ≥88cm in women). All analyses were adjusted for age, sex, blood cell distribution estimates, technical covariates, site of data collection and population stratification. We also did a replication study of previously reported EWAS loci for anthropometric measures in other populations. Results: We identified 18 DMPs for BMI and 23 for WC. For obesity and abdominal obesity, we identified three and one DMP, respectively. Fourteen DMPs overlapped between BMI and WC. DMP *cg00574958* annotated to gene *CPT1A* was the only DMP associated with all outcomes analysed, attributing to 6.1% and 5.6% of variance in obesity and abdominal obesity, respectively. DMPs *cg07839457* (*NLRC5*) and *cg20399616* (*BCAT1*) were significantly associated with BMI, obesity and with WC and had not been reported by previous EWAS on adiposity. Conclusions: This first EWAS for adiposity in Africans, identified three epigenome-wide significant loci (*CPT1A*, *NLRC5* and *BCAT1*) for both general adiposity and abdominal adiposity. The findings are a first step in understanding the role of DNA methylation in adiposity among sub-Saharan Africans. Studies on other sub-Saharan African populations as well as translational studies are needed to determine the role of these DNA methylation variants in the high burden of adiposity among sub-Saharan Africans.

1549W

Serum bilirubin levels, UGT1A1 gene expression, and risk for ulcerative colitis. *C.J. Gallagher[1], K. Schieffer[2], S. Bruffy[3], R. Rauscher[2], L. Harris[2], S. Deiling[2], G. Yochum[2], W. Koltun[2].* 1) Biochemistry, Lincoln University, Lincoln University, PA; 2) Department of Surgery, The Pennsylvania State University, College of Medicine, Hershey, PA, USA; 3) Center for Clinical and Translational Research, Virginia Commonwealth University, Richmond, VA, USA.

Chronic inflammation associated with inflammatory bowel disease (IBD) results in increased oxidative stress that damages the colonic microenvironment. A low level of serum bilirubin, an endogenous antioxidant, has been associated with increased risk for Crohn's disease (CD), but no study has tested another common IBD ulcerative colitis (UC). Bilirubin is metabolized in the liver by uridine glucuronosyltransferase 1A1 (UGT1A1) exclusively. Genetic variants cause functional changes in UGT1A1 which result in hyperbilirubinemia, which can be toxic to tissues if untreated and results in a characteristic jaundiced appearance. Approximately 10% of the Caucasian population is homozygous for the microsatellite polymorphism *UGT1A1*28*, which results in increased total serum bilirubin levels due to reduced transcriptional efficiency of *UGT1A1* and an overall 70% reduction in UGT1A1 enzymatic activity. The aim of this study was to examine whether bilirubin levels are associated with the risk for ulcerative colitis (UC). Using the Informatics for Integrating Biology and the Bedside (i2b2), a large case-control population was identified from a single tertiary care center, Penn State Hershey Medical Center (PSU). Similarly, a validation cohort was identified at Virginia Commonwealth University Medical Center. Logistic regression analysis was performed to determine the risk of developing UC with lower concentrations of serum bilirubin. From the PSU cohort, a subset of terminal ileum tissue was obtained at the time of surgical resection to analyze *UGT1A1* gene expression (which encodes the enzyme responsible for bilirubin metabolism). Similar to CD patients, UC patients also demonstrated reduced levels of total serum bilirubin. Upon segregating serum bilirubin levels into quartiles, risk of UC increased with reduced concentrations of serum bilirubin. These results were confirmed in our validation cohort. *UGT1A1* gene expression was up-regulated in the terminal ileum of a subset of UC patients. Lower levels of the antioxidant bilirubin may reduce the capability of UC patients to remove reactive oxygen species leading to an increase in intestinal injury. One potential explanation for these lower bilirubin levels may be up-regulation of *UGT1A1* gene expression, which encodes the only enzyme involved in conjugating bilirubin. Therapeutics that reduce oxidative stress may be beneficial for these patients.

1550T

Methylation-wide association study of sex-specific methylation effects on central adiposity. *A. Justice[1,2], E. Lim[3], M.L. Grove[4], N. Heard-Costa[14,15], W. Guan[7], J. Bressler[4], M. Fornage[5,6], E. Boerwinkle[4,8], Y. Li[9,10], E. Demerath[11], E.A. Whitsel[1], P. Gordon-Larsen[12,13], A. Howard[9,13], L.A. Cupples[3,14], C. Liu[3], K.E. North[1,13].* 1) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Geisinger Health Systems, Department of Biomedical and Translational Informatics, Danville, PA; 3) Department of Biostatistics, Boston University, Boston, MA; 4) Human Genetics Center, Department of Epidemiology, Human Genetics & Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 5) Center for Human Genetics, The University of Texas Health Science Center at Houston, Houston, TX; 6) Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center McGovern Medical School, Houston, TX; 7) Division of Biostatistics, University of Minnesota, Minneapolis, MN; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 10) Department of Genetics, University of North Carolina at Chapel Hill, NC; 11) Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN; 12) Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, NC; 13) Carolina Population Center, University of North Carolina at Chapel Hill, NC; 14) Framingham Heart Study, Framingham, MA; 15) Department of Neurology, Boston University School of Medicine, Boston, MA.

Stark differences exist across sexes in central adiposity and its genetic predisposition. DNA methylation is a key epigenetic mechanism that links genotypes and the environment to central adiposity. While there is clear evidence of sex-specific global and regional methylation patterns, these patterns have yet to be examined within the context of adiposity. Given the sex-specific genetic architecture of central adiposity traits, such studies are of great interest. Thus, our primary aim was to conduct a methylation-wide association study (MWAS) to identify CpGs associated with waist circumference (WC), while accounting for sex as a possible effect moderator. We included 1,669 female and 954 male African Americans of the Atherosclerosis Risk in Communities study with whole blood Illumina 450K methylation array data. CpG beta values were adjusted for batch effects and blood cell type proportions. We used linear regression for CpG association analyses with WC, stratified by sex and adjusted for age, education, body mass index, smoking, center, and 10 principal components calculated from genotype data. The stratified results were used to estimate differences in sex-specific methylation effects (CpGxSex interaction [INT]) and joint main+interaction effects (JOINT). CpGs that reached suggestive significance (P_{JOINT} or P_{INT} <1x10-6) were brought forward for validation of INT in an independent sample, the Framingham Heart Study (2,140 females and 1,847 males, European descent). We identified 8 CpGs in our discovery analyses. In the validation study, cg06500161 in *ABCG1* was nominally significant for INT, but did not meet the significance criteria after multiple test correction. Cg19693031 in *TXNIP* reached Bonferroni corrected significance for INT on WC (P<0.006 [0.05/8 CpGs]). Increased methylation at this site is associated with a decrease in WC and exhibits a two-fold greater effect on WC in females compared to males. Methylation at cg19693031 has been associated with expression of *TXNIP* and several cardiometabolic traits (i.e. type 2 diabetes [T2D], insulin, glucose, and triglycerides) for which central obesity is a known risk factor. *TXNIP* displays the highest expression in subcutaneous adipose tissue, and is involved in glucose homeostasis. Differences in fat distribution between sexes and *TXNIP* methylation effects may help to explain sex differences in downstream diseases, like T2D. These findings underscore the importance of accounting for sex in MWAS studies.

1551F

Body mass index variant protects Mexicans from obesity through long intergenic non-coding RNA on chromosome *20q13.33*. *Y.V. Bhagat[1], A. Ko[1,2], K.M. Garske[1], D.Z. Pan[1,3], M. Alvarez[1], Z. Miao[1,3], K.L. Mohlke[4], M. Laakso[5], T. Tusie-Luna[6,7], C.A. Aguilar-Salinas[7], P. Pajukanta[1,2,3].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) Molecular Biology Institute at UCLA, Los Angeles, USA; 3) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 4) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 5) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 6) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 7) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

The Mexican population has a high susceptibility to obesity-related complex cardio-metabolic disorders. Genome-wide association studies (GWASs) have identified many noncoding variants associated with measures of obesity, such as body mass index (BMI), but the underlying mechanisms through which these variants influence an individual's BMI remain elusive. A previous cross-population GWAS performed in our laboratory showed that variant rs6027281 at locus *20q13.33* is associated with lower BMI in Mexicans. We hypothesized that the protective allele of this lead GWAS variant, or its linkage disequilibrium (LD) proxies, differentially regulates *cis* genes to drive the lean phenotype in Mexicans. Our data-mining for LD proxies and overlap with known epigenetic histone mark ChIP-seq data found that variant rs11086694, in LD (r^2 = 0.87) with the lead GWAS SNP, resides in a transcription start site (TSS), known to be active in monocytes. Our *cis* expression quantitative trait locus (eQTL) analysis on RNA-sequence (RNA-seq) data from subcutaneous adipose biopsies of 335 Finnish men of the Metabolic syndrome in men (MET-SIM) cohort found that the protective allele G increases expression of long intergenic non-coding RNA (lincRNA) *MIR646HG* (p=1.4x10-6), indicating a connection with obesity in a directly relevant tissue type. The *cis*-eQTL finding was also strongly replicated in a previously reported independent whole blood microarray study (n=5,311, p=7.7x10-11), supporting a connection of variant rs11086694 to inflammatory response in obese adipose tissue. To assess its function, we performed electrophoretic mobility shift assays (EMSAs) with monocyte and pre-adipocyte extracts to find that the protective allele G of variant rs11086694 reduces binding of *HOXA9* when compared to the reference allele A. We compared the consequence of this allelic effect on local gene expression by conducting dual luciferase reporter assays in liver cells (HepG2) to show that allele G increases transcription (p=5.6x10-4). Overall, our analysis suggests that protective allele G of variant rs11086694 prevents obesity in Mexicans through disrupting *HOXA9* binding and consequent increased expression of lincRNA *MIR646HG* in *cis*, which may in turn upregulate production of miR-646 that ultimately regulates obesity-related genes in *trans*.

1552W

Transcriptome study of metabolic healthy obesity in African Americans.
A. Gaye[1], A. Doumatey[1], S. Davis[1], C. Rotimi[1], G. Gibbons[1,2]. 1) National Human Genome Research Institute, Bethesda, MD 20892, USA; 2) National Human Genome Research Institute and National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA.

Background Several clinical guidelines have been proposed to distinguish metabolically healthy obesity (MHO) from other subgroups of obesity but the molecular mechanisms that underlie this phenotype remains elusive. In this study of whole blood transcriptome, we searched for specific sets of genes that might play a role in protecting obese subjects from developing metabolic disorders or in delaying their onset. **Methods** MHO was defined as the absence of hypertension, insulin resistance, type 2 diabetes, inflammation and dyslipidemia in obese state. Total RNA samples from 8 MHO and 21 metabolically abnormal obese (MAO) subjects were sequenced. Unbiased transcriptome-wide network analysis was carried out to identify co-expressed genes whose collective expression is associated with MHO status. Machine learning technique (Random Forest) and differential expression analysis were used to respectively predict MHO status and identify differentially expressed genes. **Results** Network analysis identified a group of 23 co-expressed genes (network module), including 17 ribosomal protein genes (RPs), significantly downregulated in MHO compared to MAO individuals. Three pathways were enriched in in the network module: EIF2 signaling, regulation of eIF4 and p70S6K Signaling and mTOR Signaling. The expression of 10 of the RPs collectively predicts MHO with an area under the curve of 0.81. Triglycerides/HDL-C ratio (TRIG/HDL) is the best predictor of expression of genes in the network module. **Conclusion** The results revealed a whole-blood transcriptomic signature of MHO that points to the role of endoplasmic reticulum (ER) and ribosomal stresses, via RPs, in the onset of inflammation in obesity state. The higher TRIG/HDL (an indicator of insulin resistance) observed in the MAO subjects appears to activate ER and related-stress pathways that lead to the inflammatory state. These findings suggest that alleviating ER stress and/or ribosomal stress by downregulating RPs or controlling TRIG/HDL ratio may be effective in maintaining normal metabolic profile in obese individuals.

1553T

A multi-tissue transcriptome association analysis of BMI provides a whole-body view into the impacts of adiposity. *T. Tukiainen[1,2], T. Pers[2], F. Aguet[2], K. Pietiläinen[3], S. Ripatti[4,5].* 1) Institute for Molecular Medicine Finland, Helsinki, Helsinki, Finland; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Obesity Research Unit, Research Programs Unit, University of Helsinki, Helsinki, Finland; 4) Public Health, Faculty of Medicine, University of Helsinki, Finland; 5) Wellcome Trust Sanger Institute, Hinxton, UK.

Obesity is a world-wide epidemic with severe health consequences ranging from increased cardiometabolic risk to higher susceptibility to certain cancers. The mechanisms that link adiposity to these sequelae remain incompletely understood. GWAS findings point to the role central nervous system (CNS) in obesity susceptibility, yet recent work shows adiposity associates with epigenetic changes in tissues from other organ systems. To further elaborate the underlying biology, we investigated the association of body mass index (BMI), a common proxy for adiposity, with gene expression (GE) in 44 tissue types in 449 individuals from GTEx (v6p). We identify significant (5% FDR) BMI-GE-associations in 40% of the studied tissues with up to 2400 BMI-associated genes per tissue (in subcutaneous fat) including, e.g. the upregulation of leptin with increasing BMI in three tissue types. The GE changes are expectedly enriched insulin responsive tissues yet not limited to these. For instance, skin ranks among tissues with most BMI-associated genes. Gene set enrichment analyses implicate diverse underlying processes, many with established connection to adiposity (e.g. inflammation and mitochondrial dysfunction), yet some more debated (e.g. reduced sensory perception in skin and circadian clock in colon) and several that are tissue-dependent (e.g. infertility in testes). While BMI-GWAS discoveries point to CNS, in our analyses the brain shows only a weak signal, suggesting the BMI-associations discovered here are likely primarily consequences of adiposity. In line with recent methylation-based findings in whole blood, our Mendelian randomization analyses demonstrate that a majority of the GE-changes in several tissues are indeed driven by changes in BMI. This observation is further supported by the lack of enrichment of BMI-GWAS genes among the top associations in any of the studied tissues. GWAS genes from other conditions, in particular autoimmune disorders, however, show significant overlap with the BMI-associated GE changes. E.g. genes conferring risk to multiple sclerosis are enriched among genes upregulated with increased BMI in adipose tissue. Analysis of BMI-associated expression changes in dozens of human tissues provides an unprecedentedly extensive view into the impacts of adiposity. By pinpointing the disturbed biologic processes this work provides potential avenues for the treatment of the complications of obesity.

1554F

CLEC16A dysfunction compromises lipophagy and mitophagy and facilitates risk to autoimmunity. *R. Pandey[1], H. Hain[1], M. Bakay[1], B. Strenkowski[1], H. Hakonarson[1,2].* 1) The Center for Applied Genomics, Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Genetic variants in the 16p13 locus harboring the C-type lectin-like domain family 16A (*CLEC16*A) gene are associated with the susceptibility to multiple autoimmune disorders. *CLEC16*'s genomic location next to the suppressor of cytokine signaling 1 (SOCS1) gene and the altered cytokine /chemokine profile in *Clec16a* knockdown mice led us to hypothesize that CLEC16A exerts its effect on a wide variety of immune cells via modulating SOCS expression and regulating cytokine signaling. We hypothesize that the molecular link between CLEC16A, mitophagy, lipophagy, and SOCS1 is abnormal and leads to autoimmune disorders. **Results:** We generated *Clec16a* inducible knockdown mice to examine the mechanism underlying the link between CLEC16A expression alteration and autoimmune susceptibility. Our ubiquitous *Clec16a* KD mice display a complex phenotype similar to that observed in lipodystrophy. CLEC16A Knockdown mice exhibit severe weight loss (loss of visceral and subcutaneous fat) despite eating normal in comparison to control. Adipose tissue is a complex organ, and plays major role in the control of energy, metabolic homeostasis and insulin sensitivity. Weight loss is attributed to increased lipolysis, confirmed by immunoblot analysis of phopshorylated hormone sensitive lipase (HSL). Adipose tissue lysates of CLEC16A knockdown mice also exhibit upregulated HIF-1α in response to stress. In addition, knockdown mice exhibit altered cytokine/chemokine profile, suggesting Clec16a exerts its effect on wide variety of immune cells by modulating SOCS1 expression in response to increased lipolysis and hypoxia and dysregulated mitophagy. **Conclusions:** The derailed functional link exhibited in our ubiquitous *Clec16a* knockdown (KD) mice with dysregulated mitophagy and lipophagy (increased lipolysis) and altered cytokine /chemokine profile may explain the immune dysregulation and the risk of autoimmunity associated with specific *CLEC16A* variants. Our results establish a functional link between *CLEC16A*, mitophagy, and lipophagy in the context of autoimmunity which explains the immune dysregulation in *Clec16a* KD mice and may explain the risk of autoimmunity in humans with specific *CLEC16A* risk variants.

1555W

Promoter capture Hi-C in primary human white adipocytes identifies an interaction hub at a Mexican lipid locus. *K.M. Garske[1], A. Ko[3], D.Z. Pan[1,2], J. Boocock[1], T. Tusie-Luna[4,5], C. Aguilar-Salinas[4], P. Pajukanta[1,2,3].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 2) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 3) Molecular Biology Institute at UCLA, Los Angeles, USA; 4) Instituto Nacional de Ciencias Médicas y Nutrición, Salvador Zubiran, Mexico City, Mexico; 5) Instituto de Investigaciones Biomédicas de la UNAM, Mexico City, Mexico.

The Mexican population is highly predisposed to hypertriglyceridemia (HTG), which is an independent risk factor for coronary artery disease (CAD) and fatty liver. We previously identified a Mexican-specific triglyceride (TG) GWAS signal on chromosome 11, with the associated SNP rs139961185 exhibiting a minor allele frequency of <1% in Europeans, 20% in Mexican cases and 15% in Mexican controls. Here we evaluated the region using promoter capture Hi-C (pCHi-C) in primary human white adipocytes (HWA), a highly metabolically relevant cell type. We determined the linkage disequilibrium (LD) of the region in both the Mexican control GWAS subjects and the 1000 Genomes Mexican Ancestry subjects (MXL) subjects, which revealed 5 SNPs in LD (r²>0.8) with the Mexican lead TG GWAS SNP rs139961185. A highly replicated European cardiometabolic GWAS signal at this locus that was also replicated in our Mexican TG GWAS contains the SNP rs964184, which, along with rs139961185, tags a Mexican-specific, TG-associated haplotype. We performed pCHi-C in the primary HWA, and used the CHiCAGO software to call significant chromatin interactions. We found that the most significant interactions at the locus are with the salt-inducible kinase 3 (*SIK3*) promoter and the lead TG GWAS SNP rs139961185. Furthermore, rs961484 also exhibits a highly significant interaction with the *SIK3* promoter. We analyzed all of the interactions within the Mexican-specific risk haplotype, and found that the *SIK3* promoter interacts with numerous *Hin*dIII fragments along the haplotype, and is in the top 0.1% of interacting captured baits genome-wide. These data indicate that the *SIK3* promoter embodies an interaction hub, which have been shown to be enriched for trait-associated SNPs, as well as play an important role in cell-type-specific gene regulatory programs. SIK3 has previously been implicated in energy homeostasis as a fasting-feeding regulator. Notably, this TG GWAS locus also contains the apolipoprotein gene cluster *APOA1/C3/A4/A5* known to affect serum TG levels. The identification of this chromosomal interaction hub contained within a Mexican-specific TG-associated haplotype on chromosome 11 suggests that the Mexican-specific genetic variation contributes to the HTG in the Mexicans via this regulatory interaction hotspot.

1556T

Transcriptomic profiles of duodenal biopsies in cholesterol gallstone diseases. *E. Riveras[1], L. Azocar[1], TC. Moyano[3], M. Ocares[1], H. Molina[1], D. Romero[2], JC. Roa[2], RA. Gutierrez[3], JF. Miquel[1].* 1) Departamento Gastroenterología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago,Chile; 2) Departamento de Anatomía Patológica, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago,Chile; 3) Departamento de Genética Molecular y Microbiología. Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile.

Cholesterol Gallstone (GSD) is a common multifactorial disease characterized by an aggregation and growth of cholesterol crystals in the gallbladder. 10-20% of the global adult population has GSD. Indeed, the Chilean population has the highest prevalence of GSD in the world, being 17% in men and 30% in women. The pathogenesis and molecular mechanisms involved in GSD are still poorly understood, however, environmental and genetic factors are implicated. Although, a handful of lithogenic genes have been identified in humans, only a few of these have been validated/replicated. Additionally, emerging evidence suggests that the small intestine plays an important role in GSD pathogenesis. To understand the role of the upper small intestine in the molecular mechanism involved in GSD pathogenesis, we performed transcriptomic approach using RNA sequencing to evaluate changes in gene expression profile in GSD. Whole transcriptome data were generated on 6 control and 6 GSD duodenal biopsy samples. Presence or absence GSD was defined by abdominal ultrasonography. Sequencing was performed using the Illumina HiSeq 2500. We used HISAT2 for align to the human genome and Rsubread for the read count per transcript. To capture the genes that are differentially expressed, we used DESeq2 package with a false discovery rate adjusted P-value of 0.05. Interestingly, the differential expression analysis between control and GSD patients revealed genes involved in lipid and cholesterol metabolism, immune response and metabolic pathways. Then, we used real time PCR to validate the changes identified in the sequencing analysis in 15 control and 15 GSD patients. Among the differentially expressed genes in GSD, we found that Zone Occludens 1 (ZO-1) and Occludin (OCLN) were repressed in GSD. These genes are key in the formation and regulation of the tight junction in the intestinal barrier. Therefore, these results suggest that GSD subjects may be have an increase in intestinal permeability. Furthermore, these data suggest that the upper small intestine may be an important organ in GSD pathogenesis. Further studies are needed to confirm and expand these findings.

1557F

Allele specific chromatin signals uncover regulatory mechanisms in autoimmune and B cell related diseases. *M. Cavalli[1], N. Baltzer[5], H. Umer[5], J. Grau[2], I. Lemnian[2], G. Pan[1], O. Wallerman[1], R. Spalinskas[4], P. Sahlen[4], I. Grosse[2,3], J. Komorowski[5], C. Wadelius[1].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany; 3) German Centre for Integrative Biodiversity Research (iDiv) Hall-Jena-Leipzig, Leipzig, Germany; 4) Science for Life Laboratory, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden; 5) Department of Cell and Molecular Biology, Computational Biology and Bioinformatics, Uppsala University, Uppsala, Sweden.

Many signals for predisposition to disease and contribution to common traits are mediated by sequence variation in gene regulatory elements. Identifying such variations is a challenging task that involves three major steps: (i) identify the true, causal variant(s), (ii) identify the transcription factors (TFs) mediating the effect, (iii) identify the target gene(s) whose deregulation lead to the phenotypic effect. ChIP-seq data for TFs and histone modifications provide snapshots of chromatin landscapes that allow the identification of heterozygous SNPs with significant allele specific binding (AS-SNPs) and thus might be indicative of a putative functional role. (i) We searched for AS-SNPs in 7 B cell lines and identified 17293 unique AS-SNPs across the different B cell lines. We found 1199 (7%) variants with minor allele frequencies below 1%, indicating that rare variants may significantly affect B cells phenotypes. In order to add a layer of biological relevance to the allele-specific binding behavior, we searched for AS-SNPs in high linkage disequilibrium to GWAS and eQTL SNPs and found 237 AS-SNPs that were associated to B cells related traits and autoimmune diseases and 714 AS-SNPs to gene expression in B cells. (ii) We identified TF binding sites affected by AS-SNPs as a way to prioritize candidate functional variants using Parsimonious Markov Models that take into account intra-motif dependencies. Using stringent models and cutoffs in scanning the genome we found candidate TFs that bound to AS-SNPs associated to 98 B cells traits and 325 eQTLs. (iii) Identifying the gene(s) regulated by a distant enhancer can be difficult since they can act at distances up to hundreds of kb. Hence, we used targeted chromosome conformation capture (HiCap) to identify high-resolution promoters-enhancers interactions affected by AS-SNPs and putative target genes. We found 60 genes as candidates to mediate GWAS signals and 173 AS-SNPs likely mediating eQTLs. This combined approach led to a collection of putative AS-SNPs supported by multiple pieces of evidence that might be putative drivers of gene regulatory processes in B cells and promising candidates for experimental validation of their functional molecular mechanisms.

1558W

Comprehensive identification of differentially methylated regions associated with systemic sclerosis in dermal fibroblasts from African-American patients. *W.A. da Silveira[1], I. Atanelishvili[2], R.C Wilson[1], E.S Hazard[1], J.C Oates[2], G.S Bogatkevich[2], G. Hardiman[1,2,3], P.S Ramos[2,3].* 1) Center for Genomic Medicine, Medical University of South Carolina, Charleston, SC; 2) Department of Medicine, Medical University of South Carolina, Charleston SC 29425; 3) Department of Public Health Sciences, Medical University of South Carolina, Charleston SC 29425.

Systemic sclerosis (SSc) is a fibrotic autoimmune disorder characterized by cutaneous and visceral fibrosis, immune dysregulation, and vasculopathy. It disproportionately affects African-Americans (AA) who, despite the higher disease severity, are dramatically underrepresented in research. The role of DNA methylation in disease risk remains unclear. This analysis was conducted to comprehensive identify differentially methylated loci associated with SSc in AA. We profiled DNA methylation patterns on cultured dermal fibroblasts isolated from 15 AA SSc cases and 15 AA controls through reduced representation bisulfite sequencing (RRBS). Alignment and methylation calling were performed using Bismarck v0.16.3 and the GRCh37/hg19 reference genome. Data was filtered, normalized, and analyzed with RnBeads v1.6.1. Differential methylation analysis was conducted on CpG, promoter, gene and system level. We generated DNA methylation data for over 5 million CpGs in each sample with at least 40x coverage in promoter and CpG islands. A total of 97 CpG islands and 197 genes and 112 promoters showed significant differential enrichment (using the Combined Score – RNbeads best practices approach) in methylation levels between cases and controls including, among others, the promoter of *SERPINA1* - a protease inhibitor of elastase, plasmin, thrombin and trypsin – and *SERPINA3, SERPINA4, SERPINA5, SERPINA11* and *SERPINA13P*. The SERPIN superfamily is characterized by it function as chaperone proteins and its roles in inflammation and immune function. Enrichment analysis revealed that both hypo and hypermethylated genes and their promotor regions were enriched for immune-related and differentiation gene ontology terms (hypomethylated regions: interleukin-2-mediated signaling pathway, P=0.005 and mesenchymal cell differentiation, P = 3×10^{-4}; hypermethylated regions: type I interferon signaling pathway, P=8×10^{-4} and positive regulation of cell differentiation, P= 1×10^{-3}). In summary, we observed dramatic DNA methylation differences between cases and controls. Interestingly, most of the dysregulated genes can be placed in immune pathways, supporting the role of immune dysregulation in triggering the fibrosis characteristic of SSc. These data support a role for DNA methylation differences in mediating susceptibility to SSc in AA.

1559T

Identification of hypermethylated T- and B-cell receptor signaling molecule clusters in systemic lupus erythematosus (SLE) patients by integrative analysis. *M.B. Guo, T.Y Wang, Y.F. Wang, J.J. Shen, J. Yang, Y.L. Lau, W.L. Yang.* Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, Hong Kong, China.

Molecular mechanism underlying the pathogenesis of systemic lupus erythematosus (SLE) has yet to be found. Epigenetics, as an intermediate mediator between genetics/environmental factors and disease phenotypes, has been investigated frequently for a possible answer. Here we used a dataset of DNA methylation to identify genes with differentially methylated regions (DMR) in patient immune cells, and validated the results in an independent dataset. Many of these DMRs are *in vivo* targets of DNA methyltransferases (DNMTs), and the methylation level of a DMR gene DNMT3B was found negatively correlated with methylation of a number of DMR genes. We also found that many T- and B-cell receptor (TCR, BCR) signaling molecules were hypermethylated in SLE patients. Both important stimulatory and inhibitory genes such as CTLA-4, PD-1, BTLA, CD28, LCK and ZAP70 in T cells, BLK, CD19 and LCK in B cells were found hypermethylated. By integrating ENCODE data from transcription factor (TF) binding experiments, we found a group of TFs having enriched binding peaks in these DMR genes. STATs, IKZFs, IRFs, NF-kB, AP-1, ESR1 and several other immune-related TFs were among those enriched. We also observed a list of TF clusters possibly bound to a subset of DMR genes and regulate their expression synergistically. Lastly, we integrated DNA methylation, TFs, histone modifications and SLE differential gene expression data to investigate the important factors influencing gene expression in SLE, using the machine learning algorithm LASSO. And methylation status was found to be the most important factor driving abnormal gene expression in SLE patients. By regulating the chromatin openness and TF binding, DNA methylation played a pivotal role in the pathogenesis of SLE.

1560F

A functional SNP in 2p14 associated with rheumatoid arthritis by modulating *ACTR2* gene expression via long-range interaction. *B. Lu, M. Yang, X. Chen, H. Chen, S. Yao, Y. Guo.* School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

   Rheumatoid arthritis (RA) is a common autoimmune disease, characterized by chronic, multi arthritis synovial inflammation and the destruction of cartilage and bone, resulting in progressive disability. Genome-wide association studies (GWASs) have revealed the association between variants within the noncoding regions of chromosome 2p14 and RA, but the molecular mechanisms of these variants are unclear. Therefore, we aimed to interpret the functions and mechanisms of noncoding variants in 2p14 by integrating a series of bioinformatics analyses and functional assays. Firstly, RA-associated SNPs (rs1858037, rs934734, rs1876518) residing in the noncoding regions in 2p14 were obtained from GWASs catalog database. By using epigenetic annotation in GM12878 and osteoblasts cell lines from Roadmap Epigenomics Project and ENCODE project, we found that all of the SNPs were enriched in active histone modifications (H3K4me1, H3K4me3 and H3K27ac), and SNP rs1858037 specifically occupied for DNaseI hypersensitive site. We further performed the eQTLs analysis in 3 different kinds of samples including whole blood (Genotype-Tissue Expression project, GTEX), lymphoblastoid cell lines (1000 genome project) and peripheral whole blood of RA patients. Consistent eQTL results were observed for rs1858037 ($p = 3.20 \times 10^{-3}$/ GTEX, $p = 7.57 \times 10^{-3}$/1000 genome project, $p = 1.10 \times 10^{-15}$/RA patients) which affects expression of a distal gene *ACTR2* (about 143kb). Moreover, the SNP rs1858037 physically interacted with *ACTR2* in GM12878 cell line as demonstrated by Hi-C analysis. We further conducted dual-luciferase reporter assay and demonstrated that rs1858037-A allele significantly inhibited *ACTR2* gene expression comparing to the rs1858037-T allele in both human embryonic kidney 293T cells (HEK 293T) and human rheumatoid synoviocyte MH7A cell line (MH7A). In conclusion, our study indicates that noncoding SNP rs1858037 affects RA by modulating *ACTR2* gene expression via long-range chromosome interaction. Our finding may illuminate the dark road of regulatory mechanisms of noncoding variants leading to underlying the development and progression of RA.

1561W

Whole blood DNA methylation signatures of Crohn's disease susceptibility and progression. *H.K. Somineni[1,2], V. Kilaru[3], S. Venkateswaran[1,2], P. Chopra[2], D. Okou[1], G. Gibson[4,5], K.N. Conneely[2], A.K. Smith[3], S. Kugathasan[1,2].* 1) Department of Pediatrics, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University, Atlanta, GA; 3) Department of Gynecology & Obstetrics, Emory University, Atlanta, GA; 4) Department of Biological Sciences, Georgia Institute of Technology, Atlanta, GA; 5) Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA.

   Background: Crohn's Disease (CD) is a life-long condition characterized by ulceration, pain, rectal bleeding, loss of quality of life and a need for bowel surgery. Most CD patients present with an inflammatory phenotype (B1), however, a subgroup rapidly progress to complicated disease behaviours such as stricturing (B2). Recent evidence suggests the involvement of gene-environmental interactions in CD. Epigenetic processes play an important role in mediating environmental influences. Therefore, we set out to examine genome-wide DNA methylation differences associated with CD susceptibility and progression. **Methods:** We utilized a subset of pediatric subjects recruited under the RISK multicenter inception cohort study to elucidate the epigenetic basis of CD. We generated genome-wide DNA methylation data using the Illumina HumanMethylationEPIC 850K array in whole blood DNA samples of 59 controls and 139 newly diagnosed CD cases (B1), of which 71 progressed to B2 at a later time point within 36 months from the day of diagnosis. β-values for each CpG site were modeled as a linear function of disease status with age, gender, and cell-type proportions as covariates to profile disease-related methylation changes. **Results:** 463 differentially methylated positions (DMPs) exhibited significant association with CD susceptibility (FDR<.05). Of these, 349 DMPs were hypermethylated and 114 were hypomethylated in cases compared to controls. We observed strong CD-associated signals on chromosomes 17 and 19. CpGs in a long non-coding RNA, *LOC100996291,* showed the strongest association with CD. In addition to detecting several novel DMPs, some of the previous 450K array findings from whole blood (*TMEM49*/*VMP1*, *SBNO2*, *RPS6KA2*, *ITGB2*, and *TXK* genes) were replicated in our study. Furthermore, many of the DMPs exhibited strong correlation with disease-related plasma C-reactive protein, albumin, and hemoglobin levels. The 463 DMPs mapped to 299 unique genes, which were enriched for immune function-related TNF-alpha, IL-10, and IL-9 signaling pathways. 154 of the 463 CD-susceptibility DMPs also exhibited association with disease progression (B1 vs B2), suggesting common epigenetic processes underlying both susceptibility and progression of CD. **Conclusion:** Our findings suggest that dysregulated DNA methylation is associated with CD susceptibility and progression, and may offer new pathophysiological insights and therapeutic targets to prevent the onset of CD and delay its progression.

1562T

An epigenome-guided approach to causal variant discovery in autoimmune disease. *R.C. Pelikan[1], J.A. Kelly[1], Y. Fu[1], C.A. Lareau[2,3], G.B. Wiley[1], S.B. Glenn[1], M.J. Aryee[2,3], C.G. Montgomery[1], P.M. Gaffney[1].* 1) Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Charlestown, MA; 3) Department of Pathology, Harvard Medical School, Boston, MA.

Translating genetic findings into functional mechanisms is an inefficient, resource-intensive task. Here, we expedite causal variant discovery for autoimmune diseases by identifying genetic variants (epiQTLs) that leave an epigenomic footprint of allelic imbalance in sequencing reads recovered by chromatin immunoprecipitation of enhancer marks (H3K4me1 and H3K-27ac). Using the WASP pipeline to control for reference genome bias and the combined haplotype test to determine statistically robust allelic imbalance, we discovered a total of 6261 epiQTLs across 25 patient-derived EBV-transformed B-cell lines. We overlaid these epiQTLs onto risk haplotypes from 21 autoimmune (AI) diseases and found that 145 reported AI risk haplotypes contained one or more of our epiQTLs. A total of 14 epiQTL SNPs were the same as the AI risk SNP reported in the GWAS catalog, while 180 epiQTLs were proxies of the reported SNP. We demonstrate that epiQTLs located on disease risk haplotypes disproportionately influence gene expression variance over non-epiQTL variants in tight linkage disequilibrium. Following this, we characterized the three-dimensional chromatin interaction topography of our cell lines by HiChIP for H3K27ac-marked active enhancers, locating a majority (78%) of epiQTLs in chromatin loop anchors. Using a generalized linear model, we identified 571 epiQTLs – not associated with autoimmune disease – that modify gene expression from 68 previously established autoimmune disease risk haplotypes through chromatin looping at an FDR ≤ 5%. Expanding this analysis to include both risk haplotypes and independent eQTLs increased the number of epiQTLs with significant (FDR ≤ 5%) modifier effects on gene expression to 1062. Of these gene expression modifying interactions, 717 (68%) increased eQTL-driven gene expression by the epiQTL, while 345 (32%) suppressed eQTL-driven gene expression. Taken together, our data suggest that the epiQTL approach can facilitate the decomposition of risk haplotypes into specific regions that are highly likely to contain functional causal variants. Moreover, epiQTLs not associated with disease haplotypes function to modify gene expression from risk haplotypes and eQTLs through long-range allele-dependent epigenetic mechanisms.

1563F

Treatment-naïve multiple sclerosis cases and controls exhibit differentially methylated regions in CD4+ and CD8+ T cells. *B. Rhead[1,2], I. Brorson[3,4], C. Adams[2], H. Quach[2], D. Quach[2], S.D. Bos[3,4], H.F. Harbo[3,4], R.A. Lea[5], S. Burnard[5], V.E. Maltby[5,6], R.J. Scott[5], J. Lechner-Scott[5,6,7], L.F. Barcellos[1,2].* 1) Computational Biology Graduate Group, University of California, Berkeley, Berkeley, California; 2) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California; 3) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 4) Department of Neurology, Oslo University Hospital, Oslo, Norway; 5) Hunter Medical Research Institute, Newcastle, Australia; 6) School of Medicine and Public Health, University of Newcastle, Newcastle; 7) Department of Neurology, John Hunter Hospital, Newcastle, Australia.

Multiple sclerosis (MS [MIM 126200]) is an immune-mediated disease in which T cells damage the myelin sheath surrounding nerves in the brain and spinal cord, leaving lesions. CD4+ helper T cells predominate in new, active lesions, while CD8+ cytotoxic T cells predominate in chronic lesions. Several genetic and environmental risk factors have been identified for MS, but their mechanisms are not fully understood. DNA methylation is an epigenetic modification that can alter gene expression, and it is influenced by genetic and environmental factors. Methylation is a potential mechanism by which exposures affect disease and may play a role in MS pathogenesis. To test methylation differences between MS cases and controls, we performed an epigenome-wide association study with 96 cases and 102 controls. We generated methylation profiles in both CD4+ and CD8+ sorted T cells for all participants using Infinium BeadChips (396 profiles total). All participants were female and recruited from Norway or Australia. All cases had relapsing-remitting MS and available treatment information. Analyses for CD4+ and CD8+ cells were performed separately. Data were processed using *Minfi*, and *SVA* was used to estimate latent variables that might confound results. Regions of differential methylation were identified with *Bumphunter*, using regression models adjusted for age, recruitment site, and latent variables. Preliminary results showed several regions of differential methylation. Cases were hypomethylated in CD4+ cells in a 22-CpG regulatory region 8 kb downstream of the myelin oligodendrocyte glycoprotein (*MOG*) gene, a primary target antigen in immune-mediated demyelination. Cases were also hypomethylated in both CD4+ and CD8+ cells in a 5-CpG region in a transcription factor binding site (TFBS) ~400 bp upstream of *LCLAT1*, an acetyltransferase involved in lipid synthesis that is upregulated in diet-induced obesity. Cases were hypermethylated in CD8+ cells in an 11-CpG region in a TFBS in the 5' UTR of *SLFN12* and 2 kb downstream of *SLFN13*. Genes in this family are regulated by interferons and inhibit retrovirus protein synthesis. Since obesity and Epstein-Barr virus infection, which can activate human endogenous retrovirus expression, are known risk factors for MS, these results suggest that environmental factors can exert their influence on MS risk through methylation changes.

1564W

Functional annotation of chronic lymphocytic leukemia (CLL) risk loci.
H. Yan, S. Tian, G. Kleinstern, S. Slager. Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55902.

Genome-wide association studies (GWAS) have been extensively used to identify disease-associated single nucleotide polymorphisms (SNPs). Over 90% of the GWAS SNPs were found to reside in the noncoding regions, often distal to the transcription start sites. They are particularly enriched in regulatory regions, such as enhancers, promoters, and insulators, that can be defined epigenetically using chromatin accessibility, histone modification and transcription factor (TF) ChIP-seq data. It is widely accepted that disease risk SNPs exert their regulatory roles, individually or in a combinatorial fashion, by altering local chromatin states, TF binding, and enhancer-promoter looping interactions, ultimately leading to the expression changes of the target genes. Nevertheless, functional interpretation of GWAS SNPs is often limited by multiple correlated SNPs in linkage disequilibrium (LD) with the index SNP, a lack of comprehensive epigenetic data in the pathogenically relevant cell type, or the difficulty to link enhancer-SNPs to the target genes. We and others have identified 41 CLL risk SNPs through GWAS. To gain a deeper understanding of the underlying mechanisms of these SNPs in the CLL etiology, we performed a systematic annotation of the risk loci. We analyzed 215 publicly available epigenetic data sets generated in CLL, B cells from healthy individuals, and lymphoblastoid cell line GM12878, which include chromatin accessibility, Hi-C chromatin interaction data, and ChIP-seq for H3K4me1, H3K4me3, H3K27ac, chromatin regulators and TFs. We extracted an additional 4,376 SNPs from the 1000 Genomes Project that were in LD (R^2 >=0.5 in the EUR ethnic group) with the 41 index SNPs. Intersection with internal H3K4me1, H3K4me3 and H3K27ac ChIP-seq data in CLL and the public epigenetic data revealed that 47% of the SNPs overlapped at least 1 epigenetic mark and 16.6% overlapped marks from over 10 samples. To pinpoint the potential casual SNPs, we identified cases of allele imbalance from chromatin accessibility and ChIP-seq data. We also scanned the SNP regions for matches to five TF motif databases. Our analyses revealed 485 SNPs from 36 risk loci that provide evidence of allele imbalance or altered TF binding motifs, suggesting a role for epigenetic mechanism in CLL pathogenesis. Further analysis with HiC data in GM12878 and RNA-seq data in CLL will illustrate how the causal SNPs contribute to the alteration of transcription program.

1565T

Integrative analysis identified pervasive long-range regulation mediated by super-enhancers shared in multiple autoimmune diseases. *X.F. Chen, B.J. Lu, M. Yang, Y.Y. Duan, Y. Guo.* School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, P. R. China, China.

Genome-wide association studies (GWASs) have identified over one thousand susceptibility SNPs for autoimmune diseases, and most of them are located in noncoding regions. Currently, the greatest challenge is to decipher the functional roles of these flourishing noncoding autoimmune SNPs, and discover the shared genetic mechanisms underlying multiple autoimmune diseases. Therefore, in this study, we aimed to identify the regulatory noncoding SNPs and potential shared pleiotropic immune genes for 20 autoimmune diseases, through an integrative analysis combing score-based epigenetic annotation, expression quantitative trait loci (eQTLs) and genomic chromatin interaction (Hi-C). We firstly extracted 1350 autoimmune GWAS SNPs from GWAS Catelog and ImmunoBase. To evaluate functionality for 25,483 noncoding primary and LD (r^2 > 0.8) autoimmune SNPs, we devised a cell-specific and score-based epigenetic annotation approach, which adopted an accumulative quantitative score system using fold change as weight of particular functional element in 27 blood cells from Roadmap or ENCODE project by comparing 25,483 autoimmune SNPs and 388,210 GWAS SNPs. We further scored for 1,784,477 negative control SNPs (r^2 < 0.1) to build null distribution with score value ranked in 5% chosen as criteria to define functionality. There are totally 1,364 potential functional SNPs as determined by the custom score system. We further identified 330 potential regulatory SNPs regulating expression of 106 target genes through combining Hi-C and eQTLs analysis. 243/330 SNPs are located in super-enhancers in blood cells, with 80 SNPs annotated in over 10 blood cells. We detected a core regulatory network involving 39 SNPs and 3 immune-related target genes (*ERAP1, ERAP2, LNPEP*). All the 39 SNPs are located in super-enhancers. Our results suggest that super-enhancers play important roles in autoimmune diseases. Moreover, most of regulatory SNPs (80.3 %) affect distal target genes, suggesting pervasive long-range regulation in autoimmune diseases. We found that many target genes (59/106) were associated with two or more autoimmune diseases. Of which, 8 genes were associated with over 5 diseases, indicating prevailing shared genetic variants on different autoimmune diseases. In a summary, our study deciphered functional basis for noncoding variants for autoimmune diseases, which is characterized by pervasive long-range regulation mediated by super-enhancers shared in multiple autoimmune diseases.

1566F

Cis regulatory variation determines dynamic HLA-DQB1 allelic expression in response to T cell activation. *M. Gutierrez-Arcelus[123], S. Hannes[123], N. Teslovich[123], Y. Luo[123], H.J. Westra[123], K. Slowikowski[123], D.A. Rao[3], J. Ermann[3], M.B. Brenner[3], S. Raychaudhuri[1234].* 1) Divisions of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, BTM, 60 Fenwood Road, Boston, MA 02115, USA; 4) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

The HLA class II genes, such as DQB1, play a major role in autoimmune disease. While the prevalent view is that coding variation and antigen presentation is driving disease risk in this locus, there is also the possibility that *cis* regulatory HLA effects might play a role. Most of the non-MHC causal variants for autoimmune diseases are in non-coding genomic regions in CD4+ T cells (Trynka *et al.* 2013). Hence, it is essential to define the genetic regulatory effects acting at the specific cell-types and physiological states most relevant for each disease within and outside of the MHC region. Since allelic expression is largely driven by genetic regulatory variation in *cis* (Buil *et al.* 2015), it is a powerful approach to study regulatory effects in multiple cell states without the need of a large sample size. This study brings <u>two main contributions</u>: 1) we developed a **robust way for studying allele specific expression in the HLA locus**, leveraging its high polymorphic and LD nature. 2) We identified **dynamic condition-specific *cis* regulatory effects harboring dramatic changes in DQB1 expression**. We isolated CD4+ memory T cells from the peripheral blood of 24 healthy genotyped individuals. We then queried the transcriptome with RNA-seq after non-antigenic stimulation at 0, 2, 4, 8, 12, 24, 48 and 72 hours. Genome-wide analysis to identify heterozygous SNPs where the **relative contribution of each allele changes across activation states**, revealed most significant signals at the DQB1 gene. Targeted sequencing revealed 14 unique DQB1 alleles across the 24 heterozygous individuals. Strikingly, by mapping RNA-seq reads to the exonic sequences of **DQB1**, 16 out of 24 individuals had **strong and significant time-dependent allelic expression** ($P < 1e-18$, beta > 1). For example, in an individual carrying the alleles 0503 and 0301, the transcript coded by the 0503 allele represented 61% of overall DQB1 counts at 0 hours, and gradually increased to 96% of counts at 72 hours. Furthermore, we observed that the same DQB1 alleles in different individuals had similar expression patterns ($P < 1e-04$). Unsupervised clustering indicates there are **3 main genetic regulatory programs influencing DQB1's dynamic expression in response to T cell activation**. Notably, we did not observe regulatory activity to this extent anywhere else in the genome. This study illustrates the dramatic gene regulatory variation, mirroring the dramatic coding variation, in DQB1.

1567W

Investigation of the effect of an autoimmunity associated SNP in the 6q23 locus on enhancer function using CRISPR/Cas9. *S. Singh[1], T. Katopodi[1], K. Duffus[1], P. Martin[1], G. Orozco[1,2].* 1) Arthritis Research UK Centre for Genetics and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT, UK; 2) NIHR Manchester Musculoskeletal Biomedical Research Unit, Manchester Academic Health Sciences Centre, Central Manchester Foundation Trust, Grafton Street. Manchester M13 9NT, UK.

It is estimated that 90% of disease-associated genetic variants identified by GWAS lie in noncoding regions of the genome and, although their role in disease is not well understood, they accumulate in enhancers. This suggests that disease may be the consequence of an impaired regulatory circuitry between enhancers and their target genes. A previous study characterizing chromatin interactions between all associated genetic regions for four autoimmune diseases and their potential target genes revealed that non-coding variants on 6q23 show significant interactions with the genes *IL20RA*, *IL22RA2*, *IFNGR1* and *TNFAIP3* in T and B cell lines. Further assays suggested that one of the SNPs mapping to a predicted enhancer, rs6927172, was correlated with increased expression of *IL20RA* and increased binding of regulatory proteins. The aim of this work is to functionally validate the ability of this predicted enhancer, containing the autoimmunity associated variant rs6927172, to modulate the expression of its putative target genes using CRISPR/Cas9-based epigenome editing. We used tiled dCas9-KRAB repression constructs to target this enhancer, using seven guide RNAs (gRNAs) spanning 250bp around rs6927172. They were each cloned into the pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-GFP plasmid. Lentivirus was produced by using third generation packaging system and gRNA cloned plasmid into Lenti-X HEK293-T cells. Concentrated virus was used to perform transduction into HEK293-T and Jurkat cells. Transduced cells expressing GFP were sorted by FACS. The effect of repressing the enhancer of interest on the expression of its putative target genes (*IL20RA*, *IL22RA2*, *IFNGR1* and *TNFAIP3*) was measured by using Taqman Gene Expression assays. *TNFAIP3* expression was significantly downregulated in engineered Jurkat cells (P= 0.0003). *IL22RA2* and *IFNGR1* gene expression was also reduced but it was not statistically significant in these cells. *IFNGR1* expression was reduced in engineered HEK293-T cells (P= 0.0004). *IL20RA* gene expression was also reduced in these cells but it was not statistically significant. Our results further support that genetic variation at 6q23 may be involved in autoimmunity through impaired ability to regulate expression of immune genes such as *IL20RA*, *IL22RA2*, *IFNGR1* and *TNFAIP3* . These results demonstrate that functional evaluation of GWAS loci, using CRISPR, can aid the translation of descriptive genetic association results into molecular mechanisms underlying disease.

1568T

An atlas of immune chromatin accessibility and gene expression. *D. Calderon[1], M.L.T. Nguyen[6], J.V. Ribado[2], A. Mezger[2,8], B. Wu[2], A. Kathiria[2], L.A. Criswell[7], W.J. Greenleaf[2,4,9], A. Marson[6,9], J.K. Pritchard[2,3,5].* 1) Program in Biomedical Informatics, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Department of Biology, Stanford University, Stanford, CA; 4) Department of Applied Physics, Stanford University, Stanford, CA; 5) Howard Hughes Medical Institute, Stanford University, Stanford, CA; 6) Department of Microbiology and Immunology, University of California at San Francisco, San Francisco, CA; 7) Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, University of California at San Francisco, San Francisco, CA; 8) Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden; 9) Chan Zuckerberg Biohub, San Francisco, CA.

From GWAS we know of nearly three thousand genetic variants associated with human phenotypes, a third of which are related to autoimmune diseases. Most of these variants fall within noncoding regions expected to regulate chromatin or gene expression, however, in many cases the relevant cell type and condition-specific mechanisms connecting genotype and phenotype remain unknown. Therefore, we developed the first comprehensive immune cell type atlas consisting of ATAC and RNA-seq data from 25 differentiated cell types under resting state and stimulated conditions from peripheral blood of four healthy individuals. On average we identified 85 thousand chromatin accessible regions per cell type, additionally, stimulated cells had 43% more accessible peaks compared to their resting state counterpart (MACS2 identified peaks, $q < 1\%$). Furthermore, identified chromatin regions were significantly enriched for corresponding cell type and stimulation-specific transcription factor motifs. Comparing stimulation to resting state-specific open chromatin regions, we observed a 1.5-fold enrichment increase of genome-wide GWAS summary statistic values for rheumatoid arthritis and found a similar trend for significant fine-mapped GWAS variants from 18 other autoimmune disorders. Finally, as an example of the utility of such a data set, we leverage cell type-specific functional information to fine map the likely causal variant linking the expression of *POLB* to systemic lupus erythematosus. Overall, our atlas of chromatin accessibility and gene expression is a powerful resource for linking genetic variants to autoimmune diseases and understanding mechanisms of gene expression regulation.

1569F

Integrative fine-mapping of genetic loci affecting risk for multiple sclerosis using stimulated primary immune cells. *R. Hinch[1], J. Davies[2], C. Dendrou[2,3], L. Fugger[2], G. McVean[1].* 1) Big Data Institute, Oxford University, Oxford, UK; 2) MRC Weatherall Institute of Molecular Medicine, Oxford University, Oxford, UK; 3) Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, UK.

Genetic variation tells us about common disease susceptibility, however, rarely identifies casual biological pathways. This is because the casual variant is often unknown due to linkage disequilibrium and/or the variant affects gene regulation which is not localised in the gene of interest. Gene regulation is also highly cell-type dependent and varies with environmental changes. An example of an environmental stimulus is when immune cells are presented with pathogens which trigger key pathways to attack the pathogen. We analyse the response of primary immune cells to stimuli (e.g. LPS) to identify the genetic mechanism of the response pathways. Using this approach, we gain power by focussing on a subset of genes involved in immunological response. Taking an integrative approach, we simultaneously analyse changes in RNA expression (RNA-seq), DNA accessibility (ATAC-seq) and histone modifications (ChIP-seq), along with structural data (Hi-C and capture-C) and genetic variation (GWAS and eQTLs) to identify pathways of genetic regulation. Further, using machine-learning techniques, we analyse transcription factor binding sequences to demonstrate how individual SNPs alter gene regulation. We apply our approach to 103 loci associated with multiple sclerosis.

1570W

The role of T cell stimulation intensity in the expression of immune disease genes. *D.A. Glinos[1], B. Soskic[1,2,3], D.M. Sansom[2], G. Trynka[1,3].* 1) Wellcome Trust Sanger Institute, Wellcome Genome Campus, Cambridge, UK; 2) Royal Free Campus, Institute of Immunity and Transplantation, University College London, UK; 3) Open Targets, Wellcome Genome Campus, Cambridge, UK.

Recent studies have recognised the importance of mapping the effects of genetic variants on gene regulation in specific cell types, environmental contexts and stimulation time points. Here, we demonstrate that the titration of the stimulatory signal is a variable that can influence gene expression profiling. Genome wide association studies (GWAS) for common autoimmune diseases have identified a large number of genetic variants located near genes involved in T cell stimulation. T cells need two signals to become activated, one generated by the T cell receptor (TCR) and the second from the CD28 costimulatory molecule. However, there is limited information on the exact pathways underlying these two signals in human CD4+ T cells. Since T cell stimulation is critical for the development of autoimmunity, we tested if GWAS variants play a role in specific TCR and CD28 stimulatory conditions. We stimulated CD4+ T cells *in vitro* with varying intensities of TCR and CD28 signals across seven conditions. Following stimulation, we performed RNA-seq and ATAC-seq on sorted activated cells from four healthy individuals. We observed a distinct gene expression and chromatin accessibility profile for each cell-type stimulation combination. Clusters of genes that were co-regulated through different stimulations were controlled by costimulation in memory cells, while dominated by the TCR in naïve cells. Surprisingly, genes in the DNA replication pathway were regulated by the different stimulatory signals in the two cell types, under CD28 control in memory cells (p-value<10^{-7}) and TCR in naïve cells (p-value<10^{-3}). We confirmed this result by cell proliferation assays. Finally, we used our dataset to test if GWAS variants play a role in specific TCR and CD28 stimulation contexts. We identified that immune disease variants were enriched nearby highly expressed genes specific to cell stimulated conditions. Crohn's disease variants were only enriched in conditions with strong TCR stimulation in the presence of costimulation in memory but not naïve T cells, while celiac disease variants were specifically enriched in memory T cells under low TCR and low CD28. This study provides a primer to prioritise TCR and CD28 stimulation states for the study of disease variants in specific cellular contexts.

1571T

Blood cell type-specific genome-wide DNA methylation analysis of Chinese patients with early-onset systemic lupus erythematosus identifies loss of DNA methylation in genes related to the Type I Interferon pathway. *H.Y.B. Chung[1], K.S. Yeung[1], T. Mok[2], S. Choufani[3], Y.L. Lau[1], R. Weksberg[3].* 1) Pediatrics and Adolescent Medicine, Queen Mary Hospital, Hong Kong, Hong Kong; 2) Department of Biomedical Sciences, The City University of Hong Kong, Hong Kong; 3) Genetics and Genome Biology Program, The Hospital for Sick Children Research Institute, Toronto, Canada.

Background: About 20% of Systemic Lupus Erythematosus (SLE) is diagnosed in children under 18. These children often present with more active and severe disease than adult-onset patients. Since whole blood is comprised of different immune cells, we aimed to identify the cell type-specific DNA methylation signatures of CD4+ T cells, CD8+ T cells, B cells, neutrophils and whole blood in individuals with SLE patients who presented before 18-years of age. Method: Specific immune cells were isolated by positive selection using MACS MicroBeads (Miltenyi Biotec). We compared the DNA methylation profiles of different blood cells for 16 Chinese SLE patients to that of 13 healthy controls using the Illumina HumanMethylationEPIC BeadChip. Data pre-processing was performed to remove cross-reactive probes, probes with SNPs at the target site, and probes from sex chromosomes. For each specific region, Wilcoxon rank-sum test was used for group comparisons, and false discovery rate was used for multiple testing corrections. CpG sites with an adjusted p-value < 0.05 and a mean methylation change > 0.1 were considered to be differentially methylated. Results: After data pre-processing, 775280 probes remained and principal component analysis showed that samples clustered according to specific cell types rather than disease manifestation. Global changes of DNA methylation were not observed among different cell types. The number of differentially methylated CpG sites ranged from 46 to 160 in comparisons of different cell types, with more CpG sites showing hypomethylation than hypermethylation. Principal component analysis showed that samples could be separated according to disease manifestation. Gene ontology analysis was performed for each cell type and revealed that in all the cell types examined, hypomethylated genes identified were overrepresented in the type I interferon pathway. This suggests that the DNA methylation changes in different immune cells of SLE patients target the same biological pathway. As type I interferon has long been believed to be involved in the pathogenesis of SLE, our findings support the importance of an epigenetic mechanism in the dysregulation of type I interferon in SLE pathogenesis. Acknowledge: This work was supported by the General Research Fund (Ref: HKU 765513).

1572F

Integrative methylation/mRNA analyses identified an interferon-inducible-gene interaction network with a key gene PARP9 in rheumatoid arthritis. *S. Lei[1,2], H. Zhu[1,2], F. Deng[1,2], L. Wu[1,2].* 1) Center for Genetic Epidemiolgy and Genomics, Soochow University, Suzhou, Jiangsu, China; 2) Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, School of Public Health, Soochow University, Suzhou, Jiangsu 215123, P.R. China.

To detect the comprehensive genome-wide rheumatoid arthritis (RA) associated methylation sites, and build their downstream regulatory networks, the peripheral blood mononuclear cells of 25 RA patients and 18 healthy controls were used to detect genome-wide methylation status and gene expression by 450K Infinium Methylation arrays and lncRNA/mRNA expression Bead Chip, respectively. The regulatory networks were constructed by applying STRING, CYTOSCAPE and the Causal Inference Test. The differential DNA methylation sites and differential mRNA levels of ten candidate genes were validated in two independent samples. We also generated *PARP9* over-expressed (OE) and knockdown (shRNA) Jurkat cell strains to explore the influence of *PARP9* on cell cycle, cell proliferation and the expression of RA-related inflammatory cytokines. About 1046 differential methylation loci (DML) were identified. The methylation level of 107 DML was significantly correlated with their corresponding gene expression (P<;0.05). An interferon inducible gene interaction network was constructed in which *MX1,IFI44L,DTX3L,PARP9* were the 4 key differential methylated genes. The differential methylation locus/loci of three genes *(BLK,MX1,PARP9)*, especially for three loci of *PARP9*, were verified in independent sample. The differential gene expression level and the up-regulation or down-regulation of four genes, *BLK,IFI44L,MX1,PARP9*, in RA patients were validated in independent samples. In Jurkat cells, the methylation level of two loci of *PARP9* showed significantly negative correlation with gene expression. The proliferation of OE Jurkat cell was significantly faster than that of OE-negative controls (P<;0.05), the proliferation of shRNA Jurkat cell was significantly slower than that of shRNA-negative controls (P<;0.05). Compared with negative control, the percentage of G0/G1 and S cells was lower than that of G2/M cell in OE Jurkat cells, but higher in shRNA Jurkat cells. The expression level of *IL-2* was positively regulated while *IL-4* and IFNγ was negatively regulated by *PARP9*. This study has identified a genome-wide rheumatoid arthritis (RA) associated methylation profiles and constructed their downstream regulatory networks. The interferon-inducible-gene interaction network with *PARP9, MX1, IFI44L* and *DTX3L* as the key genes might play an important role in RA pathology. The abnormal methylation status of specific sites in *PARP9* may mediate pathogenic effects in RA.

1573W

Epigenome-wide association study of autoimmune thyroid disease by next-generation capture sequencing. *T.C. Martin[1], P.J. Campbell[2], S. Mullin[2], X. Shao[3,4], E.M. Lim[2], S.J. Brown[2], T. Pastinen[3,4], F. Dudbridge[5], T.D. Spector[1], J.P. Walsh[2,6], E. Grundberg[3,4], J.T. Bell[1], S.G. Wilson[1,2,6].* 1) Department of Twin Research and Genetic Epidemiology, King's College, SE1 7EH, London, United Kingdom; 2) Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; 3) Department of Human Genetics, McGill University, 740 Docteur-Penfield Avenue, Montreal, Quebec , Canada H3A 0G1; 4) McGill University and Genome Quebec Innovation Centre, 740 Docteur-Penfield Avenue, Montreal, Quebec, Canada H3A 0G1; 5) Department of Health Sciences, University of Leicester, Leicester, LE1 7RH, United Kingdom; 6) School of Medicine and Pharmacology, University of Western Australia, Crawley, Western Australia, Australia.

Thyroid peroxidase antibody (TPOAb) positivity is strongly associated with two major autoimmune thyroid diseases (Graves' and Hashimoto's diseases). Although the heritability of serum TPOAb levels is estimated at 66%, studies have shown that environmental factors such as stress, smoking and infection also contribute and can have opposing effects in different disease subtypes. Epigenetics, particularly DNA methylation, is seen as the link connecting environment and genetics to phenotype and thus could play a major role in thyroid diseases. In this study, we performed epigenome-wide association study (EWAS) to 98 TPOAb-positivity discordant MZ twin pairs (only female Caucasians) using the state-of-art MethylC-Capture Sequencing (MCC-Seq) approach. TPOAb levels were measured by immunoassay. MCC-Seq used a custom autoimmune panel designed for whole-blood interrogation and covering 4.9M CpGs and 1.3M SNPs. CpGs were selected in dynamic regions of the epigenome including DNaseI hypersensitive sites from the ENCODE project and blood-methylome footprints (unmethylated and low methylated regions from different blood cell-types). Analyses employed a likelihood ratio test in a generalised linear mixed model with a beta-binomial distribution for differentially methylated positions (DMPs) and a Bonett-Seire test for differential variability in DNA methylation (DVMPs); smoking, blood cell composition, age, family and analytical batch were included as covariates. We obtained on average 63% on-target rate per sample corresponding to 14x mean genome coverage. After quality control, downstream analyses comparing the level and variance of DNA methylation in the TPOAb discordant MZ pairs highlight 72 DVMPs associated with TPOAb positivity and a single DMP that was contingent on adjustment for smoking. As expected, a considerable reduction in a number of CpG sites significantly associated with TPOAb-positivity was observed when we corrected for blood cell composition and smoking. In summary, this cutting-edge EWAS of TPOAb in a discordant MZ twin study design using MCC-Seq technology identified a large number of DVMPs driven by biological and environmental factors and suggests methylation variability is a key factor in disease susceptibility in autoimmune thyroid disease.

1574T

Using clustering analysis and meQTLs to probe differential methylation in females with multiple sclerosis. *B. Reinstadler[1,2], B. Rhead[2], H. Quach[2], D. Quach[2], C. Schaefer[3], L.F. Barcellos[2].* 1) Department of Applied Mathematics and Theoretical Physics, Cambridge Computational Biology Institute, University of Cambridge, Cambridge, United Kingdom; 2) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California; 3) Kaiser Permanente Division of Research, Oakland, CA.

Multiple sclerosis (MS [MIM 126200]) is an immune-mediated demyelinating disease that exhibits substantial heterogeneity in disease type and severity. Several genetic and environmental risk factors are associated with MS, but their mechanisms are not fully understood. DNA methylation is an epigenetic modification that is influenced by genetic variants and environmental exposures and can alter gene expression; thus it may play a role in MS pathogenesis and in determining disease course. We generated genome-wide DNA methylation profiles of whole blood from 250 MS cases using Infinium MethylationEPIC BeadChips. Cases were non-Hispanic black and non-Hispanic white female never-smokers drawn from Kaiser Permanente Northern California. We used unsupervised clustering and supervised classification methods to identify MS case groups characterized by methylation differences. When we assessed cases using the most variable CpG sites, they clustered into groups based on self-reported race. Clinical differences between black and white MS cases have been observed and support a complex disease etiology; black cases are more likely to experience relapses and greater disability, have greater risk of faster progression, later age of symptom onset and are more likely to have optic nerve and spinal cord involvement (Khan, 2015). We aimed to identify methylation differences between the two patient groups that are biologically relevant to underlying MS etiology and are not due to differences in ancestry. We estimated measures of genetic ancestry and cellular heterogeneity using GLINT (Rahmani, 2017). We then evaluated ~33,000 CpG sites known to be affected by methylation quantitative trait loci (meQTLs; Moen, 2013) and tested for differential methylation in the two groups after adjusting for genetic ancestry, cellular heterogeneity, age at sampling, and batch. We prioritized CpGs with meQTLs that are also known MS risk SNPs. Preliminary results showed that cg20485607 was hypomethylated in black MS cases compared to white cases (p=0.0003). The CpG lies in a transcription factor binding site 50 kb upstream of the MS risk SNP rs483180, in an intron of *PHGDH*, a protein-coding gene expressed in most brain tissues. This finding suggests that differential methylation may contribute to disease differences in black and white MS cases. Full characterization of disease differences in MS cases will have implications for both more effective treatment through precision medicine and prevention.

1575F

A proteomic approach to identify transcription factors that selectively bind to causal polymorphisms in inflammatory bowel disease (IBD). *C.J. Cardinale, H. Fazelinia, L.A. Spruce, S.H. Seeholzer, H. Hakonarson.* Children's Hospital of Philadelphia, Philadelphia, PA.

Genome-wide association studies have identified thousands of trait-associated SNPs, but the causal variant responsible for the association signal as well as the mechanism by which the SNP modifies gene expression is generally unknown. Recently, in inflammatory bowel disease, high density genotyping on the Immunochip with imputation made it possible to fine-map association signals to single variants. Ten of the single variants are non-coding SNPs located in introns and intergenic regions. We used causal variant rs1887428 in the promoter of *JAK2* as a model system for developing a method to identify the mechanism of action of GWAS SNPs by affinity purification-mass spectrometry (AP-MS). Nuclear extracts were prepared from a relevant cell line, in this instance, Jurkat T cells, and an affinity purification was performed using desthiobiotinylated 35-mer dsDNA probes bearing either allele as baits. After competitive elution with biotin, the proteins were pixelated on an SDS-PAGE gel followed by trypsinization and label-free relative quantification on the Q-Exactive HF mass spectrometer. Spectral counts for peptides were determined in MaxQuant and analyzed in the CRAPome resource using fold-change scores and Significance Analysis of INTeractome (SAINTexpress) probabilistic scoring to identify a transcriptional repressor which binds selectively to the protective allele of this SNP. AP-MS is an effective means to identify regulatory factors for causal SNPs and can find mechanisms undetected by other genomic methods.

1576W

Functional characterization of *TNIP1* causal variants associated with Systemic Lupus Erythematosus. *S. Pasula[1], M. Wiley[1], A. Nair[1], Y. Fu[1], J. Gopalakrishnan[1,2], YY. Wu[1], PM. Gaffney[1,2].* 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; 2) Department of Pathology, University of Oklahoma Health Science Center, Oklahoma City, OK, USA.

Background: GWAS have identified more than 80 susceptibility loci associated with systemic lupus erythematosus (SLE) and other autoimmune diseases. Fine mapping of the TNFAIP3 interacting protein 1 (*TNIP1*) locus identified two independent *TNIP1* SLE risk haplotypes that reduced the expression of *TNIP1* mRNA and ABIN1 protein. In this work, we characterized SLE-associated candidate causal SNPs that influence hypomorphic expression of *TNIP1*. **Methods:** We selected eleven SLE-associated *TNIP1* SNPs with low binding scores (RegulomeDB) for EMSAs to evaluate whether SLE risk alleles affect binding of nuclear protein complexes extracted from different immune cells treated with and without PMA/Ionomycin (P/I). Affinity DNA pull-down assays and Western blotting (WB) were performed to identify proteins bound to the rs10036748 probe. Enhancer activity of rs10036748 was measured by luciferase assay. **Results:** Three of the eleven *TNIP1* risk variants exhibited reduced binding of nuclear proteins from Jurkat T cells, EBV transformed B cells and THP-1 monocytoid cells. Five risk variants demonstrated P/I stimulation dependency. Furthermore, three risk variants showed increased binding under resting conditions; one showed increased binding after P/I stimulus. Overall, the SNPs carried on both *TNIP1* SLE risk haplotypes demonstrated complex binding activity; Jurkat T cells exhibited the most activity with 8 of 11 SNPs showing differential binding. The rs10036748 variant demonstrated reduced binding of nuclear proteins to the risk allele in all cell types. Affinity pull-down followed by WB confirmed ENCODE-defined transcription factor binding predictions that the rs10036748 non-risk allele bound early growth response protein 1 (EGR-1), cyclic AMP-responsive element binding protein 1 (CREB-1), and class E basic helix-loop-helix protein 40 (bHLHe40) with relative higher affinity than the risk allele. Interestingly, the risk allele showed increased enhancer activity in HEK293T, Jurkat and EBV transformed B cells. **Conclusion:** Functional analyses of SNPs in *TNIP1* SLE risk haplotypes suggest a complex regulation at *TNIP1* locus. Further, these SNPs exhibit cell type specific, stimulation dependent and allele specific binding of transcriptional protein complexes. Regulatory insights gained will better direct future characterization of individual SLE-associated *TNIP1* variants *in-vivo* to decipher molecular mechanisms and cell states that contribute to SLE pathogenesis.

1577T

Long-range regulation of *IRF5* expression mediated by a functional SNP associated with systemic lupus erythematosus and systemic sclerosis. *H.N. Thynn, X.F. Chen, D.L. Zhu, B.J. Lu, M. Yang, Y. Rong, W.X. Hu, S.S. Dong, T.L. Yang.* School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

Systemic Lupus Erythematosus (SLE) and Systemic Sclerosis (SSc) are autoimmune diseases, which have been shown to share similar genetic backgrounds. Genome-wide association studies (GWASs) have successfully identified numerous genetic variants conferring SLE and SSc disease risk, but the functional mechanisms underlying them are still unknown. In this study, we aimed to decipher the functional mechanisms underlying rs13239597 located in the noncoding region in 7q32.1, which is the most significant SNP ($p = 1 \times 10^{-29}$) associated with both SLE and SSc, through the assimilation of various bioinformatics analyses and functional assays. We found that rs13239597 was located in a putative regulatory element (PRE) with extremely enriched enhancer-related epigenetics markers (H3K27ac, H3K4me3, H3K4me1, Pol II, P300, DNase hypersensitivity) in GM12878 cell line from Roadmap and ENCODE project. The expression quantitative trait locus (eQTL) analysis indicated that different alleles of rs13239597 significantly generated different expression levels of the distal gene *IRF5* (~118kb), which is a key regulator of immune complex formation leading to the risk of SLE and SSc, with the consistent eQTL results in three independent samples including lymphoblastoid cell lines ($p = 1.84 \times 10^{-07}$) from 1000 genome project, whole blood ($p = 9 \times 10^{-04}$) and cell-transformed fibroblasts ($p = 4.3 \times 10^{-03}$) from Genotype-Tissue Expression (GTEx) project and meta-analysis in peripheral blood samples ($p = 6.97 \times 10^{-21}$) from 5,311 individuals. We further revealed that rs13239597 physically interacted with *IRF5* promoter in K562, MCF7, IMR90, CD34 and GM12878 cell lines through Hi-C analysis. Besides, we conducted motif analysis by using multiple databases and found that EVI1, BCL6B, TAL1, FOXD2 and FOXD3 transcription factors specifically bind to rs13239597-A allele. Moreover, we also demonstrated that rs13239597-A allele significantly enhanced *IRF5* gene expression than rs13239597-C allele through the dual-luciferase reporter assay. Therefore, our effort deciphered the functional basis of rs13239597 in noncoding region which affects the risk gene *IRF5* via a long-range chromatin loop formation. Our finding might fulfill the current challenges for complete ascertainment of causal variants towards the understanding in complex genetic architecture of SLE and SSc autoimmune diseases.

1578F

Annotating the regulatory genome of CD4+ T cells: Predicting active in vivo transcription factor binding sites. *T. Amariuta[1,2,3,4], E. Davenport[2,3,4], H.J. Westra[2,3,4], N. Teslovich[2,3], S. Raychaudhuri[1,2,3,4].* 1) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA; 2) Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA; 3) Partners Center for Personalized Genetic Medicine, Boston, MA, 02115, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.

CD4+ T cells regulate the adaptive immune response by mediating inflammation via cytokine signaling. Understanding CD4+ T cell gene regulation may help us better interpret non-coding autoimmune disease risk alleles and the wide range of differentiated CD4+ T cell and memory states. Transcription factor (TF) binding is the hallmark of regulatory elements. However, TF binding cannot be determined by sequence alone. We therefore propose constructing a map of the human CD4+ T cell regulome that defines a per base pair probabilistic score, from 0 to 1, quantifying regulatory binding potential using multiple regulatory markers. We build our regulatory map by training a probabilistic binary classifier on lineage-specific TF binding sites (TFBS) using elastic net logistic regression. To stratify cell-type specific regulation, we utilized data on 9 lineage-associated CD4+ T cell regulators, including TBET (Th1), GATA3 (Th2), STAT3 (Th17), and FOXP3 (Treg). We acquired publicly available ChIP-seq data from isolated CD4+ naïve Treg, Th1, Th2, and Th17 cells of healthy donors, identified TF binding motifs genome-wide and annotated their TF binding state (active/inactive). Finally, we used a total of 65 publicly available datasets of DNase-, ATAC-, FAIRE-, TF and histone modification ChIP-seq, and Capture Hi-C from ENCODE, Roadmap, Blueprint, and NCBI as well as Phastcons conservation. Our classifier predicts TFBS with a mean AUC of 0.73 (95% CI=[0.69, 0.77]; *n*=50 repeated 10-fold cross validation). However, individual TF performance varied: active TFBS in Th1, Th2, and Th17 cells are more accurately predicted (mean AUC=0.86) than those in Tregs (mean AUC=0.59) (*p*<5.8e-4). This may be due to a more specific chromatin landscape in Th1, Th2, and Th17 cells, greater heterogeneity within Tregs, or differences in data quality. We show that our classifier also predicts the presence of active regulatory sites, independent of TF binding. With FANTOM5 data, the classifier predicts that T cell enhancers have significantly more regulatory potential than random genomic sites (mean *p*(bound) for enhancers: 0.56, for random: 0.14; *p*<6.4e-207). The classifier more accurately predicts active enhancers (AUC=0.74) than H3K4me1 and H3K4me3 annotations (AUC=0.68). Our classifier may be used to annotate the noncoding CD4+ T cell regulome for active regulatory elements. Our method already out-performs standard annotations and may be applied to any cell-type where TF ChIP-seq is available.

1579W

Annotations that capture tissue-specific transcription factor binding explain a large fraction of disease heritability. *B. van de Geijn[1], H. Finucane[1,2,4], S. Gazal[1], A. Gusev[1,3], F. Hormozdiari[1], X. Liu[1], P. Loh[1], Y. Reshef[6], A. Price[1,2,5].* 1) Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Dana Farber Cancer Institute, Harvard Medical School, Boston, MA; 4) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA; 5) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 6) Department of Computer Science, Harvard University, Cambridge, Massachusetts,.

It is widely known that regulatory variation plays a major role in complex diseases and that tissue-specific binding of transcription factors (TF) is critical to gene regulation, but TF binding information is not currently available for most (tissue,TF) pairs . Here, we construct tissue-specific TF binding annotations by intersecting sequence-based TF predictions with tissue-specific chromatin data; this addresses both the limitation that identical sequences may be bound or unbound depending on surrounding chromatin context, and the limitation that sequence-based predictions are not tissue-specific. We evaluated different combinations of sequence-based TF predictions and chromatin data from T cells by partitioning the heritability of 5 autoimmune diseases (celiac, Crohn's, lupus, rheumatoid arthritis, and ulcerative colitis; average N = 30,789) using stratified LD score regression (Finucane et al. 2015 Nat Genet) with the baselineLD model (Gazal et al. biorxiv). We determined that 100bp windows around MotifMap sequenced-based TF predictions (Daily et al. BMC Bioinformatics 2011) intersected with a union of promoter and enhancer related chromatin marks (H3K4me1, H3K4me3, H3K27ac, and DHS) performed best, with 0.9% of SNPs explaining 27% of SNP-heritability (29x enrichment; P = 7x10[-12]). This is significant conditional on the baselineLD model, which includes many non-tissue-specific functional and genomic annotations (τ*=2.20±0.38; τ* = proportionate change in per-SNP heritability associated to a 1 s.d. change in the annotation value conditional on other annotations in the model). This represents a large improvement over results obtained using either MotifMap+100bp (2.3x enrichment, τ*=0.17±0.35) or chromatin (9x, τ*=0.83±0.23) alone. We thus focused on MotifMap+100bp intersected with tissue-specific chromatin in our remaining analyses. We applied this approach to 32 diseases and complex traits (average N=74,562), using Roadmap chromatin data from 127 tissues. We observed stronger enrichments using our approach vs. chromatin data alone, including 31x enrichment for HDL cholesterol in the liver (vs. 11x for chromatin alone) and 11x enrichment for schizophrenia in fetal brain (vs. 4x for chromatin). Overall, our results show that intersecting sequence-based TF predictions with chromatin information is a promising approach to pinpointing disease-relevant regulatory regions, which will aid interpretation and fine-mapping of GWAS signals.

1580T

Integrative analysis of transcriptional regulation unveils regulatory modules that stratify SLE transcriptome. *T. Wang[1], Y.F. Wang[1], Y. Zhang[1], J.J. Shen[1], M. Guo[1], J. Yang[1], Y. Lau[1,2], W. Yang[1].* 1) Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 2) The University of Hong Kong-Shenzhen Hospital, Shenzhen, China.

Systemic lupus erythematosus (SLE) is a prototype autoimmune disease with a poorly understood pathogenesis, despite recent advances in unraveling the genetic contributions to SLE. Integrative analysis of multi-omics data is demanded to uncover its detailed mechanisms. In this study, making use of transcriptomic data on SLE, we identified 750 differentially expressed genes (DEGs) in T, B and peripheral blood cells. Using transcription factor (TF) ChIP-seq data from ENCODE project, we have analyzed TFs bound to the regulatory regions of the upregulated DEGs in SLE and identified enriched TFs that may play vital roles in SLE pathogenesis. Intriguingly, further clustering analysis of binding patterns of these enriched TFs unveiled multiple co-regulatory modules among the DEGs. The modular regulatory repertoire not only stratifies the IFN signature genes, but reveals a number of SLE-relevant pathways, including inflammation process, cell cycle regulation, NETosis, and immunoproteasome activity. By integrative analyses of the disease-associated genes (DAGs), the DEGs and relevant TFs, we also identified a hierarchical regulatory system with TFs generally regulated by DAGs and in turn modulating gene expression. Overall, the analytical strategy we used in multi-omics data analysis provided an efficient method for interpretation of large-scale datasets and facilitation of the mechanistic understanding of SLE.

1581F

Altered methylation marks and compromised spermatogenesis in human male infertility. *S.Kumar. Mohanty, Saumya Sarkar, Rajender Singh.* Endocrinology, Central Drug Research Institute, Lucknow, India.

Male Infertility or subfertility is a highly complex and multifactorial disorder that affects 15% of the couples worldwide. Aberrant methylation patterns are frequently associated with germ cell dysfunctions in subfertile and infertile men. We hypothesized that altered epigenetic mechanisms may contribute significantly towards the development of male infertility. To test this, we utilized an epigenome-wide association approach to analyzing DNA methylation levels in subfertile (n=6, Oligozoospermia) and fertile controls (n=6) using Illumina Human 450K Bead array. We identified 1640 significant and differentially methylated CpG sites in infertile individuals. The differentially methylated regions (DMRs) were specifically over-represented in gene regions such as gene body, intergenic and 3'UTR, while under-represented in the promoter region. Gene ontology and pathway analysis revealed different biological processes (reproduction, development, immune system) and pathways pertinent to modulation of male reproductive capacity. The prospective DMRs were validated using independent sequencing approach in a larger cohort of samples. These results strongly suggest that spermatogenesis specific aberrant DNA methylation patterns are involved in the pathogenesis of infertility. In depth investigation of the candidate genes would identify the signaling and developmental pathways that are critical to spermatogenesis. The study clearly suggests that the pathology of male infertility is associated with impairment of various biological pathways and advances our understanding of epigenetic regulation in the etiology of male infertility.

1582W

Comparison of X chromosome inactivation in peripheral tissues and visceral organs in females with X-linked diseases. *M. Reboun[1], D. Musalkova[1], M. Hrebicek[1], J. Sikora[2], J. Kulhanek[2], J. Vcelak[3], G. Storkanova[1], H. Vlaskova[1], L. Dvorakova[1].* 1) Institute of Inherited Metabolic Disorders, Charles University-First Faculty of Medicine and General University Hospital in Prague, Prague, Czech Republic; 2) Department of Pediatrics and Adolescent Medicine, Charles University-First Faculty of Medicine and General University Hospital in Prague, Prague, Czech Republic; 3) Department of Molecular Endocrinology, Institute of Endocrinology, Prague, Czech Republic.

   X chromosome inactivation (XCI) is considered to be one of the factors determining the phenotypic severity of X-linked diseases in heterozygous female patients. XCI analysis in peripheral tissues (e.g. circulating white blood cells (WBC), buccal or urine epithelia) is commonly used to infer the XCI ratios in the visceral organs that are affected by the diseases but cannot be sampled for testing. Herein, we present the results of XCI analyses performed by DNA-based methylsensitive and/or transcript analyses in heterozygous female patients with Danon disease (DD, OMIM 300257), ornithine transcarbamylase deficiency (OTCdef, OMIM 311250), and apoptosis-inducing factor deficiency associated with the *AIFM1* gene (OMIM 300169). XCI ratios (46:54, 33:67, 34:66, 42:68, and 27:73) in WBCs of 5 female patients with DD were random. In a single DD patient (XCI 27:73 in WBCs), the total of 28 myocardial samples were available for further testing. Interestingly, the myocardial values ranged from 19:81 to 56:44 (average 42:58, median 42:58). Similarly, a single OTCdef female patient was tested both in peripheral tissues and 37 liver samples. The values of XCI in WBCs, buccal and urine epithelis were 44:56, 60:40 and 55:45 respectively. In the liver tissue the XCI ranged from 46:54 to 87:13 (average 71:29, median 75:25) with the mutated allele being preferentially active. Last, XCI was assessed in one symptomatic and two asymptomatic *AIFM1* female patients. In all three, XCI was extremely skewed in WBCs with the mutant allele almost completely inactivated (93:7) regardless of the patients phenotype. XCI skewing in buccal swabs was less pronounced (85:15). As the AIFM1 protein contributes to apoptosis regulation, our results imply secondary skewing as a result of clonal disadvantage of the cells expressing the mutant *AIFM1* allele. To summarize – (i) XCI ratios in individual samples from the visceral organs are very variable, which suggests that small-sized bioptic sampling to establish the XCI ratios could be biased, (ii) provided there is no additional clonal selection, peripheral tissues can be used as an approximation of XCI in the visceral organs, (iii) skewed XCI is not required for the manifestation of either DD or OTC deficiency in heterozygote females. Support: GAUK No.580716, MZ CR – RVO VFN 64165, MZ CR – RVO EU 00023761, SVV No.260367, AZV CR - 15-27682A and 15-33297A,.

1583T

Transcriptome analysis of cystic fibrosis molecular signatures. *J.E. Ideozu[1,2], X. Zhang[1,2], Z. Ashrafi[1], M. Kaldunski[3], A. Pan[3], M. Kandpal[4], R.V. Davuluri[4], P. Simpson[3], M. Hessner[3], H. Levy[1,2,4].* 1) Ann & Robert H. Lurie Children's Hospital of Chicago, IL; 2) Stanley Manne Children's Research Institute, Chicago, IL; 3) Children's Hospital of Wisconsin, Medical College of Wisconsin, Milwaukee, WI; 4) Feinberg School of Medicine at Northwestern University Chicago, IL.

   Introduction: Cystic fibrosis (CF) is an autosomal recessive disease due to mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. Airway inflammation leading to chronic lung disease due to colonization of *Pseudomonas aeruginosa* (*Pa*), is the major cause of morbidity and mortality with median predicted age of survival in CF now 40 (Cystic Fibrosis Foundation). Clinical heterogeneity, even in patients with the same *CFTR* genotype, suggests modifier genes and their regulatory molecules including microRNAs (miRNAs) are involved in underlying pathogenesis. **Objective:** Characterize molecular signatures of CF airway inflammation underlying the lung disease via simultaneous total RNA and miRNA transcriptome sequencing of peripheral blood mononuclear cells (PBMCs). **Methods:** Molecular signatures were captured via whole transcriptome sequencing, and comparison made between CF subjects infected with mucoid *Pa* (n=4) and non-infected (n=5). Simultaneous total RNA and small RNA single-ended sequencing libraries were prepared from one RNA source and sequenced using Illumina Next-Seq and MiSeq sequencer respectively. We performed miRNA target prediction using the Targetscan algorithm in Integrated Pathway Analyses and correlated with the predicted miRNA-mRNA pairs to validate the target pairs. Biological pathway and gene ontology enrichment were performed for top differentially expressed genes with at least 2-fold difference (p<0.05) or more.   **Results:** We generated average of ~37.2 million (total RNA-Seq) and ~10.1 million (miRNA-Seq) single-ended reads. Preliminary analyses indicate five hundred and seventy-two genes were differentially expressed by 2-fold or greater between CF subjects infected with mucoid *Pa* compared to non-infected (p<0.05). A total of 76 miRNAs targeting 36 mRNAs in our miRNA-mRNA pairs were inversely expressed and a cluster of miRNAs interacting with inflammatory mediators- CXCL8 and IL10. Ontological analysis of differential expressed genes revealed biological processes altered including regulation of immune response, cytokine activity and response to bacterial origin (p<0.05). Upstream regulator analysis showed downregulation of NFkB proteins and cytokines including CCL5 and IL17A. **Conclusion**: Transcriptional signatures generated suggest miRNAs are key regulatory molecules likely modulating host immune response to *Pa* infection in CF.

1584F

DNA hypermethylation and other epigenetic regulatory signaling pathway genes associated with hidradenitis suppurativa (acne inversa). *D. Jhala[1], U. Ratnamala[2,3], N.K. Jain[2], F.M. Al-Ali[4], M. Naveed[4], B.C Gorijala[5], U. Radhakrishna[6].* 1) Department of Zoology, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India; 2) Department of Life Sciences, School of Sciences, Gujarat University, Ahmedabad, India; 3) Department of Pharmacology, Creighton University, Omaha, NE, USA; 4) Dermatology Centre, Rashid Hospital, Dubai Health Authority (DHA), Dubai, United Arab Emirates; 5) Oncology Department, Krishna Institute of Medical Sciences, Secunderabad, Andhra Pradesh, India; 6) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

Hidradenitis Suppurativa (HS), also known as Acne Inversa (*AI*), is a chronic skin condition characterized by swollen, painful lesions occurring in the armpit (axillae), groin, and breast regions. Mutations in γ-secretase genes have been found to cause HS in some individuals and familial cases. Genetic heterogeneity has long been suspected and a role for other molecular mechanisms such as epigenetic modifications anticipated. To identify DNA methylation biomarkers for the detection of HS, a genome-wide DNA methylation scan using the Infinium HumanMethylation450 BeadChip array (Illumina) was performed in a cohort of 24 HS subjects previously excluded from γ-secretase mutations and 24 controls matched for age, gender, and ethnic origin. We identified significant CpG hypermethylation (at least 2.0-fold) at 304 sites in 304 genes in HS subjects (false discovery rate (FDR) ≤0.00001). No hypomethylation was detected. These 304 CpG sites have a receiver operating curve area under the curve (ROC AUC) ≥0.75 and for HS detection and were spread throughout the whole-genome. A total of 23 cytosine loci had excellent accuracy (AUC ≥ 0.90) for the detection of HS. The top 10 hypermethylated genes were *CUL2* (cg15475055), *KIAA1539* (cg10455757), *ZFYVE28* (cg11332822), *SEC24B* (cg05861071), *TMEM191A* (cg08235057), *CHD1L* (cg27261837), *DCPS* (cg20032723), *LRRC8B* (cg22081084), *LYST* (cg12875892) and *PHTF1* (cg00419759). Pathway analysis of the differentially-methylated genes in HS revealed the involvement of multiple signaling pathways. CpG methylation sites that reside in promoter regions or within the genes of the Notch pathway are implicated in the development of skin and other cancers, wound healing, immunological processes, cell cycle regulation and apoptosis, and the formation of channel proteins that are required for skin development and growth.

1585W

Dissecting regulatory mechanisms altering skin pigmentation in Africans using genetic and functional genomic data. *D. Kelly[1], M. Holsbach Beltrame[1], N.G. Crawford[1], M.E.B. Hansen[1], S. Fan[1], S.L. Bowman[3,4], E. Jewett[5,6], A. Ranciaro[1], S. Thompson[1], Y. Lo[1], S.P. Pfeifer[7], J.D. Jensen[7], M. Campbell[1,8], W. Biggs[1], S.W. Mpoloka[9], G. Mokone[10], T. Nyambo[11], D.W. Meskel[12], G. Belay[12], C. Abnet[15], E. Oceana[17], Y.S. Song[5,6,18,19,20], E. Eskin[21], K.M. Brown[13], M.S. Marks[3,4], S.K. Loftus[14], W.J. Pavan[14], M. Yeager[15,16], S. Chanock[22], S. Tishkoff[1,23].* 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA; 3) Department of Pathology & Laboratory Medicine, Children's Hospital of Philadelphia Research Institute, Philadelphia, PA; 4) Department of Pathology & Laboratory Medicine and Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Department of EECS, University of California, Berkeley, CA; 6) Department of Statistics, University of California, Berkeley, CA; 7) School of Life Sciences, Arizona State University, Tempe, AZ; 8) Department of Biology, Howard University, Washington D.C; 9) Department of Biological Sciences, University of Botswana, Gaborone Botswana; 10) Dept. of Biomedical Sciences, University of Botswana School of Medicine, Gaborone, Botswana; 11) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 12) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 13) Laboratory of Translational Genomics (LTG), Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute, National Institutes of Health, Bethesda MD; 14) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 15) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD; 16) Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD; 17) Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University, Providence, RI; 18) Chan Zuckerberg Biohub, San Francisco, CA; 19) Department of Biology, University of Pennsylvania, Philadelphia, PA; 20) Department of Mathematics, University of Pennsylvania, Philadelphia, PA; 21) Department of Computer Science, Department of Human Genetics, University of California, Los Angeles; 22) Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institute of Health, Bethesda, MD; 23) Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA.

Skin pigmentation varies considerably across human populations and is correlated with latitude and levels of ultraviolet radiation (UVR), suggesting it is an adaptive trait. To gain a more complete understanding of the genetic and biological basis of skin pigmentation variation we conducted a genome-wide association study (GWAS) in ethnically diverse African individuals, revealing novel variants associated with pigmentation in or near the genes *SLC24A5, MFSD12, TMEM138[DDB1]*, OCA2 and *HERC2*. To make finer resolution predictions we performed fine mapping and predicted causal variants using *CAVIAR*. A number of candidate causal loci are located in non-coding regions. We integrated RNA-seq data from primary human melanocytes with publicly available functional genomic datasets and performed luciferase expression reporter assays to identify potential regulatory variants. At the *MFSD12* locus we identify variants overlapping CTCF and a melanocyte-specific enhancer element bound by MITF, a key regulator of the melanocyte lineage. Genetic variation at this locus is significantly correlated with expression of nearby genes and luciferase reporter assays confirm their regulatory potential. At the *DDB1* locus we identify variants intersecting predicted melanocyte enhancer elements. For example, we identify an insertion/deletion polymorphism intersecting a melanocyte-specific enhancer bound by MITF. Genetic variation in this region is significantly correlated with expression of genes important for DNA mismatch repair, vesicle trafficking, and protein sorting in vacuoles (including melanosomes) and luciferase reporter assays confirm their regulatory potential. Finally, at the *OCA2/HERC2* locus, we identify a variant disrupting an antioxidant response element (ARE), elements which, in conjunction with Nrf2, regulate genes important for cellular response to oxidative damage, a pathway known to be disrupted in individuals with vitilgo. This work identifies novel variants related to human pigmentation variation and identifies functionally important variants, genes, and molecular mechanisms for future study.

1586T

Hippocampus and blood *APOE* locus DNA methylation in Alzheimer's disease. *L. Bekris[1], M. Krestian[1], G. DAleo[1], J. Zahratka[1], J. Leverenz[2], Y. Shao[1].* 1) Genomics Medicine Institute, Cleveland Clinic Lerner Research Institute, Cleveland, OH; 2) Lou Ruvo Center for Brain Health, Cleveland Clinic Neurological Institute, Cleveland, OH.

In humans and in mouse models, *APOE* ε2 carriers have higher apoE levels, compared to ε3 and ε4 (ε2>ε3>ε4) and are protected against Alzheimer's disease (AD) pathology, such as the toxic Aβ protein. It has been suggested that if apoE2 or apoE3 protein levels are therapeutically increased, without increasing apoE4, this will allow for better clearance and less deposition of the toxic Aβ protein in AD. Many questions still remain on how *APOE* expression is regulated by distant regulatory elements at this locus and how they might be exploited to modulate apoE levels in AD. For example, our groups and others have proposed that the activity of a *APOE* exon 4 regulatory element is influenced by epigenetic factors, such as methylation, that either enhance transcription or decrease expression depending on the *APOE* ε2/ε3/ε4 haplotype and regional promoter. However, it is still unclear if other regulatory elements at this complex extended locus are influenced by epigenetic factors in human tissue. Therefore, the aim of this exploratory investigation is to characterize the DNA methylation status of *TOMM40-APOE-APOC2* gene region in both hippocampus and blood in AD compared to cognitively normal controls. The central hypothesis of this exploratory investigation is that the complex regulatory structure surrounding the *APOE* gene is differentially methylated depending on tissue type and disease status. DNA was extracted from human brain (n=12) and lymphocytes (n=85) and percent methylation was measured using the bisulfite conversion EZ DNA Methylation™ Kit (Zymo Research) and Infinium HumanMethylation450 BeadChip Kit (Illumina). Extended methylation analysis of the *APOE* locus suggests that methylation at *TOMM40* and *APOC1,* but not the *APOE* exon 4-3'UTR CpG, are different between AD and controls in hippocampus DNA. Methylation at the *TOMM40, APOE* exon 4-3'UTR CpG island, *APOC1, APOC1P1* and *APOC4-APOC2* is different primarily in MCI compared to controls in blood DNA. This exploratory DNA methylation analysis of an extended region of the *APOE* locus identified regions outside of the *APOE* gene that may be differentially regulated according to tissue type and disease status. These results suggest that other regulatory regions at this locus, in addition to *APOE,* are differentially methylated implicating tissue and disease specific regulatory mechanisms that may play an important role in neurodegenerative diseases, such as AD. .

1587F

Transcriptomic analysis of whole blood reveals potential biomarkers in African American Alzheimer disease. *S.K. Sivasankaran[1], A.J. Griswold[1], F. Rajabli[1], B. Kunkle[1], K. Hamilton[1], J. Jaworski[1], W.S. Bush[1,2], E.R. Martin[1,3], G.W. Beecham[1,3], G.S. Byrd[4], J.L. Haines[2], M.A. Paricak-Vance[1,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 4) Department of Biology, North Carolina A&T State University, Greensboro, NC.

Alzheimer disease (AD) is the leading cause of dementia in the elderly and occurs in all ethnic and racial groups. African Americans (AA) are substantially underrepresented in efforts to identify genetic variation that enhance or protect against AD, despite having a higher prevalence than non-Hispanic Whites (NHW). To address this important disparity, we are performing RNA-sequencing of peripheral blood samples on 250 AA and 250 NHW (125 cases, 125 controls in each ethnic group) all with genome-wide association study (GWAS) data as an extension to The Alzheimer's Disease Sequencing Project (ADSP). Because the majority of variants identified from WGS will be regulatory in nature, direct measurement of blood gene expression will help translate variant associations to functional outcomes. Here we present the results from the first pilot set of 35 samples (AA: 7 cases and 13 controls, NHW: 7 cases and 8 controls). Total RNA was extracted, enriched by poly-A selection, and globin and ribosomal RNA depleted. Libraries were sequenced for minimum of 40 million paired end 100bp reads on the Illumina HiSeq3000. Alignment to hg19 and gene quantification to GENCODE v.19 were performed with the STAR algorithm. Differential expression of genes was calculated using baySeq. Alignment was robust with 89% overall unique alignment of which 70% of reads map to known protein coding genes. Case-control comparison in the NHW samples did not reveal any significant genes by false discovery rate (FDR) correction. In the AA, there were eight genes FDR significant by baySeq. These genes include an increase in AD cases of *MTRNR2L12* (FDR = 0.0002, fold change (FC) = 3.5) which has been identified as a potential biomarker in blood expression studies of Down syndrome patients with dementia. Furthermore, significant increases were noted in a cell adhesion molecule *NCAM1* (FDR = 0.0003, FC = 1.9) and synaptic exocytosis regulator *SCAMP5* (FDR = 0.0447 FC = 1.5). Notably, these genes have high expression in brain tissue according to the Genotype-Tissue Expression (GTEx) project, suggesting blood may be a useful material to understand a neurodegenerative disease. While this represents preliminary data from an ongoing larger study, identification of disease biomarkers, correlation of expression to DNA sequence changes, and creation of ethnically diverse expression reference sets is a critical component of understanding the etiology of AD.

1588W

Understanding the endogenous regulation of Ataxin-1 in SCA-1. *R. Manek[1,2], T. Nelson[1,2], E. Rodríguez-Lebrón[1,2].* 1) University of Florida, Department of Pharmacology and Therapeutics Gainesville, FL; 2) University of Florida, Center for Translational Research in Neurodegenerative Diseases, Gainesville, FL.

Spinocerebellar ataxia type-1 (SCA-1) is an autosomal dominantly inherited, progressive neurodegenerative disease caused by the abnormal expansion of a 'CAG' tri-nucleotide repeat within the coding region of the *ATXN1* gene. The CAG expansion is translated into a polyglutamine (PolyQ) tract in the N-terminus of the Ataxin-1 protein. Mutant Ataxin-1 expression is necessary for both the onset and progression of this currently untreatable disease. Novel therapeutic strategies such as RNAi that down regulate overall Ataxin-1 protein levels mitigate SCA-1 phenotype in mouse models of the disease. Thus, understanding the endogenous cellular mechanisms that regulate Ataxin-1 expression, and how this regulation may be disrupted in SCA-1, is key to the development of novel, clinically relevant therapeutics. Here we seek to better define the post-transcriptional regulation of *ATXN1* by evaluating the role that the 5'UTR of *ATXN1* plays on the expression of Ataxin-1, in the context of both the WT and the CAG-expanded transcript. In order to determine if the 5'UTR ATXN1 can independently regulate Ataxin-1, we have designed highly versatile reporter-based assays. In preliminary studies, we find that the 5'UTR *ATXN1* sequence exerts a strong regulation of protein levels (GFP reporter gene). Analysis of steady-state RNA levels by quantitative PCR appear to indicate decrease in RNA levels. Since cellular post-transcriptional gene regulation is mainly exerted by altering mRNA stability and/or translational efficiency, we have designed studies to investigate the role the 5'UTR of ATXN1 plays in regulating ATXN1 mRNA stability and translational efficiency.

1589T

Haploinsufficiency models of *CHD8* in neuronal cells display alterations in chromatin landscape and regulatory consequences in Wnt signaling. *E. Kerschbamer[1], T. Tripathi[1], S. Erdin[2,3], F. Di Leva[1], J.F. Gusella[2,3], S. Piazza[1], M.E. Talkowski[2,3], M. Biagioli[1].* 1) Center for Integrative Biology, University of Trento, Trento, Italy; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, United States; 3) Department of Neurology, Harvard Medical School, Boston, MA, United States.

Autism Spectrum Disorders (ASD) is a collection of heterogeneous neurodevelopmental disorders defined by social impairment and repetitive behaviors with significant genotypic and phenotypic complexity. Mutations in several hundred loci have been associated with the disease, but the chromodomain helicase DNA binding protein 8 (*CHD8*) represents a recurrent and independently validated ASD-risk gene. All ASD mutations in *CHD8* gene have been reported to be disruptive, leading to haploinsufficiency. In physiological conditions, *CHD8* binds to the DNA at promoter and enhancers regions where it regulates transcriptional initiation. Here we investigate how chromatin landscape reacts to *CHD8* suppression by analyzing different histone modifications through Chromatin Immunoprecipitation and Sequencing (ChIP-seq) in iPS-derived neural progenitors after *CHD8* knock-down. We interrogated transcriptionally active and repressed regions as well as active and poised enhancers to explore the potential impact of loss-of-function mutations in *CHD8* on regulatory networks and pathways. *CHD8* suppression alters the overall chromatin landscape and ~4000 poised enhancers (enriched for H3K4me1) targeting genes implicated in 'cell cycle' and 'Wnt signaling' pathways. *CHD8* suppression also affected ~5000 actively transcribed gene bodies (enriched for H3K36me3) that were again associated with 'Wnt signaling' and 'RNA pol II transcription activity', as well as 2130 silent promoters (-5Kb/+2Kb of the transcription start site; enriched for H3K27me3) involved in biological processes such as 'axon guidance' and 'synaptic transmission'. The majority of loci that undergo chromatin changes are unique to one specific histone mark, while other genomic regions (enriched for H3K4me2, H3K4me3 and H3K27ac) seem to be less sensitive to *CHD8* suppression. Our results point toward both broad regulatory consequence of *CHD8* suppression as well as association with specific biological pathways of relevance to ASD pathogenesis.

1590F

Correlation of methylomic profiles between blood and cerebral spinal fluid in aneurysmal subarachnoid hemorrhage patients. *A. Arockiaraj[1], J.R. Shaffer[1], P.R Sherwood[3], E.A Crago[3], S.M Poloyac[4], D.E Weeks[1], Y.P Conley[2].* 1) Department of Human Genetics,Graduate School of Public Health,University of Pittsburgh, Pittsburgh, PA; 2) Department of Health Promotion and Development, School of Nursing, University of Pittsburgh,PA; 3) Department of Acute and Tertiary Care, School of Nursing, University of Pittsburgh,Pittsburgh,PA; 4) Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh,Pittsburgh,PA.

Aneurysmal subarachnoid hemorrhage (aSAH) is a form of stroke leading to severe patient outcomes including morality, cognitive impairment, functional disability, and emotional dysfunction. Methylomic changes in neurological tissue are hypothesized to occur in patients after aSAH, and these changes may influence the development of adverse outcomes. Consequently, the methylomic profiles of brain tissue in aSAH patients are of particular interest, both for understanding the pathophysiology of aSAH, and for identifying predictive biomarkers of adverse outcomes and/or recovery. However, assessment of methylomic profiles in relevant neurological tissue requires prohibitively invasive biopsy. Therefore, we investigated to degree to which methylomic profiles in peripheral blood, a readily available tissue, were correlated with those in cells derived from cerebral spinal fluid (CSF), a feasibly accessible neurological tissue surrogate. Blood and CSF samples were obtained from 68 patients within two days after aSAH. Whole-genome methylation data were collected using the Illumina HumanMethylation450K BeadChip. Batch effects were removed by functional normalization. CpG sites assayed by probes containing SNPs, cross reactive probes, and sex-chromosome probes, as well as CpG sites showing multimodal methylation distributions or identified via data quality diagnostics were removed from analysis. Duplicate samples were used to identify noisy CpG sites showing greater technical variation than experimental variation, separately in blood and CSF. After filtering, we retained 249,821 and 255,280 CpG sites in blood and CSF samples, respectively, with 172,278 CpG sites common to both tissues. For each of these CpG sites we calculated the correlation in methylation M values between blood and CSF. CpG sites showed a range of correlation coefficients (mean=26%, min=-69%, max=95%), with CpG islands and shores showing greater correlation than shelves. Moreover, the correlation differed with respect to the position of the CpG sites within genes, with CpG sites generally nearer the transcription start site and first exon showing stronger positive correlations than CpG sites in the 3' UTR. These results suggest that, while some CpGs are correlated between blood and CSF, globally, blood is not a suitable surrogate for inferring methylomic profiles in neurological tissue post-aSAH. More work is needed to identify the methylomic profiles relevant to aSAH recovery outcomes.R01NR013610.

1591W

Epigenomic signature of adrenoleukodystrophy predicts compromised oligodendrocyte differentiation. *A. Pujol[1], A. Schluter[1], M. Ruiz[1], S. Fourcade[1], J. Sandoval[2], A. Diaz-Lagares[2], M. Esteller[2].* 1) Neurometabolic Diseases, IDIBELL, Barcelona, Barcelona, Spain; 2) Program of Epigenetics and Biology of Cancer, IDIBELL,, Barcelona, Spain.

Epigenomic changes occur in cancer, aging and neurodegenerative disorders, and may either cause disease or modulate its expressivity, adding a layer of complexity to mendelian conditions. X-linked adrenoleukodystrophy (X-ALD) is a rare neurometabolic condition exhibiting discordant phenotypes, ranging from a childhood cerebral inflammatory demyelination (cALD) to an adult-onset mild axonopathy in spinal cords (AMN). The AMN form may occur with superimposed inflammatory brain demyelination (cAMN). All patients harbor loss of function mutations in the ABCD1 peroxisomal transporter of very-long chain fatty acids. The factors that account for the lack of genotype-phenotype correlation, even within the same family, remain largely unknown. To gain insight into this matter, here we compared the genome-wide DNA methylation profiles of morphologically intact frontal white matter areas of children affected by cALD with adult cAMN patients, including age-matched male controls. We identified a common methylomic signature between the two phenotypes, comprising i) hypermethylation of genes harboring the H3K27me3 mark at promoter regions, ii) hypermethylation of genes with major roles in oligodendrocyte differentiation such as *MBP*, *CNP*, *MOG* and *PLP1*, and iii) hypomethylation of immune-associated genes such as *IFITM1* and *CD59*. Moreover, we found higher hypermethylation in CpGs of genes involved in oligodendrocyte differentiation such as *MBP*, *CNP*, *SOX10* and *NINJ2*, and also in genes with H3K27me3 marks in their promoter regions such as *HOXA3* and *AGAP2,* in cALD compared with cAMN. We also uncovered differentially methylated and expressed genes related to immune response, such as *UNC45A*, which may be of use in discriminating between these phenotypes. The differences of methylation correlated with transcriptional and translational changes, as supported by Affymetrix expression arrays, quantitative PCR and protein expression. Further, using a penalized logistic regression model, we identified the hypermethylation of *SPG20*, *UNC45A* and *COL9A3* to be good markers capable of discriminating childhood from adult inflammatory phenotypes. We thus propose that an epigenetically controlled, altered transcriptional program drives an impaired oligodendrocyte differentiation and aberrant immune activation in X-ALD patients. These results shed light into disease pathomechanisms and uncover biomarkers for prognosis and phenotypic stratification.

1592T

Epigenetic silencing in Friedreich ataxia is caused by hypermethylation of the *FXN* promoter CpG island shore. *L.N. Rodden[1], Y.K. Chutake[2], S.I. Bidichandani[1,2].* 1) Graduate Program in Neuroscience, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Friedreich ataxia (FRDA) is caused by an expanded GAA triplet-repeat (GAA-TR) mutation in intron 1 of the *FXN* gene that results in epigenetic silencing of the *FXN* promoter. DNA methylation of CpG island shores, regions flanking human gene promoters that are embedded in CpG islands, is a known mechanism of epigenetic silencing. Deep sequencing revealed that DNA hypermethylation spreads from the expanded GAA-TR mutation to the *FXN* CpG island shore. The CpG island shore is hypermethylated in FRDA, but it remains unmethylated in the non-disease state, and becomes unmethylated when the expanded GAA-TR is reverted to the normal size in isogenic cell lines, thus functioning as a FRDA-specific differentially methylated region (FRDA-DMR). The hypermethylated FRDA-DMR was detected in various patient-derived cell types, and also in tissues from the humanized mouse model of FRDA that carries an expanded GAA-TR. Analysis of individual DNA molecules revealed a variegated pattern of DNA methylation within the FRDA-DMR, the magnitude and extent of which was dependent on the length of the expanded GAA-TR mutation. Knockdown of DNMT3A in patient-derived cells reduced methylation of the FRDA-DMR, and a CRISPR-deactivated Cas9 strategy to target DNMT3A to the *FXN* CGI shore increased methylation of the FRDA-DMR. Treatment with 5-aza-2'-deoxycytidine enhanced the ability of a class I histone deacetylase inhibitor, known to ameliorate epigenetic promoter silencing in FRDA, to further increase *FXN* transcript levels in patient-derived cells. We conclude that hypermethylation of the *FXN* promoter CpG island shore, which is mediated by DNMT3A, plays a key role in epigenetic silencing in FRDA, and is a novel therapeutic target for reactivation of the epigenetically silenced *FXN* gene.

1593F

The study of Vitamin D *effect on VDR* gene expression in multiple sclerosis patients. *Z. Shirvani-Farsani[1], M. Behmanesh[2], M.A. Sahraian[3].* 1) Department of Cellular and Molecular Biology, Faculty of Biological Sciences and Technology, Shahid Beheshti University G.C., Tehran, IR Iran; 2) Department of Basic Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 3) MS Research Center, Neuroscience Institute, Tehran University of Medical Science, Tehran, Iran.

Background: Multiple sclerosis (MS) is an autoimmune disease. Inflammatory immune responses in MS patients could be inhibited by vitamin D. Vitamin D effects are occurred via the vitamin D receptor (VDR). VDR is a transcription factor and expresses in different immune cells. It plays an important role in the stating of MS risk and disease activity. **Objective:** The aim of this study was to investigate the expression of VDR gene in MS patients before and after treatment with Vitamin D. **Methods:** Blood samples were obtained from 20 MS patients before and after Vitamin D treatment. RRMS patients received 50,000 IU vitamin D3 per week as intra-muscular injection for 8 weeks. *VDR* mRNA levels were measured by real time polymerase chain reaction and analyzed with paired t-test. Moreover, correlation analysis was performed between the expression level of gene and some clinical features. **Results**: Beside significant increase of serum 25(OH) D levels after supplementation, the expression level of *VDR* gene significantly decreased after 8 weeks Vitamin D treatment. **Conclusion:** We found some new evidences about the molecular mechanism of vitamin D effectiveness in MS treatment.

1594W

Supplemental treatment for Huntington disease (HD) with miR-132 that is deficient in HD brain. *M. Fukuoka[1], M. Takahashi[1,2], H. Fujita[3], T. Chiyo[4], A. Popiel[3], S. Watanabe[5], H. Furuya[6], M. Murata[5], K. Wada[3], T. Okada[4], Y. Nagai[3], H. Hohjoh[1].* 1) Department of Molecular Pharmacology, National Institute of Nuroscience, National Center of Neurology and Phychiatry, Tokyo, Japan; 2) Department of RNA medical Science, The Institute of Medical School, The University of Tokyo, Tokyo, Japan; 3) Department of Degenerative Neurological Diseases, National Institute of Neuroscience, NCNP, Tokyo, Japan; 4) Department of Molecular Therapy, National Institute of Neuroscience, NCNP, Tokyo, Japan; 5) National Center Hospital, NCNP, Tokyo, Japan; 6) Oomuta Hospital, Fukuoka, Japan.

Huntington disease (HD) is a dominantly-inherited neurodegenerative disorder characterized by chorea movement, muscular incoordination, cognitive decline and psychiatric problems. The responsible gene for HD is the *Huntingtin* (*HTT*) gene, and aberrantly expanded CAG repeats (translated into polyglutamine tracts) in exon 1 are closely related to the onset and severity of HD. Currently, there is no definitive treatment of HD; and many efforts have been made for searching for clues that would provide better treatments for HD. MicroRNAs (miRNAs) are 21~23-nucleotide-long small noncoding RNAs and function as mediators in gene silencing by inhibiting translation of messenger RNAs (mRNAs), and by digestion of mRNAs, or by RNA interference (RNAi). MiRNAs might provide us with clues helpful for better treatments of diseases because their disease-associated expression has been detected. A major change in the expression of miRNAs occurs in the brain during the first one month of mouse life, which corresponds to the stage of a rapid brain growth. Abnormal changes in the expression of miRNAs during the stage may cause deleterious effects on the brain formation. In this study, we investigated *miR-132* that markedly decreased in the brain of R6/2 (HD-model) mice and conducted a challenging experiment to see what would happen to HD mice if *miR-132* were supplied into the brain. To reduce the shortage of *miR-132* in the brain, we constructed a viral *miR-132*-expression system with a recombinant adeno-associated virus (rAAV) and introduced the rAAV into the striatum of HD and wild-type mice, at the age of 3weeks when a marked difference in the *miR-132* level started appearing between the mice. The results demonstrated that *miR-132* supplement into the brain of HD mice produced symptomatic improvement of the motor function and lifespan of the mice without altering the expression level of mutant *HTT*. In addition, the post synaptic density 95 kDa (Psd95) protein decreased in the striatum of HD mice recovered to a normal level as in wild-type mice after supplementation of *miR-132*. Our study shed light on a new HD pathophysiological condition involving *miR-132* deficiency, and also on a novel therapeutic approach to HD.

1595T

Decreased expression of *Beclin2* and *LC3* genes in PGRN deficiency: A CRISPR-Cas9 neuronal cell model. *S. Napoletano[1], H. Kukreja[2], A. Rendina[1], A. Riccio[2], E. Vitale[1].* 1) Institute of Protein Biochemistry (IBP), National Research Council of Italy (CNR), Naples, Italy; 2) Institute of Genetic and Biophysics, (IGB), National Research Council of Italy (CNR), Naples, Italy.

Progranulin (*GRN*) is genetically associated with frontotemporal dementia (FTD), with mutations present in 23% of patients with familial FTD. The neurobiology of progranulin (PGRN) is still unclear, although the proposed disease mechanism is linked to a deficiency of the protein. Furthermore, patients with PGRN deficiency express primarily ubiquitin-positive TAR DNA-binding protein 43 (TDP-43) aggregates in the brain. It is possible that misfolded proteins may aggregate similarly to prions, converting a native protein into a pathologic misfolded replicate able to initiate template-directed protein misfolding. We enrolled 256 FTD patients and 300 healthy age-, sex- and geographic region-matched controls and sequenced all the coding exons of *GRN* and *MAPT* genes in patients with a familial history for dementia. We identified a rare *GRN* gene exon six deletion, g10325_10331delCTGCTGT (relative to nt 1 in NG_007886.1), alias Cys157LysfsX97, in three autosomal dominant pedigrees in Southern Italy segregating FTD in three generations. Transcriptional and translational analysis of mRNA from WBC and plasma proteins from FTD patients carrying the mutation exhibited deficiency of the protein. We used mouse embryonic stem cells (E14) with the goal of generating a cell model using CRISPR-Cas9, to test regulatory functions and pathological mechanisms of PGRN in vitro. We designed single guide RNA for the *GRN* gene with high specific, on-target activity and derived pure cellular clones carrying homozygote *GRN* gene mutations causing protein deficiency. We subjected wild-type and knock-out E14 cells to neuronal differentiation, generating an efficient model system to dissect the molecular mechanisms of PGRN in pathways potentially involved in FTD-TDP43 proteinopathy. To test the hypothesis that PGRN deficiency results in an impairment of the autophagic-lysosomal pathway, we performed mRNA gene expression analysis of autophagy-related genes using quantitative RT-PCR. Preliminary results demonstrated a decrease of *Beclin2* and *LC3* genes in PGRN-deficient cells compared to wild-type E14. Our findings support the role of the *GRN* gene in the etiology of FTD and the hypothesis that PGRN insufficiency can be a predisposing factor in neuronal degeneration. We can also speculate that PGRN plays a role in the regulation of the autophagic-lysosomal pathway. Further analyses are in progress to confirm this hypothesis and to identify the connection with the onset of FTD.

1596F

An integrated genetic-epigenetic approach for assessing risk for stroke in the Framingham Heart Study. *R. Philibert[1,2], M. Dogan[2,3].* 1) Behavioral Diagnostics, Coralville , IA. 52241; 2) University of Iowa, Iowa City, IA 52242; 3) Cardio Diagnostics, Coralville IA 52241.

Stroke is the fifth leading cause of death in the United States with one in every 17 Americans dying as a result of a stroke. Uncommon genetic variants contribute to a small proportion of total risk variance. But the majority of total risk is not well explained through conventional molecular approaches but can be partially attributed to known clinical risk factors such as high blood pressure, smoking and atrial fibrillation. Unfortunately, efforts to combine the genetic and clinical assessments of these risk factors into effective screening tools for the prevention of stroke have not been fully effective. In prior work, we have demonstrated that the methylation effects of smoking, which is a major risk factor for stroke, map to gene networks associated with embolic stroke. When taken together with other evidence, this suggests the possibility that a method that incorporates other epigenetic signatures of risk variables, such as hypertension, together with genetic risk factors, may be able to predict risk for stroke. Using the Random Forest analytic approach and integrated genome wide methylation and genetic variation data sets from Framingham Heart Study subjects, we herein test whether machine learning algorithms that use a combination of SNP, methylationxSNP interaction (MethxSNP) and methylation effects can predict stroke. As a first step, using the training set of 1562 individuals (38 with stroke), genome wide methylation and SNP data was regressed against stroke status to identify the CpG residues and SNPs most associated with stroke. Then, using stratified sampling to account for class imbalance, the Gini index was used to identify important loci. 10 CpG sites and 5 SNPs were retained for prediction. The trained model had an area under the curve (AUC) of 0.89, and an accuracy of 85%. However, in a very limited test set of 18 cases and 18 controls, the model performed only modestly with an AUC of 0.60 and accuracy of 56%. In light of prior results, these results suggest the possibility that Precision Medicine tools utilizing machine learning algorithms and both genetic and epigenetic data may be able to accurately predict the occurrence of stroke. However, in order to effectively construct those tools, larger data sets from longitudinally informative cohorts for stroke will be necessary.

1597W

Isogenic iPSC-derived neurons for modeling the differential regulation of *SNCA* expression: Implication to the heterogeneity of synucleinopathies. *O. Chiba-Falek[1,2], L. Tagliafierro[1,2], M.E. Zamora[1,2], O. Glenn[1,2], H. Bergstrom[1,2], D. Lukatsky[3], R. Gordan[2].* 1) Neurology, Duke University, Durham, NC 27710, USA; 2) Center for Genomic and Computational Biology, Duke University, Durham, NC 27710, USA; 3) cDepartment of chemistry, Ben-Gurion University of the Negev, Beer Sheva, Israel.

Synucleinopathies share a common pathological hallmark of intracellular a-syn inclusion bodies, however each disease presents distinct pathological and clinical characteristics. While *SNCA* expression levels are crucial to the development of synucleinopathies, the molecular mechanisms regulating *SNCA* expression and their contribution to the phenotypic heterogeneity of synucleinopathies are largely unknown. To study the differential regulation of *SNCA* expression in the context of Parkinson's Disease (PD) compared to Dementia with Lewy Bodies (DLB), we have developed *isogenic* induced Pluripotent Stem Cells (iPSCs)-derived dopaminergic and cholinergic neurons to model PD and DLB, respectively. We established two *isogenic* systems using iPSC-lines derived from a healthy subject and a patient with *SNCA*-triplication. *SNCA*-mRNA and protein expression in the *SNCA*-triplication derived neurons exhibited two-fold increase compared to the healthy neurons, recapitulating the observations in human tissues. The *SNCA*-triplication neurons exhibited increased susceptibility to disease related phenotypes and progressive aging compared to the healthy lines, indicating the system's suitability for studying the downstream effects of *SNCA* upregulation. The *isogenic* dopaminergic vs. cholinergic neurons revealed differential vulnerability to cellular phenotypes of disease related pathways and aging. Next, we assessed *cis* and *trans* factors involved in *SNCA* regulation. Expression profiles of conserved miRNAs targeting the *SNCA*-3'UTR (TargetScan) were different in cholinergic compared to dopaminergic neurons. We then sequenced the *SNCA*-3'UTR from PD and DLB cases. While we did not detect variants that overlap with the conserved miRNAs binding sites, common variants, most significantly associated with PD (but not with DLB), were identified within the binding sites of 3 poorly conserved miRNA families. Interestingly, one of these miRNAs was expressed significantly higher (~70%) in dopaminergic neurons compared to cholinergic. Similarly, transcription factors (TF) that bind to *SNCA* locus (ChIP-seq, ENCODE) were differentially expressed in dopaminergic compared to cholinergic neurons. We evaluated the effect of proximate repeat variants on the TFs binding affinity and found considerable variation among different allele lengths. These findings suggested that neuronal-type specific mechanisms regulating *SNCA* expression contribute, at least in part, to the heterogeneity of synucleionpathies.

1598T

Interpreting regulatory effects of disease-associated variants: A lesson from *SNCA* rs356168. *O. Glenn[1,2], L. Tagliafierro[1,2], O. Chiba-Falek[1,2].* 1) Neurology, Duke University Medical Center, Durham, NC; 2) Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC.

In the post genome wide association (GWA) era the fundamental question is, which are the actual causal variants within disease-associated genomic regions and what are their mechanisms of action. The *SNCA* locus has been implicated in the genetic etiology of Parkinson's disease (PD), and the *SNCA* intronic single nucleotide polymorphism (SNP), rs356168, is among the top ranked PD-associated SNPs in large genome wide association studies (GWAS). Recently, the PD-risk allele, rs356168-G was shown to increase *SNCA*-mRNA expression using genome edited human induced pluripotent stem cells (iPSC)-derived neurons. In this study, as means of validation, we tested the effect of rs356168 on total *SNCA*-mRNA levels using brain tissues, temporal and frontal cortex, from healthy control donors. Carriers of the rs356168-G allele demonstrated significantly decreased *SNCA*-mRNA levels in temporal brain tissues (p=0.02) compared to individuals homozygous for the 'A' allele. Similar trend was observed in the analysis of frontal cortex samples, however, this analysis did not reach statistical significance. Furthermore, eQTL analysis using the available GTeX datasets suggested a similar trend (AA>GG) for several human brain regions. These results conflict with effect of *SNCA* SNP rs356168 reported previous by Soldner *et al*. Our study conveys the need to carefully interpret the precise molecular mechanism by which rs356168, or another tightly linked variant, affects the regulation of *SNCA* expression. We postulate that non-transcriptional mechanisms, such as splicing, selection of polyadenylation site, and/or message stability, may also contribute to the observed associations between PD and the *SNCA*-3' linkage disequilibrium (LD) region. .

1599F

Translation regulation in Alzheimer's disease. *A. Shieh, C. Liu.* Psychiatry, University of Illinois - Chicago, Chicago, IL.

Alzheimer's Disease (AD) is the most common cause of dementia with molecular hallmarks of abnormal proteins in the brain. Studies in human and model organisms have observed that abnormalities in protein translation frequently accompany neurodegenerative phenotypes, highlighting the importance of studying translation regulation in the brain. In this study, we used Ribosome Profiling to measure translation in postmortem brains of 16 age-matched individuals (8 AD cases and 8 controls). We found 250 differentially translated genes at 10% FDR. These genes are enriched for pathways in concordance with neuroinflammation and neuronal and metabolic dysfunctions in the pathogenesis of AD. Ribosome Profiling measures the footprints of translating ribosomes. The resulting sequencing reads exhibit a signature 3-nucleotide codon periodicity that captures the mechanism of translation, and demarks coding regions. Using this signature, we identified more than 30,000 Open Reading Frames (ORFs) that show strong evidence of translation. Among these, we recovered ORFs from 25,971 transcripts, which covers more than 80% of the annotated CCDS genes. In addition, we found 1,971 ORFs originating from transcripts that are either antisense or classified as non-protein coding RNAs (lincRNA and pseudogenes). Finally, We found evidence of active translation at the UTR of nearly 2,000 genes. Translation at the 5' UTR (upstream ORF) has been shown to regulate the expression of their associated genes (main ORF). When compared to the 5'UTR ORFs found in HEK cells, we found only 20% overlap. Importantly, while the genes with translated 5'UTR in our samples are enriched for pathways important for the brain (ion transport and neurotransmitter signaling), those found in HEK cells are involved in cell cycle and growth, as expected for a transformed embryonic cell-type. Our results demonstrate that 5'UTR ORFs and their potential regulatory roles are tissue specific. In addition to potential regulatory functions, many of the newly discovered ORFs are small and predicted to encode short peptides. These peptides are likely secreted into the bloodstream and could be potential biomarkers for AD. Further analysis of the data produced in this study would provide important understanding of brain biology and the etiology of AD.

1600W

5-hydroxymethylcytosine alterations in the human postmortem brains of autism spectrum disorder. *Y. Cheng[1], Z. Li[2], L. Lin[1], S. Manupipatpong[1], T. Xu[2], H. Wu[2], Y. Jiang[3], P. Jin[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; 2) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA;; 3) Department of Pediatrics, Duke University School of Medicine, Durham, NC 27710, USA.

Autism spectrum disorders (ASDs) comprise a group of disorders characterized by impaired language, social, and communication skills, in addition to restrictive behaviors or stereotypies. Symptoms begin to manifest in early development and persist throughout the affected individual's life. However, with a prevalence of 1.5% in developed countries and relatively high comorbidity rates, no clear underlying mechanism that unifies the heterogeneous phenotypes of ASD is known. Cytosine methylation has been shown to serve as a critical epigenetic mark in gene regulation and 5-methylcytosine (5mC) was viewed as a stable and long-lasting covalent modification to DNA. However, it has been shown that 5mC could be enzymatically modified to 5-hydroxymethylcytosine (5hmC) by Tet family proteins through Fe(II) a-ketoglutarate-dependent hydroxylation, resulting in a new perspective on the previously observed plasticity in 5mC-dependent regulatory processes. 5hmC has been reported highly enriched in the central nervous system, and is now recognized as an important epigenetic mark in developmental and neurological disorders. To explore the role of 5hmC in ASD, we utilized the genomic DNA isolated from the postmortem cerebellum of both ASD patients and age-matched controls to profile genome-wide distribution of 5hmC. Compared to adult samples, we identified many more differential hydroxymethylated regions (DhMRs) (232 vs 58; q-value < 0.05, FDR adjusted) in young group (age < 18). No significant DhMR was identified in the group over 18-year-old, suggesting the dynamic change of 5hmC could play important roles in neurodevelopment and ASD pathogenesis. Pathway and disease association analyses indicated that the DhMRs identified in young group were located in the genes that involved in axon guidance (q-value = 0.0030) and Williams Syndrome (q-value = 0.0279). These results together suggest that 5hmC dynamic alteration is associated with ASD, particularly in the early development stage, and could contribute to the pathogenesis of ASD.

1601T

Genome-wide analyses of DNA methylation in autism brains suggest epigenetic-mediated dysfunction in GABA signaling. *J.I. Young[1], A. Siladi[1], C. Brandenburg[2], J.M. Vance[1], M.A. Pericak-Vance[1], G.J. Blatt[2].* 1) Hussman Inst Human Genomics, Univ Miami, Miami, FL; 2) Hussman Institute for Autism, Baltimore MD, USA.

Autism is characterized by social deficits, impaired communication and repetitive behaviors. Ample evidence supports involvement of the striatum and cerebellum in autism pathophysiology. Emerging evidence implicates altered DNA methylation as a likely factor contributing to autism etiology. To gain understanding on the role of DNA methylation and transcriptional regulation in autism, we profiled the methylome and transcriptome of striatal and cerebellar tissue from autistic and control individuals. DNA methylation patterns were interrogated by use of the EPIC array. This analysis identified isolated CpGs with significant methylation differences between postmortem autism cases compared to age- and post mortem interval matched controls. Further, we identified significant differentially methylated regions (DMRs) in both tissues in genes previously implicated in autism through genetic studies. Pathway analysis of genes with DMRs in the striatum shows significant enrichment in the pathway "GABAergic synapse", whereas the pathway "endocytosis" is enriched in the cerebellum. Both pathways have been postulated to contribute to autism pathophysiology. RNA-seq based analysis identified differentially expressed genes (DEG) in both brain regions, including several that have been implicated in autism. Pathway analysis found that striatal DEG genes were enriched for the pathway "Neuroactive ligand-receptor interaction" and the category "GABA-A receptor activity". Thus, analysis of the methylome and transcriptome suggested that autism is associated with transcriptional derangement of GABAergic neurotransmission in the striatum leading to altered inhibitory signaling to cortical and cerebellar areas likely affecting autism-related behaviors. Comparison of the genes containing DMRs and DEG in the striatum suggests direct epigenetic regulation of transcription through DNA methylation in autism. In addition, transcription factor binding site enrichment analysis revealed enrichment of PAX6 binding to striatal DEG genes. Our data showing that PAX6 contains a DMR in the striatum in autism individuals suggest epigenetic indirect transcriptional regulation mediated by PAX6. Thus, in spite of genetic and phenotypic heterogeneity in autism, shared molecular alterations could be uncovered by smaller sample sets than those required for genetic studies. We identified tissue-specific epigenomic alterations in autism brains, including alterations in known autism-associated genes. .

1602F

A massively parallel reporter assay for variants associated with schizophrenia and Alzheimer's disease. *D. Avramopoulos, L. Myint, R. Wang, L. Boukas, K. Hansen, L. Goff.* Johns Hopkins University, Baltimore, MD.

Genome wide association studies (GWAS) have reported on numerous loci for complex disorders including schizophrenia (SZ) and Alzheimer's disease (AD), most located outside protein coding regions. We hypothesize that some of these loci act through differential enhancer activity and we test this hypothesis by systematically assaying each locus using a Massively Parallel Reporter Assay (MPRA). Specifically, we investigate all variants in high LD with 71 SZ and 8 AD loci, for a total of 1083 variants. For each variant we designed a 95 bp sequence centered on the variant and linked it to 5 distinct barcodes, including positive and negative controls. We generated a plasmid library using standard protocols, with final oligo representation of 80% (85% of alleles represented by ≥ 3 oligos) and transfected it 3 independent times into K562 and SHSY5Y cells. Following transfection we extracted RNA and DNA and assayed the transfection efficiency and technical reproducibility. Specifically we found that on average, 80% of barcodes were represented in DNA and 72% in RNA. After summing counts over barcodes, we compute the correlation of counts and activity measures (log ratio of RNA over DNA) across replicates and find high mean correlations: 0.94 - 0.99 for both DNA and RNA. Significant overlaps with open chromatin marks validate our results. We find 116 SNPs with significant differences in driving expression between the two alleles including 5 observed in both cell lines. More specific results of our assay for SZ and AD will be presented.

1603W

Gene body methylation of tyrosine hydroxylase (*TH*) in the striatum is associated with cocaine dependence in humans. *K. Vaillancourt[1,2], C. Ernst[1,3], G. Chen[1], A. Bramoulle[1], J-F. Théroux[1], L. Fiori[1], G. Maussion[1], E. Calipari[4], B. Labonté[4], E. Nestler[4], D.C. Mash[5], G. Turecki[1,2,3].* 1) McGill Group for Suicide Studies, Douglas Hospital Research Center, Verdun, Quebec, Canada; 2) Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Icahn School of Medicine, Mount Sinai Hospital, New York, NY, USA; 5) Department of Neurology, University of Miami Miller School of Medicine, Miami, FL, USA.

Background: Cocaine dependence is a chronic relapsing disorder with widespread biological consequences, including epigenetic alterations, in numerous brain areas. In particular, the striatum is a highly addiction-relevant structure that has been investigated in animal models of cocaine seeking behaviors and epigenetics. Of particular interest is DNA methylation as it may mediate the long-term effects of chronic cocaine abuse on brain cell functioning. While research has begun to identify diverse sites of hyper- and hypo-methylation in rodents, little is known about the relationship between cocaine dependence and DNA methylation in the context of the human disorder. **Methods:** We used Reduced Representation Bisulfite Sequencing (RRBS) to identify clusters of differentially methylated CpGs (DMRs), in post mortem tissue from 25 individuals with cocaine dependence and 25 drug-naïve controls. We focused on two striatal sub-regions, the nucleus accumbens (NAc) and caudate nucleus (CD), and validated a DMR within the tyrosine hydroxylase (*TH*) gene using targeted amplicon bisulfite sequencing. This method was also used on an independent cohort of human samples (n=18 per group) and on neuronal and non-neuronal nuclei separated by fluorescence activated cell sorting (FACS). In addition, we used cocaine self-administering mice, transcriptome sequencing, and *in vitro* experiments to investigate the relationship between methylation, expression and behavior at this locus. **Results:** We found numerous DMRs in both brain regions, including within exon 8/9 of *TH* that are more methylated in the cocaine group. Methylation negatively correlates with *TH* expression in the CD of the cocaine group only. We replicated this effect in the striatum of an independent cohort (N=18 per group), and of chronically self-administering mice (N=8-10 per group). This hypermethylation appears to be neuron-specific and impedes enhancer activity at this locus. **Conclusions:** Hypermethylation of *TH* is associated with chronic drug seeking behavior and may have regulatory potential. Our ongoing research uses epigenome editing techniques to understand the mechanisms through which methylation regulates *TH* transcription. Ultimately, work on this system will uncover the importance of epigenetic dysregulation of *TH* to the time course and trajectory of chronic cocaine dependence. Funded by NIDA (DA033684).

1604T

DNA methylation profiles in a cohort of Brazilian children with ADHD.
T.V.M.M. Costa[1,5], F.A. Marchi[2], C. Milani[1], J.G. Damasceno[1], F.A.R. Madia[1,3], E.A. Zanardo[1], G.M. Novo-Filho[1], M.M. Montenegro[1], A.T. Dias[1], A.M. Nascimento[1,3], M. Rocha[1], J.L. Mussolini[1], C.A. Kim[3], V. Schuch[4], C.B. Mello[4], M. Muszkat[4], L.D. Kulikowski[1,3]. 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da USP, São Paulo, Brasil; 2) Laboratório NeoGene, AC Camargo Cancer Center, São Paulo, Brasil; 3) Laboratório de Citogenômica, Unidade de Genética, Departamento de Pediatria, Instituto da Criança, HC/FMUSP, São Paulo, Brasil; 4) Núcleo de Atendimento Neuropsicológico Infantil Interdisciplinar (NANI), Departamento de Psicobiologia da Escola Paulista de Medicina, UNIFESP, São Paulo, Brasil; 5) Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA), Campus Barreirinhas, Barreirinhas, Brasil.

Attention deficit hyperactivity disorder (ADHD) is a heterogeneous behavioral syndrome characterized by inadequate levels of hyperactivity/impulsivity and/ or inattention. ADHD begins in childhood and affects around 5% of school children worldwide and, in most cases (50-80%), the disorder persists during adolescence and adulthood. High heritability (estimated at 76%) suggests that genetic factors strongly influence the etiology of ADHD. However, the genetic architecture of ADHD is not clear. In addition, some environmental factors are epidemiologically associated with ADHD. Therefore, it is possible that the actions of the environmental factors associated with ADHD are mediated by epigenetic mechanisms. In this sense, we studied the DNA methylation profiles in a cohort of Brazilian children with ADHD. Diagnosis was performed according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). DNA was extracted from blood lymphocytes of ADHD children (n=29) and children with neurotypical development (n=11). After, DNA methylation profiles of all children were performed using Infinium HumanMethylation450 BeadChip (Illumina). Arrays data were treated and analyzed using specific packages in the R environment. The probes were annotated according to the data provided by Illumina using the genome hg19 reference. Considering $p \leq 0.05$, we obtained 121 differentially methylated probes between the children groups (with and without ADHD). Also, when we compared the samples by gender, we obtained two lists of differentially methylated probes: a list with 50 hypomethylated probes and a list with 50 hypermethylated probes. The classification of the probes, in both lists, was performed according to the *p*-values. Our results highlight the importance to perform epigenetic studies in ADHD, provide novel insights to genetic architecture in this disorder and also allows to envision how these can be transformed into future effective treatments. Grants: FAPESP: 2015/05350-5, 2014/02565-8 and FINEP-CT INFRA 0160/12 SP8.

1605F

A reference map for open chromatin-associated histone methylation and acetylation landscapes in the human frontal lobe.
K.G. Girdhar[1], GH. Hoffman[1], YG. Jiang[1], MK. Kundakovic[1], LB. Brown[1], LZ. Zharovsky[1], RJ. Jacobov[1], JW. Wiseman[1], RP. Park[1], SKS. Sieberts[3], MP. Peters[3], BH. Harris[4], BL. Lipska[4], PS. Sklar[1], PR. Roussos[1,2], SA. Akbarian[1,2]. 1) Division of Psychiatric Genomic, Icahn School of Medicine at Mount Sinai, NYC, NY; 2) Department of Psychiatry and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Systems Biology, Sage Bionetworks, Seattle, WA; 4) NIMH Neuroscience Center at St. Elizabeth's Hospital, Washington DC, Maryland.

We present for the dorsolateral prefrontal and anterior cingulate cortex, two frontal lobe areas broadly implicated in cognition and neuropsychiatric disease, cell-, subject- and region-specific reference maps for two histone marks associated with transcription and open chromatin, H3-trimethyl-lysine 4 (H3K4me3) and H3-acetyl-lysine 27 (H3K27ac). Exploration of this to date largest histone ChIP-seq resource for the human brain, with 128 datasets collected from 17 individuals, revealed cell type and subject as the two key variables shaping histone modification landscapes in the human frontal lobe, but only minimal contributions by region- and sex-specific factors. We highlight regulatory mechanisms unique to neurons, including much higher (20% vs. 15%) genome-wide coverage for H3K27ac and broader histone methylation and acetylation peaks at proximal promoters and transcription start sites, and strong overlap with the genetic risk architectures of schizophrenia, depression and neuroticism. These findings highlight the utility of this PsychENCODE NIMH-sponsored resource to explore epigenomic architectures of regulatory and disease-associated non-coding DNA in the human brain.

1606W

Convergence analysis on risks for schizophrenia by integrating genomics, DNA methylation and gene expression. *D. Lin[1], J. Chen[1], J. Bustillo[2,4], N. Perrone-Bizzozero[2,4], J. Sui[1], Y. Du[1], V. Calhoun[1,2,3], J. Liu[1,3].* 1) The Mind Research Network, Albuquerque, NM; 2) Dept. of Neurosciences, University of New Mexico, Albuquerque,NM; 3) Dept. of Electronic and Computer Engineering, University of New Mexico, Albuquerque,NM; 4) Dept. of Psychiatry, University of New Mexico, Albuquerque, NM.

Schizophrenia (SZ) is a complex mental disorder with estimated heritability of ~80% and lifetime prevalence of 1%. Substantial success has been made to explore genetic risk loci of SZ by international consortium. Yet the precise causal relations and exact mechanisms remain to be elucidated. By integrating cellular level variation such as DNA methylation and gene expression with sequence risks we are able to tap into pathologic mechanism of identified risk loci, and thus shed light on the true causal relation. DNA methylation and gene expression are known to be tissue specific. To increase the confidence on risk SNPs effect, and test the surrogate ability of peripheral tissues for the hardly accessible brain tissues, we focused on tissue independent genetic effect on DNA methylation and gene expression across brain, blood and/or saliva. First we investigated genomic methylation quantitative trait loci (meQTLs), and characterized consistent cis-acting effects (<20k base pairs) cross brain, blood and saliva. Then we identified cross-tissue cis-acting (within 1M bp) expression QTLs (eQTL) from brain and blood GTEx project. Enrichment test was applied to cis-meQTLs on cis-eQTLs with consideration of minor allele frequency and Linkage disequilibrium. Finally, we matched QTLs regulating both methylation and expression with the reported SZ risk loci to identify potential pathways from risk SNPs to SZ through epigenetics and gene expression. We identified tissue independent 116,005 meQTLs, 10879 target CpGs, and 39,653 eQTLs. 7372 SNPs were both eQTLs(18.59%) and meQTLs(6.35%), showing significant enrichment of meQTLs in eQTLs (odds ratio = 16.21, P_permutation < 1e-5, P_Fisher exact < 1e-200) compared to non-meQTLs. Interestingly, the overlapped eQTLs and meQTLs, mainly located on chromosome 6p21.1 – 6p24.3 regions, showed significant SZ risk with p-value<1e-5. In particular, SNPs in genes BTN3A2, ITIH4 and HLA-DQA strongly regulate both methylation of their promoter and gene expression, suggesting potential functional pathways.

1607T

The role of DNA methylation and the 5-HTTLPR long/short variant of the serotonin transporter gene (*SLC6A4*) in antidepressant treatment response. *A.J. Lisoway[1,2], C.C. Zai[1,3,4], A.K. Tiwari[1,3], R. Harripaul[1,2], N. Freeman[1], L. French[2,3,5], Z.A. Kaminsky[6], J.L. Kennedy[1,2,3].* 1) Neurogenetics Section, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON; 2) Institute of Medical Science, University of Toronto; 3) Department of Psychiatry, University of Toronto; 4) Department of Laboratory Medicine and Pathobiology, University of Toronto; 5) Computational Neurobiology Lab, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON; 6) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD.

Background: The current process used to prescribe antidepressant medication is markedly inefficient, as more than 50% of treated patients fail to reach remission. Antidepressant treatment success has been associated with genetic variation, but studies are not well replicated and epigenetic mechanisms remain under investigated. DNA methylation may provide more information regarding antidepressant response and guide physicians in choosing the most effective medication for each patient. We investigated the influence of *SLC6A4* methylation and 5-HTTLPR genotypes on antidepressant response in our sample. **Methods:** A subset of depression patients with complete clinical data (European Caucasian, n=166), were selected from our discovery sample (IMPACT; N=8,000). Saliva samples were collected at enrollment and genetic testing was done of cytochrome P450 liver enzyme, 5HTTLPR and *5HT2A* variants to guide clinical choice and dosage of psychotropic medications. DNA was bisulfite converted and methylation profiles were interrogated using the Infinium HumanMethylation450 Beadchip array. Nine CpG probes located across the *SLC6A4* promoter region were selected, based on previous studies, to quantify methylation levels. Change in Beck Depression Inventory II (BDI-II) score was used to measure antidepressant response over eight weeks. The relationship among genotype, methylation levels, and antidepressant response was analysed using linear regression modeling. **Results:** Increased methylation in the promoter region of the *SLC6A4* gene was nominally associated with impaired response to antidepressants in our sample. The effect was driven by methylation site cg05016953, located in exon 1A. Increased methylation level as measured by CpG cg05016953 was associated with less improvement in BDI-II score over eight weeks of antidepressant treatment (F(3,163)=6.14, p_uncorr=0.012). 5-HTTLPR genotype was not associated with *SLC6A4* DNA methylation or antidepressant treatment response. **Conclusions:** *SLC6A4* DNA methylation at cg05016953 may be associated impaired response to antidepressant medications in our sample. The *SLC6A4* transcriptional control region contains a CpG island, previously shown to functionally influence mRNA levels depending on 5-HTTLPR genotypes. Typically, hypermethylation in promoter regions is associated with decreased gene expression. Further pharmaco-epigenetic studies are required to elucidate the effect of DNA methylation changes on antidepressant response.

1608F

Dynamic DNA N6-methyladenine modification in mammalian brain and implications in neuropsychiatric disorders. *B. Yao[1], Y. Cheng[1], Z. Wang[1,2], Y. Li[1], L. Chen[4], L. Huang[1], W. Zhang[3], D. Chen[3], H. Wu[4], B. Tang[2], P. Jin[1].* 1) Department of Human Genetics, Emory University, Atlanta, GA. USA; 2) Department of Neurology, Xiangya Hospital, Central South University, Changsha, Hunan, P.R. China; 3) Institute of Zoology, Chinese Academy of Sciences, Beijing, P.R. China;; 4) 4Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, GA, USA.

Chemical modifications on DNA molecules, such as 5-methylcytosine and 5-hydroxymethylcytosine, play important roles in mammalian brain. A novel DNA adenine modification, N6-methyladenine (N6mA), has been recently found present in mammalian cells. However, the presence and function(s) of N6mA in mammalian brain remain unclear. Here we demonstrate the dynamics of N6mA in mouse brain in response to environmental stress. We found that overall N6mA level was significantly elevated upon stress. Genome-wide N6mA and transcriptome profiling revealed an inverse association of N6mA dynamic changes with a set of upregulated neuronal genes or downregulated LINE transposon expression. Genes bearing stress-induced N6mA changes significantly overlap with loci associated with neuropsychiatric disorders. These results together suggest the epigenetic role of N6mA in mammalian brain as well as its potential involvement in neuropsychiatric disorders.

1609W

A multi-dimensional characterization of anxiety in monozygotic twin pairs reveals susceptibility loci in humans. *R.S. Alisch[1], C. Van Hulle[2], P. Chopra[3], R.J. Davidson[2,4], N.H. Kalin[1], H.H. Goldsmith[2,4].* 1) Department of Psychiatry, University of Wisconsin, Madison, Madison, WI; 2) Waisman Center, University of Wisconsin, Madison, WI; 3) Department of Human Genetics, Emory University, Atlanta, GA; 4) Department of Psychology, University of Wisconsin, Madison, WI.

The etiology of individual differences in human anxiousness is complex and includes contributions from genetic, epigenetic (i.e., DNA methylation) and environmental factors. Past genomic approaches have been limited in their ability to detect human anxiety-related differences in these factors. To overcome these limitations, we employed both a multi-dimensional characterization method to select monozygotic twin pairs discordant for anxiety, and whole genome DNA methylation sequencing. This approach revealed 230 anxiety-related differentially methylated loci that were annotated to 183 genes, including several known stress-related genes such as *NAV1*, *IGF2*, *GNAS*, and *CRTC1*. As an initial validation of these findings, we tested the significance of an overlap of these data with anxiety-related differentially methylated loci that we previously reported from a key neural circuit of anxiety (i.e., the central nucleus of the amygdala) in young monkeys and found a significant overlap (*P*-value < 0.05) of anxiety-related differentially methylated genes, including *GNAS*, *SYN3*, and *JAG2*. Annotation of the 183 anxiety-associated human loci to gene ontologies found several biologically relevant ontological terms that were among the differentially methylated genes, including regulation of dendrite development and neuron differentiation. Finally, sequence motif predictions of all the human differentially methylated regions indicated an enrichment of five transcription factor binding motifs, suggesting that DNA methylation may regulate gene expression by mediating transcription factor binding of these transcripts. Together, these data demonstrate environmentally sensitive factors that underlie the development of human anxiety.

1610T

Methylomic profiling and replication implicates deregulation of PCSK9 in alcohol use disorder. *F.W. Lohoff[1], J. Sorcher[1], A. Rosen[1], K. Mauro[1], R. Fanelli[1], R. Momenan[1], C. Hodgkinson[1], L. Vendruscolo[2], G. Koob[2], M. Schwandt[1], D. George[1], I. Jones[3], A. Holmes[1], Z. Zhou[1], M. Xu[1], B. Gao[1], H. Sun[1], M. Phillips[1], C. Muench[1], Z. Kaminsky[3].* 1) National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD; 2) National Institute on Drug Abuse, National Institutes of Health, Baltimore, MD; 3) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD.

Alcohol Use Disorder (AUD) is a common and chronic disorder with substantial effects on personal and public health. The underlying pathophysiology is poorly understood but strong evidence suggests significant roles of both genetic and epigenetic components. Given that alcohol affects many organ systems, we performed a cross-tissue and cross-phenotypic analysis of genome-wide methylomic variation using Illumina HM450 and EPIC chip arrays in AUD samples from 3 discovery, 4 replication, and 2 translational cohorts. The discovery samples consisted of postmortem brain tissues (n=46), bloods form a resting-state functional connectivity imaging endophenotypes (n=68) and postmortem brain tissues sorted into neuronal and non-neuronal cells (n=58). Overrepresentation analyses identified 68 significant CpG probes of which the most significantly associated probe cg01444643 was in in the promoter of the proprotein convertase subtilisin/kexin 9 (*PCSK9*) gene (p=0.002). Biological validation showed that *PCSK9* promoter methylation is conserved across tissues (brain-blood-liver) and positively correlated with expression. Replication in AUD datasets confirmed *PCSK9* hypomethylation (n=392, p<0.05) and a translational mouse model of AUD showed that alcohol exposure leads to *PCSK9* mRNA and protein downregulation (p<0.0001). Postmortem human liver tissue analyses in control (n=47) and liver transplant cases due to alcohol cirrhosis (n=50) showed marked increase of methylation at cg01444643 (P<0.0001) and significantly decreased PCSK9 expression (p<0.01). *PCSK9* is primarily expressed in the liver and regulates low density lipoprotein cholesterol (LDL-C). Our finding of alcohol-induced epigenetic regulation of *PCSK9* represents one of the underlying mechanisms between the well-known effects of alcohol on lipid metabolism and cardiovascular risk, with light alcohol use generally being protective while chronic heavy use has detrimental health outcomes.

1611F

Epigenome-wide association study of opioid dependence in European American women. *J.L. Montalvo-Ortiz[1], H.R. Kranzler[2,3], H. Zhao[4], H. Zhang[5], J. Gelernter[1,6].* 1) Division of Human Genetics, Department of Psychiatry, Yale School of Medicine, West Haven, CT; 2) Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Center for Studies of Addiction and Crescenz Veterans Affairs Medical Center; 4) Department of Biostatistics, Yale School of Public Health, New Haven, CT; 5) Departments of Psychiatry and Medicine (Biomedical Genetics) Boston University School of Medicine, Boston, MA; 6) Veterans Affairs Connecticut Healthcare Center, West Haven, CT.

Opioid dependence (OD) is currently epidemic in the US. Although OD has a higher prevalence in men, fatal opioid overdoses have increased at a higher rate among women. Epigenetic mechanisms have been implicated in the increased risk for OD, however, most studies to date have used candidate gene approaches, mainly focusing on the opioid receptor mu 1 (*OPRM1*) gene. In this study, we conducted the first epigenome-wide association study (EWAS) of OD in women. DNA was derived from whole blood samples and EWAS was assessed using the Illumina Infinium HumanMethylationEPIC array. Our sample included 111 European American women who were administered the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), which yields a DSM-IV diagnosis of OD. To identify differentially methylated CpG sites, we performed an association analysis using the 'cpg.assoc' function from the minfi Bioconductor R package, adjusting for age, estimates of cell proportions, and the first 3 principal components estimated using the Barfield et al. method to correct for population stratification. Association analysis identified a significant differentially methylated CpG at cg19642402 ($p = 4.6 \times 10^{-8}$, FDR = 0.03) located within the proline-serine-threonine phosphatase interacting protein 1 (*PSTPIP1*) gene. *PSTPIP1* is involved in inflammatory responses. When the top 10 ranked differentially methylated CpGs were used in a gene-based functional enrichment analysis, "response to stress" emerged as the top significant GO biological process ($p = 7.35 \times 10^{-6}$, FDR = 0.005). If replicated, our results would suggest that changes in DNA methylation at *PSTPIP1* may play a role in OD in women and implicates stress response as a potential mechanism of that effect. Two limitations of this study are the modest sample size and the use of blood for DNA extraction given the tissue-specific DNA methylation changes.

1612W

EGR family genes; new potential markers for etiology and symptoms' severity of schizophrenia. *M. Amini faskhodi[1], R. Mahdian[2], A. Haghighatfard[3], M. Hashemi[1].* 1) Department of Biology,Tehran Medical Branch, Islamic Azad University,Tehran, Iran, Tehran, Iran; 2) Department of Medical Biotechnology, Pasture Institute, Tehran, Iran; 3) Department of Biology, Tehran North Branch, Islamic Azad University, Tehran, Iran.

Schizophrenia (SCZ) is a major psychiatric disorder with 1 percent prevalence and unknown etiology with acute and chronic positive, negative and cognitive symptoms. Early growth response (EGR) family genes including EGR1 to EGR4 are transcription factors that are essential for development and function of central and peripheral nervous system. Previous studies were reported involvement of EGR1 and EGR3 with etiology of schizophrenia and addiction. mRNA level of EGR1, EGR2, EGR3 and EGR4 were studied in blood samples of 150 Iranian schizophrenic patients and 50 non-psychiatric subjects using quantitative Real time PCR. Also positive and negative symptom scale (PANSS) from patients and Wechsler Adult Intelligence Scale (WAISE) and NEO-FFI personality test from all subjects were obtained. Significant down regulation of EGR1, EGR2 and EGR3 were detected in SCZ vs. non-psychiatrics. No significant expression alteration was revealed in EGR4 between SCZ vs. non-psychiatrics. Correlations were found between EGR1 and EGR2 down expression and higher scores of negative symptoms in PANSS and memory deficit in digit span memory subscale of WAISE. In addition significant correlation was found between neuroticism personality factor and EGR1 down regulation. Results were suggested EGR family role in etiology of schizophrenia. In addition findings indicate down regulation of transcription factors may lead to negative symptoms, memory deficiency and higher neuroticism in patients. EGR family genes may present as biomarker for schizophrenia.

1613T

Transcriptional profiling of long noncoding RNA in PTSD patients reveals a potential early biomarker of trauma-induced alterations in the acute phase after exposure. *G. Guffanti[1], A. Wingo[2], T. Jovanovic[2], S. Maddox[1], S. Sharma[1], C. Nemeroff[6], A. Myers[3], K. Ressler[1].* 1) Harvard Medical School, Belmont, MA; 2) Emory University, Atlanta, GA; 3) University of Miami, Miami, FL.

Though relatively few long noncoding RNAs (lncRNAs) have been fully functionally characterized yet, there is growing evidence on their well-adapted and highly specialized biological role. Transcriptome analyses have demonstrated that lncRNAs are actively transcribed in a much more condition-specific manner than protein coding genes, underlying their potential as biomarkers of early detection. Finding lncRNAs that are associated with development of PTSD in the aftermath of a traumatic event is a key step toward identifying biomarkers of early diagnosis of stress- and trauma-related outcomes. We set up to test the hypothesis that lncRNA regulation might be perturbed in PTSD in a longitudinal cohort of individuals highly traumatized. The study sample derives from the larger ongoing collaboration between the University of Emory (Atlanta, GA) and Miami (FL) for the analysis of transcriptome of peripheral blood of individuals highly traumatized admitted at the Emergency Department. We decided to arbitrarily divide the overall sample in two subsamples, called discovery (n = 228) and replication (n = 85), respectively, based on the timing of the analysis of their RNA. Only ~7% of the participants developed PTSD in the aftermath of traumatic events. Raw paired end RNAseq data were processed using ad hoc pipeline PLAR to identify and annotate novel and already established lncRNAs. We tested for statistical significance of observed changes in lncRNA expression between PTSD patients and trauma-exposed controls using FPKM, correcting for demographic and batch effect variables. We used the individual level of expression of a set of genes that were significantly perturbed both in discovery and replication samples to generate lncRNA-based profile to classify cases and controls in discovery sample. We found 67 lncRNAs and other noncoding RNAs significant in both the discovery and replication samples at the significance threshold of p < 0.05. The profile generated based on the individual level of expression of these signature lncRNAs classifies cases and controls with ~98% accuracy. Our analysis shows that lncRNA-based profile provides a potential biomarker for individual "personalized" perturbation. These findings support the original hypothesis that lncRNAs play a role in PTSD and are regulated by trauma exposure. The mechanism by which signature lncRNAs perform regulatory function remains to be elucidated and is currently investigated in PTSD-related animal models.

1614F

DNA methylation as a candidate biomarker for predicting antidepressant response. *C. Ju[1], R. Belzeaux[1], L.M. Fiori[1], J.F. Theroux[1], J. Foster[2], S. Kennedy[2], CANBIND. Working group[3], G. Turecki[1,2].* 1) McGill Group for Suicide Studies, Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; 2) Department of Psychiatry, University Health Network, University of Toronto, Toronto, Ontario, Canada; 3) Canadian Biomarker Integration Network In Depression.

Purpose: Major depressive disorder (MDD) is a severe diagnosis and recognized as the leading cause of global disability. Antidepressant therapy (ADT) is the standardized first-line treatment, but an estimated 40% of patients fail to respond initially, and 20-30% do not respond after multiple interventions. Identifying a biomarker for ADT response is enticing, as it would promote evidence-based clinical guidelines. MDD has a heterogeneous phenotype, and our lab has shown the importance of epigenetic considerations when pursuing candidate biomarkers. DNA methylation, particularly at CpG dinucleotide sites, is an epigenetic mark that serves as a suitable candidate for biomarker studies given its stability and accuracy, even when retrieved from indirect tissue, such as peripheral blood. Within a well-characterized, clinical cohort, I will investigate differentially methylated positions (DMPs) between responders and non-responders of ADT, and assess if they are suitable biomarkers for predicting response. **Methods:** Peripheral blood samples were obtained from 206 MDD patients and 101 healthy controls (HCs) recruited under the CAN-BIND-1 clinical trial. MDD patients were subjected to 8 weeks of escitalopram ADT. At week 0 (pre-ADT), and week 8 (post-ADT), MDD patients received Montgomery-Asberg Depression Rating Scale (MADRS) scores for symptom severity. A patient was considered a responder (RES) if they had a ≥ 50% MADRS score decrease, and a non-responder (NRES) otherwise. Genome-wide methylation analysis was conducted using Illumina's EPIC microarray platform. Array data was pre-processed, and analyzed with the *ChAMP* Bioconductor package in R. Differential methylation analysis was performed between (i) HCs and MDD patients and (ii) ADT RES and NRES. Furthermore, week 0 genome wide expression data was acquired from the same cohort using Illumina's HT-12 Beadchip platform. DMPs found between RES and NRES that overlap with differential gene expression data will be subject to targeted validation using bisulfite amplicon sequencing. **Results:** 39,727 DMPs were identified between MDD patients and HCs, and 9,551 DMPs were identified between RES and NRES (p<0.05, FDR=0.05). 817 genes were found to be differentially expressed between RES and NRES (p <0.05). After overlap analysis, 3 genes (*CHN2, FAM24B,* and *JAK2*) were found to have both differential expression, and DMPs within their promoter regions. These findings are specific to ADT response. .

1615W

DNA methylation markers associated with injection drug use status and HIV infection among chronic injection drug users in the ALIVE study. *C. Shu[1], K.M. Bakulski[2], K.S. Benke[1], A.E. Jaffe[3], S. Wang[1], S.H. Sabunciyan[1], S.H. Mehta[1], G.D. Kirk[1], B.S. Maher[1].* 1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 2) University of Michigan, Ann Arbor, MI, USA; 3) Lieber Institute for Brain Development, Baltimore, MD, USA.

Background: Injection drug use (IDU) is associated with biological modification of the genome, i.e., epigenetics marks such as DNA methylation and histone modification. However, the biological mechanisms of how substance use and cessation affects epigenetic outcomes remain largely unknown. **Objective:** To conduct the epigenome-wide association analyses among injection drug users to identify whether injection drug use and HIV status during the past six months is associated with genome-wide blood epigenetic changes. **Methods:** In the AIDS Linked to the Intravenous Experience (ALIVE) study, blood was obtained from 288 current IDUs, resampled after cessation and then again after relapse (total samples = 774). Blood DNA methylation marks were measured using the Illumina Infinium MethylationEPIC BeadChip. Standard procedures in the minfi R package were used to preprocess raw Illumina image files into noob processed methylation beta values. Differences in DNA methylation at individual probes by current injection status were tested using generalized linear regression including age, gender, race and cell heterogeneity as covariates. DNA methylation age was estimated for study subjects using the epigenetic clock method developed by Horvath et al. 2013. **Results:** DNA methylation at individual loci (cg10801015, $p=2.47*10^{-11}$; cg11415166, $p=5.15*10^{-10}$; cg03426703, $p=2.62*10^{-9}$; cg14977491, $p=9.94*10^{-8}$) is significantly associated with current injection drug use status after correction for multiple testing. Those CpG sites were near the PDP1, NARFL, DVL2 and PFN2 genes correspondingly. Single-site epigenetic analysis of HIV status was also performed. HIV positive individual's average biological age is about 3 years older than their chronological age, but the biological age among HIV negative individuals, injection drug users and non-users is no different from their chronological age. **Conclusion:** In a preliminary study, we performed a genome-wide scan of methylation changes in a longitudinal study of injection drug use and identified genomic locations exhibiting significant changes in peripheral DNA methylation associated with injection drug use status. Individuals with HIV infection's biological age is older than their chronological age, which is consistent with the literature.

1616T

Small non-coding RNAs in major depression and antidepressant response. *R. Lin[1], J.P. Lopez[1], L. Fiori[1], R. Belzeaux[1], C. Cruceanu[1], J.F. Theroux[1], CANBIND. Working group[3], J. Foster[2], S. Kennedy[2], G. Turecki[1].* 1) Department of Psychiatry, McGill Group for Suicide Studies, Douglas Mental Health University Institute, McGill University, Montreal, Quebec, Canada; 2) Department of Psychiatry, University Health Network, University of Toronto, Toronto, Ontario, Canada; 3) Canadian Biomarker Integration Network in Depression.

  Statement of Purpose: Major depressive disorder (MDD) is a prevalent disorder treated primarily by antidepressants (ADT). Although effective, on average 30%-40% of subjects experience an inadequate response to treatment after several attempts. Thus, there is a great need to identify biomarkers associated with MDD and ADT response. Recent discoveries have pointed towards small non-coding RNAs (sncRNAs) as feasible biomarkers, as they have shown to be stably circulated in peripheral blood and may be associated with a disease state. Here, I profile sncRNAs in MDD pathology, and evaluate their ability to act as candidate biomarkers for diagnosis and/or ADT treatment response. This project addresses these questions with a unique combinatorial approach of using both post-mortem human brain tissue samples and living patient blood samples. **Methods:** Human post-mortem brain samples were obtained the Douglas-Bell Canada Brain Bank. Brain tissue will be collected from individuals who died by suicide during an episode of MDD, and psychiatrically normal individuals who died suddenly. Human peripheral blood samples were collected from a clinical trial using duloxetine (samples provided by Lundbeck). This cohort consisted of depressed subjects treated with duloxetine for 8 weeks. All subjects were assessed for depression severity based on the Montgomery-Asberg Depression Rating Scale (MADRS) at week 0 (pre-ADT treatment) and week 8 (post-ADT treatment). After concluding 8-weeks of ADT therapy, subjects were separated into responders (>50% decrease in MARDS score) or non-responders (<50% decrease in MARDS score) of ADT treatment. Small RNA-sequencing was used to generate expression data on whole sncRNA-transcriptome in post-mortem brain and peripheral blood of MDD subjects. **Results:** 37 snoRNAs and 6 piRNAs showed significant up-regulation in MDD brains compared to healthy controls. From this study snoRNAs and piRNAs do not show predictive value for ADT response. snoRNAs and piRNAs only showed differential expression across time-points (i.e. week 0 vs week 8) in responders of duloxetine treatment. 31 snoRNAs showed significant up-regulation in responders of duloxetine. 3 piRNA showed significant up-regulation and 4 piRNAs showed significant down-regulation in responders of duloxetine after treatment. Significant results were specific to responders for both snoRNAs and piRNAs.

1617F

Transcriptional and genetic changes underlying psychiatric disorders converge on a network of transcription factors and their target genes in the human brain. *S.A. Ament[1], J.R. Pearl[2,3], D. Bergey[2], C. Funk[2], L. Hood[2], C. Colantuoni[1,4], N.D. Price[2].* 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; 2) Institute for Systems Biology, Seattle, WA; 3) Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA; 4) Departments of Neurology and Neuroscience, Johns Hopkins University School of Medicine.

  Genetic and genomic studies suggest that psychiatric disorders involve changes in brain gene regulation, but roles for specific transcription factors remain largely uncharacterized. We reconstructed a transcriptional regulatory network (TRN) model for the human brain by integrating human brain-specific DNase-seq and TF-gene co-expression, predicting the genome-wide binding sites and target genes for 741 TFs. We used our TRN model to predict TFs that regulate prefrontal cortex gene expression changes in psychiatric and neurodegenerative diseases, as well as functional, disease-associated SNPs that disrupt transcription factor binding sites. Our model predicted that 69 TFs are key regulators of prefrontal cortex gene expression changes in schizophrenia (SCZ), 13 of which were also predicted to be key regulators of gene expression changes in bipolar disorder (BD). Notably, several of these predicted key regulator TFs *are* located at genetic risk loci for SCZ or BD, including *POU3F2*. We over-expressed POU3F2 in primary human neural stem cells to validate our network predictions and characterize its biological functions. We show that POU3F2 has anti-proliferative effects, involving transcriptional repression of a cluster of cell cycle genes that is highly expressed in proliferative neural stem cells and in adult astrocytes, and which is up-regulated in the prefrontal cortex of individuals with SCZ and BD. In addition, we found that a BD- and SCZ-associated SNP in the promoter of *VRK2* disrupts a putative POU3F2 binding site and alters the activity of the *VRK2* promoter. Our results suggest that genetic and transcriptional changes underlying psychiatric diseases converge on a shared transcriptional regulatory network in the brain, with key regulators linked to disease through both cis- and trans-acting mechanisms.

1618W

A direct regulatory link between microRNA miR-137 and *SHANK2* with implications for neurodevelopmental disorders. *S. Berkel[1], A. de Sena Cortabitarte[1], C. Fischer[1], G.A. Rappold[1,2].* 1) Institute of Human Genetics Heidelberg, University of Heidelberg, Heidelberg, Germany; 2) Interdisciplinary Center for Neuroscience (IZN), University of Heidelberg, Heidelberg, Germany.

Mutations in the genes coding for the postsynaptic scaffolding proteins SHANK2 and SHANK3 have been linked to a spectrum of neurodevelopmental disorders including intellectual disability, autism spectrum disorders and schizophrenia. We have carried out an *in silico* analysis of the *SHANK* gene family for microRNA binding sites which revealed a putative binding site for the schizophrenia GWAS top hit miR-137 in the *SHANK2*-3´UTR. Using human neuroblastoma cells (SH-SY5Y) and mouse primary hippocampal neurons, we now show that miR-137 targets the 3´UTR of *SHANK2* in a site-specific manner. Overexpression of miR-137 significantly lowered endogenous Shank2 protein levels without detectable influence on mRNA level; conversely, the inhibition of miR-137 resulted in the increase of Shank2 protein. We conclude that miR-137 regulates *SHANK2* expression and that this takes place by repression of protein translation rather than mRNA degradation. To further support the link between miR-137 and schizophrenia, we compared the expression of mir-137 precursor and miR-137 target genes (including *SHANK2*) in the dorsolateral prefrontal cortex (DLPFC) of schizophrenia and control individuals using RNA-Seq data from the CommonMind Consortium. Almost a third (18/63; 29%) of the validated miR-137 target genes showed significant expression differences in the DLPFC of schizophrenia individuals compared to controls and we propose that further targets (like *SHANK2*) may be regulated on protein level. Our study provides first evidence of a direct regulatory link between miR-137 and *SHANK2* and supports the finding that miR-137 signaling is altered in schizophrenia.

1619T

G-quadruplex binding chemicals may ameliorate the cognitive function of ATR-X syndrome. *T. Wada[1], N. Shioda[2], H. Sugiyama[3], H. Tanabe[4], K. Kurosawa[5], N. Okamoto[6], K. Fukunaga[7].* 1) Medical Ethics and Medical Genetics, Kyoto Univ. Grad. Sch. Medicine, Kyoto, Kyoto, Japan; 2) Department of Biofunctional Analysis Laboratory of Molecular Biology, Gifu Pharmaceutical University, Gifu, Japan; 3) Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto, Japan; 4) Biosystems Science, School of Advanced Sciences, The Graduate University for Advanced Studies, Hayama, Japan; 5) Division of Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 6) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 7) Department of Pharmacology, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan.

Background; Disorchestration of epigenetic regulation of genes can cause many mendelian diseases, one of which is ATR-X syndrome due to mutations of *ATRX* gene, encoding a chromatin remodeling protein, ATRX. One of the strategies of treatment for this syndrome is to hunt for the genes which are regulated by ATRX, and to ameliorate the abnormally expressed genes. Recently, it has been reported that ATRX protein targets tandem repeats, forming G-quadruplex (G4) structures, and regulates genes nearby. This suggests that the expression of many target genes of ATRX protein with G4 structure should be perturbed in ATR-X patients. In this presentation, we show that a porphyrin analogue, TMPyP4, with G4 binding ability, could improve the cognitive function of Atrx mouse. Method and Results; Atrx mice, lacking Atrx exon 2, shows cognitive dysfunction as well as other several phenotypes in ATR-X syndrome, with abnormal dendritic spine morphology leading to aberrant CaMKII activity. [Nogami T. Hippocampus, 2011, and Shioda N. J Neuroscience, 2011]. After TMPyP4 treatment, the mice showed ameliorated cognitive function based on memory-related behavioral tests, including Novel-object recognition test, Latency time in retention trials in a passive avoidance test, Alternations in a Y-maze test. DNA microarray analysis in fore brain of wild-type and Atrx mice showed 260 out of 23474 genes were abnormally expressed in Atrx mice, and TMPyP4 normalized their expression in 92 out of 140 genes. Discussion; Recently the connections between G-quadruplexes (G4) and mendelian diseases have been established. Our findings suggest a potential therapeutic strategy targeting G-quadruplexes could be treatment for ATR-X syndrome, as well as for other G4-related diseases. We should find other safe agents with G4 binding ability other than TMPyP4, because of its toxicity. We are now studying the effectiveness of these compounds for neurons developed from ATR-X patients-derived iPS cells. We are searching a key gene for cognitive function with ATRX-regulated G4-structure.

1620F

An epigenome-wide association study of Williams syndrome. *R. Kimura[1], T. Awaya[1], M. Nakata[1], T. Kato[2], Y. Funabiki[3], K. Tomiwa[4], T. Heike[2], M. Hagiwara[1].* 1) Department of Anatomy and Developmental Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 2) Department of Pediatrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 3) Department of Cognitive and Behavioral Science, Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan; 4) Todaiji Ryoiku Hospital for Children, Nara, Japan.

Epigenetic alterations, including DNA methylation, have been implicated in many psychiatric and neurodevelopmental disorders. Williams Syndrome (WS) is rare genetic neurodevelopmental disorder and characterized by multiple symptoms including hypersociability. WS is caused by a 7q11.23 heterozygous deletion containing 26-28 genes, which include genes associated with epigenetic regulation. In this study, we performed epigenome-wide association study (EWAS) to determine the contribution of DNA methylation to WS. We examined a total of 171 subjects, including a discovery set (34 WS and 34 controls) and an independent confirmation set (56 WS and 47 controls). We used the Social Responsiveness Scale-2 (SRS-2), the Hyperacusis Questionnaire (HQ), the revised Fear Survey Schedule for Children (FSSC-R) and the sensory profile (SP) to evaluate behavioral and neuropsychiatric symptoms in WS. DNA methylation profiles were obtained from whole blood samples from a discovery set using the Illumina HumanMethylation 450 K BeadChip. Validation analyses of the epigenome-wide findings performed in a confirmation set using site-specific methylation with pyrosequencing. After accounting for batch effects, the effects of leukocyte subsets and multiple testing, we have identified multiple differentially methylated positions (DMPs) and regions (DMRs) consistently associated with WS in a discovery set. Some of the identified genes with differential methylation (ANKRD30B) have previously been associated with WS and we replicate those findings in a confirmation set. GO and Pathway analysis revealed significant enrichment of our DMPs in Synapse organization, Glutamatergic synapse and Calcium signaling pathway. To our knowledge, this study is the largest EWAS performed in WS. Our findings provide insights into the molecular basis and pathophysiology of the various phenotypes of WS.

1621W

Epigenetic suppression of VEGF in retinal pigment epithelial cells by ascorbate. *D. Sant[1], V. Camarena[1], S. Mustafi[1], C. Gustafson[1], Y. Li[2], L. Dennison[1], B. Hampton[1], Z. Wilkes[1], D. Van Booven[1], R. Wen[2], G. Wang[1,2,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami - Miller School of Medicine, Miami, FL; 2) Dr. John T. MacDonald Foundation Department of Human Genetics, University of Miami - Miller School of Medicine, Miami, FL; 3) Dr. Nasser Ibrahim Al-Rashid Orbital Vision Research Center, University of Miami - Miller School of Medicine, Miami, FL.

Age-related macular degeneration (AMD) is the leading cause of blindness in individuals over 50 years of age in the industrialized world. Intraocular anti-VEGF therapies are effective for many but not all cases of wet AMD, (advanced AMD, characterized by choroidal neovascularization.) The anti-VEGF therapies target VEGF protein, but do not influence the expression of VEGF. Ascorbate is known to influence DNA demethylation, through which it regulates gene expression. We examined the impact of ascorbate on RPE cells. In cultured ARPE-19 cells, treatment with ascorbate (50 mM) induced 5hmC generation and differential expression of 1,915 genes, as revealed by RNA-seq. Among which, *VEGFA*, a gene encoding VEGF protein, is one of the most downregulated genes. Consequently, the translation and secretion of VEGF are decreased by ascorbate treatment. The downregulation of VEGF expression by ascorbate was validated in RPE-J cells. Furthermore, we found that supplementation of ascorbate in *Gulo$^{-/-}$* mice, which are dependent on exogenous ascorbate, also downregulated VEGF levels in the vitreous humor. The suppression of *VEGFA* is independent of HIF-1a but corroborates with increased 5hmC at promoter of the gene. These results are consistent with the notion that ascorbate suppresses VEGF expression through 5hmC-mediated epigenetic pathway. Our findings provide experimental basis for using ascorbate to treat retinal angiogenesis diseases, such as wet AMD and diabetic retinopathy.

1622T

C-to-U RNA editing of osteopontin in mouse retina with laser-induced choroidal neovascularization. *J. Chen[1], Z. Xu[1], Z. Wei[1], R. Sun[1], X. Pan[2], D. Zhou[2].* 1) Laboratory of Genomic and Precision Medicine, Wuxi School of Medicine, Jiangnan University , Wuxi, Jiangsu, China; 2) Department of Ophthalmology, Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu, China.

   Purpose: Laser-induced choroidal neovascularization (CNV) in mice is an important animal model to study exudative age-related macular degeneration, but the role of RNA editing in the retina in CNV mice remain largely unclear. In this study we reported that C-to-U RNA editing of osteopontin, a key cytokine in cell-mediated inflammation, was found in the retina of a mouse CNV model. **Methods:** Laser photocoagulation of Bruch's membrane was used to induce CNV in the right eyes (OD) of three adult C57B/L6 mice. Total RNA was extracted from the whole retina of CNV (OD) and untreated contralateral (OS) eyes on Day 3. RNA-Seq was performed on the Illumina HiSeq 2000 platform with configuration of 20 million 100 bp pair-end reads per sample. Reads were aligned to the mouse full genome (UCSC version mm10) using Bowtie and splice junctions were identified by using TopHat2. Quantification of expression level and analysis of differential expression of RNA-Seq data were done using R language package EdgeR. Candidate RNA editing sites with editing level ≥ 1% were identified using Samtools and VarScan, and verified using Sanger sequencing. Expression level of mRNA was confirmed using quantitative real-time PCR (qPCR). **Results:** We focused on C-to-U RNA editing in protein-coding genes. The highest C-to-U RNA editing was found at the 3'-UTR of osteopontin mRNA, and was only in the retina with CNV (average editing level 6.69%) but not in the contralateral untreated retina. Sanger Sequencing of the osteopontin 3'-UTR confirmed that the editing was specific to CNV induction. RNA-Seq and qPCR results showed that expression of osteopontin dramatically increased in the retina with CNV, and intriguingly correlated with a higher expression level of C->U-editing enzyme *APOBEC1*. **Conclusions:** In the current study, RNA-Seq analysis revealed a novel target of C-to-U RNA editing in osteopontin in the retina of CNV mouse model. These findings thus warrant further investigation of the role played by RNA editing in retina inflammation during CNV.

1623F

DNA hypermethylation is associated nonsyndromic cleft lip and palate. *B. Gorijala[1], U. Ratnamala[2], D.D. Jhala[3], N.K. Jain[2,4], S.S. Chettiar[3], A.K. Maiti[5], S.K. Nath[6], U. Radhakrishna[7].* 1) Oncology Department, Krishna Institute of Medical Sciences, Secunderabad, Andhra Pradesh, India; 2) Department of Life Sciences, School of Sciences, Gujarat University, Ahmedabad, India; 3) Department of Zoology, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India; 4) Department of Pharmacology, Creighton University, Omaha, NE, USA; 5) Green Cross Pathology & Molecular Laboratory, Ahmedabad, India; 6) Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, US; 7) Department of Obstetrics and Gynaecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

   Non-syndromic cleft lip with or without cleft palate (NSCL/P) is considered the most common, severe craniofacial disorder worldwide. While most studies point to genetic factors as the major cause, NSCL/P has a multifactorial complex etiology and arises through the interaction of genetic and environmental factors. However, the exact mechanisms by which these environmental factors induce NSCL/P remain unknown. Epigenetics is among the most rapidly-expanding areas of study in molecular biology. Epigenetic studies that pertain to NSCL/P development are still relatively new but represent an important area for future study. To identify epigenetic risk variants for NSCL/P, we have conducted genome-wide DNA methylation profiles using the Infinium Methylation EPIC BeadChip (Illumina) in 24 NSCL/P cases and an equal number of age, ethnic and sex-matched controls from India. The samples included were children aged between 2 to 5 years. We identified significant CpG hypermethylation (> 2.0-fold) at 3500 sites in 3500 genes (false discovery rate (FDR) ≤0.000001) with a receiver operating curve area under the curve (ROC AUC) ≥0.75 for NSCL/P detection. The methylation difference between NSCL/P and controls ≥20%, in the majority of the genes. A total of 183 CpG loci had an excellent accuracy (AUC ≥ 1.00) for the detection of NSCL/P. We identified significantly altered CpG sites that are not only located within the CpG islands in the promoter region of genes but were also distributed throughout the gene bodies. Pathway analysis identified genes with significant methylation changes are involved in several important pathways including cancer, calcium signaling, MAPK signaling, signaling pathways regulating pluripotency of stem cells, and HIF-1 signaling pathway. These epigenetic studies may give us clues towards the elucidation of the disease mechanisms and pathogenesis of NSCL/P.

1624W

Developmental *cis*-regulatory elements revealed by open chromatin landscapes in mouse fetal tissues. *Y. Zhao[1,2], H. Huang[1,2], D. Gorkin[2], S. Preissl[2], Y. Zhang[2], R. Fang[1,2], V. Afzal[3], D. Dickel[3], A. Visel[3,4,5], L. Pennacchio[3,4], B. Ren[1,2].* 1) UC San Diego, La Jolla, CA; 2) Ludwig Institute for Cancer Research, La Jolla, CA; 3) Lawrence Berkeley National Laboratory, Berkeley, CA; 4) US DOE Joint Genome Institute, Walnut Creek, CA; 5) UC Merced, Merced, CA.

Mammalian development depends on coordinated regulation of gene expression in each cell type and tissue. In order to identify and characterize the *cis*-regulatory sequences and their dynamics during development, we carried out assays of transposase-accessible chromatin followed by sequencing (ATAC-seq) to delineate the open chromatin regions during mouse fetal development. We examined 12 distinct tissue types across up to 7 developmental stages, identified over 740,000 candidate developmental *cis*-regulatory sequences. We characterized their dynamic chromatin states among different tissues and developmental stages. We showed that promoter-distal elements exhibit highly tissue-specific and stage-specific dynamics in transposase accessibility relative to promoter-proximal regions. Integrating our catalog of *cis*-regulatory elements with matched gene expression data, we predict key regulators for lineage specification. Finally, we map 70.2% of our catalog to orthologous human loci, and find that disease-associated variants are enriched in a tissue-dependent manner. Together, our results provide a general resource for understanding mammalian development and human disease.

1625T

Disruption of a remote putative novel enhancer in the *cis*-regulatory domain of *FOXL2* in a multigenerational Polynesian family with BPES. *H. Verdin[1], A. Shelling[2], D. Markie[3], A.L. Vincent[4,5], E. De Baere[1].* 1) Center for Medical Genetics, Ghent University Hospital and Ghent University, Ghent, Belgium; 2) Department of Obstetrics and Gynecology, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand; 3) Pathology Department, Dunedin School of Medicine, Otago University, Dunedin, New Zealand; 4) Department of Ophthalmology, New Zealand National Eye Centre, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand; 5) Eye Department, Greenlane Clinical Centre, Auckland District Health Board, Auckland, New Zealand.

Both loss-of-function mutations in *FOXL2* and *cis*-regulatory deletions lead to the rare, autosomal dominant, developmental disorder blepharophimosis syndrome (BPES), associating an eyelid malformation with premature ovarian insufficiency. Despite extensive genetic studies, the molecular causes remain unexplained in 12% of typical BPES patients. The goal of this study was to unravel the molecular cause in a multigenerational Polynesian family in which linkage to a 34,4 Mb genomic region on 3q23 containing *FOXL2* was shown (LOD of 3.8). First, we delineated the *cis*-regulatory domain of *FOXL2* to a 414 kb region (chr3:138,652,808-139,067,278;GRCh37) using a combination of whole-genome human–mouse–chicken multiple alignments and publically available high-resolution interaction maps enabling us to visualize the topologically associated domain (TAD) containing *FOXL2.* Next, we resequenced this putative *cis*-regulatory domain using HaloPlex enrichment followed by massive parallel sequencing on the MiSeq next-generation sequencing in seven affected and 4 unaffected family members. Data-analysis was performed using CLC bio Genomics Workbench. After mapping, variants were called, annotated and filtered using several online resources including VEP, RegulomeDB, CADD, Roadmap Epignomics, goNL and Kaviar. Using this approach, we identified a heterozygous non-coding variant, Chr3(GRCh37):g.138954755G>A, located in a genomic fragment predicted by the Roadmap Epigenomics to function as an enhancer in the ovary. In addition, this predicted enhancer is located in the shortest region of overlap of previously delineated *cis*-regulatory deletions and an interaction of this fragment with the *FOXL2* promoter has been demonstrated using chromosome conformation capture in human granulosa-like tumor KGN cells (D'haene et al. PLoS Genetics 2009). We are currently performing *in vitro* reporter assays in KGN cells to validate the regulatory potential of this predicted enhancer. In conclusion, this is the first report of a non-coding variant in a putative novel enhancer of *FOXL2* leading to BPES. Our study adds to increasing number of Mendelian developmental disorders caused by subtle genetic defects of *cis*-regulatory elements, such as the ZRS and SIMO elements in the *SHH* and *PAX6* regions respectively.

1626F

Differential expression of immunoglobulin genes in blood and lesion burden in familial cerebral cavernous malformation type 1 (CCM1) patients. *H. Kim[1,2,3], J. Nelson[1], H. Choquet[4], D. Guo[1], C.E. McCulloch[2], M. Bartlett[5], B. Hart[6], L. Morrison[5,7], L. Pawlikowska[1,3].* 1) Center for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, University of California San Francisco, San Francisco, CA; 2) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 4) Kaiser Permanente of Northern California, Division of Research, Oakland, CA; 5) Department of Neurology, University of New Mexico, Albuquerque, NM; 6) Department of Radiology, University of New Mexico, Albuquerque, NM; 7) Department of Pediatrics, University of New Mexico, Albuquerque, NM.

Background: Inflammation may contribute to the pathogenesis of cerebral cavernous malformations (CCM); increased mRNA expression of immunoglobulins has been found in CCM lesions. We previously reported common genetic variants in inflammatory and immune response genes associated with CCM1 disease severity phenotypes, including total and large (>5mm) lesion burden. We investigated whether blood mRNA expression levels of these genes were associated with total and large lesion burden in familial CCM1 cases with the Common Hispanic Mutation (CCM1-CHM). **Methods:** RNA was extracted from whole blood (PAXgene) collected from 95 Hispanic CCM1-CHM patients enrolled in the Brain Vascular Malformation Consortium CCM study RNA expression was measured using the Affymetrix HuGene ST 2.1 array, containing probes for 48,226 transcripts. We compared whole-transcript RNA expression levels between subjects with higher than expected vs. lower than expected log-lesion or log-large lesion counts given age and sex using ANOVA. Transcripts with absolute fold-change (FC) >1.2 and ANOVA p<0.05 were considered differentially expressed between the two groups. We evaluated the expression of 13 individual genes and of transcripts mapping to the IGH (chr14, 58 transcripts) and IGL (chr22, 46 transcripts) immunoglobulin clusters. **Results:** There was no difference in blood expression levels of 13 individual inflammatory and immune response genes between the high lesion vs. low lesion group for either lesion phenotype. However, there was evidence for differential expression of immunoglobulin cluster transcripts. For total lesion burden, transcript levels of 1 IGL and 4 IGH genes were up-regulated in the high lesion group (3 IGH transcripts had FC >1.5). For large lesion burden, 8 IGL transcripts and 4 IGH transcripts were down-regulated (one IGH transcript had FC >1.5). One IGH transcript (IGHM;IGHV4-31) and one IGL transcript (IGLV3-12) were consistently up-regulated or down-regulated, respectively, in both total and large lesion analysis. In transcriptome-wide analysis, IGH transcripts were amongst the top differentially expressed genes for total lesion burden (FC >1.5 and p<0.05). **Conclusions**: Blood RNA expression levels of IGH and IGL genes are associated with total and large lesion burden in CCM1-CHM patients. Specific immunoglobulin levels may serve as a potential biomarker of CCM disease severity.

1627W

NSD1 haploinsufficiency evokes DNA hypomethylation at imprinted DMRs and the increased expression of imprinted genes. *H. Watanabe[1], K. Higashimoto[1], N. Miyake[2], T. Maeda[1], H. Hidaka[1], S. Aoki[1], N. Matsumoto[2], H. Soejima[1].* 1) Division of Molecular Genetics & Epigenetics, Department of Biomolecular Science, Faculty of Medicine, Saga University, Saga, Japan; 2) Department of human genetics, Yokohama City University Graduate School of Medicine.

Purpose SS is caused by haploinsufficiency of *NSD1*, which catalyzes mono- and di-methylation of histone H3K36. In addition, mutations of the *DNMT3A* (*de novo* DNA methylase) and *SETD2* (H3K36 trimethylase) have been identified in Sotos-like over growth syndrome, which exhibits only some SS features. Because DNMT3A recognizes H3K36me3 and plays a pivotal role in the DNA methylation of imprinted loci, we hypothesized that the mutations of *NSD1* cause aberrant methylation at imprinting associated differentially methylated regions (DMRs). The purpose of this study is elucidation of relationship between *NSD1* mutation and aberrant methylation of DMRs, and expression of imprinted genes. Materials and Methods Genomic DNAs were extracted from peripheral blood of 31 SS patients with the *NSD1* mutation. We screened DNA methylation status of 32 DMRs by the MALDI-TOF MS and confirmed aberrant methylation by bisulfite-pyrosequencing. TCL-1 and HTR8 cells were treated with DNA demethylating agent 5-Aza-2-deoxycitidine (5-Aza-CdR). After the treatment, we analyzed methylation status of DMRs and the expression level of imprinted genes by quantitative RT-PCR. Results *IGF2*-DMR0 and IG-DMR-CG6 showed hypomethylation in approximately half of SS patients. 5-Aza-CdR treatment induced hypomethylation of the two DMRs in the cell lines. Expression of *IGF2* in the 5-Aza-CdR treated cells increased approximately 5 times compared to that in the untreated cells. Expression of *MEG3* and *MEG8* also increased in the cell lines with 5-Aza-CdR treatment. Discussion and Conclusions Our results suggested that the *NSD1* mutation induced hypomethylation of DMRs, leading to overexpression of the imprinted genes. Especially, since *IGF2* encodes a growth factor and its overexpression causes Beckwith-Wiedemann syndrome, which is a representative overgrowth syndrome, it is plausible that overgrowth of SS is attributed to *IGF2* overexpression. In addition, overexpression of *MEG3* and *MEG8* also may involve in the features of SS.

1628T

Genome-wide miRNA profiling in plasma of pregnant women with Down syndrome fetuses. *I. Svobodova[1], S. Pospisilova[1], B. Chylikova[1], M. Korabecna[1], P. Calda[2], M. Brestak[2,3], M. Krkavcova[4], O. Seda[1], A. Horinek[1].* 1) Institute of Biology and Medical Genetics, General University Hospital in Prague, Prague, Czech Republic; 2) Department of Obstetrics and Gynecology of the First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic; 3) Screening center, Prague, Czech Republic; 4) GENVIA genetic laboratories, Prague, Czech Republic.

Introduction: Down syndrome (DS) is the most common aneuploidy (1:700 live births). Molecular pathogenesis of the DS is still incompletely understood. Impact of epigenetic mechanisms is currently often discussed in the context of DS. We follow up on a previous study comparing miRNA expressions in normal and trisomic placentas[1], where seven miRNAs were found as being significantly upregulated in DS placentas. In the present study, we focused on plasma samples to further investigate biological functions of miRNAs. Methods: Total RNA enriched for small RNAs was extracted from 900ul of plasma using NucleoSpin miRNA Plasma kit (Mecherey-Nagel) and eluted with 20ul of elution buffer. A total of 26 plasma samples from pregnant women - 13 bearing DS fetuses and 13 normal karyotype fetuses, were included to our study. Genome-wide miRNA profiling was performed using Affymetrix miRNA 4.1 array strips. This pilot study serve to rough selection of miRNAs discriminating compared groups of pregnant women. Results were evaluated using Partech Genomics Suite software. The one-way ANOVA with a cut-off p-value ≤ 0.05 and fold change ≥ 1.8 was used for detection of differentially expressed miRNAs. Association with potentially affected biological pathways was performed using Ingenuity pathway analysis (IPA) software. Results: A total of 23 miRNAs were selected for a following validation study. The set of miRNAs consisted of three groups: 1) 12 human miRNAs which were identified as being significantly dysregulated between euploid and DS plasma samples of pregnant women according to the results of the pilot study; 2) 7 miRNAs which were verified as being overexpressed in DS placentas in our previous study; 3) 4 miRNAs which were in a wider group of differentially expressed miRNA in case of setting milder criteria (p-value ≤ 0.05; fold change ≥ 1.5) and simultaneously are mentioned in the literature as being possibly associated with Down syndrome. Conclusion: Our pilot study suggested that differences of miRNA profiles between normal and DS fetuses are detectable even on the level of maternal plasma. Many biological pathways might be potentially influenced based on pilot study results; cell communication and signalling pathways are between the most enriched ones. More accurate data will be available after the validation study. Supported by the Ministry of Health of the Czech Republic RVO VFN64165.

1629F

The NIA Aging Cell Repository: Facilitating aging research on cells in culture. *D. Requesens, A. Green.* Coriell Institute for Medical Research, Camden, NJ.

The NIA Aging Cell Repository was established at the Coriell Institute for Medical Research, an independent non-profit research institution, in 1974. This invaluable resource supports cellular and molecular research studies on the mechanisms of aging and associated degenerative processes. The NIA Repository aims to provide well-characterized, viable cultured cell lines requested by researchers in the field of aging, and to publicize the availability of these cell lines and in some cases cell-line derived DNA to the research community. The cells stored and distributed by the repository have been collected over three decades using strict diagnostic criteria and banked under the highest quality standards. Scientists use the highly-characterized, viable, and contaminant-free cell cultures from this collection for research on aging-related diseases. These samples have been used in the identification of many disease genes including the gene for Hutchinson Gilford Progeria Syndrome. Included in the repository are skin fibroblast cultures from individuals with premature aging syndromes, including Werner and Hutchinson-Guilford (progeria), cultures from clinically documented and at-risk individuals from families exhibiting familial Alzheimer's disease, differentiated cell lines, cell lines from animals, mouse embryonic stem cells (ESC) and human induced pluripotent stem cells (iPSC). The repository also contains DNA from many cell lines, available individually or in panels such as the Primate DNA panel, Aging Syndrome DNA panel, Characterized Alzheimer's disease mutation DNA panel, Early and Late Onset Alzheimer's disease DNA panels, and Aged Sib Pairs DNA panel. Scientists from over 40 countries have used samples from this collection for research. More than 1,000 papers have been generated used samples from the NIA Repository. The full online catalog can be accessed at https://catalog.coriell.org/NIA.

1630W

Transcriptional profiling of aging effects in human trabecular meshwork.

S. Ramdas[1], G. Li[1], Q. Ma[2], F. Rozsa[3], J. Richards[3], J.Z. Li[1,2]. 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor MI; 3) Department of Ophthalmology and Visual Sciences, University of Michigan, Ann Arbor, MI.

The trabecular meshwork (TM) is an area of tissue in the outer eye, consisting of a network of fibers involved in the outflow of aqueous humor from the cornea. Increasing stiffness of this tissue with age is a risk factor for glaucoma, one of the leading causes of blindness in the world. In this study, we use microarray-based gene expression profiling of post-mortem samples from 93 subjects (age: 13-88 yrs) to analyze age-related gene expression changes in the TM. We identify a transcriptional signature of age in this tissue, involving genes such as *SPOPL*, *CASC15* and *TGM2*. The *TGM2* protein is responsible for cross-linking of extracellular proteins, making them resistant to degradation. Previous studies have reported higher levels of *TGM2* (transglutaminase 2) in glaucoma samples than in controls. The age-related increase of *TGM2* therefore could be related to the increasing stiffness of the TM with age and a risk factor for glaucoma. Comparisons with published aging signatures in the rat reveal convergence at the level of biological pathways, including those related to the immune system, the extracellular space and the ribosome. Genes in these pathways tend to have higher expression in older subjects. While aging studies of other tissues show a decreased expression of genes for ribosomal function, TM appears to be one of the few tissues showing an increase with age. Further, more genes show a higher variance among older samples than those showing higher variance in the young. This greater inter-individual variation in older samples is consistent with a decreased stringency of gene regulation with advancing age, as proposed by previous studies. Lastly, we were able to develop gene expression predictors of age using subsets of our samples as training sets, achieving a performance of correlation coefficient r of 0.68 between the predicted and actual age in the test set.

1631T

Aging and subregion specific transcriptional changes in the rat hippocampus.

I.S. Piras[1], M.D. De Both[1], A.L. Siniard[1], L. Ianov[2], M.K. Chawla[3], A. Rani[2], A.J. Kennedy[4], J.J. Day[4], A. Kumar[2], J.D. Sweatt[5], T.C. Foster[2], C.A. Barnes[4], M.J. Huentelman[1]. 1) Evelyn F. McKnight Brain Institute, & Translational Genomic Research Institute, Phoenix, 85004 AZ; 2) Evelyn F. McKnight Brain Institute, & University of Florida, Gainesville, 32611 FL; 3) Evelyn F. McKnight Brain Institute, & University of Alabama, Birmingham, 35294 AL; 4) Evelyn F. McKnight Brain Institute & University of Arizona, Tucson, 85724 AZ; 5) Evelyn F. McKnight Brain Institute & Vanderbilt University School of Medicine, Nashville, 37232 TN.

Here we used next generation RNA sequencing to investigate differential changes in the rat whole transcriptome within each of three hippocampal subfields CA1, CA3 and Dentate Gyrus (DG). We investigated the genes associated with normal aging and the sub-regional specific transcriptome differences regardless of age. The Morris Water Maze was conducted in male Fisher 344 rats (n=34) of two different age groups: *young* (5-6 mos; n=10) and *aged* (17-22 mos; n=24). The experiments were conducted at University of AZ (n=16) and University of FL (n=18). The whole RNA isolated from the three subfields was library prepped with TruSeq Stranded Total RNA with Ribo-Zero Gold Kit (Illumina) and sequenced using the Illumina HiSeq2500; differentially expressed genes (DEGs) between groups and subregions were determined using DESeq2. We sequenced an average of 58.6±19.5 millions of reads for sample, detecting DEGs associated with aging ranging from 6 (CA3) to 23 (CA1). When considering all of the three subfields as one group, we detected a total of 135 DEGs. Among all of the genes observed, most of them were region specific: CA1 (n=4), CA3 (n=22), and DG (n=6). Pathway analysis suggested the enrichment of immune related pathways as "*Innate Immune System*" (adj p=5.6E-10; n=36). The subregion specific transcriptome analysis regardless of age enabled us to detect expression patterns including genes specifically over or under expressed in each subregion: CA1 (n=406), CA3 (n=540), and DG (n=963). In particular, the DG was enriched with specific pathways for high expressed genes as "*Axon guidance*" (adj p=7.4E-04) and "*cAMP signaling*" (adj p=3.2E-03). Our results highlight the association of immune response genes with aging. The subfield with most relevant changes was CA1. In this region we identified: overexpression of *Laptm5*, associated with Systemic Lupus Erythematosus and lysosomal degradation; *Trem2*, associated with Alzheimer's Disease and *Npc2*, associated with the neurodegenerative *Niemann-Pick disease*. In DG we found *Adgre1*, and the lncRNA *AABR07001734.1*. Finally, we detected specific expression signatures for each hippocampus subfield regardless of age, with enrichment for neurogenesis processes in the DG. These data may be useful in identifying transcriptional changes associated with the normal aging process in the mammalian hippocampus as well as identifying molecular approaches to limit or target gene expression to specific subregions of the hippocampal formation.

1632F

A longitudinal study of DNA methylation as a mediator of age-related diabetes risk. *C.D. Grant[1,15], J. Nadereh[2], L. Hou[3,4], Y. Li[5,6], J.D. Stewart[7], G. Zhang[8], A. Lamichhane[9,10], J.E. Manson[11], A.A. Baccarelli[12], E.A. Whitsel[13,14], K.N. Conneely[15].* 1) Genetics and Molecular Biology Graduate Program, Emory University, Atlanta, GA; 2) Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 3) Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 4) Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, IL; 5) Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC; 6) Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 7) Carolina Population Center, University of North Carolina, Chapel Hill, NC; 8) Department of Computer Science, University of North Carolina, Chapel Hill, NC; 9) Environmental Health Sciences, RTI International, Research Triangle Park, NC; 10) Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 11) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 12) Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY; 13) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 14) Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC; 15) Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA.

Background: Age is the main risk factor for many chronic diseases in humans. Age-related cardiovascular diseases, cancers, and neurodegenerative disorders are among the dominant health problems faced by the population. Aging is marked by widespread, reproducible changes in biological processes and the epigenome. The most studied epigenetic mechanism, DNA methylation (DNAm), has been found to show robust and widespread age-related changes across the genome. DNAm profiles from whole blood can be used to predict human aging rates with great accuracy, suggesting that DNAm can act as a biomarker of aging. We sought to test whether DNAm-based predictions of age are related to risk factors for diabetes, with the goal of identifying risk factors potentially mediated by DNAm. Methods: Our participants were 43 post-menopausal women, between 50 and 76 years of age, enrolled in the Women's Health Initiative (WHI). We obtained methylation data via the Illumina 450K Methylation array on whole blood samples from participants at three timepoints (covering on average 16 years per participant). We employed the method and software of Horvath, which uses DNAm at 353 CpGs to form a DNAm-based estimate of chronological age (DNAm age). We then calculated the epigenetic age acceleration, or Δ_{age} (DNAm age minus chronological age), of participants at each timepoint. This allowed us to observe longitudinal changes in participants' Δ_{age} and DNAm age. We fit linear mixed models to characterize how Δ_{age} contributed to a longitudinal model of aging and diabetes-related phenotypes. Results: For most participants, Δ_{age} remained constant over the course of the follow-up period, indicating that age acceleration is generally stable over time. We found that Δ_{age} contributed significantly to models of body mass index (BMI) (p=0.0012), waist circumference (p=0.033), and fasting glucose (p=0.0073), with the relationship with BMI maintaining significance after correction for multiple testing. Replication in a larger cohort of 157 WHI participants spanning three years was unsuccessful, possibly due to the shorter time frame covered. Conclusions: Our results suggest that DNAm has the potential to act as a mediator between aging and risk of having a high BMI, waist circumference, and fasting glucose, or alternatively, it may serve as a biomarker of these phenotypes. Further studies of the role of DNAm in mediating age-related outcomes are warranted, as is follow-up in other longitudinal cohorts.

1633W

An evolutionary perspective of DNA methylation associated with age within the primate lineage. *G. Housman[1,2], E. Quillen[3], A.C. Stone[1,2].* 1) School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA; 2) Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA.

Background: Aging is thought to be a developmentally regulated process that is controlled by epigenetic mechanisms, such as DNA methylation. Methylation patterns throughout the genome show distinct changes with age in humans. However, the evolution of age epigenetics and the degree to which epigenetic signatures of aging are conserved between species has not readily been explored. **Objective:** This project examines the evolution of aging epigenetics specifically within the primate lineage by identifying age-related DNA methylation patterns in baboons (*Papio spp.*) and by comparing these findings to what is known in humans. **Methods:** Genome-wide DNA methylation patterns were identified in femoral trabecular bone from 46 pedigreed baboons, 28 that were adults and 18 that were juveniles using the Illumina Infinium MethylationEPIC BeadChip. **Results:** Several significantly differentially methylated positions (DMPs) were found between these age cohorts. Specifically, out of the 191,631 sites examined, 17% (33,057 sites) were differentially methylated between age cohorts, and similar to other animals, adult baboons display global hypomethylation as compared to juvenile baboons. The significant age-related DMPs identified are associated with genes involved in developmental processes and pathways related to the progression of diseases of aging. Additionally, while some of these age-related DMPs overlap with and display methylation patterns similar to those previously identified in human aging studies (n=11), the majority of previously identified age-related methylation loci in humans that were also tested in baboons (n=137) either showed opposite methylation patterns in baboons (n=3) or were not significantly differentially methylated in baboons (n=123). Nevertheless, methylation levels at these human loci are still able to differentiate baboon age cohorts. **Discussion:** Overall, these results begin to clarify the evolutionary conservation of epigenetic aging processes within the primate lineage. This is the first study to specifically assess aging effects on the DNA methylation profiles of skeletal tissues from a nonhuman primate, and the findings of this research inform our understanding of DNA methylation variation in one skeletal tissue from a nonhuman primate, as well as the degree to which chronological age affects this variation.

1634T

Testing a stochastic model of epigenetic drift in longitudinal DNA methylation data. *C. Robins[1,2], D.J. Cutler[1], L. Hou[3], C.D. Grant[1], E.M. Kennedy[1], A.A. Baccarelli[4], E.A. Whitsel[5,6], K.N. Conneely[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Graduate Program in Population Biology, Ecology, and Evolution, Laney Graduate School, Emory University, Atlanta, GA; 3) Department of Preventative Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 4) Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA; 5) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 6) Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC.

Robust DNA methylation changes are known to occur with age at sites across the human genome. These age-related changes, collectively termed epigenetic drift, are suggested to be caused by environmental factors and dysfunctional epigenetic maintenance and repair mechanisms. The drift process is hypothesized to be stochastic with the gains and losses of methyl groups directly resulting from random unrepaired environmental insults and methylation maintenance failures. Here we test the stochastic process hypothesis of epigenetic drift using a stochastic Ornstein-Uhlenbeck (OU) model and longitudinal DNA methylation data. DNA methylation was measured at >480K CpG sites in blood sampled at 3 time points (timespan avg=16 y) from 43 women participating in the Women's Health Initiative Long Life Study (age range=50-76). After exclusion of probes containing SNPs, multi-modal probes, probes that bind to multiple chromosomes, and probes without a CpG site, 382K sites remained for analysis. We hypothesize age-related methylation changes to be the result of gains and losses of methyl groups that occur at random throughout the genome rather than at targeted genes or CpG sites. If these changes are truly stochastic, we expect constant rates of methyl group gain/loss across all CpG sites. To test this hypothesis we estimated the rates of methyl group gain/loss via maximum likelihood estimation for nested OU models: 1) a null model where all CpG sites have the same rate methyl group gain/loss, and 2) an alternative model where CpG sites included in a functional group (e.g. sites in a promoter) have different rates of methyl group gain/loss. Five functional groups were investigated: promoters, enhancers, insulators, CpG islands, CpG island shores, and CpG island shelves. We compared the models via likelihood ratio tests and rejected the null model for all functional groups tested ($p < 10^{-27}$). This result is consistent with a scenario where age-related methylation changes resulting from epigenetic drift are not purely stochastic at the genome-wide scale. This may suggest the presence of targeted rather than stochastic changes, or that methylation maintenance and repair mechanisms may be better maintained at select sites or in select regions based on location or function. Future analyses will include cell type proportions as covariates to remove potential confounding due to the heterogeneous and changing composition of whole blood, and to assess the robustness of our results. .

1635F

Identification, replication and characterization of epigenetic remodelling in the aging genome. *S. Li[1], L. Christiansen[2], K. Christensen[2], T.A. Kruse[1], P. Redmond[3], R.E. Marioni[4,5,6], I.J. Deary[3,5], Q. Tan[1,2].* 1) Human Genetics, Institute of Clinical Research, University of Southern Denmark, Odense, Odense, Denmark; 2) Epidemiology, Biostatistics and Biodemography, SDU, Odense, Denmark; 3) Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom; 4) Centre for Genomic & Experimental Medicine, University of Edinburgh (UE), Edinburgh, United Kingdom; 5) Centre for Cognitive Ageing and Cognitive Epidemiology, UE, Edinburgh, United Kingdom; 6) Queensland Brain Institute, The University of Queensland, Australia.

Background: Aging is a complex biological process that involves numerous changes at various levels through remodelling of multiple biological processes and regulatory mechanisms including epigenetics. Recent analysis of the DNA methylome has reported large numbers of epigenetic markers associated with human aging. Validation and characterization of the significant findings are necessary both for obtaining reliable results and for better describing the age-associated epigenetic modifications. **Methods**: Based on genome-wide DNA methylation data measured in a large collection of older individuals in the Lothian Birth Cohorts of 1921 and 1936, we performed an association study to identify age-associated CpGs and replicated the findings in two independent cohorts of Danish twins. The double-replicated CpGs were characterized by distribution over gene regions and by their location in relation to CpG islands conditional on their age-dependent patterns of methylation status. The replicated CpGs were further characterized by their involvement in gene-sets or canonical pathways to study their functional implications. **Results**: We identified 67,604 age-associated CpG sites reaching genome-wide significance of FWER<0.05. The double-replication resulted in 5,168 CpGs with more than 60% displaying decreased methylation with increasing age. The replicated CpGs were characterized by high concentration of age-methylated CpGs at 1stExon and TSS200 and a dominant pattern of age-demethylated CpGs at other gene regions, and by overwhelming age-related methylation in CpG island and demethylation at shore/shelf and open sea. Biological pathway analysis showed that age-dependent methylations were especially involved in cellular signalling activities while demethylations were particularly related to functions of the extracellular matrix. **Conclusion**: Extensive epigenetic remodelling in the DNA methylome could be involved in the aging process. The identified age-methylated and demethylated sites displayed differential distribution patterns over genomic regions and were involved in biological pathways closely related to aging phenotypes and age-related diseases.

1636W

Accelerated epigenetic aging in middle-aged African Americans and Whites. *S. Tajuddin[1], D. Hernandez[2], B. Chen[3], N. Noren Hooten[1], N. Mode[1], M. Nalls[2,4], A. Singleton[2], A. Zonderman[1], M. Evans[1].* 1) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD; 2) Laboratory of Neurogenetics, National Institute on Aging, National Institute of Health, Bethesda, MD; 3) Translational Gerontology Branch, National Institute on Aging, National Institute of Health, Baltimore, MD; 4) Data Tecnica International, Glen Echo, MD.

Epigenetic changes including DNA methylation are hallmarks of aging. Biological aging rates vary by sociodemographic, genetic, and environmental factors. Epidemiologic studies suggest ethnic minorities manifest significantly greater rates of premature aging phenotypes, but little is known on extent of the molecular aging rate. DNA methylation, which gradually changes across the life course, could be used to predict age. This DNA methylation predicted age (DNAmAge) also called the "epigenetic clock" correlates with chronological age and was shown to be a good marker of biological age. The objective of this study was to identify factors that influence the epigenetic clock in the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study. Participants were men and women African Americans and whites above and below poverty status aged between 30 to 64 years. Using the Illumina MethylationEPIC arrays, we measured methylation levels at 866,836 CpG sites in DNA extracted from PBMCs. DNAmAge was calculated in 441 participants using two methods based on 353 and 71 CpG sites described by Horvath and Hannum *et al.*, respectively. Blood cell counts of granulocytes, CD4+T, CD8+T naïve, CD8+CD28-CD45RA-, monocytes, natural killer cells, and plasmablasts were estimated using DNA methylation data to control for the potential confounding effects of blood cell abundance on DNA methylation. We used three measures of epigenetic age acceleration: (1) universal measure of age acceleration (AgeAccel) - residuals after regressing DNAmAge on chronological age; (2) intrinsic epigenetic age acceleration (IEAA) - residuals after regressing DNAmAge on chronological age and blood cell counts; and (3) extrinsic epigenetic age acceleration (EEAA) which captures epigenetic and immune system cell aging rate. DNAmAge using both 353 and 71 CpG sites were strongly correlated with chronological age (r=0.85, p=3.8E-122). In multiple linear regression models adjusted for age, sex, race, socioeconomic status (SES), and blood cell counts, African Americans had substantially lower EEAA compared to whites (β=-2.51, p=1.33E-07) while men showed significantly higher EEAA compared to women (β=0.78, p=0.04). Similarly, IEAA was higher among men (β=0.77, p=0.08) but not nominally significant. Surprisingly, the universal AgeAccel measure was not associated with race, sex, or SES. In conclusion, race was associated with EEAA while sex was associated with both IEAA and EEAA.

1637T

Better statistical methods to predict age from DNA methylation. *Q. Zhang[1], R. Marioni[2,3], L. Lloyd-Jones[1], N. Wray[1,4], I. Deary[2,5], A. McRae[1], P. Visscher[1,4].* 1) Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; 2) Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK; 3) Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK; 4) The Queensland Brain Institute, The University of Queensland, St Lucia, QLD 4072, Australia; 5) Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.

DNA methylation age of an individual has been proposed as a biomarker for biological aging and a tool for assessing associations with traits. The relationship between experimental sample size, statistical algorithm and prediction accuracy is not well understood. We used 8,561 samples in the age range of 2 to 104 years from 12 Cohorts (11 from blood and 1 from saliva) measured on Illumina HumanMethylation450 arrays and performed cross-validation prediction analyses. We compared age prediction using Elastic Net, Best Linear Unbiased Prediction (BLUP) and BayesR based on Root Mean Square Error (RMSE) and the correlation between predicted age and chronological age. Our results showed that Elastic Net has the smallest average RMSE (5.3 yrs) and highest average correlation between predictor and outcome (r=0.94), but BLUP outperforms Elastic Net when the training sample size is over 6,000. Furthermore, we confirmed that increasing the training sample size can achieve smaller RMSE and higher correlations in test data sets. Our predictors have better accuracy than the widely-used Horvath and Hannum age predictors. Our results imply that with training sample sizes of tens of thousands, predictors with an error of less than 4 years can be achieved. As this error is reduced, the scope of real biological ageing becomes revealed.

1638F

Fasting and solar time independently regulate expression of hundreds of genes in skin and fat tissue in population-level transcriptomes. *A. Couto Alves, K.S. Small.* Twins Research Genetic Epidemiology, Kings College London, London, London, United Kingdom.

   Intermittent fasting, an acute form of diet restriction, is associated with lower risk biomarkers for diabetes, CVD, cancer and ageing. However, few genes have been robustly implicated in the response to fasting in humans and even fewer in the mediation of its putative disease-protective effects. Here we report on a transcriptome-wide association study (TWAS) of gene expression and hours fasting (5-23h) in fat and skin tissue collected from 615 individuals of the TwinsUK registry, and a secondary TWAS with solar time of sample collection. All data were analysed cross-sectionally. We identified 239 and 30 genes whose expression levels associated ($P<2\times10^{-6}$) with the number of hours fasting in fat and skin respectively, with 15 genes common to both tissues. Genes associated to hours of fasting are enriched for the circadian rhythm gene ontology term (Fat $P=3\times10^{-7}$, Skin $P=5\times10^{-12}$). The associations between gene expression and hours of fasting are conditionally independent of the time samples were collected, and the enrichment for circadian rhythm genes persists after restricting the analysis to samples collected in the afternoon only (N = 390) (Fat $P=1\times10^{-9}$, Skin $P=3\times10^{-6}$) The top genes ($P<1\times10^{-12}$) associated to hours of fasting in fat and skin includes *ARNTL, C1orf51, DBP, PER1, TEF, NR1D1* which are involved in regulation of circadian rhythm. Other genes of considerable interest include *PPARG, PIK3R1, PTGER4, FOXO{1,2,3}, HIF3A and DDIT4*. A total of 97 and 5 of the genes identified by the TWAS have regulatory variants (eQTL Permutation FWER<0.01), out of which 7 variants (or proxies in LD, $r^2>0.9$) are associated ($P<5\times10^{-8}$) with cardio-metabolic and immune-related conditions in GWAS. Of the 239 and 30 genes found in the TWAS of hours fasting, 21 and 3 had GxE eQTL regulatory variants interacting with hours fasting (permutation FWER<0.05) in fat and skin. The solar time TWAS yielded 61 and 18 genes whose expression levels are associated to the time of biopsy collection ($P<2\times10^{-6}$). Solar-time associated genes were also significantly enriched for circadian rhythm as expected ($P<4\times10^{-11}$ and 9×10^{-14}). 50 and 9 genes associated with solar-time overlap with hours fasting-associated genes in fat and skin, including: *ARNTL, C1orf51, CRY2, TEF, PER1, TSC22D3, ZBTB16* but their effect is in the opposite direction to hours of fasting. This study contributes to the characterization of the molecular mechanism of response to fasting and solar time in fat and skin in humans.

1639W

Identifying causal mutations with RNA-seq in mice with Mendelian disorders. *N. Raghupathy, B. Harris, S.Y. Karst, W. Martin, H. Fairfield, D.E Bergstrom, L. Reinholdt.* The Jackson Laboratory, Bar Harbor, ME.

   Exome and whole genome sequencing has been the mainstay of identifying causal mutations in rare Mendelian disorders in human and mice. Although exome and whole genome sequencing have greatly helped us to identify causal mutations, the success rate of discovery is still less than 50%. In a previous study, we used exome sequencing in 172 distinct, spontaneously arising mouse models of Mendelian disorders and identified the causal mutation in 92 strains. Here, we explore the use of transcriptome sequencing to identify causal mutations in strains where exome sequencing failed. We sequenced transcriptomes of relevant tissues with biological replicates in 30 mouse models of Mendelian disorders. And we primarily analyzed splicing aberrations that could be causal for the Mendelian disorder. We used an integrated analysis framework that uses EM algorithm to account for multi-mapping reads and estimated read depth at known and novel splice junctions. Our preliminary analysis shows many cases of putative splicing aberrations and coding mutations that are potentially causative of rare Mendelian disorders.

1640T

Discover regulatory grammar across 127 human cell types using tree-based recurrent neural network. *Z. Zhang, M. Kellis.* EECS, Massachusetts Institute of Technology, Cambridge, MA.

The chromatin state or gene expression of a genomic region are highly governed by the regulatory code in the static DNA sequence, and the study of underlining regulatory grammar has been hindered by the lack of large-scale datasets and its complex nature. To understand the regulatory grammar, some shallow learning algorithms (e.g., SVM, RandomForest) were developed using sequence motifs or k-mers to predict chromatin state or gene expression, which only consider motifs co-occurrence statistics and ignore the relative positions and orientation among motifs in the regulatory regions. Biologically, some regulatory elements function as insulators, and the position of insulators can completely change the semantic meaning of the whole regulatory region. Here, we propose a recurrent neural network (RNN) approach, which has been successfully applied in the natural language processing. The RNN approach models the semantic role of each motif, each phase (motif-pair) and each whole sentence (regulatory region) in a unified high-dimension hidden space. More interestingly, we can explore hidden space to identify the different functional clusters of motifs. The interplay among different motif elements in different positions are naturally encoded during the parsing procedure in the RNN, which empowers to learn more complex regulatory grammars than previous approaches. Using back-propagation technique, the RNN can be trained with multiple regulatory outcomes (e.g., different histone marks), so that single regulatory grammar learned by RNN can explain them simultaneously. With simulated datasets, we show our RNN approach can identify not only activators or repressors and but also insulators, and recover regulatory outcome accurately.

1641F

Single cell methylomes: A method to assess mammalian neuron diversity. *L. Kurihara[5], C. Luo[1,2], C.L. Keown[3], R. Castanton[2], J. Lucero[4], J.R. Nery[2], E.A. Mukamel[3], T. Harkins[5], M.M. Behrens[4], J.R. Ecker[1,2].* 1) Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA, United States; 2) Genomic Analysis Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, United States; 3) Department of Cognitive Science, University of California, San Diego. La Jolla, CA, United States; 4) Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, United States; 5) Swift Biosciences Inc, Ann Arbor, MI, United States.

Neuron types have been classified by structure, electrophysiology, connectivity and RNA expression with single-cell resolution. Identifying gene regulatory elements that distinguish specific neuron types also requires single-cell resolution. We developed a method for profiling single neuron methylomes to achieve this goal. Epigenomic marks such as cytosine DNA methylation (mC) have highly diverse patterns across brain cell types. Human and mouse brains accumulate high levels of non-CG methylation (mCH) throughout the genome. In addition, neurons contains differential mCH signatures that allow distinguishing neuronal subtypes by low coverage single cell WGBS (whole-genome bisulfite sequencing). To meet the need of large-scale single cell methylome profiling, we developed a novel workflow using Adaptase, the technology underlying the Accel-NGS Methyl-Seq Library Preparation Kit. Through more efficient adapter attachment to bisulfite converted ssDNA, libraries with greater complexity are generated and single cell pooling enables high-throughput library preparation. Single cell methylomes were generated from 6,000 single neuronal nuclei isolated from human and mouse frontal cortex. From <2M reads per nucleus followed my mCH clustering analysis, we identified 16 mouse and 21 human neuronal subpopulations in the frontal cortex. The observed mCG and mCH patterns exhibited neuron type-specific distributions that recapitulated transcriptome signatures. In addition, ~500,000 putative regulatory elements were identified with differential methylation across neuron types which may be drivers of neuron type differences. When comparing mouse and human results, global mCH and mCG levels were strongly correlated between homologous clusters, suggesting evolutionary conservation. This method is applicable to other studies requiring high resolution cell-type identification based on epigenomic patterns.

1642W

DNA methylation of *TNF* decreases after an intense bout of eccentric exercise. *B. Hussey, D.J. Hunter, L. Wilson, M.R. Lindley, S.S. Mastana.* Translational Chemical Biology Research Group, School of Sport Exercise and Health Science, Loughborough University, Loughborough, Leicestershire, United Kingdom.

   Introduction: DNA methylation of CpG sites are modified by the environment, lifestyle factors, and have implications for disease progression. Investigating gene targets within inflammatory pathways may provide insight into how a disease can occur and how the burden may be reduced. Here we focused on the methylation profile of the gene encoding the inflammatory cytokine, Tumour Necrosis Factor alpha (TNFα). Methylation status of *TNF* [MIM 191160] has been associated with inflammatory diseases, such as Crohn's disease, selective neuronal vulnerability, severe systemic inflammation and age-related inflammatory diseases. The aim of this study was to investigate the short-term changes to the percentage of 5-methylcytosines (%5mC) within *TNF* in response to an intense exercise stimuli.   **Method:** Untrained healthy males (n=7) performed a bout of intense eccentric exercise (20 sets, 10 repetitions of maximal eccentric bilateral knee extensions). Venous blood samples were collected before exercise (pre), immediately post exercise (PostEx), 1.5, 3 and 48 hours PostEx. DNA was extracted from blood and bisulfite converted using EpiTect LyseAll kit (Qiagen). Methylation of four CpG sites (+202, +207, +219 and +227 bases from the transcription start site) in exon 1 of *TNF* was measured by pyrosequencing with a PyroMark Q48 Autoprep (Qiagen). This region has previously been demonstrated to correlate with circulating TNFα cytokine levels. Data is presented as mean ± standard deviation.   **Results:** *TNF* exon 1 was hypomethylated at baseline (10.13±2.15%5mC). There was a significant difference in %5mC across the time course of the study (p<0.001). *TNF* exon 1 methylation levels decreased from pre-exercise levels immediately (10.13±2.15%5mC vs 8.34±1.59%5mC) and 1.5 hours PostEx (10.13±2.15%5mC vs 6.48±3.02%5mC). Methylation levels were significantly lower than baseline at 3 hours PostEx (10.13±2.15%5mC vs 5.89±1.05%5mC, p=0.01). Methylation levels returned towards baseline 48 hours PostEx.
Conclusion: This investigation indicates a significant change in methylation post exercise and suggests that there are subtle time course responses to the hypomethyated state of the first exon of *TNF*. Further investigations should focus upon delineating this time course and interrogating the impact of reduced methylation state of the *TNF* exon 1 across a range of biological processes.

1643T

Common DNA sequence variation leads to variation in 3D genome organization. *Y. Qiu[1,2], D.G. Gorkin[1,3], M. Hu[4], A. Noor[3], A.D. Schmitt[1], J. Sebat[3], B. Ren[1,3].* 1) Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093, USA; 2) Bioinformatics and Systems Biology Graduate Program, 9500 Gilman Drive, La Jolla, CA 92093, USA; 3) Department of Cellular and Molecular Medicine, University of California, San Diego, School of Medicine, 9500 Gilman Drive, La Jolla, CA 92093, USA; 4) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA.

   The 3-dimensional (3D) organization of chromatin inside the nucleus is integral to many genomic processes including transcriptional regulation, DNA replication, and X chromosome inactivation. In addition, several recent studies demonstrate that disruption of normal 3D chromatin organization can cause disease. However, little is known about whether 3D chromatin organization varies in the human population, or what impact DNA sequence variation may have on 3D chromatin organization. To address the critical gap in current knowledge, we performed Hi-C (High-Throughput Chromatin Conformation Capture) on Lymphoblastoid Cell Lines from 20 individuals. While 3D chromatin organization is strikingly similar at megabase scale, we found extensive variations (>10% of the genome) in 3D chromatin organization at a finer resolution. Notably, variations in 3D chromatin organization coincide with variations in other genome functions including gene expression, TF binding, and histone modifications in many cases, suggesting that common mechanisms underlie variations in genome function and 3D organization. Moreover, we found several lines of evidence indicating genetic basis of variations in 3D chromatin organization. We found that SNPs associated with variation in genome activity are enriched in regions of variable 3D organization, and that variation in CTCF motifs can alter the strength of 3D chromatin interactions. Furthermore, we generated chromosome-span haplotypes with high accuracy (~98% accuracy) and high resolution (~90% heterozygous SNPs) using Hi-C. We identified thousands of regions showing allelic-specific variations in 3D chromatin organization and found that in some cases allelic-specific regions show transmission from parents to child, suggesting that 3D chromatin organization is a heritable feature of the human genome. In sum, our study addresses an important gap in our knowledge about how 3D chromatin conformation varies in the human population and provides insight into how, at the molecular level, DNA sequence variation can contribute to phenotypic variation and disease risk.

1644F

Chromatin plasticity during hematopoietic cell differentiation and stimulation. *J.V. Ribado[1], M.L.T Nguyen[2], D. Calderon[3], A. Mezger[1,4], B. Wu[1], A. Kathiria[1], J.K. Pritchard[1,5,6], L.A. Criswell[7], W.J. Greenleaf[1,8,9], A. Marson[1,2,9].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Microbiology and Immunology, University of California at San Francisco, San Francisco, CA; 3) Program in Biomedical Informatics, Stanford University, Stanford, CA; 4) Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden; 5) Department of Biology, Stanford University, Stanford, CA; 6) Howard Hughes Medical Institute, Stanford University, Stanford, CA; 7) Rosalind Russell/Ephraim P. Engleman Rheumatology Research Center, University of California at San Francisco, San Francisco, CA; 8) Department of Applied Physics, Stanford University; 9) Chan Zuckerberg Biohub, San Francisco, CA 94158.

Little is known about genome-wide changes that contribute to mechanisms of hematopoietic stem cell differentiation and plasticity. Thus, we have conducted ATAC-seq on 25 primary immune cell types in both resting and activated conditions from four healthy individuals to study the epigenetic landscape of immune cells during differentiation. We integrated these data with previously published ATAC-seq data of undifferentiated hematopoietic cells. Unsupervised clustering of chromatin accessible regions is sufficient to distinguish progenitor, T cell, and B cell lineages that are enriched for respective master transcription regulator motifs, such as PU.1, SpiB, Runx1, RORγt, Oct2 and PU.1:IRF8. We found that upon stimulation, cells within a lineage are more similar than when compared to their resting state. Additionally, stimulated suppressive regulatory T cells have a higher proportion of overlapping chromatin accessible regions between unstimulated inflammatory T cell subsets. These results suggest that differentiated cells have the flexibility to respond similarly to TCR activation and lineage-specific chromatin stability may be context specific. Understanding the regulatory structure of immune cells during differentiation and stimulation is important in elucidating the etiology of inflammatory diseases and identifying subsequent targets for therapeutic manipulation.

1645W

Pleiotropic effects of trait-associated genetic variation on DNA methylation: Utility for refining GWAS loci. *E. Hannon[1], T. Gorrie-Stone[2], M. Smart[2], M. Weedon[1], N.J. Bray[3], M.C. O'Donovan[3], M. Kumari[2], L.C. Schalkwyk[2], J. Mill[1].* 1) University of Exeter Medical School, University of Exeter, Exeter, United Kingdom; 2) University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ; 3) MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, CF24 4HQ, UK.

Most genetic variants identified in genome-wide association studies (GWAS) of complex traits are thought to act via effects on gene regulation rather than directly altering the protein product. As a consequence, the actual genes involved in disease are not necessarily the most proximal to the associated variants. By integrating data from GWAS analyses with that from genetic studies of regulatory variation, it is possible to identify variants pleiotropically-associated with both a complex trait and measures of gene regulation. In this study, we use Summary data–based Mendelian Randomization (SMR), a method developed to identify variants pleiotropically associated with both complex traits and gene expression, to identify variants that are associated with complex traits and DNA methylation. Building on our previous efforts, we increased our catalogue of DNA methylation quantitative trait loci (mQTL) in whole blood by taking advantage of the latest microarray technology, the Illumina EPIC Beadchip array, profiling DNA methylation at over 800,000 genomic loci. These data were used to prioritize genes for >40 complex traits with robust GWAS data, highlighting overlap with the results of SMR analyses performed using expression QTL (eQTL) data. We identify multiple examples of variable DNA methylation associated with GWAS variants for a range of complex traits, demonstrating the utility of this approach for refining genetic association signals, and highlighting the novel insights the EPIC array provides over previous technologies.

1646T

Fine mapping of interacting functional elements in Hi-C peaks. *A.T. Jaroszewicz, J. Ernst.* University of California, Los Angeles, Los Angeles, CA.

The advent of high-throughput chromatin conformation assays such as Hi-C has enabled genome-wide detection of long-range chromatin contacts, which are integral to various regulatory mechanisms, including enhancing transcription and demarcating topologically associating domains. However, even ultra-deep sequencing cannot robustly identify specific interacting loci or provide biological interpretation due to its coarse resolution, generally on the order of 5-25kb in most ENCODE cell types. Here, we present a novel computational method to fine map these broad interactions to their functional elements. Using transcription factor and histone mark ChIP-seq data, which has a much finer resolution, we train a deep neural network (DNN) to discern between true interactions and permuted ones. We then use a masking algorithm along with this DNN to predict which fine-resolution regions (on the order of 100bp) are most important to the Hi-C predicted interaction. This method allows us to assign each pair of subregions in the broad domains a score that corresponds to the importance of its role in the interaction. In contrast to other methods, our method is not limited to certain types of interactions such as promoter-enhancer pairs, but any protein-mediated interaction with available ChIP-seq data. These results can be used to discover novel types of interactions, provide a biological interpretation of interactions, prioritize GWAS hits, and reduce the need for expensive, deeply-sequenced Hi-C experiments. In our preliminary results, we show that our predictions are enriched for promoter-enhancer and cohesin interactions. We validate computationally by taking finer resolution Hi-C peak calls, extending them, and showing that we preferentially recover the original peak. We also compare these predictions to Promoter Capture Hi-C and other fine-resolution chromatin interaction assays.

1647F

Comprehensive functional annotation of the zebrafish genome. *T. Liu[1], H. Yang[1], Y. Wang[1], G. Gerhard[3], D. Balciunas[2], F. Yue[1].* 1) Dept. Biochemistry and Molecular Biology, ennsylvania State University College of Medicine, Hershey, PA.17033; 2) Dept of Biology, Temple University 1900 North 12th Street Philadelphia, PA 19122; 3) Dept of Medical Genetics and Molecular Biochemistry, Temple University Room 621 Kresge Hall Philadelphia, PA 19140.

Zebrafish has been widely used for the study of human diseases, as ~70% of the protein-coding genes are conserved between the two species. Moreover, zebrafish embryos are transparent and thus can serve as an ideal model for genetic studies in animal development. Surprisingly, the functional annotation of zebrafish genome itself has been severely lagging when compared with other model systems such as mouse and drosophila. Here we took a similar approach adopted by the ENCODE and Roadmap Epigenomics projects, and performed RNA-Seq and ChIP-Seq for several histone modifications to generate a comprehensive map of transcriptomes and regulatory elements in a variety of zebrafish tissues, including brain, heart, liver, skeletal muscle and kidney. We predicted over 100,000 cis-regulatory elements in the zebrafish genome, the most comprehensive functional annotation effort in zebrafish so far to our knowledge. We also identified tissue-specific and developmental stage-specific regulatory elements. By comparing the data generated by the ENCODE and Roadmap Epigenomics projects, we also defined a set of functionally conserved and species-specific regulatory sequences among zebrafish, mouse and human. In summary, we generated a great genomics/epigenomics resource for the functional annotation in the vertebrate genomes and further expanded the value of zebrafish as a model of human disease.

1648W

DNA methylation and its impact on inter-population differences in disease risk and prognosis. *M. Loh[1,2,3], F.L. Tai[1], M. Kalra[1], S. Tan[4], F.F. Gan[4], W.Y. Saw[5], R. Soong[4,6], Y.Y. Teo[5,7,8,9,10,11].* 1) Translational Laboratory in Genetic Medicine (TLGM), Singapore, Singapore; 2) Department of Biochemistry, National University of Singapore, Singapore; 3) Department of Epidemiology and Biostatistics, Imperial College London, London, UK; 4) Cancer Science Institute of Singapore, National University of Singapore, Singapore; 5) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 6) Department of Pathology, National University Hospital, Singapore; 7) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; 8) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; 9) National University of Singapore Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore; 10) Life Sciences Institute, National University of Singapore, Singapore; 11) Department of Statistics and Applied Probability, National University of Singapore, Singapore.

Epigenetic mechanisms such as DNA methylation play a key role in the regulation of numerous biological processes including gene expression and cellular differentiation, with altered methylation shown to be associated with complex human diseases such as cancer and diabetes. However, to date, little is known about variation in DNA methylation between ethnic populations.   In this study, we measured methylation in DNA extracted from peripheral blood from 287 healthy individuals (102 Chinese, 87 Malays and 98 Indians) using the Illumina Infinium MethylationEPIC BeadChip which covers >850,000 methylation sites at single-nucleotide resolution. White blood cell subsets were imputed using the Houseman method, with regression analysis used to quantify the associations between methylation and ethnicity status, and adjustment for technical and biological confounding.   A total of 1,122 CpG sites were differentially methylated between any two of the three ethnic populations at $p<1E-7$, with 619 sentinel CpG sites remaining after clumping at +/-500kb. Among these sentinel CpG sites, 99.8% were significantly differentially methylated between Chinese and Indians, 85.0% between Malays and Indians and only 1.6% were significantly different between Chinese and Malays.   Taking a global approach with unsupervised hierarchical clustering, we observed two distinct clusters, with Indians forming a cluster of its own, while the other cluster consisted of Malays and Chinese individuals. This observation was consistent with previous findings from genetic studies.   The top hit was located nearest to the gene *VKORC1*, which encodes for the enzyme Vitamin K epOxide Reductase Complex (VKORC) subunit 1. Mutations in this gene are well established to be associated with deficiencies in vitamin-K-dependent clotting factors, with resistance to warfarin previously reported to exhibit ethnic differences in maintenance dose requirements, including between Singapore Chinese, Malays and Indians.   Other top hits include genes associated with diseases known to demonstrate ethnic differences in risk and prognosis such as Type 2 Diabetes, insulin resistance and other metabolic disturbances, various cancer types (breast and lung cancer), as well as neurological disease such as Parkinson's.   Taken together, preliminary findings from our study suggest that differences in methylation profiles between ethnic populations could potentially account for differences in disease risk, prognosis and response.

1649T

X-chromosome epigenetic markers for age-prediction. *I.C.T. Mello[1], A.E. Woerner[2], J. King[2], B. Budowle[2,3], R. Moura-Neto[4], R. Silva[1].* 1) Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; 2) Center for Human Identification, University of North Texas Health Science Center, Fort Worth, Texas, US; 3) Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia; 4) Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

The epigenetic modifications play a significant role in the regulation of gene expression Epigenetic patterns can be stable, as they are often preserved during cell division, or they may be transient, and change over an individual's lifetime. With respect to the latter, recent studies have shown that many CpG regions are subject to methylation changes according to the aging process and many autosomal epigenetic markers have been proposed as an age predictor for forensic samples. In order to achieve dosage compensation, one of the two X chromosomes in females is inactivated and these difference in regulation throughout the development generate variability that could be informative as an epigenetic marker. In this study, we investigated the usefulness of methylation status of the X chromosome for age-prediction. We investigated and tested the performance of X chromosome markers using methylation data from three different sources. The first two sources were produced using Illumina Infinium Human Methylation450k Beadchip taken from two separate populations. Identification of age-associated methylation markers was performed by regularized linear regression, using the least absolute shrinkage and selection operator on the data from Hannun et al (GSE40279). We then validated the performance of the 51 markers identified in this process using robust linear regression, as a linear function of age, to predict age using the samples in Liu et al (GSE4286). Our markers were highly predictive of age ($R^2= 0.83$). Next we selected new 30 CpG islands across the X-chromosome located (1) on gene sites, (2) next to genes sites (<50.000bp of distance) and (3) far from genes sites (>50.000bp of distance) to investigate age-related methylation sites. Our results showed that 51 markers found, in 33 CpG regions have the potential to be used as age predictor in male samples. A group of 150 unrelated male subjects between 20 and 60 years of age from an admixture population were selected and will be tested with these methylation markers applying bisulfite conversion DNA treatment and multiple parallel sequencing technologies. Funding: CAPES, CNPq.

1650F

Genomic DNA methylation changes in myalgic encephalomyelitis. *L. Sarria[1], M.S. Trivedi[2], E. Oltra[3], K. Bauckman[1], M.A. Fletcher[1], N.G. Klimas[1], L. Nathanson[1].* 1) INIM, College of Osteopathic Medicine, Nova Southeastern University, Fort Lauderdale, FL, USA; 2) College of Pharmacy, Nova Southeastern University, Fort Lauderdale, FL, USA; 3) Universidad Catolica de Valencia, C/Quevedo 2, 46001 Valencia, Spain.

   Background: Myalgic Encephalomyelitis (ME) is a complex condition that is associated with a variety of symptoms including fatigue, memory problems, muscle and joint pain, gastrointestinal issues, neurological problems, hormonal imbalance and immune dysfunction. Currently, treatment relies solely on symptom management but does not address the underlying mechanisms of disease. Our previous research efforts have shown that the transcriptional regulation altered the phenotypic nature of ME. DNA methylation is one of the epigenetic mechanisms that regulates gene transcription without changes in the DNA sequence. It involves the covalent binding of a methyl group to a Cytosine-5 at a C-phosphate-G (CpG) site. Negative associations between methylation and transcription are known to be enriched particularly in promoter regions. **Objectives:** To provide insight into the biological mechanisms of ME, the main objective of this research is to identify novel mechanisms of transcriptional regulation. While previous research efforts in ME have focused on targets further down the pathways that have involved enzyme activation, our research efforts focus at the higher end of the signaling pathway. **Methods:** We evaluated levels of DNA methylation in peripheral mononuclear blood cells isolated from 13 female CFS patients (ME diagnosis was based on the Fukuda and the Canadian criteria) and 12 matching by age and BMI female healthy controls using Illumina MethylationEPIC BeadChip arrays, controlling for the probes with the low detection, invariant probes and probes overlapping polymorphic sequences. DNA methylation analysis was performed using R software with RnBeads package. The mean difference in means across all sites in a region of the two groups, the mean of quotients in mean methylation and combined p-values calculated from all site p-values in the region were used to rank the degree of the differential methylation. **Results:** We found an increased abundance of hypomethylated promoters of genes related to the immune functions, cancer, cellular bioenergetics, neuronal functions in the cells of ME patients. The results have been validated by pyrosequencing using extended cohort of 34 patients and 34 healthy controls. **Conclusion:** This data show the importance of the DNA methylation in the regulation of the changes in gene expression in ME.

1651W

Wnt signaling in neural crest development: A possible mechanism for nonsyndromic cleft lip and palate. *A. Vedenko[1,2], J.I. Young[1], K.E. Hatzistergos[2], S.H. Lang[1], D. Van-Booven[1], J.M. Hare[2], J.T. Hecht[3], S.H. Blanton[1].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, Miami, FL; 2) Interdisciplinary Stem Cell Institute, Leonard M. Miller School of Medicine, University of Miami, Miami, FL; 3) Department of Pediatrics, McGovern Medical School, University of Texas Health, Houston, TX.

   Nonsyndromic cleft lip, with or without palate (NSCLP), is the most common craniofacial birth defect. Although both genetic and epigenetic factors have been associated with NSCLP, the molecular mechanisms underlying these defects remain poorly understood. Discordant monozygotic (MZ) twins provide a unique opportunity to dissect the underlying causes of complex diseases. Applying this approach, we performed methylation analysis of 3 MZ twin pairs discordant for NSCLP using whole genome bisulfite sequencing (WGBS) on saliva DNA. Two pairs were also subjected to methylation array analysis for validation purposes. We identified significantly distinct methylation patterns within twin pairs. Notably, GO and KEGG analysis, based on differentially methylated genes, revealed an enrichment in cell adhesion-related ontologies and pathways. It is well known that cell adhesion and migration are regulated via Wnt signaling pathways and downstream effectors. Importantly, variation in Wnt genes has been associated with NSCLP. Neural crest cells (NCC) contribute extensively to the formation of head structures including the palate through migration and population of branchial arches. Thus, abnormalities of NCCs are associated with numerous craniofacial defects. These findings lead us to hypothesize that epigenetic perturbations in NCC could be involved in the development of NSCLP, potentially through Wnt signaling alterations. We utilized *Sox10::GFP* reporter human embryonic stem cell line (*Sox10::GFP* hESCs) as an *in vitro* model of human NCC development. We found that canonical Wnt signaling was important for the development of *GFP+* human-NCCs. Activation of Wnt with the GSK3 inhibitor CHIR99021 significantly induced the expression of DNMT1, and to a lesser extent of DNMT3A and DNMT3B, in a dose-dependent manner. Conversely, expression of DNMT1 was significantly downregulated in response to treatment with the porcupine inhibitor IWP2. Moreover, these changes were accompanied by significant differences in the expression of cell adhesion genes, such as CDH1. Together, these correlative findings suggest that NSCLP is associated with changes in methylation of cell adhesion genes, and that canonical Wnt signaling-mediated induction of DNMT1 may be an underlying mechanism of this process. Future studies are underway to confirm these results and determine the role of DNMT1 in epigenetic regulation of human NCCs.

1652T

Shared household environment makes an important contribution to variation in the human methylome. *Y. Zeng[1], P. Navarro[1], C. Xia[1], A. Tenesa[2], D. Sproul[1], R. Walker[3], R. Marioni[3], S. Morris[3], K. Rawlik[2], C. Amador[1], O. Canela-Xandri[2], C. Hayward[1], A. Campbell[3], R. Nagy[1], I. Deary[4], D. Porteous[3], K. Evans[3], A. McIntosh[5], C. Haley[1].* 1) MRC human genetic unit, IGMM, University of Edinburgh, EDINBURGH, United Kingdom; 2) The Roslin Institute and Royal (Dick) School of Veterinary Sciences, University of Edinburgh, UK; 3) Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK; 4) Department of Psychology, University of Edinburgh, UK; 5) Division of Psychiatry, University of Edinburgh, Edinburgh, UK.

   DNA methylation levels at individual sites have been associated with many traits and diseases. To understand the regulatory pathways through which DNA methylation modulation may affect the phenotype, it is important to understand the genetic and environmental determinants of variation in DNA methylation levels. In this study we analysed data from a large family-based cohort (Generation Scotland, N=5200) with DNA methylation in blood measured at ~790,000 sites, genome-wide genotyping of ~500,000 common SNPs, family relationships from complex pedigrees and multiple disease and phenotypic measurements for each participant. We dissected the variation in methylation level at individual sites into genetic and family environmental contributions using variance component analysis. The average heritability, measuring the whole genome contribution to variation across all methylation sites was 19.2% (2.5[th] /97.5[th] percentile: 7.1%/70.8%). The mean contribution of the environment shared by a couple to variation in methylation levels was 4.1% (2.5[th] /97.5[th] percentile: 0%/18.8%) across all measured sites. An elevated contribution from the couple-shared environmental component and a decreased contribution from the genetic component to methylation levels was found close to transcription start sites (TSS) for genes with high levels of expression in blood. Couple-shared environment explained an average of 8.2% (2.5[th] /97.5[th] percentile : 0%/ 24.5%) of the variance of methylation at sites that were within a CpG Island-located transcription factor binding site (TFBS) within 200 bp of the TSS for the 20% most highly expressed genes in blood. This study identifies important contributions from genetics and couple-shared environment to DNA methylation. Our results further suggest that regulatory regions (TSS, TFBS and CpG Islands) of expressed loci may be especially susceptible to environmental modulation of methylation which could have potential consequences for subsequent phenotypic variation. This work is supported by the UK Medical Research Council ,Welcome Trust (104036/Z/14/Z), and Chief Scientist Office funding (CZD/16/6).

1653F

Rare variants and parent-of-origin effects on whole blood gene expression assessed in large family pedigrees. *A. Brown[1], A. Viñuela[1], A. Martinez-Perez[2], A. Ziyatdinov[2], M. Sabater-Lleal[3], A. Hamsten[3], J. Souto[4], A. Buil[1], J. Soria[2], E. Dermitzakis[1].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Unit of Genomics of Complex Diseases, Biomedical Research Institute Sant Pau (IIB-Sant Pau), Barcelona, Spain; 3) Karolinska Institute, Medicine, Stockholm, Sweden; 4) Sant Pau Hospital, Hematology, Barcelona, Spain.

   Studying genetic regulation of gene expression in related individuals provides insights inaccessible when using unrelated individuals, such as heritability, rare (in population) regulatory variants commonly observed in the pedigree, imprinting, and parent-of-origin effects. We report the GAIT2 study of 935 individuals from 35 pedigrees, with whole blood RNA-seq, blood cell count and extensive phenotypic information available. We identified 11,297 eQTLs (FDR=0.01) using a cis association mapping with variance components to consider the family structure. To test whether these eQTLs were rare in the general population, we looked at their MAF in 1000 Genomes populations. Compared to eQTLs from the DGN study with 922 unrelated individuals, we see an excess of variants with MAF<0.01 (9.6% eQTL vs 0%, median MAF is 0.11 vs 0.27). We used trios within the pedigrees to find parent-of-origin effects on expression, where the effect of an allele depends on whether it is maternally or paternally inherited (parent-of-origin eQTL or poeQTL). A cis scan revealed 12 significant poeQTLs (FDR<0.05). Six of these affect known imprinted genes, implying a cis-eQTL with effect masked on the imprinted haplotype. However, for four other poeQTL genes, measures of allelic ratio from either our data or the GTEx study showed a combination of mono- and bi-allelic expression across individuals, suggesting these genes may be partially imprinted (the other two genes had no allelic ratio calls). We are currently producing higher coverage RNA-seq data for a subset of individuals, to associate partial imprinting with genetic variation and produce a genome-wide assessment of imprinting in general. Both rare variants and parent-of-origin genetic effects have been shown to be relevant for human disease. Studies such as this allow a deeper understanding of their properties, for example revealing that exclusively studying imprinted regions will miss many parent of origin effects.

1654W

Stratified comparison and network analysis of large eQTL-studies reveals factors affecting validity of cis- and trans eQTLs. *H. Kirsten[1,2], R. Burkhardt[2,3], J. Thiery[2,3], K. Krohn[4], M. Loeffler[1,2], M. Scholz[1,2].* 1) Medical Faculty / IMISE, University Leipzig, Leipzig, Germany; 2) Medical Faculty / LIFE, University Leipzig, Leipzig, Germany; 3) Medical Faculty / ILM, University Leipzig, Leipzig, Germany; 4) Medical Faculty / IZKF, University Leipzig, Leipzig, Germany.

Associations of genetic variants with gene expression levels (eQTL studies) are increasingly used in the elucidation of the human genetic architecture and causal molecular mechanisms of genetically influenced complex traits and diseases. Typically, eQTL studies are done independently from a certain GWAS that uses the eQTL information. Hence, it is important to understand factors affecting validity of reported eQTLs. Recently, the so far largest single eQTL study done with 5,257 individuals (Joehannes 2017, DOI: 10.1186/s13059-016-1142-6) done in whole blood reported replication rates for trans eQTLs down to 5.2%. However, systematic differences between studies might explain these low numbers, e.g. gene-expression measurement technology, effect size of eQTLs, sample size of the studies, and tissue type. We present novel data of two large eQTL studies (total n=6,741) done in two tissues (PBMC and whole blood) to unravel factors influencing replication rates of cis/trans effects. Factors analyzed include tissue type, technology, sample size, eQTL-effect size, known GWAS association of eQTL-SNPs (i.e. potential functional relevance) and network properties (using the framework of a bipartite eQTL-network). We identify power, namely eQTL-effect size as very critical factors especially for trans-eQTLs replication rate followed by tissue type with up to tenfold higher trans-replication rates than originally reported in the FHS-study. We observe that GWAS-related cis-eQTLs show roughly comparable replication rates whereas GWAS related trans-eQTLs, shows considerably higher replication than the respective non-GWAS related eQTLs. Interestingly, correspondence rate of the direction for eQTLs between our study and the FHS study was even higher for trans-eQTLs than for cis-eQTLs. This difference might be explained by different gene expression technologies (Affymetrix exon array vs. Illumina HT12v4 array) hinting to specific effects of different splice variants of a certain gene. Furthermore, we find similar eQTL-network properties across both studies. Our results contribute to the understanding of the genetic architecture of eQTLs and are important for the interpretation of GWAS results when using data from external eQTL studies.

1655T

Glomerular and tubulointerstitial eQTLs of patients with nephrotic syndrome. *R. Putler[1], X. Wen[2], C.E. Gillies[1], M.G. Sampson[1], Nephrotic Syndrome Study Network (NEPTUNE).* 1) Pediatrics-Nephrology, University of Michigan, Ann Arbor, MI; 2) Biostatistics, University of Michigan, Ann Arbor, MI.

Background Using intrarenal mRNA expression as molecular phenotypes in expression quantitative trait loci (eQTL) studies of nephrotic syndrome (NS) can lead to the discovery of transcripts under significant genetic control, reveal novel NS-related biology, and identify loci associated with clinically meaningful outcomes. To describe the intrarenal eQTL landscape of NS, we paired whole genome sequencing with glomerular (GLOM) & tubulointerstitial (TI) transcriptomes from patients in the Nephrotic Syndrome Study Network (NEPTUNE). **Methods** NEPTUNE is a prospective, longitudinal study of NS enrolling affected adults and children receiving a clinically indicated biopsy. Rich demographic, clinical, and genomic data is collected and GLOM and TI transcriptomes are derived from a microdissected research biopsy core. We used the Bayesian "Deterministic Approximation of Posteriors" (DAP) to fine-map eQTL signals within each locus while controlling for multiple testing. DAP uses linkage disequilibrium and annotation information to determine independent eSNPs within an eQTL locus. We included age, sex, PEER factors, & the 1st 4 principal components as covariates. We utilized MatrixEQTL for the same data to gain insight into direction of effect for these signals. **Results** At genome-wide significance, we discovered 340 independent eQTL signals in 313 unique genes in GLOM and 862 independent signals in 772 unique genes in TI. *PLCG2*, previously implicated in steroid sensitive NS via burden testing, was one of the strongest glomerular eQTLs. Five of 30 Mendelian NS genes had both a GLOM eQTL and a significant association with eGFR, suggesting that non-Mendelian variation in these genes can impact clinical outcomes. We find 14% and 28% of identified eQTLs are GLOM-specific and TUB-specific, respectively, and 58% of eQTLs are shared between the two tissues. When we compared the significance of GLOM- & TI-eQTLs with those derived from blood (from GTeX), 19% of eQTLs had more significant minimum p-value in GLOM & TI vs blood, despite smaller sample size. **Conclusions** In our GLOM & TI eQTL study of NS patients, we discovered genes under substantial genetic control and can now pursue their biologic & clinical correlates. Making this NS eQTL database publicly available could be a useful resource to inform future molecular & epidemiologic studies of NS.

1656F

Epigenetic marks at major histocompatibility complex affect male fertility. *S. Sarkar, R. Singh.* Endocrinology Division, CSIR-Central Drug Research Institute, Sector 10, Jankipuram Extension, Sitapur Road, Lucknow, Uttar Pradesh 226031.

Human leucocyte antigens are membrane glycoproteins playing salient role in immune recognition and response. The expression pattern of HLA class antigens is dubious in nature. However, quite a number of studies have reported the expression of HLA class molecules. Trophoblastic *HLA-C* gene variants have been shown to be associated with pre-eclapmsia in African and European women. Maternal *HLA-C* genetic variants are also seen to have effect on high fetal and maternal mortality and morbidity. HLA class 1 molecules were found to be expressed in spermatozoa of infertile patients but showed no significant difference against fertile groups. On the other hand, it was found that mRNA levels of *HLA-G* in normal spermatocytes were significantly higher than the testicular samples with only Sertoli cells and/or spermatogonia, which suggests its involvement in spermatogenesis. A study observed that the predisposition to anti-sperm antigen can increase the levels of some HLA class-1 antigen in vasectomized men. A whole genome epigenetic study on adolescent chemotherapy patients causing infertility also showed differential methylation pattern in *HLA-C* gene. We investigated differential methylation regions present in different classes of *HLA* in patients with different sperm count in order to correlate epigenetic patterns with sperm production ability. To achieve this objective, our experimental method included isolation of sperm DNA from individuals of various sperm count groups followed by Infinium 450K DNA methylation array. Among these groups we have indentified six different classes of *HLA* genes to be differentially methylated. *HLA* class I genes showed higher level of methylation in oligospermic group as compared to normospermics, suggesting the potential role of these genes in spermatogenesis. Similar methylation pattern was also seen in DNA from PBMCs. Out of all immune response gene only *HLA-DRB1* showed disparate methylation pattern in oligospermics (b-value=0.345) than normal men (b-value=0.679). This DNA methylation pattern suggests an important role of these immune responsive genes in sperm production. These observations opened a valuable insight into the cause of infertility which directs the focus to select best sperm for fertilization success. However, more epigenetically focused experimental designs are warranted to draw conclusive insights and theories about the role of immune response genes and reproductive health.

1657W

Low correlation observed between DNA methylation in blood measured between a majority of CpG sites measured on both Illumina 450K and EPIC BeadChips. *M.W. Logue[1,2,3], A.K. Smith[4,5], E.J. Wolf[1,2], H. Maniates[1], A. Stone[6], S.A. Schichman[6], R.E. McGlinchey[7,8], W. Milberg[7,8], M.W. Miller[1,2].* 1) National Center for PTSD, VA Boston Healthcare System, Boston, MA; 2) Department of Psychiatry, Boston University School of Medicine, Boston, MA; 3) Biomedical Genetics, Boston University School of Medicine, Boston, MA; 4) Gynecology and Obstetrics, Emory University School of Medicine, Atlanta, GA; 5) Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA; 6) Pharmacogenomics Analysis Laboratory, Research Service, Central Arkansas Veterans Healthcare System, Little Rock, AR; 7) Geriatric Research Educational and Clinical Center and Translational Research Center for TBI and Stress Disorders, VA Boston Healthcare System, Boston, MA; 8) Department of Psychiatry, Harvard Medical School, Boston MA.

BeadChips are the de-facto instrument of choice for genome-wide epidemiological investigations of DNA methylation. Many researchers are working with data generated using the Illumina Infinium HumanMethylation450 BeadChip, which measures the proportion of methylated DNA (denoted β) at nearly 500,000 CpG sites. This chip was recently discontinued and replaced with the Illumina MethylationEPIC BeadChip, which assesses methylation at nearly 850,000 CpG sites, including 90% of sites assessed in the 450K chip. We examined the concordance between 422,524 CpG sites measured by both chips in 145 whole blood DNA samples. The correlation of β values at most sites was quite low; the median correlation for a CpG was $r=0.15$, (range $r=-0.34$ to $r=0.95$). Only 23% had correlation values > 0.50. We then investigated the association between the number of beads per site, missing rate, and range of observed β values and the correlations for each CpG, with separate models for sites depending on the Beadchip Assay used (Type I or II). For Type I probes, there was a strong positive association between the β range and CpG correlation and with a quadratic effect indicating downward curvature. The other model terms were significant ($p<2.2*10^{-16}$), but had less of an effect. A comparison of the multiple $r^2=0.63$ for the full model compared to the multiple $r^2=0.0048$ for a model excluding the β range variables indicated that the majority of the variance in correlation for the CpG sites was explained by the amount of variability in β values. Results were similar for type II probes. The vast majority of sites with β range <0.05 had low correlation values (median correlation $r=0.031$), and most were nearly completely methylated or unmethylated (95.81% had a median methylation value <0.05 or >0.95). The correlation was higher for CpGs with a greater β range. For the CpGs that had a β range of greater than 0.20 (n=61,248), the median correlation was $r=0.71$. An examination of the intraclass correlation coefficient based on 11 DNA samples assessed repeatedly with the EPIC chip indicated that ICC values depended on the β range ($p<2.2*10^{-16}$), indicating that ICC values may also be reduced in single-chip studies. Low correlation observed at sites with a small range may be due to the larger proportion of measurement error compared to true signal. Filtering out probes based on the observed correlation or the low variability may increase reproducibility of BeadChip-based methylation studies.

1658T

DNA methylation of *PPARGC1A* is associated with cycling performance.
D.J. Hunter, B. Hussey, L. Wilson, M.R. Lindley, S.S. Mastana. Translational Chemical Biology Research Group, School of Sport Exercise and Health Sciences, Loughborough University, Loughborough, UK.

   Introduction: Exercise and nutritional interventions, such as n-3 polyunsaturated fatty acids (n-3 PUFAs), have been shown to alter DNA methylation and expression of genes involved in metabolism and inflammation. Peroxisome proliferator-activated receptor gamma co-activator 1-alpha encoded by the *PPARGC1A* gene [MIM 604517] regulates exercise adaption in skeletal muscle via activation of pathways for mitochondrial biogenesis and *PPARGC1A* mRNA expression is linked to inflammation by an inverse correlation with interleukin-6 mRNA. Skeletal muscle *PPARGC1A* is shown to be hypomethylated post exercise; however, it is unknown if this response is tissue specific. The aim of the investigation was to determine if hypomethylation of *PPARGC1A* occurs in blood post exercise and if n-3 PUFAs modulate the response. A second aim was to assess if DNA methylation of *PPARGC1A* associated with phenotypes related to exercise and inflammation. **Method:** Eight male cyclists completed 4 trials, each separated by 4-weeks. Supplementation of n-3 PUFA or olive oil separated the trials, using a double-blind, randomised repeated measures design. Blood samples were collected pre and post a 1h cycling performance test (45 min steady state and 15 min time trial) during each trial. DNA was extracted and bisulfite converted using EpiTect LyseAll kit (Qiagen) and methylation of a CpG -260 bases from the *PPARGC1A* transcription start site was quantified using a Pyromark Q48 Autoprep. **Results:** *PPARGC1A* -260 CpG methylation was unaltered by exercise or supplementation. After correction for cell heterogeneity, a significant decrease in methylation post exercise was detected ($p < 0.05$). Baseline and post-exercise methylation was positively correlated with mean power ($R > 0.449$; $p < 0.010$), and work completed ($R > 0.396$; $p < 0.024$) during the time trial. After correction for cell heterogeneity the correlations with post-exercise *PPARGC1A* methylation and performance remain significant ($p < 0.05$). **Conclusion:** Decreased *PPARGC1A* methylation post exercise in blood, after correction for white blood cell populations, suggests that exercise induced hypomethylation may not be tissue specific. A positive correlation shown between methylation and cycling performance indicates that methylation can affect performance. Comparative studies of skeletal muscle and blood are required to delineate any tissue specificity to determine the efficacy of using blood as a proxy for skeletal muscle in methylation studies.

1659F

Exploratory transcriptome and methylome analysis in Gilbert's syndrome. *A. Tosevska[1], A. Grindel[2], D. Doberer[3], R. Marculescu[4], C. Moelzer[5], M. Morselli[1], M. Wallner[6], K.-H. Wagner[7], M. Pellegrini[1].* 1) Department of Molecular, Cell and Developmental Biology, UCLA, 610 Charles E. Young Drive East, Los Angeles, CA 90095, USA; 2) Department of Nutritional Sciences, University of Vienna, Althanstrasse 14, Vienna, A-1090, Austria; 3) Department of Clinical Pharmacology, Medical University of Vienna, Waehringer Guertel 18–20, 1090 Vienna, Austria; 4) Clinical Institute of Laboratory Medicine, Medical University of Vienna, Waehringer Guertel 18–20, 1090 Vienna, Austria; 5) School of Medicine, Medical Sciences and Nutrition, Institute of Medical Sciences, University of Aberdeen, Ashgrove Road West, AB25 2ZD, Aberdeen, UK; 6) Institute of Dietetics and Nutrition, University of Applied Sciences, FH JOANNEUM, Alte Poststraße 149, 8020 Graz, Austria; 7) Research Platform Active Ageing, University of Vienna, Althanstrasse 14, Vienna, A-1090, Austria.

   Unconjugated bilirubin (UCB) is an endogenous metabolite in mammals; its circulating concentration is regulated by the hepatic expression of uridine diphosphate glucuronosyltransferase (UGT1A1). In humans, variants in the UGT1A1 gene lead to reduced enzyme activity, resulting in a moderate increase in circulating UCB beyond the range found in normal serum. Gilbert's Syndrome is one common that arises from homozygosity of variant-carrying UGT1A1 alleles (UGT1A1*28). While hyperbilirubinemia has been associated with a beneficial metabolic and aging phenotype, much less is known on the effects of UCB on cellular processes. The aim of this exploratory analysis is to identify changes in the epigenomic and transcriptiomic regulation that might be driven by acute and chronic hyperbilirubinemia. To this end, reduced representation bisulfite sequencing (RRBS) was performed on peripheral blood samples from an age and sex-matched case-control study (n = 60/60) on Gilbert's Syndrome. In addition, we used an Affymetrix GeneChip™ Human Gene 2.1 ST Array on a subset of samples (n = 23 cases/ 23 controls). We performed weighted gene coexpression analyses and identified gene modules related to increased UCB concentration. In addition we identified CpG sites positively or negatively associated with UCB, and the genes associated with these sites. Our preliminary results show a negative relationship between UCB and genes responsible for gene expression, RNA processing and transport as well as vitamin D3 and B3 metabolism and cholesterol biosynthesis. Genes found in close proximity of significant CpGs were associated with heme degradation, de novo pyrimidine and ATP biosynthesis, pentose phosphate metabolism and circadian rhythms, among others. There is little overlap between the two datasets, suggesting the short-term and long-term effects from increased UCB might differ. The findings can greatly contribute to our understanding the role of bilirubin in human biology and the mechanisms by which increased UCB offers protection against age-related disease.

1660W

Various relationships between DNA methylation and gene expression in different tissues and ages. *K. Wang[1], R. Dai[1], Y. Xia[1], C. Jiao[1], C. Chen[1], C. Liu[1,2].* 1) State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China; 2) Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA.

Traditionally, DNA methylation is described as a mechanism that represses gene expression by blocking the binding of transcription factors at promoter regions. However, many studies suggest that DNA methylation variation makes a number of diverse contributions to gene expression and its complex underlying mechanisms. To uncover the relationship between gene expression and DNA methylation, we collected tissues that have matched methylation and expression data from databases and publications. Totally we retained 502 adult prefrontal cortices (PFCs), 43 lymphoblastoid cells, 75 livers, 1201 monocytes, and 247 developmental PFCs across the lifespan. For each region, we defined gene and CpG pairs (GCPs) as CpG site located within the 10kb flanking region of the corresponding gene. We used Spearman's rank test to assess the correlation of GCPs and the FDR to correct for the multiple hypothesis testing. In the adult PFC, 230,087 GCPs were analyzed for the Spearman correlation between gene expression and CpG site methylation, and only 8,828 GCPs were significantly correlated after FDR correction. Of these, 5,530 GCPs were negatively correlated and 3,298 GCPs were positively correlated. The other data also exhibited a small proportion of significant GCPs. Though thousands of significant GCPs have been identified in different tissues, only 3 of them were consistent in all tissues. Additionally, there were 442 negatively and 153 positively correlated GCPs shared between the adult PFC and the developmental PFC. The CpGs that were negatively correlated with genes expression were more likely to locate in the promoter and genebody and the positive ones only preferred genebody. We also explored the epigenetic state of these CpGs using the Roadmap 15-core state annotation. Negatively correlated CpGs were enriched in active regions while positive ones were enriched in repressive regions. For the genes of GCPs, we performed the GO term enrichment analysis using WebGestalt. We found that genes in correlated GCPs in developmental PFC were enriched for neuron projection development, neuron differentiation, and synapse part. In the adult PFC data, genes were also enriched for neuron part, distinct from the GO term referring to genes in the developmental PFC. In other tissue data, the genes of significant GCPs also enriched for tissue-specific functions. Besides, the genes of significant GCPs in developmental and adult PFC showed association with psychiatric related gene sets.

1661T

Nanopore full length mRNA sequencing resolves transcript structure in single auditory hair cells. *P. Ranum[1,2], A. Goodwin[2], R. Smith[1,2,3].* 1) Interdisciplinary Graduate Program in Molecular & Cellular Biology, The University of Iowa Graduate College, University of Iowa, Iowa City, IA 52242, USA; 2) Molecular Otolaryngology and Renal Research Laboratories, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA; 3) Department of Otolaryngology, Head and Neck Surgery, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA.

Single cell RNA-Seq (scRNA-Seq) has revealed higher than anticipated levels of heterogeneity both in terms of transcript abundance and transcript structure prompting researchers to rethink assumptions about transcription. We have applied scRNA-Seq technology to auditory hair cells, the critical mechanosensory cells of the auditory system. In a dataset of over 200 individual auditory hair cell and supporting cell samples we were surprised to find unexpected variation in transcript structure of many genes. Our analysis of transcript structure was complicated by the limitations of the Illumina sequencing technologies that require library fragments between 150 and 800 bp and limit read length to 200bp or less. Analysis of Illumina reads allows the identification of splice sites but precludes the reconstruction of full length transcripts especially for long genes with many exons or isoforms. We utilized long read Nanopore sequencing technology to sequence full length mRNA transcripts from single cell cDNA libraries. This analysis revealed the inclusion of previously unreported exons, unreported splicing patterns, and by extension novel isoforms in a variety of genes. By quantifying the abundance of these full length transcripts we can establish the relative abundance of various isoforms and separate very low abundance isoforms from those that are abundantly transcribed and more likely to have biological significance. The identification of previously unappreciated transcript structural diversity at the single cell level in a dataset containing only three cell types indicate that there is likely to be a wealth of similar transcriptional diversity waiting to be characterized in other highly specialized cell types throughout the mammalian body. These findings in auditory hair cells have implications for the study, diagnosis, and treatment of deafness. Similar findings in other cell types may have implications for other diseases.

1662F

Local and systemic alterations in extracellular RNA following traumatic knee injury implicate catabolic and inflammatory biomarkers. *A.J. Griswold[1,2], J. Perez[2], N. Hofmann[1], T. Strong[1], L. Wang[1,2,3], D.D. Vance[4], T.M. Best[2], J.M. Vance[1,2,3], M.A. Pericak-Vance[1,2,3], L.D. Kaplan[2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami Miller School of Medicine, Miami, FL; 2) UHealth Sports Medicine Institute, University of Miami, Miller School of Medicine, Miami, FL; 3) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School Medicine, Miami, FL; 4) Department of Orthopedic Surgery, New York Presbyterian Hospital, Columbia University Medical Center, New York, NY.

Meniscal and anterior cruciate ligament (ACL) injuries are the most common activity-induced orthopedic traumas, and are risk factors for osteoarthritis (OA). We hypothesize that molecular changes, initiated at the time of or soon after trauma, lead to increased risk of OA. This study set out to characterize local and systemic extracellular RNA (exRNA) altered by traumatic knee injury in synovial fluid and plasma. The cohort consisted of 14 subjects undergoing ACL and/or meniscal repair surgery during which whole blood was collected and synovial fluid was aspirated. Five subjects had synovial fluid collected from the contralateral knee. Radiography found no OA, though Outerbridge assessment indicated varying levels of chondral damage. RNA was extracted from plasma and synovial fluid and analyzed via massively parallel RNA sequencing. More than 30 million reads per sample were generated, aligned to the reference genome, and exRNA abundance quantified. Plasma and synovial samples had >10,000 types of exRNA including protein coding genes and miRNAs. Plasma exRNA expression levels showed no differences when stratifying Outerbridge scores or injury duration. However, there was high expression of inflammatory (*CXCL5, CCL5,* and *IL16*), catabolic (*MMP -21, -17, and -13,* and *ADAMTS8*), and miRNA genes involved in arthritis (including *MIR16, MIR20b,* and *MIR146a*). Comparing synovial exRNA between uninjured vs injured revealed 69 differential genes (p ≤ 0.001, 44 increased and 25 decreased) including increases of three matrix metallopeptidase (*MMP3, -21,* and -1) with roles in cartilage catabolism. Injuries with higher average Outerbridge scores showed 278 differential genes (p ≤ 0.001, 87 increased and 191 decreased). The decreased genes were enriched for translational initiation factors (*EIF1, EIF2S1,* and *EIF3E –I,* and *–L*), and a matrix metallopeptidase (*ADAM12*) was increased. Finally, comparing <4 months injury duration to greater indicated 46 differential genes (p ≤ 0.001, 33 increased and 13 decreased). The increased genes include skeletal and muscular function genes including skeletal troponins (*TNNC2* and *TNNI2*), tropomyosin 2 (*TPM2*), myosin light chain 1 (*MYL1*), and skeletal muscle actin (*ACTA1*). Overall, this study demonstrates that exRNA molecules can be isolated and characterized from synovial fluid and plasma. While the mechanism of these exRNAs remains unknown, identification of altered pathways could lead to targets for early intervention against OA.

1663W

An epigenome correlation map using Infinium 450 DNA Methylation Array. *W. Guan[1], Q. Chen[1], J.S. Pankow[1], E.W. Demerath[1], J. Bressler[2], M.L. Grove[2], M. Fornage[2].* 1) University of Minnesota, Minneapolis, MN; 2) University of Texas Health Science Center at Houston, Houston, TX.

DNA methylation is an epigenetic mechanism used by cells to control gene expression and involves the addition of a single methyl group to cytosine at position 5. Linkage disequilibrium (LD), i.e., the correlation of alleles at different loci, of the human genome has been widely studied, which provides insights to the history of recombination in populations. In contrast, the correlation of DNA methylation levels among different CpG sites has only received limited attention. Since the methylation level is mitotically heritable but also sensitive to environmental exposures, both short-range and long-range correlation can exist. As part of the Atherosclerosis Risk in Communities (ARIC) study, the Illumina Infinium HumanMethylation450K (HM450) BeadChip was used to measure DNA methylation at over 480,000 cytosine-guanine (CpG) dinucleotide sites in peripheral blood obtained from ~3000 African American participants. In this project, we computed and characterized epigenome-wide correlation between all pairs of CpG sites using the ARIC data. We compared correlation patterns among CpG sites in close proximity to those in long distance or even on different chromosomes. We also evaluated the impact of adjustment for technical and environmental factors on these correlations. For example, for cg05575921, a CpG site known to be associated with smoking status, the 99th percentile of Pearson correlation coefficients between this site and CpG sites within 20 kb is 0.54, and 0.28 with CpG sites >20 kb away. After controlling for age, sex, and smoking, the 99th percentile value drops to 0.26 and 0.10, respectively. The results demonstrate stronger correlation among CpG sites over short distances, but many sites also show extensive long-range correlation. This analysis may help identify genes or biological pathways that are coordinately regulated.

1664T

The 5-HTTLPR polymorphism does not moderate the effect of sleep loss on neural responses to implicit threat and fear learning and memory. *V.C. Kodavali[1], J. Nguyen[1], R. McNamee[1], N. Rode[1], V.L. Nimgaonkar[1,2], A. Germain[1].* 1) Psychiatry, University of Pittsburgh, Pittsburgh, PA; 2) Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

Introduction: The short (s) allele variant of the serotonin transporter polymorphism 5-HTTLPR, compared to the long (l) allele variant, has been associated with heightened stress reactivity and greater vulnerability to stress-related psychiatric disorders. We explored the moderating impact of this polymorphism on neural correlates of implicit threat detection and fear-associated learning and memory following normal sleep, total sleep deprivation, or sleep restriction in healthy young adults. **Methods:** 134 participants spent 3 night in the laboratory after being genotyped for 5-HTTLPR. After a baseline night, neural responses to implicit threat cues were assessed in the evening. Participants were then randomized to total sleep deprivation (SD, n = 35), sleep restriction (SR or half of habitual time, n=40), or normal sleep (NS, n = 43). On the morning following the randomization night, participants completed a fear conditioning and extinction protocol. Fear extinction recall was tested in the evening, when neural responses to implicit threat cues were also re-tested. BOLD signals in the amygdala and ventromedial prefrontal cortex (vmPFC) were compared between the three genotypes (s/s vs. s/l vs. l/l carriers). ROI analyses were conducted using SPM8 and SPSS. **Results:** No significant Sleep Group X Genotype interaction was found for fear conditioning, fear extinction, or extinction recall for the right or left amygdala, or vmPFC. For the implicit threat task, a main effect of genotype was detected for the right amygdala (F(2,129) =3.35, p < 0.05) at baseline, where the s/l group showed lower BOLD activation in response to threatening vs. neutral cues than the l/l and s/s groups. Following the sleep manipulation, the three genotypes no longer differed. **Conclusion:** The 5-HTTLPR polymorphism does not appear to moderate the effect of sleep loss on neural responses to Implicit threat and fear learning and memory. Heterozygotes (s/l carriers) showed the greatest BOLD reactivity to threat cues in the right amygdala at baseline, but this effect was absent following the sleep manipulation. Sleep loss may override the impact of genotype on processes related to fear learning and extinction as well as neural reactivity to salient cues. The extent to which sleep loss may attenuate the effects of other polymorphisms on neural response to emotionally salient cues remains to be determined. .

1665F

Powerful and robust method for XCI-escape inference from bulk RNA-seq. *R. Sauteraud, J. James, D. Liu, L. Carrel.* Pennsylvania State University, Hershey, PA.

X chromosome inactivation (XCI) mechanism randomly silences one female X (Xi) and only expressing the alleles on the active X (Xa). ~10% of the X-linked genes escape XCI and express both copies of alleles. Xa/Xi assignment varies between cells (mosaicism), which makes it challenging to identify XCI escape genes from bulk RNA-seq data. We propose a novel binomial mixture model for identifying XCI escape gene. The method proceeds by estimating XCI skewing (i.e. the fraction of cells where one particular haplotype is Xa) from a training set of putative inacted genes. The genes with more balanced allelic expression than XCI skewing will be deemed as candidates of XCI escape genes. The model accounts for the read depths over-dispersion, while simultaneously accommodating for the possible sequencing error or imperfect training set (e.g. some putative inactivated genes may actually escape XCI). The number of mixture components is automatically determined via a backward elimination procedure, which ensures model parsimony and maximizes power. We experimentally established the validity and power of proposed method using single cell clones and male-female differential gene expression (DGE). In the single cell clone experiment, cells from a single individual with maternal or paternal Xa were isolated and separately cloned. Each clone thus has identical Xa/Xi assignment. Any genes that are mono-allelically/bi-allelically expressed in single cell clones will be identified as inactivated/escape genes. We also created mixtures of single cell clones, with skewing ranging from 50/50 to 90/10 to mimic a randomly X-inactivated sample. Type-I error and power was evaluated by comparing inference results from mixture samples to single cell clones. The type I errors across all scenarios were well controlled. The power increases with XCI skewing and is well above 80% for cell mixtures with a skewing greater than 20/80. We further demonstrate the utility of this method with DGE in the GEUVADIS dataset. Our method correctly predicts all 11 genes in PAR1 region as escape gene. Consistent with existing literature, the genes in non-PAR region predicted as escape tend to be up-regulated in females. On the contrary, genes in the PAR1 region are down-regulated in females due to dosage compensation. The method is implemented as a publicly available R package. We envision that it may be extremely useful for the analysis of broad available bulk RNA-seq data and XCI inference.

1666W

Prevalence, tissue-specificity and age-dependent heritability of skewed X-inactivation. *A. Zito, P-C. Tsai, S. Verdi, T. Martin, J.T. Bell, K.S. Small.* Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK.

   To equalize the X-linked transcriptional dosage between the sexes, one of the two X-chromosomes is inactivated in somatic tissues of female mammals. The long non-coding RNA *XIST* is uniquely expressed from the inactive X and drives the X-chromosome inactivation (XCI) process. Non-random, or skewed XCI, defined where at least 80% of the cells exhibit a preferential inactivation of one parental X chromosome, has been observed in female somatic tissues. Skewed XCI has been associated with ageing and complex traits including autoimmune diseases and cancer. We assessed the prevalence, heritability and phenotypic associations of skewed XCI in a multi-tissue sample of 850 female twins from the TwinsUK cohort (age 38-85, median age = 60). XCI was quantified in LCL (lymphoblastoid cell lines), whole blood, fat and skin biopsies via allele-specific expression (ASE) of *XIST* from paired RNA-seq and WGS data. We validated our quantification of XCI with HUMARA assays (N = 32, rho = 0.71) and repeated blood draws taken 1-3 years apart (N = 16, rho = 0.94). By defining skewed XCI when *XIST*-ASE < 0.2 or *XIST*-ASE > 0.8, we identify a high prevalence of tissue-specific XCI, with skewed XCI patterns being more common in blood-derived tissues (LCL = 35%, blood = 22%) than skin (10%) or fat (12%). Tissue-tissue correlations indicate highest similarities of XCI skewing levels between blood-derived tissues (rho = 0.8) and between skin and fat (rho = 0.47), with lower correlations observed between blood-derived tissues and either skin (rho = 0.3) or fat (rho = 0.3). XCI skew was associated to age (P = 0.001) in blood-derived tissues, with an inflection point at 55 years of age and 42% of individuals >65 demonstrating skewed XCI. No association was detected between XCI skew and age in skin (P = 0.37) or fat (P = 0.7). We confirmed previous reports of heritable XCI skew in blood, but find that heritability of XCI is limited to women >55 years-old (h^2 = 0.25). XCI skew was not heritable in blood of women <55 years-old (h^2 = 0), nor in skin and fat (h^2 = 0) at any age. XCI skew was more extreme in blood-derived tissues in individuals with an autoimmune disease (P = 0.01) and in current smokers (P = 0.02). We conclude that XCI patterns are tissue-specific and that the heritability of XCI skew in blood is dependent on age, representing a Gene x Age interaction that can shift the functional allelic dosage of an entire chromosome in a tissue-rectricted manner.

1667T

RIPK3-dependent regulation of cell death switch (live or dye) as major determinant in incontinentia pigmenti. *A. Pescatore, M. Cirigliano, L. Patanella, A. Barra, M.V. Ursini.* Institute of Genetics and Biophysics "A. Buzzati-Traverso" CNR-Via P. Castellino, 111 80125 Naples, Italy.

   Incontinentia pigmenti (IP) is a X-linked neuroecodermal disease. The only causative gene is NEMO (also called IKBKG), encoding for the crucial regulator of the NF-κB transcription pathway. Accordingly, the impaired NF-κB activation and the defective response to pro-apoptotic stimuli induced trough the TNF-α receptor1, underline the pathogenesis of this inflammatory disease. Conditional or full Knock out mouse models of the disease have confirmed these knowledges. A conundrum has been to distinguish between the two independent activities of NEMO: the one of NF-κB-mediating gene transcriptional activation and that of inhibition of cell death. The IP-associated NEMO-A323P mutant has provided an answer to that question. Cell expressing the A323P NEMO mutant have a completely abrogated TNF-induced NF-κB activation, nevertheless exhibiting resistance to TNF-induced apoptosis. We demonstrated that NEMO, independently from NF-κB activation, controls structural and functional dynamics of the different TNF-R1-induced complexes by regulating the RIP kinases activities. We observed that while a robust caspase activation is evidenced in NEMO-null cells concomitant with the RIPK3 recruitment to the apoptosis-mediating TNF receptor complex, In contrast, cells expressing the ubiquitin-binding mutant NEMO-A323P did not recruit RIPK3 to complex II, an event that prevented caspase activation. We propose RIP3 as a new crucial player in the IP pathogenesis because in the presence or in the absence of NEMO a different cellular outcome (live or die) or a different cell death process (apoptosis or necroptosis) may result. In this respect, RIP3 represents a putative therapeutic target in Incontinentia Pigmenti and in other NF-κB-related human diseases.

1668F

Identification of genetically associated changes in 3D-chromatin architecture by leveraging haplotype information across a three-generation family. *W.W. Greenwald[1], H. Li[2], P. Benaglio[3], A. Schmitt[4], Y. Qiu[1], B. Ren[2,5,6,7], M. D'Antonio[2], E.N. Smith[3], K.A. Frazer[2,3].* 1) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, USA; 2) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA; 3) Department of Pediatrics and Rady Children's Hospital, University of California, San Diego, La Jolla, CA, USA; 4) Arima Genomics, San Diego, CA, USA; 5) Ludwig Institute for Cancer Research, La Jolla, CA, USA; 6) Moores Cancer Center, University of California, San Diego, La Jolla, CA, USA; 7) Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA.

Genetic variation that affects gene expression can occur in regulatory regions that are far from the gene's locus, yet colocalize spatially through 3D-chromatin loops. However, it is unclear if this genetic variation acts by modulating regulatory region activity, or by inducing changes in the 3D-architechture of the genome itself. To study this question, we obtained paired iPSC and iPSC-derived cardiomyocyte (iPSC-CM) lines from 7 individuals in a 3-generation family with whole genome sequence data. We performed in-situ Hi-C and obtained a total of ~6 billion read pairs which we analyzed using Fit-Hi-C and HICCUPS to identify 17,567 and 19,003 significant loops in iPSCs and iPSC-CMs, respectively, 9,888 of which were common to both cell types. To identify loops associated with genetic variation, we performed allelic imbalance analyses with the HiC reads. To assign reads to either the maternal or paternal haplotype, we initially obtained long-range haplotypes from the HiC data; however, when we compared the resulting haplotypes to expected inheritance patterns, we observed a 1.8% switch error rate. We thus devised a method which leveraged family structure, enabling us to resolve 95% of these switch errors and resulting in 4.7M phased SNVs across the genome. To estimate allelic imbalance, we assigned reads within 25kb of a loop's anchor that carried an allele of a phased heterozygous SNV to the corresponding maternal or paternal haplotype of the individual, performed a binomial test, and combined the individuals' associations at each loop using Fisher's combined p-value. In total, we identified 35 genetically associated chromatin loops with FDR < 0.05 across both cell types. Across all loops, 8 overlapped 2 somatic CNVs identified in these iPSCs, and for both CNVs we observed genetic effects with either a corresponding gain or loss, supporting that our method accurately detects genetically associated changes in chromatin loop formation. These data suggest the 3D-chromatin architecture of the genome may be largely invariant and only modestly affected by genetic variation; we are currently validating this observation by testing whether the genetic variants that are located in the distal end of a chromatin loop and are associated with allele-specific RNA expression at the genic end of the loop are also associated with corresponding differences in chromatin loop formation.

1669W

HyCCAPP uncovers CALR as a novel DNA-binding protein. *H. Guillen[1], D.S. Perumalla[1], P.K. Rao[1], A.Y. Jadhav[1], J.M. Proffitt[1], M.J. Monotya[1], M.R. Shortreed[2], L.M. Smith[2], M. Olivier[1].* 1) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Chemistry, University of Wisconsin, Madison, WI.

Whole-genome sequencing analyses have greatly improved the identification of disease-associated sequence variants. Most of these associations are located in non-coding regions, suggesting likely expression-modulating mechanisms. Not surprisingly, a number of these sequence variants are strongly associated with both disease-related traits and the expression of individual genes (expression quantitative trait loci, eQTL), reinforcing a functional link between gene expression regulation and disease risk for a large number of human disorders. We recently adapted the Hybridization Capture of Chromatin-Associated Proteins for Proteomics (HyCCAPP) technology by incorporating luciferase reporter plasmid constructs commonly used to assess the impact of sequence variants on promoter activity, to better study selected eQTL variants reported in the literature. We analyzed a promoter variant, -105G/A, upstream of the selenoprotein S (*SELS*) gene. *SELS* is a gene involved in stress response and inflammation control. We tested the two alleles of the -105 variant in luciferase reporter assays introducing promoter-containing plasmids in HepG2 cells. The two variants resulted in differential activity in the reporter assay, suggesting that the two alleles likely result in altered gene expression. ChIP-seq data from ENCODE did not reveal any suggested DNA-binding proteins mediating the observed effect. Through mass spectrometry we identified CALR, PCK2 and HSPA9, among others, to be significantly enriched in the promoters containing the minor allele. The allele-specific binding of the proteins was confirmed by ChIP-qPCR. None of these proteins have been previously reported to bind DNA before in a sequence-specific manner. Even though further analyses are needed to confirm the chromatin enrichment of these proteins and to understand their mechanisms in this context, HyCCAPP is showing its potential to uncover novel DNA-protein interactions.

1670T

Furthering the GTEx project legacy through the GTEx biospecimen resource. *E. Gelfand, W. Brodeur, K. Larkin, J. Nedzel, K. Ardlie.* Broad Institute, Cambridge, MA.

The Genotype-Tissue Expression (GTEx) project is an NIH Common Fund initiative aimed at providing a resource with which to study human gene expression and regulation and its relationship to genetic variation. This project collected and analyzed multiple tissue types (up to 53) from over 950 healthy donors, to assess the role of genetic variation in regulating gene expression. By analyzing global RNA expression within individual tissues and treating the expression levels of genes as quantitative traits, variations in gene expression that are highly correlated with genetic variation can be identified as expression quantitative trait loci (eQTLs). Correlations between genotype and tissue-specific gene expression levels help identify regions of the genome that influence whether, and how much a gene is expressed. GTEx serves as a resource to study inherited susceptibility to disease, providing a database and tissue bank for many studies in the future. In addition to primary sequence data and analyses available through dbGaP, the residual banked GTEx biospecimens, including DNA, RNA, and tissue, are now housed at the Broad Institute and are available to scientists for additional research projects. The GTEx portal (www.gtexportal.org) has a sample request feature to enable searching and requesting access to samples. It is being enhanced to enable linking of samples to the derived and analyzed datasets and the many data visualization tools and views available on the portal. We will describe this resource and demonstrate ways to search it and integrate it with data features on the GTEx portal.

1671F

The components of the human epigenetic machinery are highly co-expressed and very intolerant to variation. *L. Boukas[1], H.T. Bjornsson[1,2], K.D. Hansen[1,3].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205; 2) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205; 3) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD 21205.

Recently, mutations in genes encoding for components of the epigenetic machinery have been strongly implicated in multiple types of solid and liquid tumors, complex neuropsychiatric disorders (e.g. Autism, Schizophrenia), and Mendelian disorders (e.g. Kabuki and CHARGE syndromes). Currently however, components with ascribed roles in disease constitute a small fraction of the human epigenetic machinery, which remains largely uncharacterized with respect to disease-relevant properties. Here, using known protein domain annotations, we systematically compiled a list of 305 genes containing at least one domain that classifies them as writers, erasers, or readers of DNA methylation, histone methylation or acetylation, or as chromatin remodelers. Using the pLI score, introduced by the ExAC as a metric of haploinsufficiency, we find that the components of the epigenetic machinery are highly intolerant to heterozygous loss of function variants ($p < 2.2e-16$, odds ratio = 7.2 for enrichment in the pLI > 0.9 category), even though many of these genes encode enzymes, a group usually tolerant of heterozygous variation. In fact, the components of the epigenetic machinery demonstrate higher mutational constraint than transcription factors ($p < 2.2e-16$, odds ratio = 4 for enrichment in the pLI > 0.9 category), a classical dosage sensitive group of central transcriptional regulators. We then used data from the GTEx project to systematically assess the expression patterns of these genes across different tissues, and find that in general, they are present at high levels, with half of them exhibiting very low tissue specificity. This is in contrast to the known tissue specificity of transcription factors. Finally, we use established methods (WGCNA) to construct a co-expression network of the epigenetic machinery across tissues and individuals. We show that there exist 95 genes that are each coexpressed with more than 10 other epigenetic machinery genes in more than half of the tissues, and this high degree of coexpression is strongly associated with constraint ($p = 2.3e-13$, odds ratio = 10.5 for enrichment in the pLI > 0.9 category). The existence of a co-expression network suggests either shared upstream regulation or cross-regulation between these factors, whereas the relationship with constraint implies that this network may play a role in disease pathogenesis. Our study uses a genetic approach to establish the central role of the epigenetic machinery in human homeostasis.

1672W

An integrated -omics analysis of the epigenetic landscape of gene expression in human blood cells. *E.M. Kennedy[1,2], G.N. Goehring[3], D. Mehta[4], T. Klengel[5], A.K. Smith[1,6,7], K.N. Conneely[1,2].* 1) Genetics and Molecular Biology Program, Emory University, Atlanta, GA; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 3) AT&T Big Data, Alpharetta, GA; 4) School of Psychology and Counseling, Queensland University of Technology, Brisbane, QLD, Australia; 5) Department of Psychiatry, McLean Hospital, Harvard Medical School, Belmont, MA; 6) Department of Gynecology and Obstetrics, Emory University School of Medicine, Atlanta, GA; 7) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA.

Gene expression can be influenced by DNA methylation 1) distally, at regulatory elements such as enhancers, insulators and non-coding RNAs, as well as 2) proximally, at promoters. Our current understanding of the influence of distal DNA methylation changes on gene expression patterns is incomplete. We assessed the correlation of DNA methylation at ~400K CpG sites with gene expression changes at ~13K expression transcripts, using a linear mixed model framework, in two independent datasets. Our data were derived from 333 human whole blood samples collected for the Grady Trauma Project (GTP) and 1,202 purified human monocyte samples collected for the Multi-ethnic Study of Atherosclerosis (MESA) Epigenomics and Transcriptomics Study. Our results indicate that among CpGs at which methylation significantly associates with gene expression ($p<10^{-11}$ denoted eCpGs), >50% are distal (>50kb) or trans (different chromosome) to the associated gene. Of the eCpG-transcript pairs found in GTP, 45% of cis (<50kb), 30% of distal and 27% of trans pairs were consistent in MESA results. Likewise, neighboring eCpGs within both datasets tend to associate with the same expression probe (MESA: 89% of eCpGs within 500bp, GTP: 82%), even among trans eCpGs (MESA: 88% of eCpGs within 500bp, GTP: 85%). Like other studies, we find that enhancers are substantially overrepresented in eCpGs (MESA-OR: 2.85, p<1e-183; GTP-OR: 2.54, p<1.9e-46). Contrary to many published findings, our results indicate that the majority (MESA: 68%; GTP: 73%) of eCpG-transcript associations within gene bodies are negative. Gene Ontology analysis suggests that for each gene, we have likely identified a regulatory profile that encompasses the indirect, trans effects (which could include regulatory networks) as well as direct cis effects (including cis and distal DNA methylation). Because we find many eCpGs, genome-wide, that associate with promoter genes and chromatin modifiers, our findings may include scenarios in which gene expression influences DNA methylation patterns, as well as vice-versa. This work sheds new light on the interplay between epigenetic changes and gene expression, and provides a useful tool for mining biologically-relevant results from epigenome-wide association studies.

1673T

DNA methylation changes as an exposure signature of cigarette smoking. *E. Kim[1], J. Kim[1], A. Ghantous[2], Z. Herceg[2], J. Sung[1,3].* 1) Complex Disease and Genome Epidemiology Branch, Department of Public Health Science, School of Public Health, Seoul National University, Seoul, Republic of Korea; 2) Epigenetics Group, International Agency for Research on Cancer (IARC), Lyon, France; 3) Institute of Health and Environment, Seoul National University, Seoul, Republic of Korea.

Introduction: Environment alters DNA methylation (DNAm), but little is known about whether exposure-specific changes exist. In identifying exposure-specific DNAm changes, given that DNAm itself is under genetic control, possible genetic confounding needs to be controlled for. In this study, we aimed to identify DNA methylation markers that are specifically modified by smoking, using twin-family data. **Materials and methods:** We analyzed peripheral blood DNAm data of 385 individuals from the Korean Healthy Twin Study (95 MZ twin pairs and their first-degree relatives from 97 families), using Infinium 450K chips. 147 (134 male and 13 female) individuals were ever-smokers and 26 MZ twin pairs were discordant for smoking status. We investigated >18,000 CpGs that were previously reported to be associated with smoking in the large-scale meta-analysis by Joehanes et al. (2016). We used a linear mixed model to account for familial relatedness. To detect mQTLs, we fitted a linear mixed polygenic model for each of all SNP-CpG pairs. **Results:** 49 CpGs at 28 distinct loci were replicated ($q < 0.05$), including cg05575921 and cg235768551 (*AHRR*) and cg03636183 (*F2RL3*). *AHRR* harbored the greatest number of smoking-associated CpGs (13 CpGs), all of which except for one were hypomethylated in the ever-smokers group. In the mQTL analysis, 415 (2%) of the >18,000 CpG sites tested were significantly associated with at least one SNP at Bonferroni-corrected $p < 0.05$. DNAm at two of the smoking-associated CpGs showed significant association with distal SNPs. 511 CpG-SNP pairs showed significant association (438 in *cis* and 73 in *trans*). **Conclusions:** We found that 49 CpGs were differentially methylated in relation to smoking. The mQTL analysis revealed that approximately 2% of the previously reported smoking-associated CpGs were shown to be under genetic control, primarily by *cis*-acting SNPs. This study may provide evidence of increased strength for epigenetic signatures that are specifically sensitive to exposure to cigarette smoking.

1674F

Targeted DNA methylation *in vivo* using an engineered dCas9-MQ1 fusion protein. *Y. Lei[1,2], XT. Zhang[1,2], JZ. Su[3], M. Jeong[1,2], M.C. Gundry[1,2], YH. Huang[1,2], YB. Zhou[4], W. Li[3], M.A. Goodell[1,2,3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 2) Stem Cells and Regenerative Medicine Center, Baylor College of Medicine, Houston, Texas; 3) Dan L. Duncan Cancer Center and Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas; 4) Center for Translational Cancer Research, Institute of Biosciences and Technology, Texas A&M University Health Science Center, Houston, Texas.

Comprehensive studies have shown that DNA methylation plays vital roles in both loss of pluripotency and governance of the transcriptome during embryogenesis and subsequent developmental processes. Aberrant DNA methylation patterns have been widely observed in tumorigenesis, aging, and neurodegenerative diseases, highlighting the importance of a systematic understanding of DNA methylation and the dynamic changes of methylomes during disease onset and progression. Here we describe a facile and convenient approach for efficient targeted DNA methylation by fusing inactive Cas9 (dCas9) with an engineered prokaryotic DNA methyltransferase MQ1. Our study presents a rapid and efficient strategy to achieve locus-specific cytosine modifications in the genome without obvious impact on global methylation in 24 hours. Finally, we demonstrate our tool can induce targeted CpG methylation in mice by zygote microinjection, thereby demonstrating its potential utility in early development. .

1675W

Looking for an epigenetic footprint of music: Behavioral effects of auditory stimulation and its relation to the methylation level of BDNF exon IV within the hippocampus of Wistar rats. *M.M. Velásquez Toledo, M.C. Lattig M., F. Cárdenas.* Universidad de los Andes, Bogotá, Colombia.

Musical stimulation has been reported to undermine behavioral responses in animal models and humans, including anxiety reduction. Additionally, musical stimulation induces changes in the expression of the brain-derived neurotrophic factor (BDNF, Gene ID: 24225). This gene, whose expression could be modulated by epigenetic factors, is involved in several processes such as learning, neuroplasticity, memory consolidation and emotional regulation. The aim of this work was to evaluate whether the stimulation with music at high frequencies could induce an anxiolytic effect in rats and if such effect, if any, would be related to epigenetic changes in the promoter region of BDNF. Male Wistar rats were assigned to one of four conditions: control (no auditory stimulation), music (Mozart K.448) transposed to a higher pitch, normal pitch music (Mozart K.448) or white noise. Animals were exposed to the respective stimuli for 21 consecutive days and the anxiety levels were tested in the elevated plus-maze. We measured the global methylation level and the methylation levels of eleven CpG positions within BDNF exon IV through direct sequencing of bisulfite-treated DNA (BSP) from rat hippocampus. No differences were found either in the anxiety levels of the animals or in the methylation levels. Given that this study considered the pitch transposition of the frequencies, our results suggest that the putative anxiolytic effect, reported in other studies, could reflect effects other than the musical stimulation by itself. It also reveals that the behavioral response to the auditory stimulation could be related to the methylation level within the BDNF exon IV, since the homogeneity of the behavior was reflected in similar methylation status. However, further research is needed to assess other epigenetic changes, including modifications to the chromatin structure.

1676T

Novel deep learning approaches reveal sophisticated epigenetic regulation in eukaryotes. *Z. Wei[1], F. Tan[1], T. Tian[1], H. Hakonarson[2,3].* 1) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 2) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Recently, the prevalence and significance of DNA N6-methyladenine (6mA) in eukaryotes was revealed. Tens of thousands of 6mA sites have been cataloged across several eukaryote genomes, thanks to new technologies. DNA 6mA is a dynamic process, which can be time and/or tissue specific. Therefore experimental approaches, although precise, may not provide a complete catalog of all 6mA sites. In addition, despite some sequence motifs around 6mA sites have been identified, they represent only a fraction (~10%) of the total methylated adenines. We propose advanced machine learning approaches to address the aforementioned two issues. First, our predictive models provide a complete catalog of 6mA sites for six species, namely, C. Elegans, Arabidopsis, Drosophila, E. Coli, Yeast, and Tolypocladium. Specifically, we collect 209354, 271929, 163493, 41333, 4094, 3342 6mA sites for the six species, respectively. We divide these samples into 90:10:10 for training, validation, and testing, respectively. We develop a novel fully sequence-based algorithmic framework based on deep residual learning for 6mA site prediction. We obtain AUC (Area under the Receiving Operating Curves) of 0.967, 0.930, 0.980, 0.999, 0.909, and 0.829, for the six species in testing, respectively, all outperforming the classical k-mer based learning approaches. Using the trained models, we identify a substantial amount of 6mA candidate sites that are not yet reported by existing studies. Second, we hypothesize that a more sophisticated mechanism exists for 6mA regulation, and propose for the first time a contextual sequence analysis protocol for motif discovery. It has four appealing features: (1) integrating sequence information from a wide sequence context; (2) learning sequence motifs at multiple spatial scales with a hierarchical architecture; (3) visualizing sequence motifs for easy interpretation, and (4) enabling sophisticated pattern representations but with low complexity. As control, we successfully detect the 5'-GAGG-3' motif conserved in the eukaryotic species, as well as the canonical GATC motif in E. Coli. This approach also allows us to identify new nonlinear patterns that are missed by conventional motif discovery methods. They account for a much higher fraction (>50%) of the identified 6mA sites. These motifs suggest that sophisticated mechanisms exist for 6mA epigenetics regulation in eukaryotes that may have implications for both health and disease.

1677F

The impact of arsenic exposure on whole blood DNA methylation: An epigenome-wide study of Bangladeshi adults. *B. Pierce[1], K. Demanelis[1], M. Argos[2], L. Tong[1], J. Shinkle[1], M. Sabarinathan[1], F. Jasmine[1], M. Kibriya[1], H. Ahsan[1].* 1) Department of Public Health Sciences, University of Chicago, Chicago, IL; 2) University of Illinois at Chicago, Chicago, IL.

Arsenic exposure affects >100 million people worldwide, including ~56 million in Bangladesh. Arsenic is a carcinogen and possible endocrine disruptor. Epigenetic alterations, including DNA methylation modification, are potential mechanisms of toxicity. We assessed associations between arsenic exposure and genome-wide DNA methylation among Bangladeshi adults exposed to naturally-contaminated drinking water obtained from local tube wells. We measured methylation in whole blood DNA at ~850,000 CpG sites using the Illumina EPIC array for 396 individuals participating in the Health Effects of Arsenic Longitudinal Study (HEALS). Association analyses were conducted using the reference-free algorithm proposed by Houseman et al. to adjust for cell-type composition. We attempted replication in an independent set of 400 arsenical skin lesion cases with Illumina 450K array data. HEALS participants had a wide range of arsenic exposure measured in urine (IQR: 114-350 ug/g), and 76% consumed water with concentrations above the WHO guideline of 10 ug/L. Urine arsenic concentration was associated with methylation at 182 CpGs (FDR <0.05). Of these CpGs 80% were also associated with arsenic measured in individuals' primary drinking well (FDR <0.05). Replication was tested for 83 arsenic-associated CpGs present on the 450K array, and 42 (51%) were associated with urine arsenic ($P < 10^{-4}$) with a consistent direction of association. After meta-analyses, the CpGs most strongly associated with urine arsenic were cg15108641 (*UBTD1*; $p=2 \times 10^{-20}$), cg13480898 (*C19orf66*; $p=9 \times 10^{-19}$), and cg26390598 (*B3GALT5*; $p=9 \times 10^{-17}$). Genes with multiple arsenic-associated CpG sites assigned to them included *BMF*, *USHBP1* and *PRDM16*. Methylation at 74% of these 182 CpGs decreased with increasing arsenic. Compared to all CpGs, negatively associated CpGs were enriched in non-CpG islands (p=0.007) and DNaseI hypersensitive sites (p=0.004) but depleted in promoters (among CpGs assigned to genes). We did not detect enrichment in these categories for positively associated CpGs. Among CpGs with $P < 10^{-4}$, 7 KEGG pathways were enriched, including *thyroid hormone synthesis*, *endocrine and other factor-regulated calcium reabsorption, and estrogen signalling* (P <0.05), suggesting arsenic may dysregulate endocrine pathways. The robust associations between arsenic exposure and DNA methylation observed in this work suggest that epigenetic alterations may be important mediators in arsenic toxicity.

1678W

Characterizing causal *cis*-regulatory variants using computational approaches and CRISPR/Cas9 genome editing. *M. Brandt[1,2], A. Vasileva[1,2], T. Lappalainen[1,2].* 1) New York Genome Center, New York, NY; 2) Department of Systems Biology, Columbia University, New York, NY.

Expression quantitative trait loci (eQTL) and splicing quantitative trail loci (sQTL) studies have identified thousands of common human genetic variants associated with gene expression levels and splicing patterns, respectively. Because of linkage disequilibrium, the true causal variants and therefore the mechanisms underlying the effect on gene expression or splicing have yet to be determined at many loci. We aim to both identify characteristics of causal variants genome-wide and discriminate causal variants from those in close linkage on a local scale. We analyzed eQTLs and sQTLs from the GTEx project v6 data, focusing on fibroblast cell line data from 272 individuals. We use a Bayesian hierarchical model fgwas for fine-mapping of causal variants accounting for functional genomic annotation. We find that eQTL variants are significantly enriched for annotations affecting both transcriptional regulation, such as promoter-associated variants, and post-transcriptional regulation, such as RNA-binding protein binding sites and miRNA target sites. Comparison of the same data analyzed with CAVIAR, another statistical method for identifying causal variants, reveals high concordance between the two methods. We also find that sQTLs are enriched in the ends of exons, canonical splice sites, and synonymous variants. Using these enriched annotations genome-wide, we are able to identify likely causal variants in linkage disequilibrium to further validate experimentally. So far, we have utilized CRISPR/Cas9 editing to edit several common and rare regulatory and splicing variants in HEK293 cells. We are developing a novel approach to experimentally test hundreds of common and rare putative causal regulatory variants in coding regions by pooled editing with CRISPR/Cas9 followed by targeted DNA and RNA sequencing.

1679T

Regulatory role of conserved non-coding elements. *B. Ambroise, J. Fischer, G. Elgar.* The Francis Crick Institute, London, United Kingdom.

Comparative genomics has revealed the existence of Conserved Non-coding Elements (CNEs) in vertebrates. It was found that these non-coding units were clustered around key developmental genes. In addition to this reporter assays in zebrafish embryos have showed that CNEs have an enhancer function making them good candidates for possible interaction partners with the promoter regions of genes. In light of recent developments in the study of chromatin interactions within the nucleus and more specifically with the identification of Topologically Associated Domains (TADs), we investigated the spatial organization of CNE cluster across the human genome in different cell types. We found a strong correlation between the boundaries of CNE clusters and TADs, reinforcing the role of CNEs as potential customers-regulator modules of genes. To further investigate potential direct interactions between CNEs and promoter regions, we have performed chromosome confirmation capture at selected viewpoints. Our preliminary results in tissue culture suggest that at least some CNEs directly interact with promoter regions supporting our initial hypothesis.

1680F

What we talk about when we talk about enhancers. *M.L. Benton[1], D. Kostka[2], J.A. Capra[3].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Departments of Developmental Biology and Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) Departments of Biological Sciences and Computer Science, Vanderbilt Genetics Institute, Center for Structural Biology, Vanderbilt University, Nashville, TN.

Gene regulatory enhancers are essential to transcription in all mammalian cells. Because they provide a window into the function of the non-protein-coding genome, enhancer annotations have increasingly become critical tools for interpreting candidate disease variants from GWAS and sequencing studies. However, accurate and comprehensive enhancer identification in a biological context of interest is challenging, and as result, many experimental and computational methods have been developed to identify enhancers. No gold standard has yet emerged in the field; nevertheless, it is common for studies to consider enhancers identified by only a single method. To assess the robustness of this approach, we analyzed the consistency of enhancer sets identified using eleven representative identification strategies across four diverse cellular contexts. We found significant dissimilarity between enhancer sets in genomic characteristics, evolutionary conservation, and association with functional loci. The substantial disagreement between enhancer sets within the same biological context is often sufficient to influence downstream biological interpretations and lead to disparate scientific conclusions about enhancer biology. Importantly, these differences are likely to affect conclusions about the function of candidate disease loci, because different methods disagree about enhancer function of the majority of GWAS SNPs and eQTL. Surprisingly, we find that enhancers identified by multiple methods are not more enriched for regulatory function than those identified by a single method. Consistent with the differences in genomic location we observe, functional annotations enriched among genes near enhancer sets identified by different methods in the same context also show low similarity and high uniqueness. These trends hold across all methods and cellular contexts considered. Our work identifies a fundamental challenge in the definition and use of genome-wide enhancer maps, which complicates the use of available approaches to elucidate the molecular underpinnings of disease in non-coding regions. To ensure reproducible results and robust conclusions, we argue that the diversity of current approaches should [m1] be considered until we have more completely characterized the complexity of enhancer biology. Therefore we have developed a database of grouped enhancer annotations from common contexts to facilitate the evaluation the effect of different annotation approaches.

1681W

GGmend: A Mendelian randomization method for finding gene-on-gene regulatory effects in the presence of unobserved confounders. *R. Brown[1], J. Joo[2], R. Smith[3], B. Pasaniuc[1,4,5], E. Eskin[1,3].* 1) Bioinformatics, UCLA, Los Angeles, CA; 2) Department of Computer Science Engineering, Dongguk University-Seoul, South Korea; 3) Computer Science Department, UCLA, Los Angeles, CA; 4) Department of Human Genetics, UCLA, Los Angeles, CA; 5) Department of Pathology, UCLA, Los Angeles, CA.

Motivation: Several studies have shown that correlated gene expression levels can be produced by complex biological regulatory networks. Unfortunately, confounding factors, such as shared environmental factors or trans-regulators affecting two genes, may spuriously induce correlations between the gene expressions. These false correlations prevent us from being able to make causal statements about the effect of one gene on another: gene-on-gene effects. By leveraging ideas from Mendelian randomization, it is possible to circumvent the confounding effects and estimate gene-on-gene effects. cis-eQTLs are genetic variants that regulate gene expression in nearby genes. These variants are subject to Mendelian randomization and can be used as instrumental variables to determine the direction and effect size of gene-on-gene regulating effects. We introduce a method, GGmend (Gene-on-Gene effect estimator using Mendelian randomization) that identifies gene-on-gene regulatory effects utilizing a quartet of two genes and their two cis-eQTLs. Results: We use several simulated datasets to show that our method is robust to expression correlations induced by confounding effects and that the family-wise error rate is well calibrated. For small sample sizes, the power of the method is influenced by the correlation in confounders. Further, we applied our method to GTEx data. Although our method has low power for the small sample sizes in the data, we successfully identified a gene pair that has a significant gene-on-gene effect after controlling for the false discovery rate at 0.05.

1682T

A novel computational and experimental approach for allele-specific expression analysis in high-throughput reporter assays. *C. Kalita[1], C. Brown[3], X. Wen[4], R. Pique-Regi[1,2], F. Luca[1,2].* 1) Center for Molecular Medicine and Genetics, Wayne State University; 2) Department of Obstetrics and Gynecology, Wayne State University; 3) Department of Genetics, University of Pennsylvania; 4) Department of Biostatistics, University of Michigan.

The majority of the human genome is composed of non-coding regions containing regulatory elements such as enhancers, which are crucial for controlling gene expression. Many variants associated with complex traits are in these regions, and may disrupt gene regulatory sequences. Consequently, it is important to not only identify true enhancers but also to test if a variant within an enhancer affects gene regulation. Recently, allele-specific analysis in high-throughput reporter assays, such as massively parallel reporter assays (MPRA) and STARR-seq (self-transcribing active regulatory region sequencing), have been used to functionally validate non-coding variants. However, we are still missing high-quality and robust experimental protocols and data analysis tools for these datasets. Computationally, current data-analysis methods to measure allele specific expression (ASE) are limited in their ability to account for the uncertainty on the original plasmid proportions, over-dispersion, and sequencing errors. Experimentally, methods derived from the STARR-seq protocol do not have a barcoding strategy to remove PCR duplicates. We have further developed our method QuASAR (quantitative allele-specific analysis of reads), to analyze ASE in barcoded read count data.Using a beta-binomial distribution, QuASAR-MPRA better models the variability present in the allelic imbalance of these synthetic reporters and results in a test that is statistically well calibrated under the null. Additionally, the provided allelic skew estimate and its standard error simplifies meta-analysis of replicate experiments. To develop QuASAR-MPRA we re-analyzed a recent MPRA study and identified 602 SNPs with allele-specific signal (FDR 10%). Additionally we validated computational annotations of regulatory variants, for subsequent experimental validation. To this end we modified the standard STARR-seq protocol to include the use of a UMI (unique molecular identifier) and used it to test 50,609 computationally and experimentally predicted regulatory variants. We confirmed the regulatory function of 2234 SNPs (FDR 10%). Our study shows that by having the appropriate data analysis tools and experimental approach, we can greatly improve the power to detect allelic effects in high throughput reporter assays. Furthermore, we can accelerate the discovery of causal variants by iteratively improving our computational models that prioritize genetic variants for reporter assay validation.

1683F

Identifying imprinted genes using parent of origin effects on gene expression in the Hutterites. *S. Mozaffari[1,2], D. Nicolae[1,2,3,4], C. Ober[1,2].* 1) Committee on Genetics, Genomics & Systems Biology, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL; 4) Department of Medicine, University of Chicago, Chicago, IL.

Gene expression studies typically treat alleles inherited from the mother and the father as equivalent, although variants can affect gene expression differently depending on whether they are maternal or paternal in origin. Previous studies to identify parent of origin effects (POEs) on gene expression to uncover new imprinted loci have used the gene expression from both inherited chromosomes. However, none have mapped human gene expression transcripts to specific parental haplotypes to either measure POEs on gene expression or identify imprinted genes. In this study, we used RNAseq from lymphoblastoid cell lines (LCLs) from 430 S Dakota Hutterites, who are related to each other in a single pedigree and descendants of 64 founders. Genotypes were imputed and phased, and alleles assigned as maternally- or paternally-inherited, using PRIMAL. We mapped reads to parent of origin depending on the origin of the allele for SNPs within each read for each individual. To identify genes that are expressed more on the maternally- or paternally-inherited chromosomes, we performed two tests of asymmetry. First, we randomly permuted (1000x) the maternal and paternal gene expression values for each gene among all individuals, and then compared the average maternal or paternal expression in the observed data to the randomly permuted average expression patterns to assess significance. Using this test, there were 1,530 significant genes ($P_{Bonferroni} < 2.9 \times 10^{-6}$), of which 41 were known imprinted genes: 14 maternally expressed, including *ABCC9*, *H19*, *ZNF597*, *KCNQ1*; and 27 paternally expressed, including *FAM50B*, *GNAS-AS1*, *L3MBTL1*, *LIN28B*, *MEST*, *NHP2L1 (DLK1)*, *ZDBF2*, and *PEG10*. In the second test, we calculated a binomial Z-score for each individual for each gene using non-normalized maternal and paternal gene expression. We then tested for normality of Z-scores using a Shapiro-Wilk test. 2,221 genes had asymmetric distributions significantly different from normal, including 29 known imprinted genes, such as *AIM1*, *DNMT1*, *FAM50B*, *H19*, *KCNQ1*, *NAP1L5*, *PEG10*, and *ZDBF2*. Some genes not previously characterized as imprinted were significant in both tests, e.g. *MAP2K3*, *PAPOLB*, *PCGF5*, *PSME2*, and *SSBP4*. These genes with POEs we detected are putative novel imprinted genes. This simple and unique approach to map reads to parental chromosomes provides an accurate estimate of maternal or paternal imbalances to identify POEs and novel imprinted loci in humans.

1684W

Efficient detection of trans-gene regulation via association with predicted gene expression. *J.A. Mefford[1], X.Y. Liu[2], A. Dahl[1], J. Witte[1], A.L. Price[2], N. Zaitlen[1].* 1) UCSF, San Francisco, CA; 2) Harvard T.H. Chan School of Public Health, Boston, MA.

Most genetic associations with human disease reside in non-coding regions of the genome, implicating perturbation of gene regulation. Although thousands of cis-eQTLs have been identified, the heavy multiple-testing burden of trans-eQTL scans remains a challenge. Here we present an efficient method that detects trans-genetic regulation of gene expression via gene-based tests of association. The method first predicts each gene's expression from cis-genetic variation using linear mixed models and their sparse extensions which improve accuracy by better modeling genetic architecture. Second, we use the cis-genetic predictions in place of individual SNPs in tests of association with expression of each distal gene. To match the relevant population, tissue-type, and experimental conditions, the cis-genetic predictive models are built using the same dataset in which the trans-analyses are conducted. To prevent overfitting, we developed a computationally efficient cross-validation procedure to generate out-of-sample predictions for each individual's expression level. In addition to reducing the multiple testing burden, our approach enables an efficient adjustment for conditional surrogate variables. We recently showed that naïve adjustment via PCA, PEER, and SVA could both reduce power and induce false positives in eQTL studies(Dahl et al 2017 bioRxiv). While conditional adjustment via SVA addresses these issues, it is computationally prohibitive in traditional trans-QTL studies because factors must be recomputed for each SNP. By using cross-validated gene expression predictions we efficiently adjust for conditional surrogate variables, substantially improving power while minimizing spurious associations. We applied our approach to detect trans-eQTLs in the GEUVADIS RNAseq dataset consisting of 375 samples of European ancestry (Lappalainen et al. 2013 Nature). To prevent false positive trans-associations driven by homologous genes, we removed reads that were mapped to multiple genomic locations (Battle et al 2014 Gen Res). We identified 139 trans-regulation gene pairs at a multiple testing threshold of 1×10^{-7}. We also identified master regulators, including *POLR1B* and *TNFRSF13C*, regulating more than 10 genes. In summary, our gene-based method reduces the burden of multiple testing, and efficient cross-validated in-sample prediction and association tests allow powerful detection of trans-regulation in specific cell-types and experimental conditions.

1685T

Determining blood cell-type composition using DNA methylation sequencing. *W.A. Cheung[1,2], X. Shao[1,2], T. Martin[3], D. Aïssi[4], E. Bouzigon[5], S. Bernantsy[6], M. Lathrop[1,2], S.G. Wilson[3,7], J. Bell[8], M. Moffatt[9], W.O.C.M. Cookson[10], F. Demenais[5], V. Siroux[4], I. Pin[4,11], T. Pastinen[1,2], E. Grundberg[1,2].* 1) Human Genetics, McGill University, Montreal, Quebec, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 3) Twin Research & Genetic Epidemiology, King's College London, London, UK; 4) Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Inserm U1209, CNRS, University Grenoble Alpes, Institute for Advanced Biosciences, Grenoble, France; 5) Genetic Variation and Human Diseases Unit, UMR-946, INSERM, Université Paris Diderot, Université Sorbonne Paris Cité, Paris, France; 6) Division of Clinical Epidemiolgy, McGill University, Montreal, Canada; 7) University of Western Australia, Perth, Australia; 8) Epigenomics Research Group, King's College London, London, UK; 9) Respiratory Genetics, National Heart and Lung Institute, Imperial College London, London, UK; 10) Genomic Medicine, Respiratory Sciences, Asmarley Centre for Genomic Medicine at the National Heart and Lung Institute, Imperial College London, London, UK; 11) Pédiatrie, Centre Hospitalier Universitaire (CHU) Grenoble Alpes, Grenoble, France.

Intra-sample cell-type heterogeneity is an important confounding factor when disentangling disease associations in molecular assays derived from complex biological samples. Previous work has successfully used HumanMethylation450 BeadArray data to estimate cell type composition in peripheral blood samples to identify DNA methylation changes associated with cell type rather than disease. However, predictions of cell proportions have only had limited direct biological validation, and were untested on next-generation sequencing (NGS) data. Here, we present the first large-scale analysis of high throughput bisulfite sequencing DNA methylation data, examining blood cell type specific methylation and validating against 897 whole blood samples in five NGS datasets with associated white cell count (WBC) data. Combining publicly available NGS methylation data as well as internal datasets from 19 purified blood cell types, we test up to 2.8M CpGs, identifying sites that are specifically hypomethylated (< 20% methylation in all samples for a cell type) or hypermethylated (> 80% methylation in all samples for a cell type). Hypo- and hyper- methylation remains consistent between cell types of the same lineage –myelocytes such as macrophages, monocytes, neutrophils and eosinophils all cluster by methylation similarity on one branch, whereas activated lymphocytes such as memory T-cells and B-cells formed a distinct methylation group. Hypomethylated CpGs directly correlate with measured blood cell proportions of neutrophils (R=-0.92, n=11,481), lymphocytes (R=-0.94, n=486), monocytes (R=-0.82, n=9661). We also observe strong correlations of hypermethylated CpGs for lymphocytes (R=0.97, n=12,342). Putatively eosinophil-specific hypomethylated CpGs (n=20,788) were instead highly correlated to the neutrophil (R=-0.90) and lymphocyte (R=0.95) cell proportions. However, a subset of these CpGs correlated to the eosinophil cell proportions (R=-0.88, n=5615). Finally, observed variation in correlation of cell type specific DNA methylation with WBC across the five cohorts may be linked to the DNA source. Specifically, we noted consistently lower correlation in buffy coat samples vs. whole blood (mean absolute correlation |R|=0.67 vs |R|=0.81, paired t-test *p*=0.02). Next generation sequencing provides a unique tool for discovery of cell-type specific methylation, as well as an effective method for assessing the cell-type composition of heterogeneous tissues such as whole blood.

1686F

Inter-individual variation in microbiome composition controls human gene expression. *A.L. Richards[1], A.L. Muehlbauer[3,4], A. Alazizi[1], M.B. Burns[3,4], C. Cascardo[1], R. Pique-Regi[1,2], R. Blekhman[3,4,5], F. Luca[1,2,5].* 1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan, USA; 2) Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan, USA; 3) Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota, USA; 4) Department of Ecology, Evolution and Behavior, University of Minnesota, Minneapolis, Minnesota, USA; 5) Corresponding Author.

Although there is large variation in the microbiome across individuals, it is not clear how this variation affects the host, and specifically which microbes and human genes are involved. Here, we used an experimental system to explore the effect of the gut microbiota on gene regulation in human colonic epithelial cells. We treated the colonic epithelial cells with live microbiota samples derived from five healthy individuals. We identified over 5,000 genes that change expression in the host cells following co-culturing (FDR = 10%, |logFC| > 0.25). Many genes show consistent changes in expression across microbiomes despite differences in microbial composition of the five samples, suggesting that these genes contribute to the host response to general microbiota exposure. Interestingly, we identified 409 host genes that showed varying changes in expression across the five microbiomes, suggesting that these genes are likely responsive to particular microbes that differ in abundance across individuals. To assess this directly, we examined whether the abundance of particular microbes influences expression of specific genes and found 121 genes whose expression was associated with at least one of 46 microbial taxa (FDR = 10%). Furthermore, the taxa with the strongest influence on gene expression (greater than 30 genes at FDR = 20%), alter the response of genes enriched for GWAS traits, including traits previously linked to the microbiome, such as colorectal cancer and obesity. These data demonstrate that we are able to identify specific microbes that are likely to influence host complex traits. We next validated the effect of the genus *Collinsella*, whose abundance has previously been associated with colorectal cancer, on gene expression by manipulating its abundance in the co-culturing system. We identified over 1,000 genes responsive to *Collinsella* that were also 1.5-fold enriched for GWAS traits (p<10[-10]). Our results demonstrate that inter-individual variation in microbiome composition leads to differential gene expression in host colonic epithelial cells and suggest that this is a potential mechanism by which the microbiota affects human complex trait variation.

1687W

Characterizing tissue-specific lincRNA transcription and regulatory roles. *A.D.H. Gewirtz[1], I.C. McDowell[2], A.A. Pai[3], C. Guo[4,5], C.M. Vockley[5,10], C.D. Brown[6], T.E. Reddy[2,5,9], B.E. Engelhardt[7,8].* 1) Quantitative and Computational Biology, Princeton University, Princeton, NJ; 2) Computational Biology and Bioinformatics, Duke University, Durham, NC; 3) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA; 4) University Program in Genetics and Genomics, Duke University, Durham, NC; 5) Center for Genomic and Computational Biology, Duke University, Durham, NC; 6) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 7) Department of Computer Science, Princeton University, Princeton, NJ; 8) Center for Statistics and Machine Learning, Princeton University, Princeton, NJ; 9) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 10) Department of Cell Biology, Duke University Medical Center Durham, NC.

Understanding genetic regulation is one of the fundamental challenges in biology, yet most studies only examine protein-coding genes. Long intergenic non-coding RNAs (lincRNAs) are enriched for genome-wide association study (GWAS) variants, implicating these non-coding RNAs in complex trait phenotypes. LincRNA expression patterns are more tissue-specific than mRNAs and their roles in mammalian cellular differentiation have been established through knockout studies, suggesting highly tissue-specific functions. Indeed, lincRNAs are often well connected in cell type-specific gene regulatory networks. In this study, we characterize tissue-specific lincRNA transcription and regulatory mechanisms using the NIH Common Fund's Genotype-Tissue Expression (GTEx) V6 project data, consisting of RNA sequencing data for 449 human individuals across 44 tissues. By using the largest cross-tissue human transcriptome study to date, we detect tissue-specific and tissue-shared relationships among lincRNA, mRNA transcription levels, and organismal traits. To examine differential transcription mechanisms between lincRNA and mRNA, we compare cis-regulatory element enrichment among cis linc- and cis protein coding-expression quantitative trait loci (eQTLs). We identify likely trait-associated lincRNAs with eQTLs using GWAS variants and investigate biological properties unique to these lincRNAs. Preliminary results on four GTEx tissues show that obesity-related traits show substantial enrichment of GWAS variants among linc-eQTLs versus protein coding-eQTLs. We use matched-tissue ChiP-Seq and Hi-C data to investigate modes of action for the trait-associated lincRNAs in the context of topologically associating domains and chromosomal architecture. Furthermore, we use a Mendelian randomization framework to examine the mechanistic effects of linc-eQTLs on distal mRNA expression levels in a genome-wide manner. This study provides novel insights into tissue-specific lincRNA transcription and subsequent regulatory roles of lincRNA across 44 human tissues.

1688T

The landscape of short RNAs in human cell types and tissues. *Y.Y. Leung[1,2], P.P. Kuksa[1,2], A. Amlie-Wolf[1,3], L.-S. Wang[1,2,3,4,5].* 1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) Penn Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Genomics and Computational Biology Graduate Group, University of Pennsylvania, Philadelphia, PA; 4) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 5) Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA.

Most of mRNAs are regulated by various non-coding RNAs. One of the classes is small RNAs (<200 nts, sncRNA), that can regulate gene expression through association with transcription factor complexes and binding to target mRNA, e.g., UTRs, promoters, and intronic regions. To identify all human specifically processed sncRNAs, we developed a novel ab initio approach, SPAR, and analyzed all ENCODE small RNA datasets across 62 human cell types and tissues. Unlike the standard approach of mapping sequencing reads to known sncRNA transcripts, our method characterizes sncRNA loci by separating the identification of sncRNA peaks (unsupervised segmentation) from the annotation process. To more accurately delineate sncRNA regions, we developed the following features to capture the specific processing patterns of sncRNAs: 1) read accumulation relative to the genomic background; 2) consistency of read sizes within locus; and 3) non-random specificity of 5' and 3' read ends. This ab initio model ensures that called sncRNAs are specifically processed from precursors rather than from random degradation. We identified 76,189 distinct specifically processed small RNA loci across 12 RNA classes, of which 76% were previously unannotated. The detected loci were highly reproducible (Spearman ρ=.82 across ENCODE replicates). 30% of loci were highly conserved (100-way phastCons >.8) suggesting important functional roles. Expression of the detected loci recapitulated known tissues and cell type lineage (Rand Index = .89). On average, in each tissue or cell type, 40% of the sncRNAs display tissue/cell type-specific expression patterns. We discovered that open chromatin regions are enriched for sncRNA loci (2.4 fold), with 34% of sncRNAs residing in open regions. Interestingly, UTR exons were enriched (4-5x) for sncRNA loci, followed by exonic (2.9) and promoter regions (2.2). However, intronic regions were depleted. 60% of the loci have highly specific start and end points across tissues, while the remaining display tissue-dependent alternative processing patterns with slight variability at the start and end sites. We found multiple distinct sncRNAs derived from the same precursors, with 4 distinct loci per snoRNAs, 5 per tRNAs, and 2 per 5' and 3' 50nt flanking regions of tRNAs. The results have been integrated into DASHR 2.0 (http://lisanwanglab.org/DASHR/v2). The compendium of sncRNA loci characterized here is valuable for biomarker development and studying sncRNA biology.

1689F

Silencing of transposable elements may not be a major driver of regulatory evolution in primate induced pluripotent stem cells. *M.C. Ward[1,2], S. Zhao[1], K. Luo[1], B.J. Pavlovic[1], M.M. Karimi[3,4], M. Stephens[1,5], Y. Gilad[1,2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 2) Department of Medicine, University of Chicago, Chicago, IL, USA; 3) The Biomedical Research Centre, University of British Colombia, Vancouver, Canada; 4) Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 5) Department of Statistics, University of Chicago, Chicago, IL, USA.

Transposable elements (TEs) comprise a substantial proportion of primate genomes. The regulatory potential of TEs can result in deleterious effects, especially during development. It has been suggested that, in pluripotent stem cells, TEs are targeted for silencing by KRAB-ZNF proteins, which recruit the TRIM28-SETDB1 complex, to deposit the repressive histone modification H3K9me3. TEs, in turn, can acquire mutations that allow them to evade detection by the host, and hence KRAB-ZNF proteins need to rapidly evolve to counteract them. To investigate the short-term evolution of TE silencing, we profiled the genome-wide distribution of H3K9me3 in induced pluripotent stem cells from ten human and seven chimpanzee individuals. We performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) for H3K9me3, as well as total RNA sequencing. We focused specifically on cross-species H3K9me3 ChIP-seq data that mapped to four million orthologous TEs. We found that, depending on the TE class, 10-60% of elements are marked by H3K9me3, with SVA, LTR and LINE elements marked most frequently. We found little evidence of inter-species differences in TE silencing, with as many as 80% of orthologous, putatively silenced, TEs marked at similar levels in humans and chimpanzees. Our data suggest limited species-specificity of TE silencing across six million years of primate evolution. Interestingly, the minority of TEs enriched for H3K9me3 in one species are not more likely to be associated with gene expression divergence of nearby orthologous genes. We conclude that orthologous TEs may not play a major role in driving gene regulatory divergence between humans and chimpanzees. .

1690W

Map of the functional domains of the long non-coding RNA XIST generated using CRISPR mediated deletions. *T. Dixon-McDougall, C. Brown.* Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

XIST is a nearly 20kb long non-coding RNA (lncRNA) that coats the X chromosome and triggers the stable inactivation of the nearly 1000 genes along its length, making it one of the most powerful regulatory factors in the epigenome. Like other regulatory lncRNAs, XIST is thought to act as a modular scaffold, binding factors that reorganize chromatin marks, silence gene expression and restructure its chromosome of origin. Many XIST interacting proteins have been identified, yet how these factors coordinate with XIST to inactivate a chromosome is still a mystery. To develop a map of the functional domains of XIST and their contributions to chromosome inactivation, the functionality of inducible full-length and partial XIST constructs were compared to determine the effect of removing various domains of XIST. As the X-chromosome's unique architecture contributes to X-chromosome inactivation, confounding variables were reduced by integrating and expressing the XIST cDNA constructs from an FRT site located on chromosome 8p in a male cell line. Most partial XIST constructs were generated from the full-length construct using CRISPR to remove specific domains, with a focus on the tandem repeats. A minority of the partial constructs were generated by separately integrating custom XIST cDNA sequences into the 8p FRT site. To compare the constructs, we used immunofluorescence and RNA FISH to examine heterochromatin mark recruitment and the chromosome of origin's localization to the perinucleolar compartment; as well as allele-specific pyrosequencing to measure XIST induced gene silencing. The region 3' of XIST's exon 1 and in particular the E repeat is crucial to the recruitment of heterochromatin marks, including H3K27me3, as well as distal gene silencing. Analysis of the XIST D-repeat suggests that even when autosomally expressed it is necessary for maintaining XIST RNA levels. In addition, numerous other internal regions of XIST were found to be necessary for establishing silencing and recruitment of specific heterochromatin marks while also demonstrating which regions are dispensable for these functions, creating the first completed map of the functional domains of human XIST. This research opens the door to identifying novel long non-coding RNA functional domains, the factors that bind these domains, and ultimately the epigenetic pathways employed by regulatory lncRNAs to control the epigenome.

1691T

Recognition of human elements regulating escape from X-chromosome inactivation in mouse. *S.B. Peeters[1], A.J. Korecki[1,2], E.M. Simpson[1,2], C.J. Brown[1].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada.

X-chromosome inactivation (XCI) epigenetically silences one X chromosome in every cell in female mammals; however, approximately 12–20% of genes escape from XCI in humans and continue to be expressed from both the active and inactive X chromosomes, while only 3–7% of X-linked genes escape in mouse. In light of known species differences between escape gene number and distribution on the X chromosome, we tested whether mouse is a suitable model for studying human escape from XCI through integration of a human BAC at the X-linked *Hprt* gene. Subsequent expression and DNA methylation analysis of transgene activity on an inactive X chromosome in adult female mice demonstrated expression and correspondingly low promoter DNA methylation of human genes *RPS4X* and *CITED1*. Therefore, either the mechanisms of escape are conserved between mouse and human, or at the very least the mouse system is capable of recognizing human elements necessary for escape from XCI. Silencing of a normally subject gene (*ERCC6L*) on the BAC, as well ongoing silencing of genes near the integration site, also suggests retention and recognition of boundary elements between subject and escape regions on the BAC. Preliminary data from developmental time points suggests that transgenic *RPS4X* is expressed from the inactive X in females as early as embryonic day 9.5, arguing for continuous expression from the transgene rather than initial silencing and reactivation. As human embryonic stem cells remain an epigenetically unstable model for studying XCI, we have generated a female mouse embryonic stem cell line with an *Hprt* "docking site" similar to the transgenic mouse system, allowing more rapid assessment of additional human genes for intrinsic escape ability; as well as facilitating further investigation and characterization of the critical regions responsible for continued expression from an inactive X chromosome. Expression of genes escaping from XCI may contribute to the phenotypes of X aneuploidies as well as influence male and female susceptibility to disease, therefore studying their mechanisms may offer insight into sex differences in addition to exploring the gene regulatory mechanisms involved in XCI.

1692F

Interaction of exocrine and endocrine pancreas in obesity ciliopathies. *T. Hostelley, J. Nesmith, S. Lodh, N. Zaghloul.* Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA.

Ciliopathies are rare genetic disorders characterized by dysfunction in primary cilia. Two ciliopathies, Alström Syndrome and Bardet-Biedl Syndrome (BBS), are characterized by high rates of obesity. However, they have strikingly different rates of diabetes. The rate is much higher in Alström patients relative to BBS. To gain insight into genetic mechanisms for this discrepancy we performed whole transcriptome sequencing on zebrafish depleted of the Alström gene, *alms1,* or the BBS gene, *bbs1*. We identified gene expression changes and found 8 genes that were differentially expressed in opposite directions between the two models. Among these, several exocrine pancreas enzymes were downregulated in Alström and/or upregulated in BBS. Based on this, and our previous demonstration of β-cell mass changes in models of both disorders, we hypothesized that the expression of exocrine pancreas enzymes affects endocrine pancreas function. This might suggest that impaired β-cell function in Alström results from reduced expression of exocrine pancreas enzymes, while the opposite may be true in BBS. These findings also support a non-ciliary role for BBS and Alström genes as we found them to be expressed in the non-ciliated acinar cells. To test our hypothesis, we overexpressed each enzyme in transgenic zebrafish embryos, in which β-cells can be visualized. In control animals, overexpression of the exocrine pancreas enzymes significantly increased the number of β-cells, suggesting that these enzymes play a role in regulating β-cell mass and function. Overexpression could also rescue the loss of β-cells observed in animals depleted of *alms1*. We have also found that the inactive, zymogen, form of these enzymes are selectively taken up by β-cells *in vitro* and induce proliferation in cultured *MIN6* β-cells. This suggests a direct interaction between the exocrine and endocrine pancreas. Taken together, these data support the possibility that exocrine pancreatic enzymes may play an important role in the modulation of β-cells in diabetes.

1693W

Down-regulation of *SRSF3* mRNA expression in mice with *KIF23* c.2747C>G mutation known as a cause of congenital dyserythropoietic anemia type III (CDA III): Expression study. *A.-L. Vikberg, S. Malla, I. Golovleva.* Medical and Clinical Genetics, Medical Biosciences, Umeå University, Sweden.

Congenital dyserythropoietic anemia type III (CDA III) is a rare hereditary disease initially described in a large Swedish family. Affected individuals suffer from mild to moderate hemolytic anemia as a result of cell division failure in erythroblasts. Retinal angioid streaks and monoclonal gammopathy of unknown significance or myeloma are additional features present in some of the CDA III individuals. Recently, we showed that CDA III is caused by a mutation c.2747C>G, p.P916R in kinesin family member 23, *KIF23*, affecting the final step of cell cycle, cytokinesis, and resulting in bi-nuclear cells. Previously, two *KIF23* transcripts, one full-length and another one lacking exon 18 have been described while we reported the third transcript with absence of exon 17 and 18. The aim of this study was to understand the tissue-specific effect of *KIF23* c.2747C>G by investigating an mRNA expression of *KIF23* in different tissues. *KIF23* expression of all three transcripts was studied by reverse transcription of RNA extracted from blood, bone marrow and CFU-E of a CDA III patient and commercially available RNA from 17 normal human tissues and 11 normal mouse tissues followed by digital droplet PCR. We also used a knock-in mouse model with *KIF23* c.2747C>G mutation and quantified expression of *KIF23* and its splicing factor *SRSF3* in blood, bone marrow, CFU-E, brain, heart, kidney, liver, lung, spleen and testis. We showed large variations in *KIF23* expression between different human tissues with the highest level in bone marrow and testis. Full-length transcript was most prevalent in testis, while a transcript lacking exon 18 was dominant in bone marrow. Notably, that a transcript without exon 18 was most prevalent in CFU-E of the CDA III patient, compare to CFU-E of healthy individual where full-length transcript was the most dominant. In mice, the highest *KIF23* expression was in testis, and unlike humans, the shortest transcript lacking both exon 17 and 18 was most prevalent in most tissues. We also examined how knock-in of *KIF23* c.2747C>G mutation in mice affects expression of a splicing factor SRSF3 known to regulate expression and alternative splicing of *KIF23*. Preliminary results show that homozygous wild type mice exhibit highest levels of *SRSF3* compared to homozygous carriers of c.2747C>G mutation. We conclude that not only SRSF3 regulates *KIF23* expression but also *KIF23* c.2747C>G down-regulates mRNA expression of splicing factor *SRSF3*.

1694T

Haploinsufficiency of a histone modifier, Kmt2d, in a mouse model of Kabuki syndrome leads to widespread defects in the B cell lineage. *G. Pilarowski[1,2], T. Cazares[3], L. Zhang[1], A. Lindsley[3], H.T. Bjornsson[1,4].* 1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Predoctoral Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Division of Allergy & Immunology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Department of Pediatrics, Johns Hopkins University, Baltimore, MD.

Kabuki syndrome (KS) Type 1 is an autosomal dominant Mendelian disorder of the epigenetic machinery caused by loss of function mutations in the histone modifying protein KMT2D. KMT2D adds histone 3, lysine 4 (H3K4) mono- and trimethylation, which are thought to play a major role in the activation of enhancers and promotors, respectively. Individuals with Kabuki syndrome demonstrate intellectual disability, postnatal growth retardation and craniofacial abnormalities, all of which are recapitulated in a mouse model of Kabuki syndrome (*Kmt2d*[+/βgeo] mice). Additionally, KS patients have been described to have immune dysfunction with decreased serum IgA although the mechanistic basis is unknown. Our study of *Kmt2d*[+/βgeo] mice suggests that defects in B cell mucosal immunity may underlie the phenotypes seen in KS patients. *Kmt2d*[+/βgeo] mice have splenomegaly (p<0.03) and slightly elevated serum IgM (p<0.02) levels and a 4 fold decreased serum IgA levels (p<0.0002) compared to WT littermates. IgA is a major component of mucosal immunity that is primarily produced in the gut. Therefore, to investigate the mechanism of observed IgA defects, we examined the secondary lymphoid tissue of the gut, the Peyer's patches (PP). Though the PP compartmental organization was grossly normal, PP's in Kmt2d[+/βgeo] mice are strikingly smaller and there are significantly fewer PP's compared to WT littermates (Kmt2d[+/βgeo] mean=2.4; WT mean=7.2; p<0.0001). We detected significantly decreased levels (50%; p<0.02) of a post-IgA-class-switch-recombination transcript in the mesenteric lymph nodes in *Kmt2d*[+/βgeo] mice compared to WT littermates. We also observe defects in antigen-specific humoral responses to mucosal vaccination. For further examination, we targeted the B1 cell population of the peritoneal cavity, which has been hypothesized to contribute to IgA-producing plasma cells in the gut. We discovered a decreased B1a cell population in *Kmt2d*[+/βgeo] mice compared to WT littermates (24% vs 42% of B1 parent population, p<0.0005). Thus, our data suggest defects in mucosal immunity and widespread defects of the B cell lineage in *Kmt2d*[+/βgeo] mice. Epigenomic study of B1a cells may potentiate mechanistic understanding regarding how a primary mutation in an epigenetic modifier can lead to immunological dysfunction through epigenetic dysregulation. Furthermore, these findings will stimulate further studies of the B cell lineage in individuals with Kabuki syndrome.

1695F

Models of human disease available from The Jackson Laboratory. *S. Rockwood, C. Lutz, The Genetic Resource Science Team.* The Jackson Laboratory, Bar Harbor, ME.

The mouse continues to provide researchers a means by which to address questions related to basic disease mechanisms. The emergence of disruptive genetic engineering technologies have proven to be very applicable to the mouse, spawning a new generation of promising preclinical models. To facilitate access to this flourishing resource, The Jackson Laboratory (JAX) Mouse Repository offers one of the largest sets of well-characterized genetically engineered mutant mouse strains. Hundreds of new mouse lines generated by both the international scientific community and JAX researchers are added to the Repository each year. A growing selection of models with applications as hosts for cancer xenograft modeling, primarily on the NSG platform, is readily available. Strains have been optimized for specific purposes: support of human and murine hematopoietic cell engraftment, reduced xenogeneic graft-versus-host disease response, engraftment of human hematopoietic stem cells without irradiation and stable engraftment of primary human hepatocytes. Mouse strains that recapitulate aspects of specific diseases, such as Alzheimer's disease, Amyotrophic Lateral Sclerosis (ALS), Friedreich's Ataxia and Huntington's disease are available. Exploiting the potential of CRISPR/Cas9 technology to generate precise models of disease is made easier by a large variety of Cas9-expressing lines, including both constitutive expressing and *cre*-inducible lines. In response to the threat posed by the spread of Zika virus, colonies of *Ifnar1* deficient mice have been expanded. Several recent studies demonstrate that *Ifnar1* knock-out mice represent a model that will facilitate the study of Zika virus pathogenesis and should prove useful in the development related therapies and vaccines. In order to promote reproducibility in research, the Repository maintains a robust quality control and assurance program. Newly arriving strains are assayed to confirm the presence of expected mutations. Equally important, strains are also screened for undesired contaminating alleles and genetic backgrounds. Donating a strain to the Repository is an easy way to fulfill the NIH's requirements for sharing mice. Researchers wishing to have strains considered for inclusion in the Repository are encouraged to submit their strains: www.jax.org/donate-a-mouse. This work is supported by NIH, The Howard Hughes Medical Institute, and private charitable foundations.

1696W

Circulating cells protect against radiation-induced intestinal injury in a murine parabiosis system. *J. Sung[1], C. Sodhi[2], H. Jia[2], Y. Yamaguchi[2], P. Lu[2], L. Voltaggio[3], D. Hackam[2].* 1) Institute of Genetic Medicine, Johns Hopkins University, School of Medicine, Baltimore, MD; 2) Division of Pediatric Surgery, Johns Hopkins Children's Center and Department of Surgery, Baltimore, MD; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD.

   Background The small intestine is one of the most radiosensitive organs due to the rapid rate of intestinal stem cells (ISC) division, although the mechanisms providing protection against radiation-induced ISC death remain unknown. We hypothesize that circulating cells participate in healing after radiation-induced intestinal injury in mice, utilizing a model of parabiosis to test this. Specifically, we hypothesize that a parabiotic partner will be protected against radiation-induced enteritis by translocation of healthy bone marrow-derived cells (BMDCs) through the shared vasculature to the site of injury. We propose that the repair mechanisms of recruited BMDCs are due to (1) direct replacement of damaged intestinal epithelial cells by the BMDCs, (2) trans-differentiation of BMDCs into intestinal stem cells, and/or (3) release of biomolecules from BMDCs via the paracrine system.   **Methods** Parabiont mice were constructed by suturing the skin and muscle wall on the flank. Whole-body gamma irradiation was applied to the one partner. Injury and recovery were examined via assessment of body weight, mucosal injury, enterocyte apoptosis (TUNEL), and qRT-PCR determination of cytokines. To tract migration of BMDCs across the parabiosis pair, pairing of fluorescent and non-fluorescent mice were used. To ascertain recruited circulating stem cells are bone marrow origin, non-parabiosis mice were irradiated prior to peripheral injection of tdTomato expressing BMDCs.   **Results** After 1-month of pairing, a shared circulation was confirmed by tdTomato positive cells in the blood of the wild-type partner. Significantly decreased body weight loss and injury markers (*TNF-a, IL-1b),* and increased proliferation (*Lgr5, Ki67*) and epithelial differentiation (*Sucrase isomaltase, MucII, Chromagranin A*), and significant histological improvement in the injured partner confirm protection by the parabiotic pairing. Detection of tdTomato positive cells in injured intestine suggests that circulating BMDCs in peripheral blood could home to the injury site. Strikingly, the recruited BMDCs were observed in lamina propria, but not in the intestinal epithelial tissue, suggesting that the protective effects occurred via paracrine effects of the recruited BMDCs, as opposed to trans-differentiation.   **Conclusion** We reveal that circulating stem cells can facilitate intestinal healing using a parabiotic system, suggesting novel cellular approaches to the treatment of radiation induced intestinal injury.

1697T

***Xenopus* as a model of precision medicine: Application of CRISPR to mimic the mutations of human CSBS syndrome patients.** *S. Cha.* Cincinnati Children Hospital Medical Center, Cincinnati, OH.

   Xenopus research has provided significant clarification of both early developmental events and later processes of organogenesis. I have developed a new methodology for homology-driven repair (HDR), going beyond simple gene editing, which typically causes deletions that inactivate or impair gene function. Because HDR is performed by injecting oocytes, which are uniquely accessible in *Xenopus* system and where recombination enzymes are highly active, very high HDR rates (typically 20%) are seen in developing embryos. This allows efficient insertion of specific sequences at target sites, e.g. to insert tags into protein-coding regions in vivo, to follow protein localization during development, or to insert specific mutations into the genome (e.g. single base mutations emulating a human congenital disease). In this study, we made several *X. tropicalis* lines to model human CSBS patients carrying clinically verified mutations in CLMP and FLNA. With this model animals, we could follow morphogenetic events of the developing intestinal epithelium, aiming to understand the mechanisms of gut tube elongation and differentiation in the context of various major signaling pathways. These studies are highly relevant to understanding human diseases such as congenital short bowel syndrome (CSBS) and intestinal malrotation, which have severely impaired gut tube elongation/rotation. Overall, this study project will propel the *Xenopus* system to a new level, setting in place the potential for significant insights for basic research and greatly improved utility for modeling human disease.

1698F

Gene expression in the developing mouse pituitary gland. *A.S. George, M.L. Brinkmeier, S.A. Camper, R.E. Mills.* University of Michigan, Ann Arbor, MI.

Combined Pituitary Hormone Deficiency (CPHD) is a common cause of endocrine referral in children, yet most familial and sporadic CPHD cases have no known genetic cause. In 2009, full-length cDNA libraries were made from mouse pituitaries collected at embryonic time points e12.5 and e14.5 to create a comprehensive developmental pituitary mouse transcriptome. Single pass end sequencing was carried out and the analysis revealed a high degree of novelty among the clones. We sought to identify transcripts that mapped to unannotated genomic regions by analyzing the cDNAs using updated current reference databases. In this analysis of the clone sequences, we assigned ten transcription factors and 351 long non-coding RNA transcripts to cDNAs that were previously marked as unidentified during the initial analysis. We created a publically available online database to house the results of the curated cDNAs and their assigned genes. End sequences that were not assigned to known genes underwent further analysis. Overlapping cDNA end sequences were combined into clusters and groups that did not have any known transcribed regions were compared to RNA-seq of six wild-type postnatal pituitaries. Five clusters were revealed as potential novel transcripts and molecular analysis of the clusters is ongoing. The results of this analysis serve as a resource of embryonic mouse pituitary gene expression that can be used to provide candidate genes for CPHD of unknown etiology and for studying basic pituitary gland development. It also reveals the potential for lncRNA involvement in the developing pituitary.

1699W

The role of FREM2 and FRAS1 in the development of congenital diaphragmatic hernia. *V. Jordan[1], T. Beck[2,8], A. Hernandez-Garcia[2], P. Kundert[2], B.J. Kim[2], S. Jhangiani[2,3], T. Gambin[2,9], M. Starkovich[3,1], J. Punetha[2], I. Paine[2], J. Posey[2], A. Li[4], D. Muzny[2,3], A. Lashua[5], X. Sun[5,10], C. Fernandes[6], M. Dickinson[1], K. Lally[7], R. Gibbs[2,3], E. Boerwinkle[4], J. Lupski[2,3,6], D. Scott[1,2].* 1) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 4) Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA; 5) Department of Medical Genetics, University of Wisconsin-Madison, Madison, WI, USA; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 7) Department of Pediatric Surgery, McGovern Medical School at UT Health, Houston, TX, USA; 8) American Society for Radiation Oncology, Arlington, VA, USA; 9) Warsaw University of Technology, Institute of Computer Science, Warsaw, Poland; 10) Department of Pediatrics, University of California-San Diego, San Diego, CA, USA.

Congenital diaphragmatic hernia (CDH) can occur in isolation, with other birth defects, or as part of a genetic syndrome. In ~20% of human CDH cases, the herniated viscera are encapsulated by a membranous sac. We have shown that isolated sac CDH in humans and mice can be caused by recessive pathogenic variants in FREM1. In the extracellular matrix, FREM1 forms a self-stabilizing ternary complex with FREM2 and FRAS1. CDH has been described in two individuals with a clinical diagnosis of Fraser syndrome which can be caused by mutations in FREM2 and FRAS1. To determine if FREM2 and FRAS1 are bona fide CDH genes, we examined FREM2- and FRAS1-deficient mice and embryos for evidence of CDH. We found sac hernias in 8.2% (5/61) of FREM2-deficient mice and 1% (1/102) of FRAS1-deficient mice. These hernias were indistinguishable from those seen in FREM1-deficient mice. We then screened a cohort of 69 individuals with CDH for changes in FREM2 and FRAS1. Four individuals in this cohort (6%) carried rare, putatively deleterious compound and/or double heterozygous changes in these genes. Three individuals carried a FREM2 c.4031G>A, p.Arg1344His change whose frequency among Caucasians in our CDH cohort was statistically greater than that seen in ethnically matched control populations (P < 0.0001). Next, we sought to determine the morphogenetic and molecular mechanisms underlying this susceptibility to CDH. We found that sac hernias in FREM1-deficient mice form when the migrating muscular components of the diaphragm fail to meet at the anterior midline. We also found an absence of mature vasculature and a dramatically increased level of apoptosis in the anterior diaphragm that may contribute to an increased susceptibility to herniation. We went on to demonstrate that GATA4, a retinoic acid responsive transcription factor implicated in the development of sac CDH, modulates the transcriptional expression of Fras1 in the developing diaphragm. This is consistent with previously published data showing that a block in retinoic acid signaling leads to decreased expression of FREM1, FREM2 and FRAS1 in the developing diaphragm. We conclude that deficiency of FREM1, FREM2, and possibly FRAS1, are associated with an increased risk of developing CDH caused by defects in diaphragmatic muscularization. Our data also suggest that retinoic acid may modulate the expression of these proteins in the developing diaphragm, at least in part, through its effects on GATA4.

1700T

Identification of novel candidate genes for 46,XY disorders of sex development (DSD) using a C57BL/6J-Y[POS] mouse model. *E. Vilain[1,4], A. Symon[3], M. Zadikyan[1], M. Almalvez[1,4], E. Segura[1], A. Eskin[1], M. Bramble[1,4], V. Arboleda[1], R. Baxter[1], S. Nelson[1], E. Delot[1,2], V. Harley[3], H. Barseghyan[1,4].* 1) Departments of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA; 2) Departments of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA; 3) Department of Brain and Gender, Hudson Institute of Medical Research, Clayton, Victoria, 3168, Australia; 4) Center for Genetic Medicine Research, Children's National Health System, Washington D.C.

Disorders of Sex Development (DSD) have an estimated frequency of 0.5-1% of live births encompassing a variety of urogenital abnormalities ranging from mild hypospadias to a discrepancy between sex chromosomes and external genitalia. In order to identify the underlying genetic etiology, we performed exome sequencing in a subset of DSD cases with 46,XY karyotype and were able to identify the causative genetic variant in 35% of cases. While the genetic etiology was not ascertained in more than half of the cases, a large number of variants of unknown clinical significance (VUS) were identified in those exomes. To investigate the relevance of these VUS in regards to the patient's phenotype, we utilized a mouse model in which the presence of a Y chromosome from the *poschiavinus* strain (Y[POS]) on a C57BL/6J (B6) background results in XY undervirilization and sex reversal, a phenotype characteristic to 46,XY DSD cases. We assessed gene expression differences between B6-Y[B6] and undervirilized B6-Y[POS] males at E11.5 and identified 515 genes that were differentially expressed (308 underexpressed and 207 overexpressed in B6-Y[POS] males). We identified 15 novel candidate genes potentially involved in 46,XY DSD pathogenesis by filtering patient exome VUS using the B6-Y[B6] gene list. Using a different animal model with *Sox9* repression in XY gonads we also show that 7 of the candidate genes are under regulation of the a well-known sex determination gene *Sox9*. The use of a DSD specific animal model improves variant interpretation by correlating human sequence variants with transcriptome variation. This method can be universally applied to different diseases for which an animal model is present.

1701F

Kisspeptin and Kisspeptin receptor may be involved in the regulation of adrenocortical development and steroid hormone secretion. *N. Settas, A. Berthon, A. Giannakou, A. Delaney, F.R. Faucz, C.A. Stratakis.* Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, United States.

It has been shown that knockout mice for Kisspeptin receptor, *Kiss1r*, (KRKO) and its ligand Kisspeptin, *Kiss1*, (KissKO) replicate the phenotype of isolated hypogonadotropic hypogonadism (IHH) that was described in association with these genes in humans. Puberty and adrenal function are tightly linked in humans. We studied adrenal function and morphology at 5 months of age in both male and female KRKO and KissKO mice that do not go through normal puberty. Two fetal markers, 20αHSD and Cyp17, were expressed in eosinophilic cells, which are accumulated at the border between the cortex and the medulla. These cells may be derived from the fetal zone that should disappear at puberty in male and during the first pregnancy in female mice. Interestingly, KRKO and KissKO females (but not males) showed corticosterone and aldosterone hypersecretion. Although hypercorticosteronism corrected overtime, hyperaldosteronism persisted at age 14 months and correlated with overexpression of the transcription factor *Star*. To determine if *KISS1* and *KISS1R* genes are involved in the development of primary aldosteronism (PA) and hypercortisolism (Cushing syndrome-CS) in humans, we sequenced them in a cohort of 65 patients. In the *KISS1* gene, we identified two polymorphisms (c.-89G>A and c.417_417delA) with a higher frequency in our cohort compared to the 1000GP (14% vs 8% and 33% vs 22%, respectively). Interestingly, we found a mutation (p.H90D, rs201073751) previously reported for causing central precocious puberty (CPP) in a patient with CS without CPP. The analysis of this patient's adrenal tumor did not reveal any somatic mutations. Moreover, we discovered three rare variants in *KISS1R* gene, two missense and one synonymous, in three different patients with PA: rs141767649 (p.C95W), rs73507527 (p.A189T) and rs115335009 (p.R229R). The two missense variants have been previously associated with IHH and Kallmann syndrome, respectively. Our findings suggest that *KISS1* and *KISS1R* may play a role in adrenal function in mice and possibly adrenocortical steroid hormone secretion abnormalities in humans. This study is ongoing, as more data are needed to elucidate the potential role of *KISS1* and *KISS1R* in adrenocortical function and development.

1702W

Genome-wide association study of infantile hypertrophic pyloric stenosis identifies four new loci and highlights the importance of embryonic *NKX2-5/BARX1* pathways. *L. Skotte[1], J. Fadista[1], F. Geller[1], J. Bybjerg-Grauholm[2], S. Gørtz[1], H. Matsson[3], H.A. Boyd[1], D.M. Hougaard[2], A. Nordenskjöld[3,4], M. Melbye[1,5,6], B. Feenstra[1].* 1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Danish Centre for Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 3) Department of Women's and Children's Health, and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 4) Department of Pediatric Surgery, Astrid Lindgren Children´s Hospital, Karolinska University Hospital, Stockholm, Sweden; 5) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 6) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA.

Introduction: Infantile hypertrophic pyloric stenosis (IHPS) is the most common cause of intestinal obstruction in infancy. IHPS is caused by hypertrophy of the smooth sphincter muscle that surrounds the pylorus, which connects the stomach to the duodenum. Materials and Methods: We performed a genome-wide association study of 1,395 Danish surgery-confirmed cases and 4,438 controls in order to detect novel loci associated with IHPS. The analysis was based on 6.6 million variants imputed from the Haplotype Reference Consortium panel, using logistic regression on imputed SNP dosages for association testing. Results: In addition to variants at the known *MBNL1, ARHGEF26, NKX2-5* and *APOA1* loci, we found four new loci with variants reaching genome-wide significance ($P<5 \times 10^{-8}$). The most strongly associated variant (minor allele frequency; MAF=2%) had an odds ratio (OR) =2.5 ($P=5.0 \times 10^{-14}$) and is intronic to *MTA3*, a gene that when knocked out in mice leads to decreased levels of circulating LDL cholesterol. The second signal came from a common variant (MAF=26%, OR=1.3, $P=2.7 \times 10^{-9}$), which is an eQTL for *SLMAP*, a gene with highest expression in smooth muscle. The third variant (MAF=9%, OR=1.5, $P=4.6 \times 10^{-9}$) is located 1kb downstream of *BARX1*, a gene highly enriched in stomach expression. Like *NKX2-5*, *BARX1* is a homeobox gene and it has been shown that *Barx1* knock-out mice have abnormal stomach morphology and fail to develop the pyloric sphincter. Furthermore, familial syndromes involving duplication of the 9q22.1–q31 region, which harbors *BARX1*, often include IHPS as part of the clinical presentation. Finally, the fourth variant (MAF=12%, OR=1.4, $P=3.2 \times 10^{-9}$) is in a gene desert at 13q21.31. Conclusion: Our findings, undergoing replication genotyping in independent samples at time of submission, identify novel genetic associations with IHPS, and highlight putative functional target genes.   Grant support: Danish Medical Research council (DFF 4004-00512), Novo Nordisk Foundation, Oak Foundation, Carlsberg Foundation.

1703T

Novel pathogenic variant in OFD1 results in male lethal oral facial digital syndrome type 1 with pituitary aplasia. *D. Aljeaid[1], R. Lombardo[1], D. Witte[2], R. Hopkin[1].* 1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Division of Pathology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

We report live birth of a male patient with OFD Type I secondary to a novel mutation in *OFD1*, c.515T>C, p.Leu172Pro, and complete pituitary aplasia with severe hypoplasia of peripheral endocrine glands. The infant was first identified with severe brain malformations and multiple congenital anomalies at 22 weeks gestation on routine ultrasound. Subsequent fetal MRI demonstrated polymicrogyria, gray matter heterotopia, ventriculomegaly, corpus callosum agenesis, and abnormalities of the posterior fossa and brainstem concerning for Dandy-Walker malformation and molar tooth sign. Recurrent episodes of hypotension and hypoglycemia after birth prompted evaluation of the adrenal glands which could not be visualized on ultrasound. Postnatal brain MRI confirmed fetal findings and was unable to identify the pituitary. Extensive endocrine work up was consistent with panhypopituitarism. Autopsy confirmed absence of pituitary tissue. Careful analysis revealed two areas of tissue near the kidneys that was identified to be hypoplastic islands of adrenal cortex under the microscope that was consistent with severe adrenal hypoplasia. The thyroid and testes demonstrated similar degrees of hypoplasia and only one parathyroid gland was identified. The pancreas was normal in size with normal histology. OFD type 1 is an X-linked dominant condition with near universal lethality in early gestation in male infants. Live birth is exceedingly rare and has been reported in only 5 patients to date. Although a spectrum of brain malformations have been associated with mutations in *OFD1*, this is the first report of pituitary aplasia. It is likely that most if not all of the other abnormalities in endocrine organs were secondary to the lack of pituitary stimulation. Maternal testing was performed and confirmed carrier status. Extended family history identified a maternal male sibling that died at approximately 30 hours of life from severe brain malformations. The exact function of *OFD1* gene is not fully understood, but it's been shown that *OFD1* gene encodes a protein that localizes to the centromere and basal body of primary cilia, which plays a role in cilia formation and left-right axis specification. It's also been noted that there's an inverse correlation between *OFD1* mutant protein length and the severity of phenotype. It has been suggested that *OFD1* might be involved in the *SHH* signaling pathway via interaction with other genes such as *HOX*, thus affecting pituitary gland development.

1704F

Human knockout of a cell cycle gene *CDC25B*: a novel disease-causing gene for an Alström-like syndrome with cataract, dilated cardiomyopathy and multiple endocrinopathies. *C. Lam, C. Law.* The University of Hong Kong, Hong Kong, China.

The cell division cycle 25 (CDC25) phosphatases families are a key player regulating the cell cycle phases during cell divisions and maintaining the genetic stability, and an overexpression of the proto-oncogenes, *CDC25A* or *CDC25B* had been shown in various cancers. Here, we report a loss-of-function mutation of *CDC25B* in a girl who presented with a new syndrome, involving multiple systems. The patient was an 11-year-old Chinese girl born to consanguineous asymptomatic parents with history of one fetal and one infancy death with unknown causes. The proband suffered from intrauterine growth retardation, delayed development, bilateral cataract (at 5 yrs), primary hypothyroidism, growth hormone deficiency and heart failure (at 9 yrs). Echocardiogram showed dilated cardiomyopathy involving mainly the right heart and progressively the left heart. The patient had no intellectual disability. Extensive biochemical workups were unremarkable. Karyotyping was normal. Magnetic resonance imaging of the pituitary gland was also unremarkable. Given the clinical presentations, provisional diagnoses of Alström syndrome and *POLG*-related mitochondrial disorders had been made but later being excluded after a negative sequencing result of *ALMS1* and *POLG* genes respectively. This undiagnosed case was subsequently referred to the Clinics for Rare Diseases Referral (*CRareDr*) and the Undiagnosed Diseases Program, to end the diagnostic odyssey. Clinical whole exome sequencing (WES) was performed in the proband and we identified a novel homozygous nonsense variant in the *CDC25B* gene, NM_021873:c.313G>T (p.Glu105*). This novel variant was located in a runs of homozygosity (ROH) of 25 Mb on chromosome 20. Her parents and two asymptomatic sisters were confirmed to be carriers, and the asymptomatic brother was homozygous for the wild-type allele. The presence of c.313G>T in the proband is expected to produce a truncated protein, terminated at codon 105 and a loss of downstream phosphorylation sites. Importantly, the human phenotypes has not been reported in *Cdc25b* knockout mice. Because *CDC25B* and *ALMS1* are both involved in maintaining centrosome integrity, we propose a loss-of-function *CDC25B* in humans will result in a new autosomal recessive syndrome, mimicking Alström syndrome.

1705W

Gene expression profiling of single oocytes reveals pathways and regulators involved in follicle activation. *Y. Lyu[1,2,3], G. Lu[1,3], Z. Xiong[2,3], J. Ma[2,3], W. Chan[1,3].* 1) CUHK-SDU Joint Laboratory on Reproductive Genetics, School of Biomedical Sciences, the Chinese University of Hong Kong, Hong Kong, China; 2) Center for Reproductive Medicine, Shandong University, Jinan, Shandong, China; 3) National Research Center for Assisted Reproductive Technology and Reproductive Genetics, Jinan, Shandong, China.

Introduction The primordial follicle activation is a crucial process for female fertility which means the quiescent primordial follicle activated to the primary follicle. Under normal circumstances, the majority of primordial follicles stay in a quiescent state, and only a few numbers of primordial follicles could be activated gradually. However, when it comes to some pathological conditions, this balance would be broken and depletion of primordial follicles might be accelerated, diseases such as premature ovarian failure (POF) will be an occurrence. However, the mechanism of follicle activation has remained poorly understood. **Purpose** In this research, we try to find the novel pathways and regulators related to follicle activation by performing the gene expression profiling of the single oocytes derived from primordial follicles and primary follicles. **Methods** 91 candidate genes of follicle activation and 5 endogenous genes were selected to customize a TaqMan array card. 40 oocytes of primordial follicles and 40 oocytes from primary follicles were picked up from four neonatal mice based on their morphology. Then the single oocytes were prepared using single cell-to-ct Kit (ThermoFisher) and loaded to TaqMan array card for qPCR. The gene expression data was analyzed using statminer single cell qPCR model and then inputted to the IPA for core analysis. **Results** Among the selected genes, 14 undetermined genes were excluded firstly, and the remaining 82 genes were used for data analysis. As expected, a majority of selected genes expressed actively in primary follicle oocytes compared to the primordial follicle oocytes. The hierarchical clustering and principal component analysis were performed, which showed that the two stages of oocytes were separated obviously. Interestingly, subgroups were observed among the same stages oocytes but not due to the source of different animals. Following that, the core analysis was conducted using IPA software, numbers of related pathways and upstream regulators were predicted which might relate to the follicle activation. **Conclusion** In this research, we first conduct gene expression profiling of two early stages oocytes using single cell qPCR. Subgroups were found might exist among the primordial follicle oocytes. Moreover, numbers of predicted upstream regulators including genes and drugs were founded, and their function needs validation in the further study.

1706T

Investigating phosphatidylserine metabolism in Lenz-Majewski syndrome. *C. Demetriou[1], E. Peskett[1], A. Kapustin[2], D. Jenkins[1], M. Seda[1], F. Vaz[3], K. Takov[4], O. Kuge[5], D. Bryant[1], P. Clayton[1], K. Mills[1], A. Barnicoat[6], M. Bitner-Glindzicz[1,6], D. Wattanasirichaigoon[7], K. Chrzanowska[8], M. Simandlová[9], L. Maldergem[10], S. Davidson[4], C. Shanahan[2], S. Sousa[11], G. Moore[1], P. Stanier[1].* 1) Genetics and Genomic Medicine, Institute of Child Health, UCL, London, United Kingdom; 2) Cardiovascular Division, James Black Centre, King's College London, London, UK; 3) Laboratory of Genetic Metabolic Disorders, Academic Medical Center, Amsterdam, The Netherlands; 4) The Hatter Cardiovascular Institute, University College London, London, UK; 5) Department of Chemistry, Faculty of Sciences, Kyushu University, Fukuoka, Japan; 6) Clinical Genetics Department, Great Ormond Street Hospital, London, UK; 7) Department of Pediatrics, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 8) Department of Medical Genetics, Children's Memorial Health Institute, Warsaw, Poland; 9) Department of Biology and Medical Genetics, University Hospital Motol and Second Faculty of Medicine, Prague, Czech Republic; 10) Centre de Génétique Humaine, Université de Franche-Comté, Besançon, France; 11) Serviço de Genética Médica, Hospital Pediátrico de Coimbra, Coimbra, Portugal.

   Background: Lenz-Majewski syndrome (LMS) is a rare sclerosing bone dysplasia associated with intellectual disability and multiple congenital anomalies. It is characterized by progressive skeletal hyperostosis particularly of the cranium, vertebrae and diaphysis of the tubular bones leading to growth restriction and fusing of fingers and toes, together with tooth, skin and brain defects. The condition is caused by an anomaly in the production of phospholipids, which are an essential component of all human cell membranes. We previously identified the genetic basis as de novo gain-of-function (GOF) mutations in *PTDSS1*, which encodes phosphatidylserine synthase (PSS1). **Objectives:** The GOF mutations activate the enzyme and reduce end product inhibition, however, the exact mechanism leading to hyperostosis is not yet understood. Current studies seek to clarify the key events at a cellular and biochemical level. **Methods:** LMS patients and control cell lines as well as a Chinese hamster ovary (CHO) cell line with or without the LMS R95K mutation have been investigated using electrospray ionization mass spectroscopy techniques to generate their phospholipid profiles. To investigate mineralisation, LMS and control fibroblasts have been used to investigate propensity to calcification using a cresolphthalein assay. Matrix vesicles (MV) were also isolated from cell culture media by ultracentrifugation and analysed for PS content using flow cytometry. **Results:** Lipid studies in CHO cells confirmed a generalised overproduction of PS species. Bone mineralization is initiated by MV containing nucleation sites consisting of PS and Annexin A5. We isolated MV from the cell culture media of human control and LMS fibroblast. Notably, LMS4 fibroblast-derived MVs were enriched with PS in comparison to control fibroblasts-derived MV. Next, we stimulated fibroblasts with elevated calcium (2.7mM) and phosphate (2.0mM) and observed rapid cell calcification *in vitro* as detected by Alizarin Red staining and cresolphthalein assay. **Conclusions:** Our findings show that LMS mutations result in a disturbance to the normally tightly regulated production of PS. Furthermore, we found that fibroblasts secrete MVs which were enriched with annexin/PS complexes that could trigger early mineralization. We intend to go on to use our model systems to investigate potential new therapies for disordered bone metabolism, which may have implications for osteopetrosis to osteoporosis and osteoarthritis.

1707F

Skeletal dysplasias in Saudi Arabian population. *M. Faden[1], A. Sheereen[2], S. Massadeh[2], M. Al khawashki[2], M. Alaamery[2].* 1) King Saud Medical City, Genetics Department/Childrens Hospital; Riyadh,Saudi Arabia; 2) Developmental Medicine Department /King Abdulaziz Medical City(KAMC)/Ministry of National Guard Health Affairs(MNG-HA), King Abdullah International Medical Research Center (KAIMRC)/King Saud bin Abdulaziz University for Health Sciences(K-SAU-HS), Riyadh, Saudi Arabia.

   Skeletal dysplasias, or osteochondrodysplasias, are a heterogeneous group of more than 430 disorders characterised by abnormalities of cartilage and bone growth, resulting in disproportion of the long bones, spine, and head. However, these disorders were reclassified in 2015, into 42 groups and the number of genes increased to 364. The prevalence of these disorders was reported as 1 in 5000 live births worldwide.   In Saudi Arabia, the prevalence of many autosomal recessive disorders is higher than in other known populations because of many factors, including the high rate of consanguineous marriages (56%). On the other hand, an absence of a countrywide registry for skeletal dysplasias is impeding further insight into the matter, as such a registry is rudimentary for determining the incidence and prevalence of these conditions in the Kingdom, which in turn is essential to the success of future prevention campaigns in the nation. Around 93% of all skeletal dysplasias reported in Saudi Arabia are autosomal recessive.   The most skeletal dysplasia disorders which have been identified in Saudi Arabia, including those with a novel mutation that has not been reported before in any other population. It includes Spondyloepimetaphyseal dysplasia (Faden-AlKuraya type), Autosomal recessive split hand and foot malformation, the Recessive form of Marshall Syndrome, Osteogenesis Imperfecta Type XIV, a Novel mutation causing Kniest dysplasia, Wolcott–Rallison syndrome, Desbuquois Dysplasia, and Novel mutations in Joubert syndrome affected individuals whom have features of Jeune asphyxiating thoracic dystrophy. .

1708W

Structural optimization of TransCon CNP: Development of a sustained-release prodrug of CNP for achondroplasia. *K. Sprogøe, C. Rasmussen, A. Bernhard, F. Faltinger, T. Woods, S. Schröder, A. Brennauer, J. Zettler, T. Wegge, U. Hersel, V. Miller Breinholt.* Ascendis Pharma, Hellerup, DK, Denmark.

   Background Achondroplasia is caused by an activating mutation in the fibroblast growth factor receptor 3 (FGFR3) gene, thereby inhibiting chondrocyte proliferation and differentiation. C-type natriuretic peptide (CNP) inhibits the effect of FGFR3 signaling via activation of the natriuretic peptide receptor B (NPR-B) pathway. Developing CNP as a therapeutic is challenging, as CNP is rapidly cleared and degraded by natriuretic peptide receptor C (NPR-C) and neutral endopeptidase (NEP).   **Objective** Development of a safe and efficacious long-acting CNP candidate requires simultaneously minimizing clearance and degradation by NPR-C and NEP, while preventing overstimulation of vascular NPR-B receptors, which may lead to hypotension. Prodrug candidates fulfilling these requirements were synthesized, in which CNP was transiently bound to a TransCon carrier via a proprietary linker. After auto-hydrolysis of the linker from the prodrug, fully active, unmodified CNP is released.   **Methods** In cells expressing either NPR-B or NPR-C, and in a NEP protease incubation assay, the relative receptor binding and stability of TransCon CNP was compared to unmodified CNP. TransCon CNP was then administered as a single subcutaneous (SC) dose of 146 µg/kg to cynomolgus monkeys (n=3) and the pharmacokinetic profile investigated. Subsequently a single SC dose of TransCon CNP 100 µg/kg vs. dose-equivalent daily CNP was administered to telemeterized cynomolgus monkeys (n=4/group) and the systolic arterial blood pressure (SAP) was assessed for up to 48 hours. **Results** The lead TransCon CNP candidate demonstrated an EC_{50} of <1% of parent CNP in the potency assay expressing NPR-B. The NPR-C binding affinity was reduced to approximately 1% of parent CNP. NEP stability was increased, with degradation half-life of 4 days compared to 12 hours for parent CNP. Reduced TransCon CNP clearance was confirmed in cynomolgus monkey by a half-life of 79 hours. No decrease in SAP was detected in TransCon CNP-treated monkeys, whereas monkeys receiving an equivalent dose of daily CNP experienced a decrease in SAP.   **Conclusion** The TransCon CNP prodrug had minimal CNP binding to the NPR-B and NPR-C receptor and improved NEP stability *in vitro*. *In vivo* TransCon CNP demonstrated the desired half-life extension without adverse hemodynamic effects. These data support clinical development for weekly dosing and suggest TransCon CNP may be a safe and efficacious option for children with achondroplasia.

1709T

Missense and splicing mutations in the retinoic acid catabolizing enzyme *CYP26C1* in idiopathic short stature. *G. Rappold[1], L. Juergensen[2], C. Thiel[3], R. Roeth[1], T. Ogata[4], M. Fukami[5], D. Hassel[2], A. Montalbano[1].* 1) Dept. Human Molecular Genetics, Heidelberg Germany, Heidelberg, Baden-Württemberg, Germany; 2) Department of Internal Medicine III – Cardiology, Heidelberg University, 69120 Heidelberg, Germany; 3) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany; 4) Department of Pediatrics, Hamamatsu University School of Medicine, Hamamatsu, 431-3192, Japan; 5) Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, 157-8535, Japan.

   Human height is a complex quantitative trait with a high heritability. Short stature is diagnosed when height is significantly below the average of the general population for that person´s age and sex. We have recently found that the retinoic acid degrading enzyme CYP26C1 modifies SHOX deficiency phenotypes towards more severe clinical manifestations. Here, we asked whether damaging variants in *CYP26C1* alone could lead to short stature. We performed exome and Sanger sequencing to analyze 706 individuals with short stature where SHOX deficiency was previously excluded. Three different damaging missense variants and one splicing variant were identified and their functional significance tested *in vitro* or *in vivo* using Zebrafish as a model. Although incomplete penetrance was observed in one case, the genetic and functional data here reported strongly suggest that *CYP26C1* represents a novel gene underlying growth disorders and that damaging variants lead to short stature.

1710F

The role of p63 isoforms in the epidermal development as replicated in cellular models for normal human skin and genetic skin diseases. *I. Barragán Vázquez[1,2], D. Lima Cunha[1,3,4], S. Hockney[1], K.Y. Tang[1], S. Thomforde Kitson[1], K.M. Eckl[5], H.C. Hennies[1,3,4,6].* 1) Biological Sciences, University of Huddersfield, Huddersfield, United Kingdom; 2) Faculty of Chemical Sciences, Universidad Complutense de Madrid, Spain; 3) Genetics and Genomics, Medical University of Innsbruck, Austria; 4) Cologne Center for Genomics, University of Cologne, Germany; 5) Dept. of Biology, Edge Hill University, Ormskirk, UK; 6) CECAD, University of Cologne, Germany.

Transcription factor p63 is a key regulator of epidermal development. *TP63* codes for two major isoforms, TAp63 and ΔNp63. Mutations in *TP63* are associated with severe diseases such as ectrodactyly/ectodermal dysplasia and cleft lip/palate (EEC) and ankyloblepharon/ectodermal defects and cleft lip/palate (AEC) syndromes. These diseases and knockout and transgenic mice pointed to a major role of p63 in ectodermal development. Details about the importance of the isoforms are not clear so far; TAp63 is present early in ectodermal development, ΔNp63 is involved in epidermal maturation. In order to study the importance of p63 for epidermal differentiation, we have analysed the expression of *TP63* and its major isoforms in keratinocytes *in vitro*. Quantitative real time PCR revealed a minor increase early in Ca^{2+} induced differentiation of primary human normal epidermal keratinocytes and a decrease during terminal differentiation. Isoform-specific analysis showed that ΔNp63, the predominant isoform in epidermis, was present during differentiation with decreasing levels. TAp63, in contrast, showed an increase late in differentiation after 10d in culture. These results are in accordance with an association of ΔNp63 with basal epidermal keratins 5 and 14 and a declining expression in suprabasal epidermal layers as shown previously by immunohistochemistry. *KLF*, a transcription factor gene important for gene expression in terminally differentiating keratinocytes, showed a significant increase in parallel with the increase of differentiation markers such as transglutaminase 1 and filaggrin, which is in line with the supposed induction of *KLF* by p63. Keratinocytes of patients with autosomal recessive congenital ichthyosis (ARCI), a disorder of epidermal keratinisation, showed a similar pattern for the expression of *TP63*, indicating that the disturbed keratinisation seen in ARCI is restricted to proteins of terminally differentiating keratinocytes but does not influence their transcription factors. Our findings underpin the importance of switching p63 isoforms during epidermal development and confirm an additional role for TAp63 in terminal differentiation. Moreover, these results corroborate the importance of cellular models based on primary keratinocytes for the study of epidermal keratinisation. Studies with 3D human skin models will give further insight into the regulation of gene expression profiles during epidermal differentiation.

1711W

***TP63* is expressed in adult epidermal and iPSC-derived melanocytes supporting the role ofΔNp63 in ectodermal gatekeeping and cell migration to the epidermis.** *D. Cunha[1,2,3], H. Tahir[4], K. Golding[4], I. Barragán Vázquez[1,5], T. Saric[6], C. Ploner[7], H.C. Hennies[1,2,3,8], K.M. Eckl[4].* 1) Department of Biological Sciences, University of Huddersfield, Huddersfield, United Kingdom; 2) Div. of Human Genetics, Innsbruck Medical University, Innsbruck, Austria; 3) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 4) Dept. for Biology, Edge Hill University, Ormskirk, Lancashire, UK; 5) Faculty of Chemical Sciences, Universidad Complutense de Madrid, Spain; 6) Inst. for Neurophysiology, Univ. Hospital Cologne, Cologne, Germany; 7) Univ.-Clinics for Plastic, Reconstructive and Aesthetic Surgery Medical University of Innsbruck, Innsbruck, Austria; 8) CECAD, Univ. of Cologne, Cologne, Germany.

TP63 encoding p63, is a member of the transcription factor p53 family, which is expressed differentially with six isoforms. ΔNp63 is today understood to be a master regulator of epidermal morphogenesis and an essential gatekeeper in ectodermal development. p63-null mice lack stratified epidermis and are born with only one layer of non-proliferating keratinocytes, resulting in massive trans-epidermal water-loss and premature death. Tremendous progress has been made in the understanding of p63 signalling and pathways in keratinocytes and epidermal commitment, but only little is known about the role of p63 in melanocytes. Pre-melanocytes (melanoblasts) arise either from trunk neuronal crest cells (NCC) or from peripheral Schwann cells. During Carnegie stage 13 melanoblasts emigrate from the NCC and arrive at the basal layer of the epidermis during the third and fourth month of foetal development, where they mature into melanocytes expressing melanin and interacting with surrounding keratinocytes during epidermal development. In this study we analysed ΔNp63 expression in melanocytes isolated from skin from a variety of healthy control persons (NHEM) , as well as from induced pluripotent stem cell-derived melanocytes (iPSC-derived hM) from patients and controls. NHEM were isolated from surplus skin following plastic surgery and cultured in defined medium, iPSC-derived hM were differentiated from iPSC following our in-house protocol for melanocyte differentiation, all melanocytes were then extensively characterised using immuno-cytochemical methods and expression profiling techniques. ΔNp63-isoform specific expression analysis in NHEM and iPSC-hM showed clear evidence of nuclear expression of transcription factor p63 in both, adult, skin-derived melanocytes and stem-cell derived melanocytes. Gene expression analysis with probe-based assays highly specific for isoform ΔNp63 revealed expression profiles for NEHM and iPSC-hM similar to basal keratinocytes from healthy donors. All qPCR expression data were normalised to housekeeping genes and to relative expression of ΔNp63-isoform found in iPSCs. Preliminary results suggest further that *SATB1*, a p63-regulated genome organiser, essential for remodelling chromatin architecture at the EDC locus, is similarly regulated and expressed in NHEM and iPSC-derived hM as previously reported for keratinocytes, suggesting an even more important role for p63: not only an ectodermal gatekeeper but also an epidermal guidepost.

1712T

Familial dysautonomia: The regulation of *IKBKAP* in the nervous system and therapeutic approaches. *S. Yannai, J. Zonszain, N. Gonceer, S. Naftelberg, M. Donyo, G. Ast.* Human Molecular Genetics & Biochemistry, Tel Aviv University, Tel Aviv, Israel.

Familial Dysautonomia (FD) is an autosomal recessive congenital neuropathy that results from a point mutation at the 5' splice site of intron 20 in the *IKBKAP* gene. Alteration in this gene splicing occurs in a tissue-specific manner, reducing the formation of an intact IKAP protein in the nervous system. Although the exact function of the IKAP protein is not clearly understood, poor development and progressive degeneration of the sensory nervous systems and depletion of dorsal root ganglia (DRG) neurons that characterize FD patients indicate a possible role of IKAP in the development of the peripheral nervous system. Phosphatidylserine (PS), an FDA-approved food supplement, was shown by us to elevate *IKBKAP* levels. However, low permeability across the blood-brain barrier led us to pursuit new approaches to deliver PS to the brain and the spinal cord. The main goal of my research is to better understand the regulation of *IKBKAP* in the nervous system, and to explore therapeutic approaches which will improve the quality of life for FD patients. For this purpose I will use two approaches: (i) increasing permeability of therapeutic agents in a mouse model for FD and (ii) examining synergistic effects of treatments that can increase *IKBKAP* levels. Moreover I intend to examine the splicing mechanism involved in mediating the splicing regulation of exon 20 following the IVS20+6T>C mutation in FD. My results indicate that intra-nasal administration of PS increased IKAP levels in the DRG, and protected against weight dysfunction, which is observed in FD patients. Further attempts to improve nasal delivery showed a delivery device enables PS transfer to the brain and spinal cord. These results suggest that innovative delivery devices may improve the quality of life for FD patients. In addition I evaluated several potential therapeutic agents that could be combined with PS; such compounds either have effect on splicing or gene expression. The HDAC inhibitor TSA and Kinetin, a plant cytokinin that affects splicing, were examined combined with PS. Both kinetin and TSA successfully elevated IKAP protein level when combined with PS.

1713F

iPSC-derived neurons from patients with idiopathic ASD show deficits in neuronal differentiation and synaptic function. *C. Garcia-Serje[1], J. el Hokayem[1], B.A. DeRosa[2], E. Artimovich[3], D. Van Booven[1], J.E. Nestor[3], M.L. Cuccaro[1], J.M. Vance[1], M.A. Pericak-Vance[1], H.N. Cukier[1], M.W. Nestor[3], D.M. Dykxhoorn[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL., Select a Country; 2) Department of Medical & Molecular Genetics, Oregon Health & Science University, Portland, Oregon, 9723; 3) Hussman Institute for Autism, Baltimore, Maryland, 21229.

Numerous candidate genes have been associated with autism spectrum disorder (ASD), with many of these genes known to have important roles in synaptic function and the development of neural circuits. This suggests that certain neurobiological processes could be commonly altered in ASD. Therefore, although there is a great deal of clinical and genetic heterogeneity in ASDs, there may be convergent deficits in key molecular mechanisms that underlie autism. Lack of appropriate human-based models of genetically complex neurodevelopmental disorders has hindered investigations of nonsyndromic ASD. induced pluripotent stem cells (iPSCs) offer the opportunity to further unravel the intricate biology underlying ASD. To that end, patient-specific iPSC from idiopathic ASD individuals and iPSCs from neurologically normal controls were differentiated into cortical neurons. RNA-seq analysis was performed on these neuronal cultures at an early (day 35) and late (day 135) time point. We then sought to identify common cellular and molecular signatures that distinguish iPSC-derived neurons from our cohort of ASD individuals from neurons derived from controls. We approached this by identifying expression differences both in a single genes and also in a higher-order network level. Transcriptional analyses of the ASD and control neurons showed ASD-specific molecular phenotypes affecting networks involved in neuronal differentiation, the cytoskeletal matrix structure formation, patterning, DNA and RNA metabolism, and synaptic function. We then applied functional analyses to identify disease-specific, cellular phenotypes. We found deficits in calcium transients and synaptic function, as well as, impaired migration in the ASD-specific neurons compared to controls. In conclusion, we have shown that patient-specific iPSCs can be used to model brain region-specific neuronal development permitting the identification of common molecular mechanisms disrupted in ASD opening the door for potential therapeutic development. .

1714W

Multi-system contributions to *Gabrb3*-related neurodevelopmental risk in utero. H. Moon[1], P. Carpentier[1], V. Saravanapandian[1], U. Haditsch[1], J. Su[1], M. Chin[1], K. Muench[1], A. Moore[1], A. Bormann[1,2], N. Nima[1,2], G. Subramanyam[1,3], M. Rivera[1], T. Palmer[1]. 1) Neurosurgery Department, Stanford University, Stanford, CA; 2) 5The Bridges to Stem Cell Research program, California State University Fullerton, Fullerton, CA; 3) The Bridges to Stem Cell Research program, San Jose State University, San Jose, CA.

Autism spectrum disorder (ASD) is a neurodevelopmental disability defined by social interaction deficits and stereotyped behaviors but the underlying pathophysiological mechanisms are not known. ASD heterogeneity may stem from dynamic interplay between genetic and environmental risk factors. Maternal immune activation (MIA) induced by prenatal infection is the ASD environmental risk factor that increase placental and fetal vulnerability. Here we test the novel hypothesis that *Gabrb3* deficiency combined with prenatal bacteria-like lipopolysaccharide (LPS) challenge may accentuate placental dysfunction-mediated complications and increase neurological impairments in offspring. The present study highlights gene-environment interactions between ASD-related genetic susceptibility and MIA in the *in vivo* mouse models. The GABA type A receptor (GABA$_A$R) subunit beta 3 (*GABRB3*) located in the human chromosomal loci 15q11-13 in which copy number variation was associated with ASD, Angelman Syndrome, Prader-Willi Syndrome, and epilepsy. Peripheral GABA$_A$Rs are expressed in immune cells and modulates immune signaling. Here we show that *Gabrb3* deficiency combined with LPS-mediated MIA accentuates placental vulnerability and exacerbates neurodevelopmental abnormalities in the developing fetus. *Gabrb3* heterozygous pregnant dams injected with LPS showed elevated maternal cytokine/chemokine relative to wild-type controls. Increased placental necrosis and impaired proliferation of neocortical progenitor cells are found in LPS-treated *Gabrb3* heterozygous placenta and fetal brain, respectively. Alterations in neuronal subtype specification and abundance were also notable in *Gabrb3*-deficient fetal brain. GABA$_A$R-selective agonist and antagonist treatment altered cytokine production in macrophages *in vitro*, suggesting a possible convergence of immuno-modulatory signaling pathways between GABA$_A$R/GABRB3 and LPS-induced MIA response in the placenta and the developing fetus. Together, our data provide important evidence that *Gabrb3* deficiency and MIA may synergistically increase placental and fetal brain vulnerability by altering the immune microenvironment at the maternal-fetal interface. We anticipate that additional ASD genetic risk factors may also elevate placental and/or fetal brain vulnerability to MIA and these gene-immune interactions may significantly contribute to the spectrum of phenotypes noted in ASD, schizophrenia, and other neurodevelopmental disorders.

1715T

A *Drosophila* model of essential tremor. L.N. Clark[1,2], P. Smith[3], I. Santa-Maria[1], S. Sonti[1], B.D. McCabe[4], K. Tsaneva-Atanasova[5,6], J.L. Hodge[3], E.D. Louis[7]. 1) Department of Pathology and Cell Biology, College of Physicians and Surgeons, Columbia University, New York, NY, 10032, USA; 2) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, NY, 10032, USA; 3) School of Physiology, Pharmacology and Neuroscience, University of Bristol, University Walk, Bristol BS8 1TD, UK; 4) Brain Mind Institute, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland; 5) Department of Mathematics and Living Systems Institute, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK; 6) EPSRC Centre for Predictive Modelling in Healthcare, University of Exeter, Exeter, EX4 4QJ, UK; 7) Department of Neurology, Yale School of Medicine, Yale University, New Haven, CT, 06520, USA; Center for Neuroepidemiology and Clinical Neurological Research, Yale School of Medicine, Yale University, New Haven, CT, 06520, USA; Department of Chronic Disease Epidemiology, Yale School of Public Health.

Essential Tremor (ET) is one of the most common neurological diseases, with an estimated 7 million affected individuals in the United States. The characteristic clinical feature of ET is a kinetic tremor in the hands or arms, which can progress to become more severe and anatomically widespread (e.g. head or neck). Recently, we identified a mutation (*Kv9.2*, c.1137 T>A, p.(D379E) in an electrically silent voltage-gated K$^+$ channel a subunit, Kv9.2, in a family with early onset, autosomal dominant ET, that is highly and selectively expressed in brain, and that modulates the activity of the Kv2.1 and Kv2.2 channels by heteromultimerization. Voltage gated potassium (Kv) channels play a major role in regulating membrane excitability and synaptic transmission in many central neurons and at the neuromuscular junction and Kv2 channels encode the majority of delayed rectifier current. To study the disease mechanism we have produced transgenic *Drosophila* lines that express pan-neuronally either the wild type human Kv9.2 (hKv9.2) subunit or the ET causing mutant human Kv9.2 (hKv9.2-D379E) subunit. Here, we show that the hKv9.2 subunit can modulate the channel activity of endogenous Shab (Kv2 subfamily). Behavioral manifestations of nervous system dysfunction consistent with a hyperexcitable phenotype were observed in flies expressing the hKv9.2 or hKv9.2-D379E subunit. The effects of hKv9.2 and hKv9.2-D379E on adult neuron activity, were determined using whole cell voltage-clamp. A Shab specific toxin was used to isolate the voltage-sensitive non-inactivating K$^+$ current. The mutant hKv9.2-D379E subunit showed significantly higher levels of inactivation than the wild-type hKv9.2 subunit (*p*=0.0462) and a significantly higher frequency of spontaneous firing rate of action potentials (*p*=0.0102) consistent with neuronal hyperexcitibility. A biophysical model of the electrical activity of the cell, combining the individual ionic currents in flies expressing hKv9.2 subunits, was in agreement with experimental data. This *Drosophila* transgenic model recapitulates several features of ET, and can now be used to advance our understanding of ET disease pathogenesis.

1716F

A mutation in *MAL* is associated with a neurodevelopmental condition characterized by central hypomyelination, cerebellar atrophy and developmental delay. *M. Elpidorou[1], J.A. Poulter[1], J.H. Livingston[2], E. Sheridan[1], C.A. Johnson[1].* 1) Leeds Institute of Biomedical and Clinical Sciences, University of Leeds, Leeds, United Kingdom; 2) Paediatric Neurology, Leeds General Infirmary, Leeds, United Kingdom.

One in every thousand children has abnormalities of the white matter of the brain. Many of these disorders are genetically determined and progressive, associated with a fatal outcome. Here we report a seven-member multiply consanguineous UK family of Pakistani origin. Two affected individuals presented with significant developmental delay, cerebellar developmental defects and cortical white matter abnormalities. The pedigree structure was compatible with autosomal recessive inheritance, and analysis was done under this model. Whole exome sequencing (WES) was performed on DNA from two affected boys and an unaffected sibling. SureSelect^QXT Target Enrichment for Illumina Multiplexed Sequencing technologies was used for the library preparation and samples were sequenced on the HiSeq3000 platform. Regions of concordant homozygosity, shared by the affected and not by their unaffected sibling, were identified from WES data using the AgileMultiIdeogram analysis program (http://dna.leeds.ac.uk/agile). Non-synonymous biallelic variants were filtered by frequency (MAF<1%) and predicted pathogenicity, prioritizing those within homozygous regions. Filtering revealed variants in four possible candidate genes, but only a single homozygous missense variant c.326C>A (p.Ala109Asp) in the myelin and lymphocyte protein *MAL* (RefSeq NM_002371, GRCh38) segregated within the family. The variant had a CADD score of 28.7, was absent from ExAC and gnomAD, and had deleterious pathogenicity predictions from SIFT (0.02), Polyphen (0.839) and Condel (0.621). *MAL* encodes a tetraspanning transmembrane protein that is implicated in the biogenesis and function of myelin. MAL regulates the trafficking of PLP, the major component of myelin, from the Golgi body to the basolateral myelin membranes by transcytosis. We present functional studies in polarized cell-lines to confirm the pathogenicity of the c.326C>A (p.Ala109Asp) variant by co-localization and biochemical interaction studies. Our data suggests that the MAL p.Ala109Asp variant disrupts the insertion of the third transmembrane helix into membranes, preventing a conformational shift in PLP that is required for correct intracellular trafficking to basolateral membranes.

1717W

Defective myoblast differentiation in human muscle dysferlin-deficient cells. *R. Ishiba[1], A. Bigot[2], A.F. Ribeiro Jr[1], V. Mouly[2], M. Vainzof[1].* 1) Human Genome and Stem-cell Research Center, Biosciences Institute, University of São Paulo, São Paulo, Brazil; 2) Sorbonne Universités, UPMC Univ Paris 06, INSERM UMRS974, Center for Research in Myology, 47 Boulevard de l'hôpital, 75013 Paris, France.

Limb girdle muscular dystrophy type 2B (LGMD2B) is an autosomal recessive disorder caused by mutations in *DYSF* gene. Dysferlin is a 237 kDa sarcolemmal protein known to interact with several proteins to allow the aggregation and fusion of vesicles to the membrane injury site. Partial or total absence of dysferlin protein is responsible for defective membrane repair. Recent evidences suggest that dysferlin plays a role also in myogenesis, but the underlying mechanism remains poorly understood. Using a model of dysferlin-deficient human muscle cells, we observed the presence of thinner myotubes and reduced frequency of myonuclei per myotube, suggesting a delayed progression of myotube formation in Dysf-/- cells. mRNA expression of the myogenic factors showed an increase of MYOD levels after 12 hours of differentiation in both lineages, compatible with a normal early muscle differentiation in Dysf-/- cells. Myogenin expression, however, occurred earlier in Dysf+/+ than in Dysf-/- cells relative to undifferentiated cells, suggesting a change in its kinetics, which could well explain the smaller myotubes observed in Dysf-/- cells. Dysferlin deficiency would therefore be less involved in the first events of formation of small myotubes in early differentiation, but seems to regulate later stages of differentiation, which involves growth and elongation of myotubes. We extended our study to investigate the effects of dysferlin deficiency on the expression of the components of the FAM65B-HDAC6-DYSF tricomplex, in the earlier stages of myoblast differentiation. We observed that mRNA levels of *FAM65B* did not statistically differ between control and dysferlin deficient cells, whereas *HDAC6* presented a transient peak, statistically significant, after 24 hours and *DYSF* increased from 48h of differentiation, only in control cells. We, therefore, suggest that the expression of dysferlin and *HDAC6* are increased in the early stages of differentiation in normal cells, and this would be important for the subsequent step of muscle formation, acting on deacetylation of α-tubulin and on its important role for the elongation of myotubes. Taken together, these results provide interesting targets for further investigation of how dysferlin deficiency affects myogenic regulators during differentiation. FAPESP-CEPID, CNPq-INCT.

1718T

Evolutionary conserved ARX-regulatory pathway in mammals and nematode to find a convergent druggable pathway damaged in neurodevelopmental disorders. *L. Poeta[1], A. Padula[1], B. Attianese[1], M. Valentino[1], S. Filosa[2], H. vanBokhoven[3], L. Altucci[4], E. Di Schiavi[2], M.G. Miano[1].* 1) Institute of Genetics and Biophysics "ABT", CNR, Naples, Italy; 2) Institute of Bioscience, CNR, Naples, Italy; 3) Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; 4) University of Campania "Luigi Vanvitelli", Caserta, Italy.

The X-linked *ARX* gene encodes the Aristaless-related homeobox protein, which belongs to a subset of morphogenetic transcription factors with a crucial role in cerebral development and patterning. Mutations in *ARX* cause a wide spectrum of X-linked neurodevelopmental disorders including lissencephaly, a severe cortical malformation, and a catastrophic epileptic encephalopathy with recurrent and resistant seizures. The health problem to find efficient treatments for *ARX*-related diseases highlights the need for understanding the components and pathways that regulate brain development. Because the neuronal functioning is maintained by a complex regulatory network, the use of powerful genetic models like mouse and worm can complement studies on human genetics and physiology by offering new opportunities to dissect complicated and evolutionary conserved regulatory circuits. Here we describe for the first time the conservation of an ARX-dependent disease-pathway among human, mouse and worm establishing a gene-phenotype association from one organism to another. Indeed, starting from the homologous gene relationships between *ARX* and its murine (*Arx*) and worm (*alr-1)* counterparts, we established a neuronal phenolog relationship among the human phenotypes associated to *ARX* mutations, *Arx* disease models in mouse, and *alr-1* mutants in *C. elegans*. Expanding our previous study on the identification of disease-related targets of ARX in human, we proved that the homologous counterparts of *KDM5C* in mouse (*Kdm5C*) and worm (*rbr-2*) are under the control of ARX/Alr-1 transcriptional activity. In line with these findings, we established a robust downregulation of *KDM5C/rbr-2* transcript in human, murine and worm mutants defective for ARX/alr-1. Given the role of KDM5C/rbr2 as chromatin JmjC eraser with H3K4me2/3 demethylase activity, we tested the H3K4me3 levels establishing a functional defect in chromatin remodelling. We will also present results on molecular and phenotypic rescue obtained by drug repositioning in each models tested upon *in vivo* epi-treatment. In conclusion, our data allow us to define the ARX/alr-1 phenolog-disease pathway leading to better understanding how to correct perturbation of neuronal physiology. Therefore our findings constitute a valuable basis for applying mechanism-based therapies aiming to treat neurodevelopmental diseases caused by defects in transcriptional regulators.

1719F

Emerging role and clinical spectrum of DNM1 in intellectual disability and epilepsy. *F. Bolduc, O. Caluseriu, R. He.* Dept Pediatrics, Univ Alberta, Edmonton, Alberta, Canada.

BACKGROUND: Dynamin-1 (DNM1) is a conserved, large guanosine triphosphatase (GTPase) essential to synaptic vesicle recycling via clathrin-mediated endocytosis. De novo missense mutations in the *DNM1* gene have been identified in patients with severe epileptic encephalopathy, but its role in intellectual disability and cognitive functions are not as well understood. METHODS: We performed exome sequencing on patients with intellectual disability. Clinical characterization in addition to neuropsychological testing were performed. Electroencephalogram investigation was also performed. RESULTS: We identied DNM1 mutations in 2 patients from our cohort. The first patient harbored a previously reported de novo missense mutation c.709C>T (p.Arg237Trp) in one patient. Interestingly, the other patient with a similar but less severe phenotype (both at the cognitive and epilepsy level) was found to have a novel in frame deletion c.647_655del (p.Asp215_Arg217del) in exon 5 of *DNM1*. SIGNIFICANCE: The initial reports on DNM1 indicated very severe epilepsy. Our results show that patients can have variable degree of severity in epileptic symptoms without infantile spasms, epileptic encephalopathy or very early onset. We would like to create a registry for DNM1 patients in order to describe more fully its clinical spectrum.

1720W

c-fos transcript profile in adult zebrafish brain after prolonged pentylenetetrazole exposure. *K. Brito[1], P.G. Barbalho[1], J.E. Cavazos[2], C.V. Maurer-Morelli[1].* 1) Department of Medical Genetics, UNICAMP, Campinas, SP, Brazil; 2) Department of Neurology and Physiology, University of Texas Health Science Center at San Antonio, TX, USA.

Purpose: Zebrafish has many advantages for genetic investigation and since 2005 it has been used for seizure studies. The *cfos* gene is recognized as a marker for neuronal activity, including in zebrafish brain. Kindling is a classical model by which a subconvulsive stimulation becomes progressively convulsive, offering a great opportunity to investigate the epileptogenic process. The present study aimed to investigate the kindling development in the zebrafish model by analyzing seizure behavior and *cfos* mRNA levels after application of subconvulsive dose of pentylenetetrazol (PTZ). **Methods:** Wild-type adult zebrafish were separated in kindling (KG) and control (CG) groups. Animals from KG were individually exposed to a subconvulsive dose of PTZ (7.5mM) for 2 minutes repeated daily for up to 60 days (5 days per week). Animals from CG were handled for 2 min/day, 5 days/week, but in PTZ-free water. The animals were sacrificed at 5th, 15th, 30th and 60th days and their brains were collected for total RNA extraction (n=5 per time point/group). Reverse transcriptase quantitative-PCR amplifications were carried out in triplicates with *cfos* and *ef1α* as endogenous control using TaqManTM System. The relative quantification (RQ) was calculated by the equation $RQ = 2^{-\Delta\Delta CT}$. Data are represented as mean ± Standard Error of Mean (SEM). Each time point was analyzed by Mann-Whitney test and significance was considered when $p \leq 0.05$. **Results:** During 60 days, no induced or spontaneous seizure-like behavior was achieved in the animals exposed daily to PTZ; only a slightly increased swim activity was observed. RT-qPCR showed an up-regulation of the *cfos* mRNA levels in the KG at 5th, 15th, 30th and 60th day compared to their respectively CG. The mean ± SEM of *cfos* gene and the *p* value obtained for comparisons between CG and SG were the following: CG05 1.49 ± 0.28 vs. KG05 5.13 ± 1.60 (p=0.004); CG15 0.43 ± 0.23 vs. KG15 0.29 ± 0.06 (p=0.34); CG30 0.96 ± 0.12 vs. KG30 0.98 ± 0.14 (p=0.05); CG60 0.64 ± 0.2 vs. KG60 1.14 ± 0.2 (p=0.008). **Discussion:** So far there is no chronic model of epilepsy described in the literature for zebrafish. Such a model would be of high impact. While we have not yet established a chronic model, our results show a significant difference in *cfos* mRNA levels between the KG and CG groups. This indicates that neurons were more prone to stimulation in the kindling group. Support: FAPESP 14/15640-8, CEPID-BRAINN 13/07559-3.

1721T

Functional evaluation of rare variants in glutamate receptor *GRIN*, *GRIA*, *GRIK*, and *GRID* genes reveals a diversity of effects on receptor activity. *S.J. Myers[1,2], G. Shaulsky[1,2], S. Kim[1,2], H. Yuan[1,2], S.F. Traynelis[1,2].* 1) Center for Functional Evaluation of Rare Variants; 2) Department of Pharmacology, Emory University School of Medicine, Atlanta, GA.

Whole exome and genome sequencing has identified numerous rare mutations in the glutamate receptor family of *GRIN*, *GRIA*, *GRIK*, and *GRID* genes that are associated with neurological syndromes and phenotypes such as epilepsy (including infantile spasms, Lennox-Gastaut, Landau-Kleffner, and West syndromes), intellectual disability, developmental delay, language disorders, hypotonia and movement disorders, autism and schizophrenia. At the Center for Functional Evaluation of Rare Variants (CFERV), we test specific glutamate receptor gene variants identified in patients in a series of electrophysiological and biochemical studies. Briefly, variant glutamate receptors are expressed in *Xenopus laevis* oocytes and in HEK cells, then used to measure agonist potency (L-glutamate and glycine EC_{50}), agonist deactivation kinetics, receptor sensitivity to inhibition by the endogenous modulators Mg^{2+}, Zn^{2+}, and H^+ (IC_{50}, % inhibition), and cell-surface receptor trafficking. Approximately two hundred specific missense, frameshift, or splice-site variants in the *GRIN* gene family (primarily *GRIN1*, *GRIN2A*, and *GRIN2B*) are reported with association to a neurological disorder/disease phenotype, of which over seventy have been functionally characterized by CFERV to date. Approximately 90% of the variants we evaluated possess altered function in at least one functional parameter (up to 8 parameters/mutation measured) with the majority (~65%) of mutations affecting multiple parameters of function. Disease associated variants are not evenly distributed across the gene with 46% residing in the agonist binding domains (ABD), 23% in membrane-spanning (M) domains, and 14% in linker domains connecting ABD to M domains. Most gene mutations (~75%) result in a gain or loss of function of less than 10 fold compared to the wild-type receptor, however several mutations altered L-glutamate agonist affinity by over 100 fold. We present data on over 20 novel glutamate receptor variants and describe the effect of these mutations on receptor function. The goal of CFERV is to expand our understanding of the relationship between genetic variation and health consequences. <u>Funding:</u> Supported by NIH awards NIH-NINDS R24NS092989, R01NS036654, and R01NS065371 to SFT, and the NICHD award number R01HD082373 to HY. <u>Competing Interests:</u> SFT consults for Janssen Pharmaceutical, Boehringer Ingelheim Pharma GmbH&Co. KG, is on the Scientific Advisory Board Sage Therapeutics, and a co-founder of NeurOp, Inc.

1722F

De novo NMDAR *GRIN* mutations in M2 channel pore-forming domain associated with neurological diseases. *H. Yuan[1,2], J. Zhang[1], W. Tang[1], H. Kusumoto[1], S.J. Myers[1,2], S.F. Traynelis[1,2].* 1) Department of Pharmacology; 2) Center for Functional Evaluation of Rare Variants (CFERV), Emory University School of Medicine, Atlanta, GA, 30322.

N-methyl-D-aspartate receptors (NMDAR) bind synaptically released glutamate and mediate a slow component of excitatory synaptic transmission in the brain that plays an important role in brain development. Genetic mutations in multiple NMDAR subunits are associated with various neurological diseases. Here, we summarized clinical information and evaluated functional changes for 12 *de novo* missense mutations in NMDAR subunit gene, *GRIN1*, *GRIN2A*, and *GRIN2B*, located in the M2 loop that lines the NMDAR channel pore. These mutations were identified in children with various neurological diseases including epilepsy syndrome, developmental delay, intellectual disability, hypotonia, autism, or speech disorder. The mutant was introduced into human NMDAR GluN cDNA construct using QuikChange protocol. cRNAs were synthesized from cDNA and injected into *Xenopus laevis* oocytes. Two-electrode voltage clamp current recordings were performed to evaluate agonist (glutamate and gycine) potency, sensitivity to negative modulators (magnesium and proton). Current-voltage relationships were also compared to evaluate changes in voltage-dependent Mg^{2+} block. Compared to WT receptors, the mutant NMDARs exhibited a modest or no changes in potency (EC_{50} values) for both glutamate and glycine, as well as a modest or no changes in proton inhibition (percentage current inhibition at pH 6.8 compared with the pH 7.6). Mg^{2+} inhibition was significantly reduced in all 12 mutants, with significantly increase in IC_{50} values (reduced inhibition) and decrease in the degree of Mg^{2+} inhibition on agonist-evoked current response. Current-voltage curves revealed marked reduction in Mg^{2+} inhibition in all the mutants. Overall, these results suggest these M2 variants are gain-of-function mutations through reduced Mg^{2+} channel blockage, which is consistent with its location in the M2 loop of the channel pore and may underlies the patients' phenotypes. <u>Funding</u>: This work was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD) of NIH under award number R01HD082373 to H.Y., by the National Institute of Neurological Disorders and Stroke (NINDS) of the NIH under award numbers NIH-NINDS R01NS036654, R01NS065371, and R24NS092989to S.F.T. <u>Competing Financial Interests</u>: S.F.T. is a consultant of Janssen Pharmaceuticals, Inc., Boehringer Ingelheim Pharma GmbH & Co. KG, a member of the Scientific Advisory Board for Sage Therapeutics, and co-founder of NeurOp Inc.

1723W

Intracellular mislocalization of mutant proteins as a screen for therapeutic agents to treat genetic diseases. *T. Kouga[1], H. Osaka[1], C. Ohba[2], E. Jimbo[1], T. Yamagata[1], H. Saitsu[3], N. Matsumoto[2].* 1) Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 2) Department of Human Genetics, Graduate School of Medicine, Yokohama City University, Yokohama, Japan; 3) Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu, Japan.

[;Background& Objective];Mutant proteins coded by pathogenic gene sequence variants may not fold properly into their normal tertiary structures. A misfolded protein may not be able to target the proper membrane or organelle in which it resides and functions. Thus, confirmation of the distinct localization of a mutant protein can validate the pathogenicity of a particular gene variant. Moreover, this miss-localization can be utilized as a probe to identify chaperon molecules. The objective of this study was to assess the feasibility of visually localizing mutant proteins to identify agents that may correct folding and organelle targeting. [;Method];In a previous exome sequencing of pediatric patients with cerebellar ataxia, 9 mutations in 5 genes were found (Ca channel a-1A subunit, *CACNA1A*; folate receptor family 1, *FOLR1*; peroxisomal membrane protein 16, *PEX16*; catalytic subunit of mitochondria DNA polymerase, *POLG*; and tripeptidyl-peptidase 1, *TPP1*) (Ohba et al. 2013). Their cellular locations are membrane, membrane, peroxisome, mitochondria, and lysosome, respectively. For each mutant gene, we prepared wild-type and mutant FLAG-tagged plasmids and transfected these plasmids into HEK293 cells. We selected plasmids that showed a difference in cellular localization between those with wild-type and mutant gene sequence by immunostaining the transfected cells with anti-FLAG antibody. We also tagged the plasmids with *green fluorescent protein (GFP)* to express as fusion protein for each plasmid and confirmed the localization differences observed with FLAG-tagged proteins. [;Result];Of the 9 mutations tested, *PEX16* and *POLG* showed differences in cellular localization between wild-type and mutant plasmid-transfected cells. We isolated stable cell lines that express these mutant proteins fused with GFP. [;Conclusion];We demonstrated the usefulness and limitations for visualizing the intracellular localization of mutant proteins found in patients with cerebellar ataxia. The stable cell lines expressing mutant proteins fused with GFP can be used for localization-oriented drug screening to identify agents that may restore activity to misfolded proteins in patients with genetic diseases. .

1724T

Genetics of congenital megacolon in East Asians. *C. Tang[1,2], X. Zhuang[1], S. Cherny[3,4], P. Sham[3,4], M. Garcia-Barceló[1], P. Tam[1,2].* 1) Department of Surgery, The University of Hong Kong, Pokfulam, Hong Kong; 2) Dr Li Dak-Sum Research Centre, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 3) Department of Psychiatry, The University of Hong Kong, Pokfulam, Hong Kong; 4) Centre for Genomic Sciences, The University of Hong Kong, Pokfulam, Hong Kong.

Hirschsprung disease (HSCR), also known as congenital megacolon, is characterized by the absence of enteric ganglia in the hindgut. HSCR is a multigenic neurocristopathy with a two-fold increase in the incidence among East Asians. The major HSCR gene, *RET*, has both rare and common variants contribute to the disease risk. In particular, ~80% of Asian HSCR patients carry the high risk common enhancer allele. When *in trans* with the recently discovered, Asian-specific missense variant, the *RET* enhancer SNP increases the risk of HSCR by more than 20-fold. Thus far, over 10 HSCR genes were found to have common and/or rare variants attribute to disease susceptibility. To discover novel HSCR-associated gene(s), we performed a whole genome sequencing (WGS) of 443 HSCR cases and 493 controls of East Asian ethnicity. Rare variant association analysis was carried out and, with subsequent follow-up study, we identified a significant excess of rare protein-altering mutations in *BACE2* (p=2.9x10^{-6}) in HSCR cases (4%) compared with controls (0.6%). The relevance of *BACE2* in HSCR is highlighted by the facts that i) the encoded protein has recently been proposed as a target for drug-based treatment of HSCR; and ii) *BACE2* maps to the HSCR-Down Syndrome (DS) critical chromosomal region, with DS being the most frequent chromosomal abnormality associated with HSCR. The *BACE2*-HSCR association may represent the missing link between these two disorders. Results of the common variant association analysis will also be presented.

1725F

Pathogenic U2-type 5' GC donor splice site in the *FOLR1* gene causes cerebral folate deficiency with autism and attention deficient hyperactive disorder in three affected from a large consanguineous family. *M. Alaamery[1], A. Sheereen[1], S. Massadeh[1], W. Eyaid[2].* 1) Developmental Medicine Department, King Abdullah International Medical Research Centre (KAIMRC), Riyadh, Saudi Arabia; 2) Department of Pediatrics, King Abdulaziz Medical City (KAMC) / Ministry of National Guard-Health Affairs (MNG-HA), King Abdullah International Medical Research Center (KAIMRC) /King Saud bin Abdulaziz University for Health Sciences (KSAU-HS), Riyadh, Saudi Arabia.

Background Cerebral folate deficiency (CFD) is an inherited brain-specific folate transport defect that is caused by mutations in the folate receptor 1 (FOLR1) gene (MIM*136430) coding for folate receptor alpha (FRα). This genetic defect gives rise to a progressive neurological disorder with late infantile onset. There are several other diseases associated with CFD and hence expands the clinical spectrum of CFD. The majority of the CFD cases are associated with Autism spectrum disorder (ASD), which is the result of autoantibody against FRα. However, this is the first report of the patient with CFD with ASD and Attention deficient hyperactive disorder (ADHD) caused by the mutation in the FOLR1 gene. **Aim** We aim to identify the genetic cause of CFD with ASD and ADHD in a large consanguineous Saudi family with three affected members. **Methods** Clinical whole-exome sequencing was performed on the proband and his parents. Later Sanger sequencing was done to validate and confirm segregation in the two affected and other family members. **Results** A novel homozygous splice site variant (c.493+2T>C) at 5' donor splice site in intron 6 of the FOLR1 gene was identified in three individuals with CFD in addition to ASD and ADHD. Software analysis Alamut (v.2.7.1) and Human Splice Finder (HSF3.0) predict this change to have an effect on splicing and is classified as the pathogenic variant. The splice site variant segregates in the family in autosomal recessive fashion. **Conclusions** This is first case study reporting the mutation in the FOLR1 causing CFD with ASD and ADHD, thereby expanding the clinical heterogeneity observed in individuals with CFD. The constitutive 5'donor splice site GT-AG in introns possess strong signals at their donor sites for splicing reaction, however, in this patients, the GT-AG is replaced by GC-AG. Although, GC-AG is a rare canonical splice site occurring in 1% of total introns, but could not rescue the phenotype as it does in its constitutive form in the introns. This mutation shows that GC-AG in the intron possesses weak consensus sequences at their donor sites in addition to the surrounding nucleotides around the 5' donor splice site. Thus this change depicted by several in silico method causes in altered spliced site, by generating a new cryptic site in an intron to produce a pathogenic variant.

1726W

Impact of defective protein N-glycosylation on the developing mouse cerebellum. *V. Cantagrel[1], D. Medina-Cano[1], M. Nicouleau[1], C. Guerrera[2], L. Colleaux[1].* 1) INSERM UMR 1163 - Imagine Institut, Paris Descartes - Sorbonne Paris Cité University, France; 2) Proteomic Platform Necker, PPN-3P5, SFR Necker US24, Paris Descartes - Sorbonne Paris Cité University, France.

N-linked glycosylation is the most frequent modification of secreted and membrane-bound proteins in eukaryotic cells. It relies on the synthesis of a precursor called the lipid-linked oligosaccharides (LLO) located in the endoplasmic reticulum (ER). Disruption of the LLO synthesis results in a reduced occupancy of protein N-glycosylation sites (i.e. hypoglycosylation), and is the basis of some congenital disorders of glycosylation (CDG) that are frequently associated with a multi-systemic presentation including motor and cognitive impairments. How this abnormal protein glycosylation affects the development or function of specific organs is often totally unknown. We aim to understand why the developing brain is frequently affected in CDG patients and to describe the underlying global protein glycosylation defect. For this purpose, we generated a conditional knockout (cKO) mouse for the *steroid 5alpha-reductase type 3* gene (*Srd5a3*). This gene codes for an enzyme involved in the synthesis of dolichol, the lipid used to build the LLO precursor. Mutations in this gene cause CDG associated with intellectual disability, ophthalmologic and cerebellar symptoms. The cerebellum specific cKO mouse shows motor coordination impairment associated with a reduced cerebellar volume and histological anomalies affecting granule cell precursors. Transcriptomic and proteomic analyses excluded a significant disruption of the ER homeostasis or a general protein glycosylation defect as the cause of the *Srd5a3* cKO mouse phenotype. These investigations point to a limited set of glycoproteins including mainly one protein family that is down-regulated. Cerebellum cell culture experiments show that the hypoglycosylation of the previously identified protein family significantly impairs neuronal development.

1727T

PEA15 deficiency is associated with striking neurologic and motor abnormalities in *Felis catus*. *J.N. Cochran[1], E.C. Graff[2], C.B. Kaelin[1], K. Day[1], H.L. Gray-Edwards[2], R. Watanabe[2], J.W. Koehler[2], R. Falgoust[2], B.L. Brunson[2], J.W. Prokop[1], B.N. Lasseigne[1], R.C. Ramaker[1], K.L. Engel[1], C. Henegar[1], J.B. Benito[1], R.M. Myers[1], N.R. Cox[2], D.R. Martin[2], G.S. Barsh[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA; 2) College of Veterinary Medicine, Auburn University, Auburn, AL, USA.

A recessively inherited developmental abnormality was recognized in a feline colony. Affected cats have perinatal generalized tremors and are ataxic as they begin to ambulate. The animals partially compensate for motor abnormalities by 6 months of age, but develop aggression and sensory abnormalities (fly biting and stargazing) at approximately 1.5 years of age. At necropsy, affected cats exhibit gross abnormalities including generalized microcephaly, reduced brain weight (~50% reduction in brain weight), and striking cerebral cortex gyrification deficits. These abnormalities are also apparent by MRI. On histologic evaluation, affected animals have less myelin, fewer oligodendrocytes, and disorganized neuronal processes in regions that typically have laminar organization. Through whole genome sequencing of 5 cats, RNA-Seq of cerebral cortex from 16 cats, and subsequent targeted genotyping of a candidate region in 96 cats, we identified a 1.3 MB haplotype on chromosome F1 associated with the phenotype. Within this region, two variants are in 100% agreement with presumed genotype based on phenotype (affected – homozygous alternate, obligate carrier – heterozygous alternate, and normal or possible carrier – not homozygous alternate) in all cats tested (LOD score > 10). One of the variants is a missense variant not predicted to be damaging by PROVEAN in SLAMF1, a gene implicated in immunity and not expressed in the brain. The putatively causative variant is a frameshift variant in PEA15, a gene involved in ERK signaling that is highly expressed in brain. Both count data from all genotype groups and strand bias in heterozygous alternate cats from RNA-Seq suggest that PEA15 is undergoing NMD, and westerns confirm complete ablation of PEA15 in affected cats. In conclusion, we propose that PEA15 may be critical for normal large animal CNS development and function.

1728F

Pathologic characterization of a neurodevelopmental abnormality in the cerebral cortex of domestic cats with loss of PEA15. *E.C. Graff[1], J.N. Cochran[2], C.B. Kaelin[2], K. Day[2], H.L. Gray-Edwards[1], R. Watanabe[1], J.W. Koehler[1], R.A. Falgoust[1], B.L. Brunson[1], J.W. Prokop[2], B.N. Lasseigne[2], R.C. Ramaker[2], K.L. Engel[2], C. Henegar[2], N.E. Morrison[1], J.B. Benito[2], R.M. Myers[2], N.R. Cox[1], G.S. Barsh[2], D.R. Martin[1].* 1) College of Veterinary Medicine, Auburn University, Auburn, AL; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA.

A developmental abnormality with an autosomal recessive inheritance pattern was recognized in a feline breeding colony. Through whole genome sequencing of 5 cats, RNA-Seq of cerebral cortex from 16 cats, and subsequent targeted genotyping of a candidate region in 96 cats, we identified a 1.3 MB haplotype on chromosome F1 associated with the phenotype. Within this region, the putatively causative variant is a frameshift variant in PEA15. Perinatal affected kittens have generalized tremors and neurological motor and sensory abnormalities that become apparent as the cats begin to ambulate. As affected cats approach adulthood, they are able to compensate partially for neuromotor deficits. At approximately 1.5 years, affected cats develop sensory abnormalities (fly biting and stargazing), and often become aggressive. Clinicopathologic data, including CSF analysis, suggests a non-progressive, stable disease process. At necropsy, lesions are limited to the cerebral cortex, with an absence of craniofacial or cerebellar abnormalities. Affected cats have marked microcephaly (brain weight ~50% less than normal) and a poorly formed cerebral cortex with smooth, shallow gyri. There is reduced to absent formation of major gyri, including the ectosylvian and sylvian gyri. MRI analysis also identified gyrification abnormalities, including a reduced to absent anterior cingulate gyrus, and flattening of the parietal and temporal lobes. On histologic evaluation of affected cats, the layers of the cerebral cortex are markedly disorganized with a distinct loss of cortical layering. In some cats, polymicrogyria are present. Affected animals have decreased myelin and fewer oligodendrocytes. Cell type deconvolution from RNA-Seq reveals that transcripts specific to oligodendrocyte precursor cells have more counts per million in affected animals, while oligodendrocyte specific transcripts have fewer counts per million, which is consistent with the histological myelin and oligodendrocyte observations. Furthermore, there is an enrichment for transcripts changing in the same direction as transcripts from RNA-Seq of brain samples from a human psychiatric disease study. In conclusion, we propose that PEA15 may be critical for normal large animal CNS development, and functional and neuropathological characterization of loss of PEA15 will lead to a better understanding of neuroembryology and cerebral cortex development.

1729W

Investigation of synergistic interactions among genes in the 15q11.2-q13.1 region using *Drosophila melanogaster*. *K.A. Hope, D. Johnson, D. Kakhniashvili, L.T. Reiter.* University of Tennessee Health Science Center, Memphis, TN.

Duplication 15q syndrome (Dup15q) is an autism-associated copy number variation (CNV) disorder resulting from maternally inherited duplications of 15q11.2-q13.1. The E3 ubiquitin ligase *UBE3A* is located within 15q11.2-q13.1 and increased dosage of *UBE3A* is believed to cause Dup15q phenotypes. A second E3 ubiquitin ligase, *HERC2*, is also within the typical duplication boundaries and individuals with Dup15q have elevated dosage of *HERC2* in addition to *UBE3A*. In vitro biochemical work demonstrated a physical interaction between UBE3A and HERC2, with HERC2 stimulating the ubiquitin ligase activity of UBE3A. Using the GAL4/UAS system in *Drosophila melanogaster*, we investigated the effects of neuronal overexpression of *dHERC2* alone, *Dube3a* alone (the fly homologs of *HERC2* and *UBE3A*), or both *dHERC2+Dube3a*. Overexpression of *dHERC2* reduced Dube3a protein levels, indicating that dHERC2 stimulates the ubiquitin ligase activity of Dube3a (Dube3a ubiquitinates itself). We performed quantitative mass spectrometry and differential expression analysis to identify global changes in the proteomic landscape of the brain under various Dube3a/dHERC2 combinations. We reliably identified and quantified approximately 5,000 total proteins, with 1,980 differentially expressed proteins in *Dube3a* overexpression, 3,037 differentially expressed proteins in *dHERC2* overexpression, and 3,224 differentially expressed proteins in *dHERC2+Dube3a* overexpression. Using Protein Set Enrichment Analysis we observed a significant enrichment in the *dHERC2+Dube3a* overexpression group for RNA such as mRNA binding, mRNA splicing, mRNA metabolic process, mRNA processing, ribosome, among others. One particular mRNA binding protein, FMR1 (fly homolog of *Fragile X Mental Retardation 1*), was significantly upregulated in *dHERC2+Dube3a* overexpression. We are currently identifying other proteins whose expression is regulated by *dHERC2* and *Dube3a* in a synergistic manner, and validating the protein changes observed in FMR1 and other proteins regulated by FMR1. Additionally, we are searching for direct ubiquitin ligase substrates of Dube3a. This study demonstrates the importance of approaching CNV disorders holistically, as elevated HERC2 levels alone appear to influence the proteomic landscape, the activity of UBE3A, and the phenotypes observed in Dup15q individuals.

1730T

IBGC mouse model with *SLC20A2* mutation and potential prevention and therapeutics. *J.Y. Liu, C. Wang, X. Xu, Z. Cao, T. Niu.* College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China.

Idiopathic basal ganglia calcification (IBGC, OMIM 213600) is an inheritable neuropsychiatric disorder characterized by mineral deposits in the basal ganglia and other brain regions. Affected individuals could present a wide variety of symptoms that include movement disorders, cognitive impairment and psychiatric signs. So far, no therapeutic drug has been developed for the treatment of IBGC. Loss-of-function mutations in *SLC20A2*, which encodes the type III Na/Pi transporter PiT2, have been reported to account for 40% IBGC cases, suggesting it's a major causative gene of IBGC. In this study, we established an IBGC mouse model with *Slc20a2* p.S602W mutation, and investigated the effect of Chinese herbal medicine SYM on brain calcification. The *Slc20a2*$^{S602W/S602W}$ mice developed brain calcification since 3 months old, and the number and volume of calcified nodules increased with age. There was marked elevation of phosphate (Pi) in the cerebrospinal fluid (CSF) of the *Slc2a2*$^{S602W/S602W}$ mice relative to the WT mice, whereas serum levels of Pi were nearly identical between groups. Motor incoordination and impaired learning/memory have been observed in *Slc2a2*$^{S602W/S602W}$ mice during behavioral tests. The *Slc20a2*$^{S602W/S602W}$ mice that drinking SYM decoction since 3 weeks old did not develop brain calcification until 5 months old, and they also showed less brain calcifications at 6 and 7 months of age relative to the controls. We found that CSF protein X, an osteogenic maker protein, was reduced in *Slc2a2*$^{S602W/S602W}$ mice from the SYM group. These results suggested that the SYM decoction may produce preventive effect on the formation of brain calcifications, which might be useful for the treatment of IBGC caused by *SLC20A2* mutations.

1731F

CRISPR/Cas9 overcomes the challenges of microsatellite knockin development. *R. Oliveira, J. Thomas, F. Ivankovic, M.S. Swanson.* Molecular Genetics and Microbiology, Center for NeuroGenetics, University of Florida, Gainesville, FL.

Myotonic dystrophy (DM) is the most common muscular dystrophy in adults, affecting 1 in 8,000 individuals. There are two types of DM, both caused by microsatellite expansion mutations. Myotonic dystrophy type 1 (DM1) is the most severe form and is caused by a CTG expansion (CTG^{exp}) in the 3' untranslated region (3' UTR) of the *DMPK* gene on chromosome 19q13.3. In unaffected individuals, this locus contains 5-37 CTG repeats, but in DM1 these repeats expand into the thousands. Transcription of the expanded *DMPK* allele produces toxic RNAs that sequester multiple RNA-binding proteins, including alternative splicing factors from the muscleblind-like (MBNL) family. MBNL sequestration results in misplicing of developmentally regulated genes, which is a characteristic feature underlying DM1 pathogenesis. Transgenic mice expressing mutant *DMPK* alleles and *Mbnl* knockout models present many of the key molecular and physiological manifestations of DM1, including abnormal alternative splicing, myotonia, heart conduction defects and REM sleep disturbances. Despite these successes, no model has recapitulated the entire spectrum of disease symptoms, a deficiency that has hindered the development of therapeutic strategies that target all tissues affected by DM1. A much needed DM1 model is the knockin of large CTG^{exp} at the endogenous mouse *Dmpk* gene, which is expected to express toxic RNA molecules with a spatiotemporal distribution mirroring that observed in affected individuals. Until recently, development of such a model has been hindered by the instability of targeting vectors carrying large CTG^{exp} in bacterial strains and the small repeat numbers capable of generating viable mouse ES cells for blastocyst injections. Using a bacteria-free cloning approach based on rolling circle amplification, together with CRISPR/Cas9 genome engineering, we generated a *Dmpk* knockin model carrying 202 CTG repeats, the largest microsatellite knockin developed to date. These mice express wild-type and mutant alleles at similar levels and accumulate toxic RNAs in several tissues following the correct *Dmpk* spatiotemporal distribution. Currently, we are developing additional *Dmpk* knockin mice with even larger CTG repeats, models that will enhance our understanding of earlier and more severe disease onsets, and provide robust platforms to develop effective therapies for myotonic dystrophy. This study was supported by grants from the NIH (NS058901, NS98819) and the MDA (RG 480539) to M.S.S.

1732W

Clec16a knockdown mice develop a neuronal phenotype with ataxia. *M. Bakay[1], H. Hain[1], E. Lancaster[2], J. Li[2], R. Pandey[1], M. Romer[2], B. Strenkowski[1], J. Grinspan[2], S. Scherer[2], H. Hakonarson[1,3].* 1) Center for Applied Genomics, Children's Hosp Philadelphia, Philadelphia, PA; 2) Department of Neurology the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA.

CLEC16A SNPs are associated with multiple autoimmune and inflammatory diseases, including type 1 diabetes, multiple sclerosis, primary adrenal insufficiency, Crohn's disease, primary biliary cirrhosis and rheumatoid arthritis, however, the primary function of this gene and the protein its encodes remains largely unknown. Several studies in different model systems revealed that CLEC16A play an important role in autophagy/mitophagy. Recent publication suggested that CLEC16A is critical autolysosome function and purkinje cell survival. We have generated an inducible whole body (WB) knockdown (KD) mouse model to look into the role of CLEC16A. We hypothesized that CLEC16A plays an important role in the nervous system and UBC-Cre-Clec16a[loxP] mice - *Clec16a* WB KD - will display a phenotype similar to other multiple sclerosis (MS) models. CLEC16A knockdown was induced in adult mice of 10 weeks of age. Clec16a KD mice exhibited significant weight loss (P<0.001) as compared to control and a neuronal phenotype including tremors, impaired gait, and spasticity with an increase in severity over time. Over 40 days observation period 100% of the UBC-Cre-Clec16a[loxP] (KD) mice exhibited some disability. About 63% of the mice reached a disability score of '3' and 58% the highest severity score of '4'. No controls showed any impairment. No difference in activity level was noted. 100% of the KD mice exhibited impaired hindlimb splay with both hindlimbs affected as compared to controls. 100% of the KD have many abnormal neurons in the dorsal root ganglia including neurons with vacuoles, myelin debris or myelin-like structures in comparison to control mice. Nerve conduction study revealed statistically significant reduction of neuronal amplitude (P=0.016) and conduction velocity (P=0.021) in CLEC16A KD mice when compared to control littermates. In addition, the UBC-Cre-Clec16a[loxP] KD mice exhibited cerebellar Purkinje cell loss similarly to previously published by Redmann et al, 2016. Taken together, these results demonstrate that CLEC16A is required for normal sensory neuronal function and that CLEC16A risk variants associated with decreased CLEC16A expression may contribute to the pathogenesis of MS and neurodegeneration in humans. Further studies are necessary to understand a precise mechanism of CLEC16A action and to determine if intervention at the CLEC16A level can prevent neurological disorders such as MS in individuals with risk associated CLEC16A SNPs.

1733T

The binding of RNA regulates the formation of nuclear membraneless structures by Matrin 3 and TDP43 in myocytes. *M.C. Gallego Iradi, R. Davila, A. Crown, H. Brown, J. Lewis, D. Borchelt.* University of Florida, Department of Neuroscience Center for Translational Research in Neurodegenerative Disease Gainesville, FL.

Mutations in Matrin 3 (MATR3) can cause familial amyotrophic lateral sclerosis (fALS) or familial distal myopathy. MATR3 is structurally similar to TDP43 and other RNA binding proteins associated with fALS, but differs in lacking a recognizable prion-like domain (PLD). RNA binding proteins with PLDs have the capacity to undergo liquid phase transition to form membraneless organelles. Similar to wild-type (WT) TDP43, WT MATR3:YFP adopted a diffuse intranuclear distribution with interspersed puncta. To understand the basis for the distribution of WT MATR3, we expressed MATR3:YFP fusions with deletions of DNA and RNA binding motifs [Zn finger domains 1 and 2, and RNA recognition motifs (RRM) 1 and 2]. Deletion of either Zn finger domain had no effect on the distribution of MATR3:YFP. By contrast, deletion of the RRM2 RNA-binding domain (ΔRRM2) caused the MATR3:YFP to form intranuclear droplets that fused to form larger structures. Some large droplets were vacuolated and appeared to possess multiple phase-compartments. Deletion of the RRM1 domain of MATR3 also produced droplets, but this case they were smaller and less stable. MATR3:YFP ΔRRM2 with the fALS mutation F115C could form large droplets similar to WT MATR3:YFP, but the myopathic S85C mutation produced much smaller and more dispersed droplets. MATR3:YFP ΔRRM2 co-expressed with TDP43:mCherryΔRRM2 produced large droplets containing both proteins. Our findings demonstrate that MATR3 is similar to TDP43 in possessing an ability to undergo liquid-phase separation, and suggest that these proteins have tendencies to interact through a phase-separation mechanism.

1734F

Neuronal inflammation and dysregulated mitophagy features in ubiquitous *Clec16a* knockdown mice. *H.S. Hain[1], R. Pandey[1], M. Bakay[1], J. Li[2], M. Romer[2], B.P. Strenkowski[1], J.B. Grinspan[2,3], S.S. Scherer[2], H. Hakonarson[1,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Genetic variants in the 16p13 locus harboring the *CLEC16A* gene have been implicated in twelve different autoimmune diseases including multiple sclerosis (MS). However, the role of the gene and variants in the physiology and pathogenesis of autoimmune disorders remains largely unknown. Studies in different model systems revealed that CLEC16A plays an important role in autophagy/mitophagy. Ubiquitous inducible knockdown (KD) mice were generated to investigate the role of CLEC16A in autoimmune diseases. It was hypothesized that UBC-Cre-Clec16a[loxP] KD mice would display a phenotype similar to other MS or neurodegenerative mouse models with inflammatory and autoimmune features. *Clec16a* KD mice exhibited immune dysfunction, weight loss, and a neuronal phenotype with an increase in severity over time leading to morbidity. Accelerated mitophagy in cerebellum, cortex, trigeminal ganglia, dorsal root ganglia, and spinal cord was observed in *Clec16a* KD mice as compared to control mice. This was noted by changes in autophagy and mitophagy-related proteins as measured by Western blots and aberrant mitochondria as observed by transmission electron microscopy. Increases in plasma inflammatory cytokines seen in autoimmune disorders accompanied the behavior and mitophagy. Inflammation rose locally in the central nervous system as exhibited by inflammatory cytokine differences and astrogliosis measured by Western blots and immunohistochemistry. The results establish a connection between CLEC16A, dysregulated mitophagy, and inflammation contributing to neurodegeneration in the central nervous system. These results suggest that *CLEC16A* risk variants associated with decreased CLEC16A expression may contribute to the pathogenesis of MS and neurodegeneration in humans. Further studies are necessary to understand the precise mechanism and interactions of CLEC16A. We conclude that drugs modulating mitophagy/autophagy, particularly in the JAK/ STAT pathways associated with cytokine regulation, may be effective to treat or prevent neurological autoimmune disorders such as MS in individuals with risk associated *CLEC16A* variants.

1735W

Impact of rare variants in genes that encode components of the endocannabinoid system. *D. Smith, C. Stanley, T. Foss, K. McKernan.* Courtagen Life Science, Woburn, MA.

Rare genetic variants in the core endocannabinoid system genes *CNR1, CNR2, DAGLA, MGLL* and *FAAH* were identified in molecular testing data from up to 5,979 patients with a broad spectrum of neurological disorders including epilepsy, mitochondrial dysfunction, and developmental disorders. The numbers of subjects harboring rare variants were 22 for *CNR1*, 11 for *CNR2*, 35 for *DAGLA*, 34 for *MGLL* and 52 for *FAAH*. Fisher exact tests were performed to evaluate the possible association of phenotypes present in cases with rare variants, compared to controls from the same data set without such variants. Heterozygous rare coding variants in *CNR1* (which encodes the cannabinoid receptor, type 1) were found to be significantly associated with headache or migraine, anxiety, sleep and memory disorders (alone or in combination) compared to controls. Similarly, heterozygous rare variants in *DAGLA* (which encodes diacylglycerol lipase alpha) were found to be significantly associated with seizures and developmental disorders, including abnormalities of brain morphology, compared to controls. Rare variants in *MGLL*, *FAAH* and *CNR2* were not found to be associated with neurological phenotypes in the patients tested. The phenotypes associated with rare *CNR1* variants are similar to those observed in *CNR1* knockout mice, and are also reminiscent of those implicated in the theory of clinical endocannabinoid deficiency syndrome. The severe phenotypes associated with rare *DAGLA* variants underscore the critical role of rapid 2-arachidonoyl glycerol synthesis and the endocannabinoid system in regulating neurological function and development. Mapping of the variants to the 3D structure of the type 1 cannabinoid receptor, or the primary structure of diacylglycerol lipase alpha, reveals clustering of variants in certain regions and is consistent with impacts to function.

1736T

Genetic analysis of Japanese patients with neurofibromatosis type 1 and the neuronal complications. *K. Fujita[1], K. Yamashita[1], T. Hoshina[1], N. Hikita[1], S. Sakuma[1], H. Shintaku[1], K. Fukai[2], T. Takenouchi[3,4], T. Uehara[3], K. Kosaki[3], H. Saya[5], T. Seto[1].* 1) Pediatrics, Graduate School of Medicine, Osaka City University, Osaka, Japan; 2) Dermatology, Graduate School of Medicine, Osaka City University, Osaka Japan; 3) Center for Medical Genetics, Keio University School of Medicine; 4) Pediatrics, Keio University School of Medicine; 5) Institute for advanced medical research, Keio University School of Medicine.

Neurofibromatosis type 1 (NF1) is an autosomal dominant disorder. Tumor-like lesions on brain magnetic resonance imaging expressed as unidentified blight object (UBO) rarely develops to pilocytic astrocytoma, optic glioma. We present patients with moderate-to-severe NF1 since we performed genetic analysis. Methods: We analyzed 5 Japanese families including 8 patients with NF1. Patient(P) 1, a proband with NF1 in this family, was a 3-year-old girl who had optic glioma, UBOs in her brain, mild intellectual disability, and autistic spectrum disorder (ASD). P2 was a 21-year-old man who had plexiform neurofibroma. P3 was a 30-year-old man who had obvious neurofibroma. P4 was a 9-year-old boy who had attention deficit/hyperactivity disorder. P5 was his mother with NF1. P6 and P7 were 5-year-old sister and 2-year-old brother. P6 had developmental delay, UBOs in her brain, plexiform neurofibromatosis and ASD. P7 also had UBOs in his brain, optic glioma. P8 was their mother with NF1. Genetic analysis was performed with the patients and their parents after genetic counseling by a medical geneticist and documented informed consent. DNA was purified from their blood samples, and the target region was concentrated by using TruSight One Sequencing Panel that contains coding regions of 4,813 genes and was sequenced on a Miseq platform. This study was approved by the institutional review boards of the participating universities. ; Results: P1 was identified to have c.2350T>C in the 20th exon of *NF1* that was missense mutation in the exon thought to be pathogenic , already reported same mutation. P2 did not show any pathogenic mutation in the exons except for the introns, even after performing comparative genomic hybridization array analysis. P3 was c.3527 delG in the 27th exon, which is thought to be novel mutation. P4 and his mother showed pathogenic nonsense mutation in the 14th exon of *NF1* as c.1603 C>T change causing exon termination, as previously reported. Finally, P5 and P6 and their mother were identified to have c.888+2T>A, which was located in the donor site of the intron, generated owing to splicing causing a pathogenic effect to produce the mRNA, thought to be a novel mutation. Conclusion: We determined the pathogenic mutation in 7 out of 8 patients. Two of them are thought to have a novel mutation. A larger sample size is required in future studies to accurately predict their prognosis, especially regarding the development of malignant tumors in the brain.

1737F

Disruptions to the miRNA regulatory pathway may cause an increased rate of schizophrenia in individuals with 22q11.2 DS. *W. Manley, G. Davis, S. Siecinski, S. Ryan, V. Coulibaly, L. Brzustowicz.* Rutgers, Piscataway, NJ.

Schizophrenia is a complex and poorly understood disease caused by the interplay of environmental and genetic factors. The 22q11.2 deletion syndrome (DS) is a disorder that is caused by the microdeletion of part of chromosome 22 leading to a 25-fold greater chance of developing schizophrenia versus the general population. The missing 22q11.2 region contains DGCR8, which is required for the initial step of miRNA biogenesis. However, the 22q11.2 deletion itself is not directly the cause of schizophrenia, since 75% of individuals with this deletion do not develop the disease. We hypothesize that the 22q11.2 deletion increases the risk of schizophrenia through alterations in miRNA regulatory networks via depletion of several miRNAs. This may serve as a buffer against the accumulation of deleterious mutations at other schizophrenia risk loci. We have generated human neural stem cells (NSCs) from six individuals with the 22q11.2 deletion and schizophrenia from iPSCs using Neural Induction Media (Life Technologies). In order to ensure the presence of the 22q11.2 deletion, a FISH probe for 22q11.2 (TUPLE) was used (Cell Line Genetics). DGCR8 was quantified using the 7900HT Fast Real-Time PCR System (Applied Biosystems) along with Taqman Copy Number Assays to ensure DGCR8 reduction in 22q11.2 DS NSCs relative to otherwise healthy control NSCs. Unexpected chromosomal abnormalities associated with iPSC culture methods have made routine detection of the deletion essential. We have characterized disruptions to the miRNA regulatory network in the NSC lines using Taqman Microarray Cards Version 3.0 and Illumina Small RNA-Seq. Here we will present the miRNAs that we have identified to be differentially expressed in otherwise healthy control NSCs versus NSCs with 22q11.2 DS and schizophrenia. Our data shows a subset of miRNAs that have significantly altered expression patterns consistently between the arrays and sequencing results. This list has further been refined to a list of 25 miRNAs previously implicated in having roles in neurological disorders or development including the very well characterized miR-185. We predict that these miRNAs could be involved in the elevated risk of developing schizophrenia in individuals with 22q11.2 deletion syndrome versus the general population. Future studies will focus on further characterizing miRNAs found to be differentially expressed as a result of the 22q11.2 deletion that may be important for schizophrenia. .

1738W

Modeling limb-girdle muscular dystrophy with in vitro human skeletal myotubes. *J.L. Marshall[1,2], Q. Wang[1,2], N. Habib[1,3,4], I. Avraham-Davidi[1], L. Caplan[1], F. Barthelemy[5,6], Y. Li[1,3,4], I. Armean[1,2], N. Khanlou[7], D.A. O'Connell[1,2], E.D. Douine[6,8], B.B. Cummings[1,2,9], F. Zhang[1,3,4,10,11], M.C. Miceli[5,6], J. Wanagat[12], S.F. Nelson[6,7,8,13], A. Regev[1,14], D.G. MacArthur[1,2,15].* 1) Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 2) The Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) McGovern Institute of Brain Research, MIT, Cambridge, MA, USA; 5) Microbiology Immunology and Molecular Genetics, University of California, Los Angeles, CA, USA; 6) Center for Duchenne Muscular Dystrophy, University of California, Los Angeles, CA, USA; 7) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA, USA; 8) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 9) Harvard Medical School, Program in Biomedical and Biological Sciences, Boston, MA, USA; 10) Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA; 11) Department of Biological Engineering, MIT, Cambridge, MA, USA; 12) Division of Geriatrics, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 13) Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, CA, USA; 14) Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, MIT, Cambridge, MA, USA; 15) Center for Mendelian Genomics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

The limb-girdle muscular dystrophies (LGMD) are a heterogeneous group of incurable muscle wasting disorders. LGMD is caused by about 30 different genes, whose encoded proteins function in diverse, poorly understood pathways critical for normal muscle function. Animal models of muscular dystrophy have been studied for decades but do not faithfully replicate human disease, making *in vitro* models of human skeletal muscle critical for the development of better therapies. Here we describe the development of *in vitro* models of skeletal myotubes to confirm the diagnosis of novel disease genes, facilitate mechanistic understanding of the downstream cellular processes leading to muscular dystrophy, and to develop screening assays to identify therapeutics. We first systematically compared three different protocols for modeling human *in vitro* skeletal myotubes: 1. Transdifferentiation of fibroblasts via MyoD overexpression, 2. Small molecule based 50 day differentiation of human pluripotent stem cells, and 3. Primary human myotubes differentiated from myoblasts. To determine which experimental protocol results in the most mature skeletal myotubes we performed immunofluorescence imaging-based computational analyses, transcriptional profiling, and immunoblotting. We will perform single nucleus RNA sequencing (sNuc-seq) to directly assess and compare the heterogeneity of cell populations within each culture system and the progression towards mature muscle gene expression. We intend to use the best *in vitro* model to systematically characterize the transcriptional and phenotypic changes in LGMD patients across several disease-causing genes, which will illuminate screening targets for identifying therapeutic agents and can also provide a functional validation for novel genes discovered in NGS studies. Ultimately, we will expand this pipeline to all identified disease-associated alleles.

1739T

The chr14.232.a pseudogene in the 14q21.2 region regulates the expression of the contiguous *LRFN5* gene and is deleted in a patient with autism. *G. Cappucio[1,2], M. Alagia[1], S. Attanasio[2], R. Borzone[2], R. Genesio[3], A. Mormile[3], L. Nitsch[3], E. Del Giudice[1], N. Brunetti-Pierri[1,2].* 1) Department of Translational Medicine, Federico II University, Naples, Naples, Italy; 2) Telethon Institute of Genetics and Medicine, Pozzuoli, Naples, Italy; 3) Department of Molecular Medicine and Medical Biotechnology, Federico II University, Naples, Italy.

Autism spectrum disorder (ASD) is a neurodevelopmental disorders with impaired social communication and is frequently associated with intellectual disability (ID). Recent advances have led to the identification of causative mutations in up to 30% of children with ASD. We report a 16 year-old boy with ASD bearing a microdeletion at chromosome 14q21.2 inherited from the father who has borderline cognitive impairment. The deletion affects a '*gene desert*' and *LRFN5* is the closest non-deleted gene. *LRFN5* encodes a protein involved in synaptic plasticity that has been heavily implicated in neurobehavioral phenotypes. We hypothesized a positional effect of the pseudogene chr14.232.a included within the deleted region that is predicted to bind the miRNAs targeting the 3'-UTR of the *LRFN5* gene. In agreement with this hypothesis, we found decreased expression of both *LRFN5* gene and chr14.232.a pseudogene in the proband's fibroblasts compared to controls. Interestingly, transfection of the chr14.232.a pseudogene in the patient's fibroblasts resulted in increased *LRFN5* expression. Transfection of pseudogenes chr14.232.a bearing a mutated miRNA binding site resulted in milder increased of *LRFN5* expression. Taken together, these data suggest that the chr14.232.a pseudogene functions as a miRNA decoy to regulate *LRFN5* expression through sequestration of *LRFN5* specific miRNAs. In conclusion, this study may unravel a novel mechanism of gene regulation involved in neurodevelopmental disorders.

1740F

Modeling the effects of autism-related *TBR1 de novo* mutations on human cortical development. *B.A. DeRosa, R.A. Barnard, S.A. Evans, L.A. Wourms, B.J. O'Roak. Molecular & Medical Genetics, Oregon Health & Science University, Portland, OR.*

   Recent studies have identified many important brain development genes that are recurrently disrupted by *de novo* mutations in children with autism. Furthermore, many of these, now high confidence, autism risk genes are co-expressed as a network during midfetal human corticogenesis, suggesting a developmental place and time where at least one common autism pathophysiology emerges. Understanding how single mutations in these diverse risk genes alter this network is key to advancing our knowledge of the molecular mechanisms underlying autism. To this end, patient-specific induced pluripotent stem cells (iPSCs) provide a unique opportunity to examine the impact of specific mutations on cortical development*,* as the *in vitro* differentiation of iPSCs into cortical neurons has been shown to mimic the natural sequence of corticogenesis. We have implemented a method that utilizes iPSCs and mini-bioreactors to generate 3D forebrain-like organoids. These organoids recapitulate cortical lamination patterns and, transcriptionally, closely matches human early and midfetal periods (Qian *et al*. 2016). We have modified the protocol to incorporate AggreWell800 plates to further increase the uniformity of the organoids by seeding ~10,000 cells into individual microwells. We have completed differentiations at 30, 50, and 70 days *in vitro*. Although we see some expected variability in the organization of organoids (e.g. number of ventricular zone (VZ)-like regions per organoid), similar to Qian *et al*., we observed strong and consistent expression of forebrain, VZ, subventricular zone (SVZ), and deep layer neurons markers; and similar relative thickness of the VZ. Using this approach, we are modeling the effect of three *de novo* mutations in *TBR1*. Experiments are also being replicated in lines where CRISPR/Cas9 genome editing has been used to correct mutations in patient iPSCs. *TBR1* encodes a transcription factor that serves as a "master regulator" of cortical development. Additionally, it has now been demonstrated that the TBR1-regulated network contains ~1/3 of high-confidence autism risk genes (Chuang *et al*. 2015). Specifically, we are examining how these mutations effect the acquisition of cellular identity in deep layer cortical neurons and the migration of specific neuronal subtypes to their proper location in the cortex. These studies will revolutionize our understanding of the impact of autism-related mutations on cortical development.

1741W

Does lack of X-inactivation for *SLC6A14* explain the very high male/female ratio in nonsyndromic autism? *F.R. Jimenez-Rondan, A. Ye, Q. Sun, K. Mohan, J. Ge, A.L. Beaudet. Molecular and Human genetics, Baylor College of Medicine, Texas Children' Hospital, Houston, TX.*

   Background: We have proposed a hypothesis that brain carnitine deficiency can cause autism with a non-Mendelian, non-dysmorphic (NoMeND) phenotype characterized by an extremely high male/female ratio, normal physical examination, normal structural brain imaging, and onset between 6 and 36 months of age (*BioEssays*, in press). We have searched for a carnitine-related gene that might explain the high male/female ratio in NoMeND autism. The *SLC6A14* gene in on the X-chromosome, encodes a blood-brain barrier carnitine transporter, and lacks the epigenetic hallmarks of random X-inactivation, suggesting that transport of carnitine across the BBB might be more effective in healthy female mice or humans compared to healthy male counterparts based on higher expression in females. **Methods:** DNA methylation across the *SLC6A14* gene was analyzed in human postmortem brain using a 450K Illumina methylation array. *Slc6a14* null mice from the Knockout Mouse Project (KOMP) and wild-type mice were bred on the C57BL/6NJ background. L-[N-methyl-^{14}C]-carnitine hydrochloride was injected by tail vein in young adult mice (6-10 week old) and mice were sacrificed 2 to 6 hours later and radioactivity was quantified in the brain. Free carnitine, γ-butyrobetaine, and trimethyllysine were measured in plasma and breast milk of wild-type and null mice. **Results:** There is no CpG island or differential DNA methylation of the active and inactive X chromosome for the *SLC6A14* gene in human brain. Genome browser data indicate that the same is true for all mammalian species for which data are available. After tail vein injection, transport of [^{14}C]-carnitine across the BBB into cerebellum is 35% greater in wild-type female mice compared to male mice likely due to lack of X-inactivation of *Slc6a14*. Findings are similar in brain excluding cerebellum. In *Slc6a14* null mutants compared to wild-type, transport of [^{14}C]-carnitine across the BBB is reduced by 68% in females and by 38% in males. The carnitine content of breast milk is substantially reduced in *Slc6a14* null mothers compared to wild-type. **Conclusion:** We propose that the extremely high male/female ratio in milder, nonsyndromic autism may be mediated by *SLC6A14* via susceptibility to autism caused by brain carnitine deficiency. If the brain carnitine hypothesis is correct, it is possible that 10-20% of all autism and a larger fraction of NoMeND autism could be prevented by fortification of infant diets with carnitine.

1742T

Further evidence supporting the involvement of *ERC1* gene variation in ASD.
S. Raskin[1], C.C.V. da Silva-Camargo[1], P. Liberalesso[2], V.S. Sotomaior[1]. 1) Group for Advanced Molecular Investigation (NIMA), Pontifícia Universidade Católica do Paraná, Curitiba, PARANÁ, Brazil; 2) Department of Neuropediatrics - Hospital Pequeno Principe - Curitiba - PR - Brazil.

Autism spectrum disorder (ASD) is reported in about 1% of children, with the highest heritability among common neuropsychiatric disorders. Chromosomal copy number variations (CNVs), that cannot be detected by conventional cytogenetic analysis, significantly contribute to the development of ASD, and new candidate genes have been associated to this condition through microarray-based comparative genomic hybridization (aCGH) analysis and Whole-Exome and Genome sequencing. We analyze two patients by aCGH and fluorescence in situ hybridization (FISH) confirmation, both patients have an interstitial heterozygous microduplication in 12p13.33. One is a 19-year-old male with ASD, epilepsy, long face and large ears, whose duplication includes the entire *ERC1* and other 11 genes. The other patient has a small 389 kb *ERC1* duplication, a 3-year-old female with ASD, lack of eye contact since 5 months, speech and psychomotor development delay, attention deficit, stereotypies, anxiety, irritability, tendency to social isolation and lack of facial expression. This is the smallest duplication ever observed in this region associated with a phenotype, and resulting in a partial duplication that encompasses only the *ERC1* gene. CNVs in this gene are reported here in patients with ASD for the second and third time, as we previously published a case where a 8-year-old male patient had a 1.5 Mb terminal deletion at 12p13.33, including *ERC1* gene. *ERC1 is* related to neuronal functions, and like most genes related to ASD, encodes a member of a family of RIM-binding proteins, a synaptic factor that regulate neurotransmitter release. The large duplication involves 12 genes but *ERC1* is the only one already related to ASD, and the small involves only the *ERC1* gene. These three cases implicates *ERC1* as a new putative ASD gene, but further functional analysis of *ERC1* is needed to improve the understanding about its possible role as a new candidate gene in this disease.

1743F

Evaluation and co-expression of marker genes of cell types in brain.
R. Dai[1], C. Jiao[1], Y. Chen[1], C. Chen[1], C. Liu[1,2]. 1) The State Key Laboratory of Medical Genetics, Central South University, Changsha, Hunan, China; 2) Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA.

Purpose: We want to use different types of data to evaluate the specificity of marker genes of brain cells and unravel the co-expression of cell type-specific genes in psychiatric disorders. **Methods:** The transcriptome and proteomics data of brain cells from human and mouse**,** by RNA-Seq and microarray from acutely isolation cell and primary culture cell were collected. We collected 543 "putative" marker genes for ten cell types from literature, in situ hybridization databases and antibody companies. We defined general marker gene as their expressions in claimed target cells is the highest across all cell types tested. The expression difference for the rigorous marker gene defined as dividing the second highest expression level across all the other cell types by the expression level in target cell type. Parietal cortex tissue from human brain included schizophrenia, bipolar disorder, and control samples were used for Weighted Gene Co-Expression Network Analysis. **Results:** With the datasets collected and the criteria provided, we found that 44 general marker genes showed stable specificity across all data collected. That included 29 neuron markers, eight astrocyte markers, seven oligodendrocyte markers. The averaged correlation values of general marker genes between human and mouse, transcriptome and proteome, RNA-Seq and microarray, acutely isolation and primary culture were 0.50, 0.58, 0.51 and 0.43, respectively. According to the criterion for rigorous marker genes, 23 of the 44 marker genes showed more than two-fold changes in at least seven data sets evaluated. The most specific marker gene (FC=710.31) is *RELN*, a neuron marker gene. It is not clear whether the non-specificity of rest of more than 400 claimed marker genes is related to data quality or technical artifacts in the data we can use for the assessment. In the WGCNA analysis, we found general marker genes of astrocyte were significantly enriched (p<2.2e-16) in disease associated module. The enriched six astrocyte marker genes (*ALDH1L1, ALDOC, CLU, GJA1, SLC1A3, and SLC4A4*) showed co-expression with significant GWAS loci of schizophrenia and bipolar disorder in previous studies. **Conclusions:** We confirmed 44 genes to be cell type-specific in the brain while many other commonly-used marker genes will require additional studies to verify their specificity. Moreover, astrocyte marker genes enriched in disease-associated co-expression module with known GWAS loci for schizophrenia and bipolar disorder.

1744W

Stem cell models for studying the role of epigenetic machinery in abnormal neurogenesis. *N. Kommu, S. Saxena, P. Naik.* Biological Sciences, BITS Pilani Hyderabad Ca,pus, Hyderabad, India, India.

In addition to genetic mutations, the role of epigenetic abnormalities is widely appreciated in the development of disorders such as schizophrenia and autism. In support of this notion, elevated levels of DNMT1 and DNMT3A have been reported in post-mortem brain samples from patients with schizophrenia. However, the mechanisms by which elevated levels of DNMT1 or DNMT3 result in formation abnormal neurons is not clear. Towards this goal, we developed a transgenic mouse embryonic stem cell line that overexpresses DNMT1 and observed that neurons obtained from the mutant line shows abnormal branching of dendrites and elevated levels of NMDA receptor activity. To gain more insights into the abnormal effects of increased levels of DNMT1, we performed genome-wide methylation comparisons between wild-type and mutant ES cells to identify loci undergoing aberrant methylation in the mutant ES cells. Results of these experiments will be presented.

1745T

Changes of open chromatin regions reveal stage-specific transcriptional network dynamics in human iPSC-derived neurons. *W. Moy[1,5], S. Zhang[1,2,5], H. Zhang[1], H. McGowan[3], J. Shi[4], C. Leites[1], A.R. Sanders[1,2], P.V. Gejman[1,2], J. Duan[1,2].* 1) Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL 60201, USA; 2) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, IL 60637, USA; 3) Department of Neuroscience and Cell Biology and Child Health Institute of New Jersey, Rutgers University, New Brunswick, NJ 08901, USA; 4) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; 5) Equal Contribution.

The application of human induced pluripotent stem cells (iPSCs)-differentiated neurons has served as a promising model for gaining insights into the molecular and cellular mechanisms of genetic risk for mental disorders. Multiple factors determine the fate and trajectory of neuronal cell differentiation, of which transcriptional regulation plays a major role. Chromatin accessibility (openness) to transcription factors (TF) strongly influences gene transcription and cell differentiation. However, the dynamic changes of open chromatin and the TF networks during neuronal differentiation from iPSCs are poorly understood. Here, we performed a global mapping of open chromatin regions (OCRs) using the Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-Seq) at different cell stages of glutamatergic neuronal differentiation, examined the correlation of changes of open chromatin and mRNA abundances (assayed by RNA-Seq), and constructed the neuronal stage-specific TF networks through a genome-wide inference of TF-binding footprints in OCRs. We found that OCRs were highly dynamic during neuronal differentiation, and more importantly that OCR accessibility at core promoter regions was positively correlated with mRNA abundances at their respective cell stages. Furthermore, we found that the dynamic changes of OCRs during neuronal differentiation were accompanied by the binding events of cell stage-specific TFs. As expected, binding footprints of *OCT4* and *NANOG*, genes that encode pluripotent stem cell-specific markers, are most enriched in iPSCs. Intriguingly, binding footprints of *NEUROD1* and *NEUROG2*, the two TFs that have been reported to rapidly induce high efficient conversion of iPSCs to excitatory neurons, are among those most enriched in the relatively mature neurons. Further TF network analysis based on the inferred TF binding footprints in OCRs successfully identified core TF networks and their master regulators for different neuronal stages. Interestingly, both *NEUROD1* and *NEUROG2* are in the same core TF network specific to more mature neurons, suggesting their pivotal role in epigenetic control of neuronal differentiation and maturation. Our study provides novel insights into the epigenetic control of glutamatergic neurogenesis and reveals neuronal stage-specific TF networks, which may facilitate the iPSC modelling of mental disorders by generating more homogenous neuronal populations in an accelerated manner.

1746F

Identifying pathogenic genes associated with autism and other developmental disorders by in-depth analysis of chromosome microarray studies. *V. Goitia[1], R.H. Peretz[1], G. Bhoojhawon[1], V. Ragoonanan[1], C.H. Wang[2].* 1) Department of Pediatrics, Driscoll Children's Hospital, Corpus Christi, TX; 2) Department of Neurology, Driscoll Children's Hospital, Corpus Christi, TX.

Chromosome Microarray (CMA) is a diagnostic test that scans the entire genome to identify chromosomal abnormalities in patients suspected to have multigenic disorders such as congenital anomalies, autism, and other developmental disorders. However, interpreting results of CMA is often problematic due to the large genomic intervals involved and difficulty knowing whether any or all genes detected in these intervals are related to the pathogenesis of the clinical diagnosis. We previously analyzed the results of CMAs among patients at Driscoll Children's Hospital (DCH) and found the following: high positive report rates (56%) in children with both autism and congenital anomalies, hotspots of genomic variation on chromosomes 10, 15, and 16 for autism, and high correlation between size of genomic variation and autism severity. Our current research goal was to identify pathogenic genes associated with autism and other developmental disorders. We included 1069 CMAs that were obtained at DCH from 2008 to 2016. There was an overall positive rate of 26.5% that included all reports which deviated from normal male and normal female. However, a high proportion of these were reported as variants of unknown significance (VUS). To further determine the pathogenicity of these genomic variations, we conducted a genome search to identify genes within each interval of genomic variation for all patients with a positive report. Those genes isolated from the intervals that occurred in multiple individuals with the same clinical diagnosis were considered pathogenic. We isolated many pathogenic genes that were previously reported as VUS. Our findings confirm that a whole genome approach such as CMA, combined with detailed characterization of clinical phenotypes, may be the best way to identify the genetic basis of multigenic disorders such as autism and other developmental disorders.

1747W

Rare family with partial duplication in 7q11.23 link four genes associated with intellectual delay and autistic phenotypes. *J.R. Korenberg[1], L. Dai[1], M.C. Burback[1], A. Ramirez[1], J.S. Anderson[2], M.B. Prigge[3].* 1) Department of Pediatrics, Brain Institute, University of Utah, Salt Lake City, UT; 2) Department of Radiology, University of Utah, Salt Lake City, UT; 3) Department of Pediatrics, University of Utah, Salt Lake City, UT.

Parsing the gene(s) associated with human cognition and behaviors is critical yet remains scientifically elusive. Full and partial deletion or duplication of ~28 genes in 7q11.23, has been linked to Williams syndrome (WS) and Autism Spectrum Disorder (ASD) respectively, providing unique opportunities for linking single gene(s) to specific cognitive and behavioral phenotypes. In this report, we show a rare family (one female patient and her two sons) with partial duplication of only four genes (*STX1A, ABHD11, CLDN3, CLDN4*). All three patients present moderate intellectual and behavioral abnormities. The duplication region was determined using Illumina OMNI5 microarray analysis on DNA isolated from subject derived blood samples. Cognition and social behaviors were also measured with this family. Opposite to the cognitive deficits in WS (verbal IQ > performance IQ), the family presents moderate cognitive delays (IQ=77 to 94) particularly in speech and language development (verbal IQ < performance IQ). They also have mild motor delays, hyperactive distractible behavior and autistic phenotypes. One of the sons has a history of abnormal EEG suggestive a seizure disorder. Compared to a rare family with partial deletion of overlapping genes (from ABHD11 to RFC2), the family with partial duplication reported here show similar mild intellectual disabilities but less tendency to approach strangers as meassured by the Salk Institute Sociability Questionnaire (SISQ). This is the first report of partial duplication of genes in 7q11.23. The results implicate that these four genes are associated with cognitive development, mainly in speech and language, and that copy number variant may affect the social behaviors in human being. Future work will include brain structural and functional analyses using magnetic resonance imaging and integration of genetics, cognition and behaviors.

1748T

Behavioral characterization in a mouse model of Bohring–Opitz syndrome. *K. Walz, C. Abad, J. Pascual, J.I. Young.* Hussman Inst Human Genomics, Univ Miami, Miami, FL.

Bohring–Opitz syndrome (BOS) is a rare neurodevelopmental condition characterized by intrauterine growth restriction, feeding difficulties, failure to thrive, developmental delay and distinctive facial features and posture. *De novo* frameshift and nonsense mutations in the *additional sex combs like 1* (*ASXL1*) gene, have been detected in the majority of BOS patients. *ASXL1* encodes a polycomb group protein that assemble multi-protein complexes on DNA to mediate transcriptional repression. Somatic mutations in this gene are also associated with myelodysplastic syndromes and chronic myelomonocytic leukemia. It is not yet known whether BOS is caused by haploinsufficiency of *ASXL1* or by gain of function of truncated ASXL1. The objective of this study is to characterize phenotypically *Asxl1* heterozygous mutant mice (*Asxl1*+/-) to understand BOS pathogenesis. Adult male *Asxl1*+/- animals were evaluated with a set of tests oriented to characterize their neurodevelopmental phenotype. All procedures were approved by the IACUC at the Miller School of Medicine, UM. *Asxl1*+/- did not have an overt phenotype and were indistinguishable from wild type littermates when superficially observed. However, the open field test revealed significant hypoactivity and on the rotarod test we observed impaired performance. In addition, *Asxl1*+/- mice were less sociable as uncovered by the three chamber sociability test, and were deficient in the auditory-cued Pavlovian fear conditioning (despite having normal hearing). Thus, these data suggest perturbed locomotor abilities or alterations in motivation as well as abnormal social behaviors and learning and memory deficiencies. Analysis of expression of *ASXL1*, as detected by β-galactosidase activity from the *ASXL1* locus revealed that it is expressed throughout the brain. Our data demonstrates that *ASXL1* haploinsufficiency results in behavioral alterations and suggest that the heterozygous *Asxl1*+/- mouse model resembles behavioral aspects of BOS. Further studies of this mouse model will advance our understanding of the abnormal brain development underlying BOS □associated behavioral and cognitive symptoms. .

1749F

Multimodal MRI and DTI reveal common systems mechanism underlying Downs syndrome and Alzheimer's disease. *L. Dai[1], O. Abdullah[1], J.S. Anderson[2], M.B. Prigge[3], M.C. Burback[1], A. Ramirez[1], J.O. Edgin[4].* 1) Department of Pediatrics, Brain Institute, University of Utah, Salt Lake City, UT; 2) Department of Radiology, University of Utah, Salt Lake City, UT; 3) Department of Pediatrics, University of Utah, Salt Lake City, UT; 4) Department of Psychology, University of Arizona, Tucson, AZ.

Down syndrome (DS) is a major genetic imbalance causing brain disease throughout the lifespan; from intellectual disability in the young to premature aging and Alzheimer's disease (AD) in the old. The partial and full trisomy 21 provide a unique opportunity to link genes to brain architecture, cognitive and behavioural phenotypes. We report a high resolution multimodal map of brain structural Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI) in a cohort of 26 subjects with full trisomy 21 (age 22±6),17 age and gender matched controls (age 23±5) and five subjects with partial trisomy 21 (age 22±3). Results show decreased volume, increased cortical thickness and decreased surface area in the most significantly changed brain regions in DS. Fractional anisotropy (FA) also differs significantly in the external capsule, fornix, anterior limb of internal capsule, sagittal stratum, inferior and superior cerebellar peduncle, posterior corona radiate and uncinate fasciculus out of 48 white matter tracts (JHU DTI-based white matter atlases). Patterns of white matter changes in DS are highly overlapped with those in AD, particularly in the external capsule, corona radiate, fornix and uncinate fasciculus, suggesting critical white matter metrics disturbed in development also in early onset of AD in DS. Most decreases in FA result from increased radial diffusivity (RD), suggesting abnormal myelination in these white matter tracts. Our results also show a significant positive correlation (R=0.6, P<0.02) between CANTAB Paired Associates Learning test and FA, RD in the external capsule and uncinate fasciculus, indicating a neural basis of deficits of memory, executive function and processing speed in DS. The external capsule and uncinate fasciculus comprise the capsular division of the lateral cholinergic pathway originating in the nucleus basalis of Meynert, a region that frequently degenerates in AD. Further studies are needed to discern if FA reduction in cholinergic circuits including external capsule and uncinate fasciculus in DS is associated with the early onset of AD in DS. Finally, we demonstrate genetic dissection using partial trisomy 21 implicating critical gene(s) associated with brain substrates and cognitive and behavioural functions. Together, our results implicate a system mechanism for the genetic and neural basis of developmental deficits and early onset of AD in DS and provide insightful information for therapeutic treatments.

1750W

Comprehensive catalog of cell types in the developing brain using single-cell transcriptional profiling. *J.M. Simon[1,2,3], L. Loo[2,4], J. Niehaus[2,4], J.D. Welch[5], E.S. McCoy[2,4], M.J. Zylka[2,3,4].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Understanding complex processes such as differentiation, development, and disease often relies on characterizing the baseline transcriptional landscape of tissues or cells. These experiments are typically performed in bulk on mixed populations of cells or heterogeneous tissues and therefore molecular complexity can be lost as the signal represents the average of that population. The advent of single-cell sequencing has enabled the categorization and characterization of heterogeneous cell types in complex adult tissues, however, the cellular composition of the developing brain has not yet been elucidated in this manner. We have utilized Drop-seq to transcriptionally profile ~20,000 single cells from the cerebral cortex of embryonic and neonatal mice. Due to the plastic nature of the developing tissue and scarcity of certain cell types, existing methods to identify cell clusters failed to resolve mixed populations of distinct but similar cells into their respective independent groups. We therefore developed a new computational procedure that iteratively identifies and refines cell-type classifications to maximize robustness of each intrinsic cell type. Using this approach, we identified 22 distinct cell types in both the embryonic and neonatal cortex, including multiple sub-classes of excitatory and inhibitory neurons, radial glia, and other neuronal and glial precursors, many of which were developmental timepoint-specific. These cell types could be reliably distinguished by the combined expression of known and novel marker genes. Further, within specific cell types, we discovered sub-clusters of cells that corresponded to various cellular states, including phases of the cell cycle and transient activation. Together, these data will serve not only as a novel resource of the different cortical cell types and their molecular markers during development, but also as a platform on which to study numerous neurological diseases and etiology.

1751T

Cytoplasmic FMRP-Interacting Protein 2 (*CYFIP2*) causes syndromic intellectual disability. *A. Begtrup[1], K. McWalter[1], M.T. Cho[1], E. Heise[1], K.J. Wierenga[2], J. Chen[2], G. Purcarin[2], C.R. Finnila[3], G.M. Cooper[3], J. Screiber[4], M. Meltzer[5], R. Mostafavi[6], D. Goh[7], B. Baskin[1], M.J. Guillen Sacoto[1], K.G. Monaghan[1], J. Juusola[1].* 1) GeneDx, Inc., Gaithersburg, MD; 2) Oklahoma University Healthy Sciences Center, Oklahoma City, OK; 3) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 4) Children's National Medical Center, Washington, DC; 5) Children's National Healthy System and Pediatric Specialists of Virginia, Fairfax, VA; 6) Le Bonheur Children's Hospital, Memphis, TN; 7) National University Children's Medical Institute, Singapore.

Syndromic intellectual disability is a common indication for genetics consultation, and due to the large degree of genetic heterogeneity, a diagnosis is often made using exome and genome sequencing. During the past four and a half years our clinical diagnostic laboratory has performed exome sequencing on 10,692 probands with intellectual disability or developmental delay as the primary indication for clinical exome sequencing. Utilizing trio-based testing, we identified *de novo* variants in the candidate disease gene *CYFIP2* in six individuals submitted for clinical exome sequencing and an additional individual tested at an outside laboratory, connected via GeneMatcher. Two unrelated probands shared the same *de novo* variant, p.R87C, while a third case had a splice site variant, c.3594+1G>T. The remaining individuals had unique missense variants with damaging *in silico* predictors and were absent from ExAC and our internal database of unaffected individuals. CYFIP2, *C*ytoplasmic *F*MRP *I*nteracting *P*rotein 2, was previously identified in yeast-two hybrid screening to interact with FMRP, the protein responsible for fragile-X syndrome (FXS) (Schenck *et al.*, 2001). Mice heterozygous for a null *Cyfip2* allele have abnormal behavioral and dendritic spine phenotypes, suggesting a significant role for this novel gene in the pathogenesis of FXS (Han *et al.*, 2015). Common clinical features observed in probands in our cohort, ranging in age from 10 months to 11 years, included developmental delay (7/7), abnormal tone (6/7 with hypotonia, 3/7 with hypertonia in the extremities), seizures (5/7), intellectual disability (4/7), microcephaly (2/7), and gastrointestinal/growth issues (6/7). Two cases, one of whom harbors the splice site variant, have autism. Brain MRI results were normal for 4/7 probands, and revealed atrophy (2/7), and periventricular leukomalacia (1/7) in others. While further characterization of the phenotypes associated with *de novo CYFIP2* variants is needed, our study, combined with the functional link of this gene to FMRP, provides additional evidence that *CYFIP2* is a new disease-associated gene causing syndromic intellectual disability.

1752F

Truncating *de novo* mutations in *DLG4* responsible for intellectual disability with Marfanoid habitus. *S. Moutton[1,2], AL. Bruel[2], M. Assoum[2], M. Chevarin[2], E. Sarrazin[3], C. Goizet[4], AM. Guerrot[5], P. Charles[6], D. Héron[6], N. Houcinat[1,2], A. Vitobello[2], F. Tran-Mau-Them[2], C. Philippe[2], Y. Duffourd[2], C. Thauvin-Robinet[1,2], L. Faivre[1,2].* 1) Reference center for developmental anomalies, Dijon University Hospital, Dijon, France; 2) INSERM 1231, LNC UMR1231 GAD, Brugundy University, F-21000, Dijon; 3) Caribean reference center for rare neurological and neuromuscular diseases, Fort de France Universitary Hospital, Fort de France, France; 4) Reference center for developmental anomalies, Bordeaux University Hospital, Bordeaux, France; 5) Department of Genetics, Rouen University Hospital, Normandy Centre for Genomics and Personalized Medicine, Rouen, France; 6) Reference center for rare intellectual disability disorders, Pitié-Salpêtrière Hospital, Paris, France.

Marfanoid habitus (MH) combined with intellectual disability (ID) is a genetically and clinically heterogeneous group of overlapping disorders. The association of various chromosomal disorders and variations in genes responsible for Marfan syndrome explain no more than 20% of subjects, with examples of digenic inheritance. Different genes responsible for ID have occasionally been described in patients with marfanoid habitus. To identify new genes specific to MH with ID, we performed a combination of trio-based (33 subjects) or single proband (17 subjects) exome sequencing in 50 affected individuals, and searched for variations in non-OMIM genes in at least 2 unrelated patients of the cohort, with a compatible segregation in a given family. We identified 3 patients carrying a potential truncating variant (1 splicing and 2 frameshift variants) of the *DLG4* gene: c.1534+2T>C, c.1009_1016delTTTATCCT and c.1705delG according to NM_001128827.1 transcript. In 2 patients, the variant was *de novo*, whereas in one patient, the father was not available for testing but it was absent from the mother. The patients were 35, 21 and 23 years of age and shared mild to moderate ID and similar marfanoid features including long face, high arched palate, dolichostenomelia, long and thin fingers, pectus excavatum and scoliosis. They did not exhibit ophthalmological or cardiac manifestations. The *DLG4* gene encodes Psd-95, a protein expressed in various tissues including the brain. In neurons, Psd-95 is localized at the post-synaptic density, associated with NMDA receptor signaling. It was hypothesized to participate in glutamatergic synapse maturation. Depletion of DLG4 is thought to change the ratio of excitatory to inhibitory synapses in hippocampal neurons. The variants were absent from databases, the gene was found to be intolerant to loss-of-function variants and it was previously hypothesized to cause ID, as *de novo* truncating variants were significantly enriched in a series of 2,104 trio-based WES, but no clinical description was provided. By describing 3 additional patients, our study emphasizes the role of *DLG4* as a novel post-synaptic associated gene involved in ID. Further descriptions will be needed in order to conclude whether *DLG4* is specific to an ID-MH phenotype.

1753W

Effect of an intronic mutation in the CLIP1 Gene (CLIP-170) in a patient with autosomal recessive intellectual disability. *A. Rincon[1,2], J. Rojas[1,2], JC. Prieto[1,3], P. Ayala[1], A. Barreto[1], K. Prieto[1], C. Serrano[4], FS. Cabrera-Solano[4].* 1) Pontifical Xavierian University, Bogota, Colombia; 2) San Ignacio University Hospital; 3) Hospital la Victoria SDS, Bogotá, Colombia; 4) Gimnasio Velmont School.

The etiological diagnosis of intellectual disability (ID) remains a challenge. Approximately 80% of ID cases do not have an identified cause. Introduction of new high-throughput DNA sequencing techniques has discovered new genes involved in autosomal recessive ID (ARID) disorders. CLIP1 gene (OMIM 179838), located in 12q14, has been recently linked to ARID. CLIP1 codes for a cytoplasmic protein, member of the microtubule plus-end tracking protein family, associated with the ends of growing microtubules within the cytoskeleton of axons and neuronal dendrites. A previous study in consanguineous families found loss-of-function homozygous mutations in the coding region of CLIP1 in patients with ARID. Functional studies demonstrated absence of CLIP1-protein in patients with ARID. Therefore, CLIP1 was proposed as a novel causing gene for ARID. However, the effect of a homozygous intronic mutation on CLIP1-protein and its correlation with ID has not been analyzed. A patient with ARID, daughter of healthy consanguineous parents underwent genetic testing to identify the genetic cause of ID. A c.2418+1G>T intronic mutation in CLIP1 gene was identified. mRNA and CLIP1-protein expression was tested in fibroblasts and blood of the patient, her mother and two healthy controls. RT-PCR showed no difference in mRNA expression levels in the patient's cell lines compared to the mother and controls. Immunofluorescence analysis of fibroblasts demonstrated no morphological differences and no changes in the localization of the CLIP1-protein among the patient, the mother and controls. However, ROI (region of interest) analysis showed significant decrease of CLIP1 fluorescence in the patient compared to the mother, who had mild decrease in CLIP1 fluorescence respect to the control. These results suggested that CLIP1 mutation may cause a reduction in the amount of CLIP1-protein. Western Blot revealed a lighter protein in the patient compared to the mother and controls, confirming our hypothesis. Predicting models suggest that exon-skipping, caused by a mutation in the donor site of splicing in the intron 11 of CLIP1 gene, yields a shorter protein, affecting the coiled-coil region of CLIP1-protein. Abnormally spliced mRNA was subsequently confirmed by amplification of the patient's cDNA. Our data suggest that c.2418+1G>T intronic mutation in the CLIP1 gene produces an altered protein with inadequate function in microtubular terminations in neuronal cells, which can lead to ARID.

1754T

MeCP2 AT-hook1 mutations disrupt DNA binding and chromatin compaction in patients with intellectual disability and schizophrenia. *T.I. Sheikh[1,2], M. Ayub[3], J.B. Vincent[1,2,4].* 1) Molecular Neuropsychiatry & Development (MiND) Lab, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 3) Department of Psychiatry, Queen's University, Kingston, Ontario, K7L 3N6 Canada; 4) Department of Psychiatry, University of Toronto, Toronto, ON, Canada.

Mutations in the Methyl-CpG-binding protein-2 gene (*MECP2*) are most commonly associated with Rett syndrome in females. However, it has long been appreciated that there exists a broader spectrum of neuropsychiatric phenotypes associated with *MECP2* variants. Most Rett causing mutations are either missense mutations located in either the methyl-CpG-binding domain (MBD) or the transcription repression domain (TRD), or lead to premature truncation of the MECP2 protein. Clear clinical roles for mutations in other known domains such the intervening domain (ID) or in AT-hook domains have yet to be determined. Here, we report on a R190H *MECP2* missense mutation, located in AT-hook1 within the ID, segregating in a large Pakistani family with childhood onset cognitive decline and schizophrenia. We studied the effects on DNA binding and chromatin clustering of this mutation, as well as a second mutation, R190C, previously reported as a *de novo* mutation in a female subject with schizophrenia. We show that both mutations affect the ability of MECP2 to bind to AT-rich DNA as well as the brain-derived neurotrophic factor (*BDNF*) promoter, with the more drastic effects seen for R190C. Both mutations show similar effects on nuclear chromatin clustering. These data indicate a clear molecular link between *MECP2* AT-hook1 mutations and a major psychiatric disorder, and support a more general link between *MECP2* mutations and psychosis.

1755F

De novo IRF2BPL pathogenic variants cause severe precocious neurodegenerative disease. *F. Tran Mau-Them[1,2,3], A. Vitobello[1,3], L. Duplomb[3], K. Lindstrom[4], I. Marey[5], R. Spillmann[6], MJ. van den Boogaard[7], N. Stong[8], C. Nava[5], B. Keren[5], A. Masurel[2], T. Jouan[3], F.E. Jansen[9], Y. Duffourd[3], K. van Gassen[7], C. Pilippe[1,2], L. Faivre[2,3], K. Riley[6], J. Friedman[10], L.D.M. Pena[6], C. Thauvin-Robinet[1,2,3].* 1) UF Innovation en diagnostic génomique des maladies rares, Laboratoire de Cytogénétique-Génétique Moléculaire, CHU Dijon Bourgogne, Dijon, France; 2) Centre de Référence maladies rares « Anomalies du Développement et syndrome malformatifs » de l'Est et Centre de Génétique, Hôpital d'Enfants, FHU TRANSAD, CHU Dijon Bourgogne, Dijon, France; 3) Inserm UMR1231 GAD, F-21000, Dijon, France; 4) Division of Genetics and Metabolism, Phoenix Children's Hospital, Arizona, USA; 5) Département de Génétique, Hôpital Pitié-Salpêtrière 47-83 Boulevard de l'Hôpital 75013 Paris, France; 6) Department of Pediatrics/Division of Medical Genetics, Duke University Medical Center, North Carolina, USA; 7) Department of Genetics, University Medical Center Utrecht, The Netherlands; 8) Institute of Genomic Medicine, Columbia University, New York, N.Y; 9) Department of Child Neurology, Brain Center Rudolf Magnus, University Medical Center, Utrecht, The Netherlands; 10) Department of Neurosciences and Pediatrics UCSD/Rady Children's Hospital San Diego, Rady Children's Institute for Genomic Medicine, California, USA.

Thanks to whole exome sequencing and international datasharing, we report six individuals with *de novo* heterozygous frameshift or nonsense variants and one additional patient with a *de novo* heterozygous missense variant in the *IRF2BPL* (Interferon Regulatory Factor 2 Binding Protein-Like) intronless gene, encoding a zinc finger/RING finger transcriptional modulator with a predicted ubiquitin ligase activity. The majority of these variants are expected to result in a truncated protein missing the C-terminal RING finger domain. The cases with *IRF2BPL* truncating variants exhibit a progressive neurodegenerative disorder with infantile onset, associated with severe intellectual disability and epilepsy in more than half individuals. Five of seven cases have abnormalities on EEG irrespective of clinical seizures. The function of the *IRF2BPL* gene product is currently unknown, but the gene appears to be highly intolerant to loss of function variants (pLI=0.97), with no nonsense variant in gnomAD. Statistical calculation of the probability that five of the cases had mutations in *IRF2BPL* due to chance had p<0.05. Neurodegenerative diseases (NDDs) are a heterogeneous group of progressive disorders characterized by dysfunction and loss of defined neuronal populations, usually associated with additional neurological features, such as seizures, ataxia and microcephaly. NDDs are triggered by abnormal or misfolded proteins that alter the physiological function of neuronal and or glial populations. Aberrant ubiquitination or degradation of ubiquitinated proteins through the proteasome, as well as epigenetic control of gene expression through transcription factors recruitment, seem to play a relevant role in the molecular pathways leading to NDDs. Functional studies of *IRF2BPL* are ongoing to better characterize the involvment of *IRF2BPL* into neurodevelopment. This clinical cohort of individuals has a number of overlapping symptoms, the majority with *IRF2BPL* truncating mutations. We propose a role for this protein in a new neurodevelopmental disorder.

1756W

DDX3X: Robust phenotype-genotype correlations from recurrent de novo mutations in DDX3X in patients with global developmental delay and intellectual disability. *R. Jiang[1], L. Suit[1], S. Floor[2], B. Fregeau[1], A.J. Barkovich[3], E.H. Sherr[1].* 1) Department of Neurology, University of California San Francisco, San Francisco , CA; 2) Department of Cell and Tissue Biology, Helen Diller Family Comprehensive Cancer Center, UCSF, San Francisco, CA; 3) Department of Radiology, University of California, San Francisco, San Francisco, CA.

DEAD-box helicase 3 X-linked (DDX3X), is an RNA helicase that regulates translation and other aspects of RNA biology. Mutations in DDX3X have been reported in 1-3% of intellectual disability cases in females, yet critical unanswered questions about the relationship between these mutations and disease pathogenesis remain. Here we demonstrate that there are many recurrent de novo mutations in DDX3X, using the largest cohort of DDX3X patients assembled to date (n=55). These mutations correspond to consistent clinical and imaging phenotypes that are linked to distinctive biochemical abnormalities. Of the 55 patients, 53 are female (1-22 years old) and 2 male. Of the 51 distinct mutations in DDX3X this cohort, 7 are recurrent and 5 are distinctive mutations but occur at the same amino acid. Clinically, this cohort of 55 patients shows consistent and significant findings including mild (32/50, 64%) to severe (18/50, 36%) global developmental delay, and neurological abnormalities such as microcephaly (17/45, 38%), epilepsy (12/45, 27%), and hypotonia (41/45, 91%). Brain MRI findings were notable for a thin corpus callosum (55%) and enlarged ventricles (79%). A smaller but significant subset, 15%, have polymicrogyria (PMG). The high preponderance of recurrent mutations led us to ask whether these mutations have consistent clinical and imaging characteristics. We found that two recurrent de novo mutations, R326H and T532M, are consistently associated with a more severe clinical phenotype and with PMG as seen by MRI. Whereas, the mutation R376C is found recurrently in 4 mildly affected patients with less impaired MRI scans. We asked whether this range of clinical & imaging severity correlated with the biochemical activity of DDX3X *in vitro*. We employed two widely used assays, one that quantifies the rate of unwinding of dsRNA and one that measures ATP hydrolysis as a function of RNA concentration. We found that for reference alleles, full unwinding occurred by 1-2 min with a V_{max} 0.27 uM±.076 μM ATP/min; mutations that were associated with a milder phenotype (e.g. R376C) had an intermediate helicase unwinding (40 min) and V_{max} (0.06±.0045 μM ATP/min); mutations that resulted in PMG and agenesis of the corpus callosum had no activity in both assays. These findings demonstrate that functional DDX3X is essential for normal brain development, that there are many recurrent mutation hotspots, and that the site of the mutation has a clear impact on degree of pathogenesis.

1757T

Regulating transcriptional activity by phosphorylation of the intellectual disability and seizure associated ARX homeodomain transcription factor. *C. Shoubridge[1,2], M.H. Tan[1], O. Dearsley[1], C.S. Hii[3], J. Gecz[1,2], T. Mattiske[1,2].* 1) Intellectual Disability Research, Adelaide School of Medicine, University of Adelaide, Adelaide, Australia; 2) Robinson Research Institute, University of Adelaide, Adelaide, Australia; 3) Department of Immunopathology, SA Pathology, South Australia, Australia.

Aristaless-related homeobox (*ARX*) gene encodes a paired-type homeodomain transcription factor with critical roles in forebrain, pancreas and testes development. Mutations in *ARX* give rise to intellectual disability, epilepsy and brain malformations. Here we have identified that ARX protein is phosphorylated and using mass spectrometry and *in vitro* kinase assays we identify serine at position 37, 67 and 174 as sites of post-translational modification. We demonstrate that phosphorylation is required for correct transcriptional activity of the ARX protein. We chose a transcriptome wide approach to analyse gene expression using phosphoablative mutants (alanines replacing serines) compared to ARX wild-type (ARX-WT) overexpressed in pancreatic alpha TC cells. When ARX-WT was overexpressed compared to non-transfected cells, 70 genes had significantly altered expression (Log2FC >+/-1.0, P-value <0.05). There was a loss of significantly regulated gene expression compared to non-transfected cells with overexpression of the double phosphoablative mutant Ser37Ala+Ser67Ala (26%) and Ser174Ala (39%), respectively. The Ser174Ala mutant significantly altered expression of an additional 45 genes (82% with reduced expression) that were not impacted by ARX-WT overexpression, including growth factors (*Hdgfrp3* and *Igf2*), transporters (*Slc2a4*), kinases and signalling molecules as well as cell proliferation and cytoskeletal reorganisation (*Abi3*, *Ablim2*). Using yeast-2-hybrid (confirmed by CoIP) we identified a novel ARX-interacting protein, PICK1 (Protein interacting with C kinase 1) binding with the C-terminal region of ARX. PICK1 is a scaffold protein known to facilitate phosphorylation of protein partners by protein kinase C alpha (PKC-alpha). We confirm that ARX is phosphorylated by PKC-alpha (*in vitro* kinase assays and specific kinase inhibitor) and in particular, demonstrate this kinase is responsible for phosphorylation at serine 174. In conclusion, we show that ARX is phosphorylated at several sites and that this modification is important for aspects of transcriptional activity. Phosphorylation at serine 174 occurs via PKC-alpha suggesting the binding of the specific protein partner PICK1 with the C-terminal region of ARX is required. We contend that missense mutations in the C-terminal region of *ARX* that cause severe early onset seizure phenotypes in affected children may be due to deficient phosphorylation and subsequent alterations in transcriptional capacity.

1758F

Molecular and biochemical analyses to understand the genotype-phenotype correlation in patients with the maternally inherited MELAS disorder. *A. Gropman[1], M. Uittenbogaard[2], C.A. Brantner[3], A. Chiaramello[2].* 1) Department of Neurology, Children's National Med Center, Washington, DC; 2) Department of Anatomy and Regenerative Medicine, George Washington University School of Medicine and Health Sciences, Washington, DC; 3) George Washington University, Nanofabrication and Imaging Center, Washington, DC.

Mitochondrial respiratory disorders are characterized by a chronic energy deficit due to mutations in the nuclear or mitochondrial genome affecting the oxidative phosphorylation (OXPHOS) system responsible for ATP synthesis. The most common mitochondrial respiratory disorder is MELAS (Mitochondrial Encephalopathy Lactic Acidosis with Stroke-like episodes). MELAS is an intractable childhood-onset progressive neurodegenerative disease that results in devastating multi-organ failure, strokes, and premature death. This orphan disease is mainly caused by a maternally inherited mutation (A3243G) in the mitochondrial-encoded tRNA$^{Leu/UUR}$ gene, which only affects a subset of the multi-copy mitochondrial genome. Thus, MELAS cells harbor a mixed population of functional (healthy) and dysfunctional (diseased) mitochondria, a state known as heteroplasmy. Patients harboring the MELAS mutation become symptomatic only when the mutant load of diseased mitochondria exceeds a certain threshold. However, the heteroplasmic load is not the sole determinant of the phenotypic variability among MELAS patients, as the nuclear background influences clinical manifestations, pathogenesis, and severity of the MELAS disorder. As a first step toward understanding the genotype-phenotype correlation, we recruited a cohort of eight MELAS patients that were genetically diagnosed with the A3243G MELAS mutation and underwent a skin biopsy from which we derived dermal fibroblasts. We performed an in-depth mitochondrial morphometric analysis by transmission electron microscopy combined with functional mitochondrial bioenergetics studies. Although the patients harbor the same MELAS mutation, they exhibited a heterogeneous mixture of abnormal mitochondrial ultrastructural features in terms of shape, number of cristae, abnormal cristae junctions and altered mitochondrial dynamics. Furthermore, the overall number of mitochondria varies among the MELAS patients. Finally, our mitochondrial bioenergetics studies revealed a heterogeneity in terms of their respiratory capacity, bioenergetics reserve and capacity to respond to stress. Our collective results demonstrate the functional impact of the nuclear background on the molecular pathogenesis of the MELAS disorder.

1759W

A recessive variant in forkhead box domain of *FOXF2* is associated with profound hearing loss and inner ear anomaly. *G. Bademci[1], O. Diaz Horta[1], A. Incesulu[2], F.B. Cengiz[1], C. Abad[1], C.J. Sineni[3], J. Foster II[1], D. Duman[4], S. Fitoz[5], S.H. Blanton[1,6,7], Z. Lu[3], K. Walz[1], M. Tekin[1,6,7].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, FL; 2) Department of Otolaryngology, Head and Neck Surgery, Eskisehir Osmangazi University, Eskisehir, Turkey; 3) Department of Biology, University of Miami, Miami, FL; 4) Division of Pediatric Genetics, Ankara University School of Medicine, Ankara, Turkey; 5) Department of Radiology, Ankara University School of Medicine, Ankara, Turkey; 6) Department of Otolaryngology, Miller School of Medicine, University of Miami, Miami, FL; 7) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, FL.

While astonishing progress has been made in identifying causative genes for hereditary hearing loss (HL), molecular mechanisms leading to inner ear anomalies in humans remain largely unknown. As part of our ongoing study of hearing loss associated with inner ear anomalies, we ascertained a Turkish child with consanguineous parents who was born with profound sensorineural HL and a cochlear anomaly (incomplete partition type I). Using whole exome sequencing, we identified a homozygous *FOXF2* c.325A>T (p.Ile109Phe) variant. This variant is not present in public databases or in over 1,000 Turkish controls. *FOXF2* is a member of the FOX protein family of transcription factors that contain a conserved forkhead box sequence of ~100 amino acids. The FOX domain forms a motif that binds to DNA to regulate the expression of genes mainly involved in embryogenesis. p.Ile109 is a highly-conserved residue of the FOX domain of FOXF2. By comparing the quantitative expression of wild type and mutant proteins in a mammalian cell line, we show that the abundance of variant FOXF2 is reduced, which suggests that the variant decreases the stability of the protein. Pathway analysis in RNA-Seq data from the HEK293 cell line shows that knocking down *FOXF2* leads to differential expression of genes involved in DNA methylation, histone deacetylation, and RNA Polymerase I Promoter Opening, which supports its role in embryonic gene regulation. Foxf2 is expressed in the cochlea of developing and adult mice and is localized to the nucleus of hair cells as well as Reissner's membrane cells, suggesting a role in hearing. We conclude that FOXF2 plays a major role in the developing cochlea, and its dysfunction leads to an isolated inner ear anomaly and deafness.

1760T

Custom capture high-throughput sequencing for mutation detection: Results from 217 coloboma subjects across 196 genes identifies novel mutations in genes associated with ocular coloboma. *V.K. Kalaskar[1], R.P. Alur[1], L.A. Li[2], R.B. Hufnagel[1], T. Cogliati[1], B.P. Brooks[1].* 1) National Eye Institute, NIH, Bethesda, MD; 2) David Geffen School of Medicine at UCLA, Los Angeles, CA.

Statement of Purpose: Uveal coloboma is a potentially blinding congenital ocular malformation, accounting for up to 10% of childhood blindness and caused by failure of the optic fissure to close during the fifth week of human gestation. Mutations in several genes have been reported, however, the total yield of mutation detection across known coloboma-associated genes has not been well-established. In this study, we report on novel mutations in known coloboma-associated genes in a large cohort of patients with uveal coloboma and on zebrafish and cell culture studies on a novel mutation in *RARB* gene. **Methods Used:** Custom capture sequencing of 196 genes from 217 study subjects from 66 families was performed on Illumina HiSeq2000 platform. Data were analyzed and variants filtered using Varsifter and confirmed by Sanger sequencing. Mutagenesis and Western blotting were employed to study the effect of mutation on RARB protein formation in cell culture. Zebrafish morpholino studies were performed on *rarga* gene, which is the nearest homologue for human *RARB*. For rescue experiments, wild type and mutant human *RARB* mRNA were used. **Summary of Results:** Custom capture sequencing identified potential variants in *BMP7* and *RARB* (missense), *TFAP2A* (frameshift) and *CHD7* (nonsense) in 4 unrelated families. These variants were predicted deleterious and not reported in HGMD, ClinVar and in population databases such as 1000 genomes, EVS and ExAC. Sanger sequencing segregated these mutations in the affected individuals. The RARB mutation was located in the conserved DNA binding domain and predicted to result in p.Arg114Gln substitution. Preliminary studies revealed higher protein levels in cells transfected with *RARB*-mutant compared to *RARB* expression constructs as detected by Western blotting with DYK-Flag and RARB antibodies. Moreover, morpholino knockdown of *rarga* gene resulted in a coloboma phenotype in zebrafish embryos and rescue experiments suggests that *RARB*-mutant mRNA may rescue the phenotype better compared to *RARB* mRNA. While further experiments are needed, our study suggests that *RARB* mutation resulted in altered protein levels causing ocular coloboma consistent with previous findings. We conclude that custom capture high-throughput sequencing can identify novel mutations in known genes, however, extensive methods such as whole exome or genome sequencing can better contribute to the identification of novel genes with mutations in individuals with coloboma.

1761F

Characterizations of *NMNAT1* mutants and mouse model of *NMNAT1*-LCA. *X. Feng[1], SY. Ji[2], C. Weng[3], JH. Sun[1], XW. Cao[4], J. Lin[1], S. Han[1], ZJ. Xia[1], HY. Fan[2], M. Qi[1].* 1) Department of Cell Biology and Medical Genetics, Zhejiang university,Zhe Jiang, China; 2) Life Sciences Institute and Innovation Center for Cell Signaling Network, Zhe Jiang university,Zhe Jiang, China; 3) Department of Genetics and Genomics, Case Western Reserve University, Cleveland, USA; 4) School of Optometry and Ophthalmology and Eye Hospital, Wenzhou Medical College, Zhe Jiang, China.

Background: The nuclear nicotinamide mononucleotide adenyltransferase *NMNAT1* catalyzes the formation of NAD^+ from nicotinamide mononucleotide (NMN) and ATP, and acts as a chaperone that protects against neuronal degeneration. We previously reported *NMNAT1* mutations cause a serve dystrophy of the retina-Leber congenital amaurosis 9 (LCA9) because of its central role in metabolism. **Methods:** In this study, LCA-associated *NMNAT1* mutants were tagged with GFP on N-terminus. HELA cells were transfected with recombinant plasmids, and used to confirmed cell localizations of wildtype and mutant *NMNAT1*. Enzymatic activities of wildtype and mutant *NMNAT1* were tested *in vivo* and *in vitro* experiments. Circular dischroism (CD) was used to monitor the secondary structure and stability of wildtype and mutant *NMNAT1*. In addition, CRISPR/Cas9 was used to construct the *NMNAT1*[E257K]-LCA9 mouse model. Electroretinogram (ERG), fundus examination and optical coherence tomography (OCT) were used to analyze its phenotype. **Results:** Two nonsense mutations alter the nuclear localization but no missense mutants disrupt the normal nuclear localization of the protein. Glu257Lys and Asn273Asp mutants have relatively normal enzymatic. Met35Thr, Val98Gly, Val151Phe and Leu153Val reduce the catalytic activity of NMNAT1, but only Val98Gly reduce the production of NAD^+ significantly. In contrast, the secondary structure of *NMNAT1* mutants was relatively less stable. Moreover, E257K/E257K mouse appear pigmentary retinopathy around 2 months, and retinal function deteriorates with increasing age. **Conclusion:** These results suggest that prolonged insufficient volume of NAD^+ can result in LCA9 but cannot account for all the pathogenesis, and the pathogenesis of most common mutation Glu257Lys needs further research.

1762W

Novel genes associated with optic nerve hypoplasia in 6 family trios: A clinical and exome study. *P. Bitoun[1,2], A. Boland Auge[3], M. Bacq-Daian[3], E. Pipiras[4], B. Benzacken[4], S. Kuzbari[5], V. Renault[6], B. Parfait[7], JF. Deleuze[3].* 1) Medical Genetics, SIDVA 91, Juvisy s/o France, IDF, France; 2) Génétique Médicale, Groupe Médical Jarente, Paris, France; 3) Centre National de Genotypage, Institut de Genomique, CEA, Evry, France; 4) APHP Laboratoire d' Embryo Cyto Genetique et Biologie de la Reproduction, Hopital Jean Verdier , Bondy, Université Paris Nord 13, Sorbonne Paris Cité, UFR SMBH, Bobigny; Inserm, U1141. France; 5) APHP Banque d'ADN , Hôpital Robert Debre, Paris, France; 6) CEPH, Hôpital Saint Louis, Paris, France; 7) Centre de Ressources Biologiques, Hôpital Cochin, APHP, Hôpitaux Universitaires Paris Centre, Paris, France.

Purpose Optic nerve hypoplasia (ONH) is a rare blinding malformation, ONH is often syndromic with MRI pituitary anomalies. We explore genetic causes of isolated ONH Material and Methods: Probands (2F/4M) 0.5-6 yrs with normal pituitary MRI were enrolled after consent with Pediatric, dysmorphologic and ophthalmologic endocrine examinations SNP aCGH, WES and Sanger validation sequencing. Ophth. exam: VEP, ERG, fundoscopy, color & acuity testing. Male patients tested for HESX. DNA extracted from lymphoblastoid lines or blood. DNA sequenced 30X in > 92% on Illumina platform; QC used Sam, Picard, Bed tools, variant call GATK, annotation snpEff/snpSift; filter using Autosomal Dominant hypothesis, absence in 1k, 65k 19 genomes,Exac databases. Variants present in subjects but not parents were selected and tested using polyphen2, Sift and mutation taster. 2nd approach for variant filtering using ANOVAR yielded 197 variants with strong deleterious effect; variants present in at least 2 pedigrees were then prioritized **Results:** Patients followed 2-15 yrs had normal intelligence, eye examination showed normal ERG with abnormal VEP, dyschromatopsia VA approx. 20/200. Three subjects had learning difficulties, One female severe dyspraxia, and one male Growth Hormone, Cortisol and testosterone deficit and micropenis. Three subjects had sleeping disorders treated with melatonin. No HESX mutations were identified. All parents were healthy but 3 fathers were unavailable for blood sampling.SNP array CGH revealed a paternally inherited a single 410,4 kb deletion of the 2p26.1 including the *VRK2* gene described in complex intellectual disability. Five genes showed deleterious missense mutations identified and being confirmed by Sanger sequencing **Conclusion:** A 2p16.1 inherited deletion variant of unknown significance was identified on array CGH. WES analysis identified 4 candidate genes confirmed by Sanger sequencing which will need confirmation in a different dataset.

1763T

Integration of whole exome sequencing, expression profiling, and pathway analysis for the identification of novel genes in familial exudative vitreoretinopathy. *M.-Y. Chung[1,2], J.-H. Wu[3], S.-J. Chen[4].* 1) Inst Genome Sci, Natl Yang-Ming Univ, Taipei, Taiwan, Taiwan; 2) Medical Research, Taipei Veterans General Hospital; 3) Pediatrics, Taipei Veterans General Hospital; 4) Ophthalmology, Taipei Veterans General Hospital.

Familial exudative vitreoretinopathy (FEVR) is a group of retinal diseases characterized primarily by incomplete development of blood vessels in the retina. Clinical manifestations of FEVR are highly variable with inter- as well as intra- familial variability, ranging from little or mild visual loss to blindness in early childhood, and overlap with other similar retinal vascular diseases, e.g. Coats disease, persistent hyperplastic primary vitreous, and Norrie disease. Currently six FEVR genes have been identified and four of them have been confirmed to be involved in the *wnt* singling pathway, accounting for mutations in about 50% of the affected individuals. While using whole exome sequencing (WES) to identify genes with disease causing mutations, locus heterogeneity has become a major obstacle to find mutations in the same gene in at least two affected individuals from different families. We sought to prioritize the WES variants with expression profiles. Unique variants were identified in three genes with high priority. They were further analyzed by quantitative PCR and tube formation assay to validate the involvement in *wnt* signaling pathway as well as whether the gene might play a role in angiogenesis. Through this strategy, private mutations may be identified in FEVR and may apply to rare diseases with high locus heterogeneity.

1764F

Evaluation and treatment of nystagmus in a Brazilian boy with septo-optic dysplasia. *L. Gabriel[1], D. Girotto[1], P. Maya[2], M. Avila[1].* 1) Ophthalmology, Federal University of Goias, Goiania, Goiania, Brazil; 2) Neurosurgery, Santa Monica Hospital, Goiania, Brazil.

Herein we present a case of a 7-year old brazilian boy with a clinical diagnosis of septo-optic dysplasia. He was born with severe hypoglycemia, and with 8 months of age, low levels of cortisol, thyroid-stimulating hormone, thyroxine, growth hormone, and testosterone were identified, characterizing congenital panhypopituitarism. Due to these hormonal imbalances he also had lower height and inguinal testicles. Additionally he presented with low visual acuity and nystagmus which led to an ophthalmology investigation revealing bilateral optic disc hypoplasia on the fundoscopy and optic nerve hypoplasia, reduction of the sella turcica, and pituitary parenchyma thinning on the magnetic resonance imaging. Visual evoked potentials were decreased. In order to ameliorate his vision the nystagmus was measured by videonystagmography and treated with gabapentin showing a visual acuity improvement.

1765W

LCA9-associated NMNAT1 mutant protein study in Drosophila. *J. Sun[1], C. Weng[2], X. Feng[1], J. Lin[1], C. Tong[3], M. Qi[1].* 1) Department of Genetics and Genomics, School of Medicine Zhejiang University, Hangzhou, China, Zhejiang, China; 2) Department of Genetics and Genomics, Case Western Reserve University, Cleveland, USA; 3) Life Sciences Institute and Innovation Center for Cell Signaling Network, Zhejiang University, Hangzhou, China.

Leber congenital amaurosis (LCA) is an autosomal recessive retinal dystrophy that manifest with genetic heterogeneity. Previously, our team have identified 8 mutations in NMNAT1 could cause LCA9 in human. Study by Hoopfer ED shows that overexpression of Nmnat protein can protects axons from degeneration in drosophila. So we choose the fly as the animal model. As Homozygous Nmnat mutation in Drosophila melanogaster is lethal, we construct MARCM mutation system in drosophila eyes. To identify whether the mutant Nmnat1 protein can rescue the retina degeneration or not in drosophila eyes we have created transgenic drosophila by microinjecting the constructed plasmid. First, we cloned 3 homo-nmnat1 mutatants and wildtype Nmnat genes into pUASTattB-HA plasmid. These mutants are C.817A>G; C.457C>G; C.507G>A, respectively. Second, these plasmids have been injected into the embryo of flies to express the mutant nmnat1 protein. Use eyFLP system allows us to construct the conditional lethal mutation of Nmnat in drosophila visual system for the study of eye structural integrity. Genotype-phenotype study has been conducted by viewing the integrate rhabdom architecture of ommatidium. The intact rhabdom number in one ommatidium in drosophila eyes for wildtype overexpression of Nmnat, Nmnat knockout, C.817A>G mutation, C.457C>G mutation and C.507G>A mutation are 7, 7, 0.8, 3.4, 3.8 and 2, respectively. Neither point mutations (C.817A>G; C.457C>G) of nmant1 could rescue the phenotype caused by defected nmnat. Prematured stop codon C.507G>A mutation generates the truncated nmnat1 protein and leads more severe eye cell degeneration than the others. These results are consistent with the LCA9 disease phenotypes in human. This drosophila will provide valuable animal models for further study the molecular mechanism of LCA9.

1766T

Bardet-Biedl syndrome, postaxial polydactyly, Shh signalling pathway and a founder effect in a Libyan extended consanguineous family from Tarhunah Berber tribe. *N.Bouayed. Abdelmoula, S. Aloulou, W. Ben Rom-dhane, R. Louati, W. Smaoui, F. Abid, N. Abdellaoui, O. Bedoui, H. Fourati, K. Chaker, Kh. Trigui, J. Dammak, Y. Kaabi, A. Kaabi, S. Kammoun, H. Denguir, S. M'rabet, H. Ben Ameur, M. Ben Aziza, W. Ltaif, S. Kammoun, J. Jaber, M.N. Mhiri, N. Moalla.* UR17ES36, Medical University of Sfax, Sfax, Tunisia.

BBS is relatively frequent in some isolated populations or areas like the negev bedouins in Israel with confirmed mutations in different BBS genes. Here, we report a Libyan tribe from Tarhunah in which several individuals were diagnosed with BBS. Tarhunah is a Libyan town in which the population generally belongs to a Berber tribe. The extended consanguineous family reported here, is named Bousitta tribe because of the syndromic recurrence of polydactyly ("six" fingers = "6: Sitta" fingers in Arabic) for more than four generations. The primary case who attended our genetic counseling at 2012 was a 34 years old infertile male with cryptozoospermia and failure of ICSI attempts using intra-testicular spermatozoa (implantation failure). Physical examination revealed cardinal features of BBS with middle cognitive impairment, obesity, visual disorders (strabismus and nystagmus) and retinal dystrophy, tetramelic postaxial polydactyly and hyperlipidemia. Craniofacial dysmorphism, convulsive crisis and psychiatric problems were also noted. Genital exploration revealed unilateral right cryptorchidism and normal penis with normal hair distribution, normal level of FSH and spermatozoa in testicular biopsy. Cytogenetic evaluation revealed a normal male 46,XY karyotype. There was a family history of polydactyly with and/or without cardinalsigns of BBS, mainly retinal dystrophy, obesity and learning difficulties. The study of BBS illustrates the value of using isolated inbred populations for the study of human genetic diseases and suggests strategies for facilitating the study of complex diseases and traits.

1767F

Acquired ventriculamegaly in a case with *SOX 9* mutation. *A. Matsumoto[1], E. Imagawa[2], N. Miyake[2], N. Ikeda[1], M. Kobashi[1], M. Goto[1], T. Yamagata[1], N. Matsumoto[2], H. Osaka[1].* 1) Pediatrics, Jichi Medical University, Shimotsukeshi, Tochigi, Japan; 2) Department of human genetics, Yokohama City University graduated school of medicine.

SOX 9 is responsible for campomelic dysplasia (CMD). Symptoms of CMD include recurrent apnea, facial features, and shortening of the lower extremities. Acampomelic CMD (ACMD), a variant, lacks long bone curvature. In mice, Sox 9 is associated with differentiation to oligodendrocytes and astrocytes, but its contributions to neuronal differentiation are unclear in humans. We encountered a two-year-old boy with ACMD presenting with tracheal malasia and macrocephaly. He also had acquired ventriculamegaly, hydrocephalus, and hypoplasia of the corpus callosum, indicating the failure of white matter growth after birth. **Case report** The patient was the second child born to nonconsangious parents. His growth parameters were normal, but his occipito-frontal circumference (OFC) was 2.8 standard deviations (SD) at birth and 3.9 SD at 18 months of age. He showed macrocephaly and minor anomalies such as hypertelorism, palpebronasal fold, small mandible, high-arched palate, a cleft soft palate, and micropenis without long bone curvature. His developmental milestones were delayed, as he could not control his head and could not comprehend simple words. His neurological evaluation revealed hypotonia and normal deep tendon reflex. He showed severe hearing loss (100 dB). Cranial magnetic resonance imaging findings were normal at one month of age but showed ventriculomegaly, hydrocephaly, and hypoplasia of the corpus callosum at two years of age. G-banding was 46, XY. Array CGH showed no abnormalities. Exome sequencing revealed a *de novo* novel mutation of c. 236A>C, p (Q79P) in *SOX9*. **Discussion** Sox 9 is thought to be crucial in neural stem cell development in the central and peripheral nervous system along with Sox 8 and Sox10 in mice. Mice with *Sox 9* ablation have defects in specification and a decreased number of oligodendrocytes and astrocytes in the spinal cord in early development. Oligodendrocyte progenitor cells from later stages of development were shown to be compensated by Sox 8 and Sox 10, but astrocytes were not, instead showing a reduced number. In humans, neuronal abnormalities in CMD and ACMD include relative macrocephaly in 11 of 22 and mild lateral ventriculomegaly in 2 of 22. We described for the first time an ACMD patient showing acquired macrocephaly with ventriculamegaly, indicating the failure of oligodendrocyte/astrocyte development postnatally. This phenotype suggests that SOX 9 plays a crucial role in human central nervous system development.

1768W

The role of WNT regulatory variants in nonsyndromic cleft lip and palate. *L. Maili[1,2], K. Westerman[3], Q. Yuan[1], B.T. Chiquet[1,3], A. Letra[3], G.T. Eisenhoffer[4], J.T. Hecht[1,3].* 1) Pediatrics, McGovern Medical School, University of Texas Health Science Center, Houston, TX; 2) Graduate School of Biomedical Sciences, University of Texas Health Science Center, Houston, TX; 3) School of Dentistry, University of Texas Health, Houston, TX; 4) University of Texas MD Anderson Cancer Center, Houston, TX.

Nonsyndromic cleft lip with or without cleft palate (NSCLP) is a common birth defect with a prevalence of 1 in 700 that affects approximately 4000 newborns in the United States annually. Despite improvement in treatments, NSCLP continues to pose major medical, psychosocial, and financial burdens. We and others have shown that perturbation of genes at the receptor level of the WNT pathway causes severe craniofacial anomalies. For example, 1) knockdown or overexpression of *fzd6* causes craniofacial abnormalities in zebrafish, 2) loss of *lrp6* significantly reduces Wnt signaling in mutant mouse embryos resulting in orofacial clefts, 3) knockdown or mutant forms of *lrp5* cause abnormal craniofacial cartilage phenotypes and cleft palate in zebrafish, and 4) the facial region completely fails to form in mice deficient for *dkk1*. The goal of this study was to determine whether variation in regulatory variants upstream of *FZD6, LRP5, LRP6* and *DKK1* contribute to NSCLP. Prioritization of potential regulatory variants in the upstream noncoding region of each gene was based on bioinformatic analysis using species conservation, epigenetic signatures, and functionality algorithms. Next, three different *in silico* transcription factor binding programs were used to filter variants that had a unique protein binding signature for each allele. Twenty-one variants, 6 in *FZD6*, 5 in *LRP5*, 5 in *LRP6* and 5 in *DKK1* were prioritized using these criteria. These variants were further assessed regarding differential allele-specific protein binding patterns using electrophoretic mobility shift assays (EMSAs). Nine variants resulted in differential protein binding patterns: 1 in *FZD6*, 3 in *LRP5*, 3 in *LRP6* and 2 in *DKK1*. These variants were cloned upstream of luciferase promoter constructs to assess their effect on gene expression. The variant in *FZD6* did not alter luciferase activity in 293T cells and is being evaluated in other embryonic cell lines. Luciferase reporter results for the other variants are pending. Further, variants that have an effect on luciferase expression will be assessed using a dual reporter approach in zebrafish to show the effect of each variant allele on expression *in vivo*. The results of this study will provide important information about the effects of noncoding variants on WNT gene expression during development.

1769T

A dog model of non-syndromic cleft palate. *B. Schutte[1,2], J. Plassais[3], B. Stanley[4], B. Davis[3], E. Ostrander[3], J. Fyfe[4].* 1) Microbiol and Molec Gen, Michigan State Univ, East Lansing, MI; 2) Pediatrics and Human Development, Michigan State University, East Lansing, MI; 3) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Small Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, MI.

Cleft palate is a common human birth defect, affecting ~1/1500 live births. The standard of care is surgical repair, which provides satisfactory resolution of facial disfigurement and feeding problems. However, post-operative complications are common, and the long-term outcomes of various surgical approaches are unknown until craniofacial growth is complete when the patient is at least 18 years old. The goal of this work is to generate a clinically-relevant dog model of cleft palate to compare surgical outcomes in 18 months, not 18 years. Toward this goal, we developed an IACUC approved (AUF 11/14-207-00) breeding colony segregating cleft palate at MSU. In a single sibship of mixed genetic background 17% of pups (11 of 63) were born with a uniform cleft of the hard and soft palate. Morphological analysis and CT scans failed to detect any other obvious abnormality of the face. In addition, the cleft of four pups, 1 male and 3 females, was repaired and all are thriving, consistent with a cleft palate that is non-syndromic. To assess the genetic etiology of the cleft, we performed whole genome sequencing in five affected individuals. We identified a single variant at the *PITX2* promoter that was homozygous in all five affected individuals sequenced and found in only 1 of 121 control dog breeds. To test for association, we genotyped 10 affected and 15 unaffected offspring of the sibship. The observed genotypes were not statistically different from Mendelian ratios (p = 0.7), suggesting that the variant near *PITX2* is not associated with cleft palate in this sibship. In addition, we performed genome wide linkage analysis on 8 affected and 35 unaffected dogs and identified three large regions (CFA 10, 20 and 32) with suggestive linkage (LOD scores>2.4). To reduce the genetic complexity of this phenotype, we mated the repaired affected male to his dam, and 2 of 8 pups were affected. We also mated the affected male with a repaired affected female, and 6 of 6 pups had cleft palate. In conclusion, we have a breeding colony of dogs with high incidence of non-syndromic cleft palate. We are using this model to better understand the complex genetics of cleft palate, to evaluate long-term outcomes of surgical repair and to develop new therapies. This work supported in part by the College of Veterinary Medicine, Michigan State University.

1770F

Syndromic and nonsyndromic congenitally missing teeth: Prevalence, clinical manifestations and patterns. *E. Severin[1], R. Baltag[2].* 1) Dept Human Genetics, Carol Davila Univ Med Pharm, Bucharest, Romania; 2) Department of Orthodontics and Dentofacial Orthopedics, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania.

Background: Tooth agenesis is the most common developmental anomaly in humans. The reported prevalence of permanent dentition tooth agenesis excluding the third molars was 15.7% for European population. Syndrome and isolated forms of tooth agenesis have been described and can occur sporadically or in families showing etiologic heterogeneity. Dental agenesis does not only represent an individual clinical problem but also a public health problem. **Objectives:** Study aims were to describe the variability of clinical phenotypes, to observe similarities/differences of familial tooth agenesis forms among affected relatives, to characterize the mode of inheritance and to identify distinct groups of patients for further molecular investigations. **Patients and Methods:** A total of 2305 subjects were included in the study. Clinical and radiographical assessments were performed, followed by detailed investigation into the personal and family medical history of the patients with tooth agenesis. If necessary chromosome or molecular tests were performed. **Results:** Tooth agenesis was observed in following two forms, hypodontia and oligodontia. More severe expression of both syndromic hypodontia and oligodontia were noticed comparing with non-syndromic tooth agenesis. All teeth were affected but not equally. Tooth agenesis was found in conjunction with other dental anomalies. In many genetic syndromes there is a typical pattern of dental agenesis but comparative analysis of patterns of dental agenesis revealed common, rare and individual dental findings. **Conclusions:** Predictive genetic testing has potential for accurate risk assessment in families with a history of tooth agenesis that may be inherited. Using the results of these tests, the dentist will be able to develop a targeted prevention plan to minimize the functional complications and esthetic and psychological stress.

1771W

A PITX1 variant in a large pedigree with dominant lower extremity anomalies. *Y. Guo[1], W. Sankar[2], M. Deardorff[3,4], H. Hakonarson[1,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA19104, USA; 2) Department of Surgery, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 4) Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Here we present a large autosomal dominant pedigree with variably penetrant lower extremity anomalies. The severe end of the clinical spectrum includes preaxial polydactyly. Intermediate phenotypes include metatarsus adductus and pes planus. The mild end of the spectrum comprises mild hip dysplasia, resulting in short femoral neck or trochanteric overgrowth, and shortened first toe. All individuals with hip abnormalities demonstrate abnormal toes, but not all individuals with toe findings have hip problems. This suggests that toe and hip findings share a dominant etiology with variable penetrance. No congenital clubfoot, upper limb or other skeletal abnormalities were noted. All individuals were cognitively normal with no other known organ system involvement. Based on a presumed autosomal dominant inheritance pattern, whole exome sequencing (WES) of two affected siblings and an unaffected child was performed. Through bioinformatics analysis and variant filtering steps, we identified a nonsense variant c.454C>T:p.Q152X in *PITX1* (paired like homeodomain 1) in both affected individuals but absent from the unaffected one. Subsequent Sanger sequencing results showed that the variant is seen in the affected father and all affected siblings in the immediate family, and is absent in unaffected family members. Additional analysis of the extended family is underway. Previously, a single large pedigree with variably penetrant preaxial foot polydactyly and clubfoot was reported with a *PITX1* missense mutation in the homeodomain that disrupted DNA binding (Gurnett, 2008). The *PITX1* nonsense variant in our family resides in the terminal exon, suggesting that it could escape nonsense-mediated decay to lead to a PITX1 protein that is truncated after the homeodomain protein, which could act in a dominant negative manner, but perhaps less damaging than a homeodomain mutation. In contrast to Gurnett et al, 2008, we did not observe severe clubfoot in our family. We also did not note upper extremity involvement that has been reported in a family with a complex rearrangement of this region (OMIM 186550).

1772T

Multidisciplinary assessment of 49, XXXYY, a rare X and Y chromosomal variation (XYV). *P. Lasutschinkow[1], C. Cappello[1], A. Gropman[2,3], F. Mitchell[1], G. Harmon[1], A. Ganges[1], M. McLeod[1], C. Samango-Sprouse[1,2,3,4].* 1) The Focus Foundation, Davidsonville, MD; 2) Dept of Neurodevelopmental Disabilities and Neurogenetics, Children's National Medical Center, Washington, DC; 3) Dept of Pediatrics, George Washington University School of Medicine & Health Sciences, Washington, DC; 4) Dept of Human and Molecular Genetics, Florida International University, Miami, FL.

   Introduction: We present the profile of a 9 year-old Italian Caucasian male with 49, XXXYY. To our knowledge, there have only been 7 reported cases of this XYV, including only one since 1995. **Statement of Purpose:** We present this boy's phenotypic profile and compare it to prior cases of 49, XXXYY. The neurodevelopmental profile for this disorder is expanded because this boy was evaluated from a multisystemic perspective. **Methods:** Neurodevelopmental testing included Receptive One Word Picture Vocabulary Test (ROWPVT-4) and Leiter International Performance Scale (LIPS-III). The Child Behavior Checklist (CBCL) was completed by the mother. The multidisciplinary care team that evaluated this patient included pediatric specialists in: neurodevelopment, neurogenetics, immunology, speech language pathology, and physical therapy. **Results:** He had increased height (76th percentile) and weight (93rd percentile). Neither parent's age was advanced. Ambulation occurred at 29 months and first words were spoken at 3 years of age. He scored below the first percentile on ROWPVT-4 (SS=57). Of the 3 subtests completed on LIPS-III, his scores ranged from slightly below average to severely delayed (SS=7, 5, 1). He was in the clinical range for 18 of 24 subscores on the CBCL, including Total, Externalizing, and Internalizing Problems (T-Scores=78, 79, 70 respectively). **Discussion:** Test scores on the ROWPVT-4 and LIPS-III support the presence of cognitive impairment. CBCL scores reflect his significant behavioral issues including low frustration tolerance, which has been reported in two other cases. His scores in Anxious/Depressed, Anxiety Problems, Attention Problems, and ADHD Problems (T-scores=70, 70, 79, 80 respectively) support anxiety disorder and ADHD. Salient features that he shares with past cases include: prognathism, prominent cheek bones, strabismus, cryptorchidism, radio-ulnar synostosis, clinodactyly, low set ears, hypotonia and epicanthal folds. Ambulation and speech were delayed, and pronounced dysfluency persists. Expressive speech problems have been reported in prior cases, and this boy likely has Childhood Apraxia of Speech (CAS). He also presents with pes planus, left torticollis, and a leg length discrepancy. His immunological profile is unremarkable besides a neonatal pneumonia. He has not had testosterone treatment, which could help manage his complex deficits to some degree.

1773F

Mutations in *NAA10* and *NAA15* are associated with a range of cardiac and neurodevelopmental phenotypes. *G.J. Lyon[1], M. Lee[2], H.Y. Kweon[2], M. Doerfel[1], A. Klimas[3], K. Rivera[1], J. Crain[1], T. Papazyan[1], Y. Wu[1], M.K. Eldomery[4], Z.C. Akdemir[4], J.E. Posey[4], M. Longoni[5], F.A. High[5,6], M.A. van Slegtenhorst[7], C. Finnila[8], C. Ruivenkamp[9], S. Naidu[10], S. Desai[10], R. Fundt[11], G. Mancini[12], J.R. Lupski[4], G.M. Cooper[8], H.A.F. Stessman[13], D. Pappin[1], W.K. Chung[14], E. Entcheva[3], G.T. Oh[2].* 1) Cold Spring Harbor Laboratory, New York City, NY; 2) Department of Life Sciences, Ewha Womans University, Seodaemoon-gu, Seoul 120-750, Korea; 3) Department of Biomedical Engineering George Washington University 5000C, Science and Engineering Hall 800 22nd Street NW, Suite 5000 Washington, DC 20052; 4) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; 5) Massachusetts General Hospital, Boston, MA; 6) Boston Children's Hospital, Boston, MA; 7) Erasmus Medical Center Postbus 2040, 3000 CA Rotterdam; 8) HudsonAlpha Institute for Biotechnology 601 Genome Way Huntsville, AL 35806; 9) Laboratory of Diagnostic Genome Analaysis (LDGA) Leiden University Medical Centre (LUMC), S-6-P PO box 9600 2300RC Leiden The Netherlands; 10) Department of Neurogenetics, 801 N. Broadway | Rm 564 | Baltimore, MD 21205; 11) RadboudUMC Nijmegen; 12) Department of Clinical Genetics Erasmus University Medical Center, room Ee2077 Wytemaweg 80. 3015 CN Rotterdam, The Netherlands; 13) Creighton University Medical School 2500 California Plaza Omaha, NE 68178; 14) Columbia University Medical Center, New York, NY, USA.

   We previously identified the genetic basis of a X-linked, infantile lethal Mendelian disorder, involving a p.S37P missense mutation in *NAA10*, encoding the catalytic subunit of NatA, involved in amino terminal-acetylation (NTA) of proteins. To date, there are >20 individuals with damaging mutations in *NAA10* or *NAA15,* with the latter gene encoding the dimeric binding partner for Naa10. Both *NAA10* and *NAA15* are relatively intolerant of variation. Phenotypic features of patients carrying disruptive variation in NatA include craniofacial anomalies, hypotonia, global developmental delays, cardiac arrhythmia, and/or cardiomyopathy. Functional analysis of the *NAA10* missense mutations has demonstrated impaired biochemical activity and/or a reduced capacity to form a stable NatA complex, although there is no proteome-wide decrease in NTA, thus implying tissue-specific and/or temporal substrate-specific effects. Attempts to knockdown or knockout *NAA10* in human cells results in cell death, making it difficult to assess proteome-wide NTA in human KO cells. Mice with decreased or absent *Naa10* survive throughout embryogenesis, but display a range of phenotypes that are variably expressive and include increased early neonatal lethality (likely related to congenital heart defects), supernumerary thoracic rib and vertebrae, piebaldism, and/or urogenital and renal abnormalities. Some of these phenotypes overlap with that seen with the missense mutation in humans, but survival of the knockout mice was unexpected, given the mutation intolerance of this gene in humans. Extensive proteomic analysis of mouse embryonic fibroblasts has not shown any proteome-wide decrease in NTA in the KO cells, with one possible explanation being our discovery of a previously unannotated *NAA10*-like gene in mice, which lacks a human ortholog, and which we refer to as *NAA12*. A rabbit polyclonal antibody raised with specificity to Naa12 shows expression of Naa12 in many mouse tissues. Given the potential redundancy in mice, fibroblasts from one affected human proband control were reprogrammed into induced pluripotent stem-cells and differentiated into cardiomyocytes (CMs). Electrophysiological experiments show rate-dependent prolongation of the effective refractory period in patient CMs, suggesting a cell-autonomous defect in the CMs. Overall, our data show that the process of NTA, carried out in part by the NatA complex, is important for normal development in mice and humans.

1774W

Protective mechanisms in Cornelia de Lange patients with early truncating variants in *NIPBL* generate an N-terminal truncated protein that is able to mediate cohesin loading in the absence of MAU2. *I. Parenti[1], J. Pozojevic[1], S. Ruiz Gil[1], B. Brouwer[2], F. Diab[3], V. Dupé[3], E. Mulugeta[2], W. van Ijcken[2], E. Watrin[3], KS. Wendt[2], FJ. Kaiser[1].* 1) Institute of Human Genetics, University of Lübeck, Lübeck, Germany; 2) Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands; 3) Faculté de Médecine, Institut de Génétique et Développement de Rennes, Rennes, France.

Cornelia de Lange syndrome (CdLS) is a rare developmental disorder. The majority of CdLS patients show mutations in *NIPBL*. The NIPBL protein interacts via its N-terminus with MAU2 to generate a complex named kollerin, engaged in loading the cohesin complex onto DNA. In addition, NIPBL is known to mediate transcriptional regulation independently from its role in cohesin loading. By analyzing a very large cohort of CdLS patients with causative mutations in *NIPBL,* we noticed that some patients with early truncating variants presented with milder phenotypes and a lower frequency of severe limb reductions as compared to those with truncating variants affecting the more C-terminal part of the protein. To investigate this particular genotype/phenotype correlation, we performed CRISPR/Cas9 genome editing targeted to exon 2 of *NIPBL*; by this, we were able to isolate three independent viable clones carrying a homozygous out-of-frame insertion in *NIPBL*. Analyses of these cells on protein level revealed the presence of an N-terminal truncated NIPBL, translated from an aberrant but stable mRNA by use of an alternative ATG start codon. Interestingly, the truncated NIPBL is incapable to form heterodimers with MAU2. Quantitative RNA analyses showed an increase of *NIPBL* transcript and unchanged level of *MAU2*. However, the inability of the kollerin subunits to interact with each other leads to a decrease of NIPBL on protein level and to a total loss of MAU2. Next, we evaluated the ability of the N-terminally truncated NIPBL to mediate cohesin loading. Fractionation experiments revealed an equal amount of total chromatin-bound cohesin compared to wild type cells, indicating that the NIPBL/MAU2 heterodimer is not essential for overall cohesin loading onto chromatin. Consistently, determination of cohesin binding sites by ChIP-sequencing showed slight changes in cohesin binding sites. Taken together, our data strongly indicate the existence of protective mechanisms preventing a total loss of NIPBL gene product by the use of alternative start codons in *NIPBL* transcripts with early truncating mutations to ensure the synthesis of the essential C-terminus of NIPBL. Furthermore, our data give new insights into molecular mechanisms responsible for the pathogenesis of CdLS, suggesting a CdLS-relevant function of NIPBL not related to cohesin loading.

1775T

11q terminal deletion associated with mild phenotype of Jacobsen syndrome. *C. da Silva-Camargo[1], H. Salomão[1], F.R. Faucz[1], A. Bonalumi[1], J. Souza[1], P. Grossfeld[2], J. Rosenfeld[3], G.S. Rigoni[4], B.L. Haliski[4], L.S. Benatti[4], V.S. Sotomaior[1], S. Raskin[1,4].* 1) Group for Advanced Molecular Investigation (NIMA), Pontifícia Universidade Católica do Paraná, Curitiba, PARANÁ, Brazil; 2) Division of Pediatric Cardiology, Department of Pediatrics/Rady Children's Hospital of San Diego, California, USA; 3) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 4) Medical School, Universidade Positivo, Curitiba, Paraná, Brazil.

Jacobsen syndrome (JBS) is caused by a partial terminal deletion at 11q23.4-q25 and demonstrates variable phenotypic expression. The incidence of deletions in this region is estimated to be less than 1 in 100,000 in live births with a female/male ratio of 2:1 and with few cases with molecular analysis reported. The characteristic clinical picture includes mainly pre- and postnatal physical growth retardation, psychomotor retardation, intellectual disability, skull deformities, characteristic dysmorphic face, congenital cardiopathy and thrombocytopenia (Paris-Trousseau syndrome) or pancytopenia. In this research deletions were identified by microarray-based comparative genomic hybridization (aCGH) between 11q23.4 - q25 in four patients, 3 females and 1 male, with clinical features overlapping JBS, generating a mild phenotype compared to classical cases. Two patients also carry other deletions in chromosomal regions not significantly associated with the clinical picture typically presented of this syndrome. The four patients do not present some of the typical clinical alterations, such as thrombocytopenia, platelet dysfunction and cardiac abnormalities, although they have JBS similar phenotypic components such as speech difficulties, facial dysmorphism, neuropsychomotor development delay, cognitive deficit, and intellectual disability. Previously published papers implicates *ETS-1* and *JAM3* as putative causatives genes for cardiac abnormalities, and there is evidence that the genes that play a fundamental role in thrombocytopenia are *FLI1*, *ETS1*, *NFRKB*, and *JAM3*. Two of our female patients have alterations in *JAM3*, and one of them also has alterations in *NFRKB*, questioning the involvement of these genes in the etiology of cardiac abnormalities and thrombocytopenia. The others two patients have no changes in these genes. Further research is still needed to define the platelet function critical genes in JBS. Our results reiterate the benefits of aCGH for the description of new phenotype/genotype associations and refinement of previously established ones, in addition, this reports demonstrates the importance of thorough phenotypic characterization of patients with JBS, thereby contributing to a better understanding of the spectrum of this syndrome.

1776F

Deciphering the mechanisms of developmental disorders (DMDD): Shedding light on human genetic disease using embryonic lethal knockout mice. *A. Galli[1,2], A. Green[1,2], DMDD Consortium.* 1) Wellcome Trust Sanger Institute, Cambridgeshire, United Kingdom; 2) dmdd.org.uk.

   Statement of purpose   Around one third of all mammalian genes are essential for life, their removal resulting in embryonic or perinatal lethality. Studying these genes is therefore a powerful way to gain understanding of the genetic mechanisms that underlie both normal development and the mutations that lead to congenital birth defects.   **Methods**   The DMDD consortium studies embryonic and perinatal lethal knockouts in mice. Multiple embryos from each line are imaged in 3D at near-histological resolution to allow comprehensive morphological phenotyping. Parallel screens identify placental abnormalities and provide gene expression profiles for each line. All DMDD data is freely available (dmdd.org.uk).   **Results**   3D image analysis of morphological defects identified in 70 lines (more than 550 embryos) examined at E14.5 (the middle of organogenesis) shows that almost all embryos exhibit multiple phenotypes, affecting many tissues and organ systems. Prevalent abnormalities include subcutaneous oedema, malformations of the forebrain and eye muscles, and loss or abnormality of the hypoglossal nerve. Remarkably, around 80% of the lines have cardiovascular phenotypes. The dataset is a particularly valuable source of information for those researching the genetic basis of cardiovascular defects.   Surprisingly, the phenotypes of individual embryos from the same line form distinct but overlapping sets. Across all lines the phenotype penetrance is highly variable and typically incomplete.   Data mining studies have shown that the phenotypes scored in eleven of the lines replicate the detailed phenotypic spectra of human genetic disorders, including Meckel and Joubert syndromes. This suggests that data from the remaining lines could be valuable in identifying and linking gene mutations with currently undiagnosed rare diseases.   Systematic placental analysis has revealed a much higher prevalence of placental phenotypes than previously appreciated. Around two thirds of embryonic lethal lines have a defective placenta. In particular, around one third of lines with a cardiovascular defect also have a placental phenotype, suggesting a link between abnormalities in the two organs.

1777W

On the significance of craniosynostosis in a case of Kabuki syndrome with a concomitant *KMT2D* mutation and 3.2 Mbp de novo 10q22.3q23.1 deletion. *A. Topa[1,2], L. Samuelsson[2], L. Lovmar[2], G. Stenman[1,2], L. Kölby[3].* 1) Department of Pathology and Genetics, University of Gothenburg, The Sahlgrenska Academy, Gothenburg, Sweden; 2) Department of Clinical Pathology and Genetics, Sahlgrenska University Hospital, Gothenburg, Sweden; 3) Department of Plastic Surgery, University of Gothenburg, The Sahlgrenska Academy, Gothenburg, Sweden.

   Craniosynostosis has rarely been described in patients with Kabuki syndrome. We report here a boy with facial asymmetry due to combined premature synostosis of the right coronal and sagittal sutures as well as several symptoms reminiscent of Kabuki syndrome (KS). Our case supports previous observations and suggests that craniosynostosis is a part of the KS phenotype. The uniqueness of our case is the sporadic co-occurrence of two genetic disorders, that is, a de novo frameshift variant in the *KMT2D* gene and a de novo 3.2 Mbp 10q22.3q23.1 deletion. Our findings emphasize the importance of the initial clinical assessment of children with craniosynostosis and that genomic and monogenic disorders, such as Kabuki syndrome, should be considered among the differential diagnoses of syndromic forms of craniosynostosis.

1778T

Expanding the phenotypic spectrum of *de novo KAT6A* mutations and their impact on biological pathways through functional genomics. *V. Arboleda[1], J. Kennedy[2], P. Goddard[1], D. Goudie[3], E. Blair[4], K. Chandler[5], S. Joss[6], C. Deshpande[7], V. McKay[8], A. Green[9], R. Armstrong[10], M. Lees[11], S. Tomkins[2], B. Kamien[12], B. Hopper[12], T. Yang Tan[13,14], P. Yap[13,14], Z. Stark[13], N. Okamoto[15], N. Miyake[16], N. Matsuoto[16], E. MacNamara[17], J. Murphy[17], E. McCormick[18], D. Li[19], H. Hakonarson[18], M. Falk[18], P. Blackburn[20], E. Klee[20,21], D. Babovic-Vuksanovic[20,21], S. Schelley[22], L. Hudgins[22], S. Kant[23], B. Isador[24], B. Cogne[24], K. Bradbury[25], M. Williams[26], C. Patel[26], H. Heussler[26], B. Pasaniuc[1,27], DDD Study[28], R. Newbury-Ecob[2], S. Nelson[1,27].* 1) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, USA; 2) Clinical Genetics, University Hospitals Bristol, University of Bristol, Bristol, UK; 3) Clinical Genetics, Ninewells Hospital & Medical School, Dundee,UK; 4) 1.Department of Clinical Genetics, Churchill Hospital, Old Road, Headington, Oxford, UK; 5) Manchester Centre for Genomic Medicine, St. Mary's Hospital, Central Manchester Foundation NHS Trust, Manchester Academic Health Science Centre (MAHSC), Manchester, UK; 6) West of Scotland Genetics Service, Queen Elizabeth University Hospital, Glasgow, UK, G51 4TF; 7) Clinical Genetics 7th floor, Borough Wing Guy's Hospital Great Maze Pond London, UK; 8) Clinical Genetics, Liverpool Women's Hospital, Liverpool, UK; 9) Department of Clinical Genetics, Our Lady's Children's Hospital, Crumlin, School of Medicine and Medical Science, University College Dublin, Ireland; 10) East Anglian Medical Genetics Service, Addenbrooke's Hospital, Cambridge, UK; 11) Clinical Genetics, Great Ormond Street Hospital NHS Trust, London, UK; 12) Hunter Genetics. Newcastle, Australia; 13) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Melbourne, Australia; 14) Genetic Health Service New Zealand, Auckland, New Zealand; 15) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health; 16) Department of Human Genetics, Yokohama City University Graduate School of Medicine Yokohama, Japan; 17) Undiagnosed Disease Program, National Human Genome Research Institute, NIH, USA; 18) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 19) Center for Applied Genomics, The Children's Hospital of Philadelphia, PA, USA; 20) Centre for Individualized Medicine, Mayo Clinic, Rochester, Minnesota, USA; 21) Department of Clinical Genomics, Mayo Clinic, Rochester, Minnesota, USA; 22) Division of Medical Genetics, Department of Pediatrics, Stanford University, Stanford, CA; 23) Clinical Genetics, Leiden University Medical Center, The Netherlands; 24) Service de génétique médicale, CHU Nantes, Nantes, France; 25) Clinical Genetics Guys and St Thomas' NHS Foundation Trust, Guys Hospital; 26) Genetic Health Queensland, Herston, Brisbane, Queensland, Australia; 27) Department of Human Genetics, David Geffen School of Medicine, UCLA, USA; 28) The Deciphering Developmental Disorders study, Wellcome Trust Sanger Institute, Cambridgeshire, United Kingdom.

D*e novo* truncating mutations in the *KAT6A* gene cause a clinically distinct congenital syndrome characterized by neurodevelopmental impairments, global developmental delay, cardiac and other congenital malformations. KAT6A is a multifunctional lysine acetyltransferase protein that controls gene expression through histone lysine acetylation and post-translational acetylation. Through an international collaboration, we have clinically characterized over 63 patients with *de novo* and rare mutations in *KAT6A*, 45 which have never been published in the literature significantly expanding the mutational and phenotypic spectrum of *KAT6A* syndrome. Early studies identified primarily *de novo* truncating mutations in the last exon of *KAT6A*, however newer cases have been identified throughout the protein, including a number of missense mutations in highly conserved functional regions of the gene. The degree of intellectual disability and delay, as characterized by the treating clinician, is most severe for patients with truncating mutations in the last exon and least severe in those with mutations earlier in the gene. In combination with the previously reported cases, there is expansion of the phenotypic variability and newly identified features include behavioral difficulties, sensory issues, high frequency of recurrent infections and several patients with proven immune or hematological dysfunction. To explore the underlying biology of the KAT6A protein, we performed RNAseq and differential expression (DE) analysis on fibroblasts derived from 5 patients with KAT6A truncating mutations and compared the gene expression profile to 5 unaffected individuals. DE genes are overrepresented in GO biological processes pertaining to systemic development, nervous system development, visual perception and skeletal development, as well as adhesion and protein glycosylation pathways. The most DE genes (qval < 10^{-6}) are involved in neurite migration and fibrosis. These data provide insight into the pathology of late-exon KAT6A truncations and further implicate aberrant p53 mediated senescence pathways in KAT6A phenotypic variation. Our work in fibroblasts suggests that KAT6A is a key protein underlying cellular processes in fetal and childhood brain development and DNA damage response. Future directions include creating an induced pluripotent stem cell model system to investigate the role of KAT6A in early cell development and tissue-specific cell types.

1779F

The IMPC: A global research infrastructure for understanding the role of genes in human development and disease. *V. Munoz Fuentes[1], P. Flicek[1], A.-M. Mallon[2], T. Meehan[1], H. Parkinson[1], D. Smedley[3] on behalf of the IMPC consortium.* 1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK; 2) Medical Research Council Harwell (Mammalian Genetics Unit and Mary Lyon Centre), Harwell, Oxfordshire OX11 0RD, UK; 3) Clinical Pharmacology, William Harvey Research Institute, School of Medicine and Dentistry, Queen Mary, University of London, Charterhouse Square, London EC1M 6BQ, UK.

   The International Mouse Phenotyping Consortium (IMPC) is building a functional catalogue of the mammalian genome by producing and phenotyping a knockout mouse strain for every protein-coding gene. To date, over 4,000 knockout mouse lines, many for poorly understood genes, have been characterized and made available to the research community. The IMPC is a coordinated effort involving more than a dozen research centers and a dedicated publicly-available online resource. Using a standardized adult phenotyping pipeline, centers measure each mouse for more than 250 phenotypic parameters including metabolite content in blood, hearing capacity, and skeletal malformations. In addition, over 1,000 embryonic lethal mouse lines are analyzed in a specialized embryonic development pipeline that uses high-resolution 3D imaging. All data is quality controlled and analyzed by a dedicated informatics consortium and all abnormal phenotypes automatically compared to clinical features of human disease populations to identify robust mouse models of disease. We will present our discoveries into the enrichment of human Mendelian disease genes among the embryonic lethal strains, the wide-ranging sexual dimorphism of phenotypic traits in both wildtype and mutant mice, and the over 300 new mouse models of human disease now available for further studies. We will also discuss the latest findings in relation to metabolism, hearing and aging. The plethora of new genetic disease models as well as the basic and translational knowledge that has arisen from our analysis has recently earned the IMPC recognition by the G7 for being the first global research infrastructure for the life sciences.

1780W

Associated anomalies in cases with esophageal atresia. *C. Stoll, Y. Alembik, B. Dott, MP. Roth.* Faculté de Médecine, Strasbourg, France, France.

   Esophageal atresia (EA) is a common type of congenital anomaly. The etiology of esophageal atresia is unclear and its pathogenesis is controversial. Infants with esophageal atresia often have other non-EA associated congenital anomalies. The purpose of this investigation was to assess the prevalence and the types of these associated anomalies in a defined population. The associated anomalies in cases with EA were collected in all livebirths, stillbirths and terminations of pregnancy during 29 years in 387,067 consecutive births in the area covered by our population-based registry of congenital malformations. Of the 116 cases with esophageal atresia, representing a prevalence of 2.99 per 10,000, 54 (46.6%) had associated anomalies. There were 9 (7.8%) cases with chromosomal abnormalities including 6 trisomies 18, and 20 (17.2%) nonchromosomal recognized dysmorphic conditions including 12 cases with VACTERL association and 2 cases with CHARGE syndrome. Twenty five (21.6%) of the cases had multiple congenital anomalies (MCA). Anomalies in the cardiovascular, the digestive, the urogenital, the musculoskeletal, and the central nervous systems were the most common other anomalies. The anomalies associated with esophageal atresia could be classified into a recognizable malformation syndrome or pattern in 29 out of 54 cases (53.7 %). This study included special strengths: each affected child was examined by a geneticist, all elective terminations were ascertained, and the surveillance for anomalies was continued until 2 years of age. In conclusion the overall prevalence of associated anomalies, which was close to one in two cases, emphasizes the need for a thorough investigation of cases with EA. A routine screening for other anomalies may be considered in infants and in fetuses with EA.

1781T

Defining requirements for cleavage of prelamin A by the zinc metalloprotease ZMPSTE24. *T. Babatz[1], E. Spear[2], S. Michaelis[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, MD.

Lamins are a family of intermediate filament proteins that polymerize to form the nuclear lamina, a structural network underlying the nuclear envelope. The nuclear lamina plays a role in nucleus structural integrity, heterochromatin organization, and nuclear pore complex distribution. The lamin A precursor, prelamin A, undergoes a multi-step maturation process which concludes with C-terminal cleavage of a 15 amino acid farnesylated tail, yielding mature lamin A. When this final cleavage is disrupted either by mutations of the gene encoding the substrate, prelamin A, or the protease, ZMPSTE24, accumulation of the persistently farnesylated form of lamin A causes a spectrum of genetic disorders including the premature aging disease Hutchinson Gilford Progeria Syndrome (HGPS) and the related progeroid disorders mandibuloacral dysplasia (MAD-B) and restrictive dermopathy (RD). The structure of ZMPSTE24 has recently been solved, revealing a novel and fascinating integral membrane CAAX protease. The seven membrane spans form a hollow chamber enclosing an interior active site with narrow side portals through which substrates access the chamber. Prelamin A is the only known mammalian substrate for ZMPSTE24, however, the basis of this specificity remains unclear. To further define the sequence requirements for recognition of prelamin A by ZMPSTE24, we are performing a comprehensive mutagenesis experiment of residues flanking the cleavage site, using degenerate oligonucleotide mutagenesis to mutate each of the ten residues flanking the scissile bond (VTRSYêLLGNS). These mutants are tested in an *in vivo* humanized yeast assay for prelamin A cleavage by ZMPSTE24. Prelamin A and ZMPSTE24 are exogenously expressed in yeast and cleavage efficiency is visualized on an SDS-PAGE gel. Already we can begin to discern a pattern for one position, L647, of which mutation to arginine has long been known to be "uncleavable." After assaying all 19 possible mutants of L647, charged residues, aromatics, and proline disrupt cleavage, while hydrophobic residues are well-tolerated. We expect a comprehensive analysis of the 10 residues surrounding the cleavage site will yield a "heat map" of cleavage efficacy for ZMPSTE24, provide mechanistic insights into the highly specific recognition of prelamin A by the protease, and define a consensus motif with predictive value to identify other ZMPSTE24 substrates.

1782F

A mouse model of Proteus syndrome. *M.J. Lindhurst[1], L.R. Brinster[2], H.C. Kondolf[1], M.R. Yourick[1], G. Elliott[3], C. Rivas[3], L. Garrett[3], J. Gomez-Rodriguez[3], P.L. Schwartzberg[3], L.G. Biesecker[1].* 1) MGMGB, National Human Genome Research Institute, Bethesda, MD; 2) Division of Veterinary Resources, Office of Research Services, NIH, Bethesda, MD; 3) GDRB, National Human Genome Research Institute, Bethesda, MD.

Proteus syndrome (PS) is characterized by progressive, mosaic, sporadic overgrowth that can affect any tissue of the body and is caused by the somatic activating mutation *AKT1* c.49G>A, p.(E17K). Common manifestations include bony overgrowth, cerebriform connective tissue nevi and vascular malformations that are often a mix of venous, capillary and lymphatic vessels and contain enlarged vascular channels. Many types of cysts, benign tumors and hyperplasia have been seen in patients with Proteus syndrome as well as overgrowth due to expansion of extracellular matrix in the tissue. Due to the observational nature of human studies, we have only been able to describe the natural history of the disorder in patients and survey tissues collected by biopsy, surgery and occasionally autopsy, for abnormal growth, histology and mutation burden. Therefore, we developed a mouse model that allows us to investigate the role the PS mutation plays in many different tissue types. Using two different methodologies, we have been able to generate mice mosaic for activated AKT1. As in the human disease, each mouse had a unique constellation of abnormalities with affected and phenotypically normal tissues containing both mutation-positive and -negative cells. Average mutation levels in mosaic mice ranged from 5 to 40%. The most common finding was vascular malformations (VM) of mixed type that contained large ectatic vessels very similar to what is seen in patients with PS. VM were often found in lymph nodes, though they also occurred in the spleen, kidney and adipose tissue. Cysts, hyperplasia and ductal ectasias were also noted. Mammary hyperplasia with ductal ectasia was found in 9 of 10 mice. However, there were some important differences. In human PS, manifestations of the disorder are usually detected in infancy or early childhood, usually between six and 18 months of age. In most of the mice, abnormalities were not detected until between 12 and 18 months, well into adulthood. In addition, very little expansion of extracellular tissue was observed in the mice. Finally, unlike in patients where the mutation is absent in peripheral blood samples, nearly all mosaic animals were positive at levels that ranged from 15 to 47%. Currently we are using immunohistochemistry to identify mutation-positive cells in several types of tissue to determine which cell types are most frequently positive in affected tissue and therefore are likely candidates to be driving the overgrowth.

1783W

Sex differences in mtDNA content and its relationship to mitochondrial enzyme activities in the context of human skeletal muscle aging. *M. Tesarova, M. Hůlková, J. Křížová, K. Beránková, H. Kolářová, H. Hansíková, J. Zeman.* Dept. of Pediatrics and Adolescent Medicine, Charles University, Faculty of Medicine and General University Hospital, Prague, Czech Republic.

The mitochondrial functional state is assessed mainly by respiratory chain enzyme activities. The coupling of mitochondrial DNA (mtDNA) content with enzyme activity could provide a more integrated view of the state. However, the context of age and gender is necessary to include in the view. The aim of this study was to analyze mtDNA content during aging, to compare its differences between males and females, and to characterize relation between mtDNA content and enzyme activities of COX, CS and SQR in set of 205 human skeletal muscle samples (age from 2 weeks to 78 years). The Spearman´s correlations between age and mtDNA content were significantly positive in both genders, although the more distinct correlation was observed in males (r = 0.626; p = 0.000) compared to females (r = 0.275; p = 0.008). Significantly positive correlation was found between COX, CS and SQR activities and mtDNA content in isolated mitochondria and COX activity and mtDNA content in muscle homogenate. The CS, COX and SQR activities were not found as age-group dependent but the activities normalized on mtDNA content depended significantly on the age-group (Kruskal-Wallis test: p = 0.002; p = 0.014; p = 0.023). Considering gender, the normalized activities of all three enzymes were significantly dependent on the age-groups only in males (Kruskal-Wallis test: p = 0.010, p = 0.005, p = 0.013). MtDNA content is changing during aging with gender specific scale and it may be considered an important normalization factor of mitochondrial enzymes especially in context of age and gender. *Supported by grants AZV 17-30965A, RVO-VFN64165, Progres Q26/LF1, and GACR 14-36804G..*

1784T

Mechanistic insight of inactivation of mouse chitinase-like protein Ym1. *F. Oyama, M. Sakaguchi, Y. Sugahara, K. Okawa.* Dept Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Japan.

Chitinase-like proteins (CLPs) are structurally homologous to chitinases but lack the ability to degrade chitin, a widespread environmental biopolymer of *N*-acetyl-D-glucosamine. Several CLPs have been identified in mice and humans. Mice express primarily breast regression protein-39 (BRP-39), Ym1 and Ym2, whereas humans produce YKL-40, the human homologue of BRP-39, but do not synthesize Ym1 and Ym2. Based on sequence similarities, CLPs belong to the family 18 of the glycosyl hydrolases. Family 18 of the glycosyl hydrolases includes two catalytically active mammalian chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase). It is generally assumed that the lack of chitinase activity in CLPs is due to the mutation of crucial residues within the conserved catalytic domain during evolution from a common ancestor. Ym1, one of the CLPs, is expressed in mouse and is overexpressed during inflammation caused by parasitic infections or allergic pulmonary inflammation, but its biological functions remain unknown. Ym1 is highly homologous to AMCase, but lack chitin hydrolyzing activity. To understand the structure and function relationship of Ym1, we introduced amino acid substitutions (N136D and Q140E), which compose the active site of family 18 chitinases. Although the mutated Ym1 possessed the same sequence of the conserved active site with AMCase, it remains inactive. This result indicates that lack of chitinase activity in Ym1 is not due to merely the simple amino acid substitution in the active site.

1785F

Gene expression profiling of puberty-associated genes reveals abundant tissue and sex-specific changes across postnatal development. *H. Hou[1,2], L. Uusküla-Reimand[1,3], M. Makarem[1], S. Saleh[1,4], C. Corre[1], A. Metcalf[1], A. Goldenberg[1,4], M. Palmert[1,5,6], M. Wilson[1,2].* 1) Sickkids Research Institute, Toronto, ON, CanadaGenetics and Genome Biology Program, SickKids Research Institute, M5G 04A, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, M5S 1A8, Toronto, Ontario, Canada; 3) Department of Gene Technology, Tallinn University of Technology, 12616 Tallinn, Estonia; 4) Department of Computer Science, University of Toronto, M5S 2E5, Toronto, Ontario, Canada; 5) Division of Endocrinology, Hospital for Sick Children, M5G 1X8, Toronto, Ontario, Canada; 6) Departments of Paediatrics and Physiology, University of Toronto, M5S 1A8, Toronto, Ontario, Canada.

The timing of human puberty is highly variable, sexually dimorphic, and associated with adverse health outcomes. However, we still know little about the spatial, temporal or postnatal expression patterns of the majority of these candidate puberty-associated genes. Rare mutations have been identified for over 20 pubertal disorders, many of which disrupt critical components of the hypothalamic-pituitary- gonadal (HPG) axis. More than 200 candidate genes at loci associated with age at menarche or voice breaking in males have been identified by genome-wide association studies (GWAS). To gain insight into the tissue and cell-type specific expression of candidate puberty associate genes, we first leveraged large-scale human and mouse gene expression data sets including microarray, bulk and single cell RNA-seq data. We found enrichments for pituitary gland, pineal gland, hypothalamus, and several additional brain regions. To gain insight into the temporal and sex specific expression of candidate puberty genes we used a high-throughput and sensitive microfluidic quantitative PCR strategy, to profile the gene expression patterns of the mouse orthologs of more than 175 puberty-associated genes in male and female mouse HPG axis tissues, the pineal gland, and the liver at five postnatal ages spanning the pubertal transition. The most dynamic gene expression changes were observed prior to puberty in all tissues. We detected known and novel tissue-enhanced gene expression patterns, with the hypothalamus expressing the largest number of the puberty-associated genes. While few hypothalamus genes showed sex-biased gene expression, over 40 puberty-associated genes in the pituitary gland showed sex-biased gene expression, most of which occurred peri-puberty. These sex-biased genes included the orthologs of candidate genes at GWAS loci that show sex-discordant effects on pubertal timing. Overall, our findings provide new insights into the expression of puberty-associated genes and support the possibility that the pituitary plays a role in determining sex differences in the timing of puberty.

1786W

SRY potentially regulates early dopaminergic differentiation from male hiPSCs. *D.D. Cao, W.N. Law, K.K. Miu, J.J. Tu, T.Y. Ha, H.H. Chueng, W.Y. Chan.* School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong, Hong Kong.

Protein of Sry, the sex-determining gene, was found to co-localize with Tyrosine Hydroxylase (TH) positive neurons in male rat substantia nigra and in postmortem male human brains. Knockdown of Sry in male rats caused motor dysfunction making it a candidate capable of exerting male-specific effects on brain sexual dimorphism. However, it is unknown if SRY is a gene that affects neurodevelopmental process in male. To answer this question, we utilized human induced pluripotent stem cells (hiPSCs) to establish an in vitro floor plate (FP) based dopaminergic (DA) neuron differentiation system. We also generated stable male hiPSC line with ectopic activation of SRY. Our results showed that endogenous SRY was upregulated during the early DA neuron differentiation stage. When we ectopically overexpressed SRY during early differentiation, several midbrain DA neuron progenitor markers FOXA1, FOXA2, LMX1A, and MSX1 were significantly upregulated which indicated a possible enhancing role of SRY in early DA neuron differentiation. In addition, from published CHIP-seq datasets of SRY, we found that SHH and FOXA1 (important mediators in the development of DA neurons) are two potential downstream targets of SRY. In the future, knockout of SRY will be conducted utilizing CRISPR/Cas9 technology to examine if early DA neuron differentiation was inhibited after SRY knockout. Furthermore, RNA-seq will be performed to examine global gene expression changes under different SRY expression status. Finally, CHIP-qPCR will be used to testify the candidate downstream targets bound by SRY in the early DA neuron differentiation process. We hope, through our study, we can understand better the molecular regulatory function of SRY in male-specific neurogenic process.

1787T

An iPSC approach to examine the molecular mechanisms underlying *SRCAP* mutations in Floating-Harbor syndrome. *R.L. Hood[1,2], K.M. Boycott[2,3,4], W.L. Stanford[1,5,6], D.E. Bulman[2,4].* 1) Department of Biochemistry, Microbiology and Immunology, University of Ottawa; 2) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa; 3) Department of Genetics, Children's Hospital of Eastern Ontario; 4) Department of Pediatrics, University of Ottawa; 5) The Sprott Centre for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute; 6) Department of Cellular and Molecular Medicine, University of Ottawa.

Floating-Harbor Syndrome (FHS; OMIM# 136140) is characterized by short stature, delayed osseous maturation, cognitive deficit, and unique dysmorphic facial features. We previously employed an exome-sequencing approach and identified heterozygous truncating mutations of *SRCAP* as the cause underlying FHS. More recently, we identified a specific DNA methylation "epi-signature" in the peripheral blood of FHS individuals by high resolution genome-wide DNA methylation analysis. SRCAP has been described as having coactivator roles in CREB and CBP-mediated and nuclear (steroid) hormone receptor signaling pathways and has also been shown to co-precipitate as part of a large chromatin remodeling complex which catalyzes ATP-dependent displacement of the histone variant H2A by H2A.Z. Based on the described roles for SRCAP and our finding that FHS-causing *SRCAP* mutations result in global methylation alterations, we expect that these mutations result in widespread gene dysregulation. Furthermore, the existence of individuals with deletions encompassing all of *SRCAP* who do not have FHS, leads us to postulate that the mechanism of disease is not due to haploinsufficiency. To examine the functional consequences of *SRCAP* mutations in cell lineages implicated in the FHS phenotype, we have utilized an induced pluripotent stem cell (iPSC) disease-modeling approach. FHS patient and gender-matched control fibroblast samples were reprogrammed into iPSCs using an episomal vector approach and subsequently characterized by pluripotency cell surface marker expression, in vitro embryoid body differentiation, and in vivo teratoma formation assays. We are currently differentiating these cells into: chondroprogenitors, to examine the growth deficiency; neural crest cells, to investigate the facial dysmorphology; and a neural lineage, to examine the cognitive deficiency aspect of FHS. RNAseq will be employed to examine global transcript expression and identify downstream targets and/or pathways which are perturbed within the specific cell lineages of FHS patients. These results will rationalize the various phenotypic aspects seen in this disease in a precise tissue specific manner and provide insight into the disease mechanisms seen in FHS.

1788F

ROS induced oxidative stress up-regulates DNA repair gene uracil DNA glycosylase, a potential anti-leishmania drug target. *A. Mishra[1], M.I. Khan[1], P.K. Jha[1], P. Das[1,2], K.K. Sinha[1].* 1) Biotechnology, National Institute of Pharmaceutical Education and Research, Hajipur, Bihar, India; 2) Rajendra Memorial Research Institute of Medical Sciences Agamkuan, Patna, Bihar-800007, India.

Introduction Currently used anti-leishmanial drugs are experiencing resistance and toxicity concerns, making identification of novel drug targets of seminal importance. *Leishmania donovani* is an intracellular parasite that able to survive and replicate concomitantly under lethal oxidative environment of host macrophages. Between different survival strategies for parasite, DNA repair mechanisms have its own exquisite importance due their ability to repair different type of oxidative lesions. Base excision repair (BER) is the predominant in correcting oxidative damage, but details of BER have not been explored in *L. donovani*. Uracil DNA Glycosylase (UNG) is one of the essential enzymes involved in BER pathway, solely for its upfront capability of detecting and isolating the lesions. **Methods** Initially, *Leishmania* cell extract and recombinant purified LdUNG was checked for the glycosylase activity using synthetic 47 base pair oligonucleotide as a substrate at different physiological conditions. The effect of overexpression of UNG on survival and growth of *L. donovani* was checked by MTT and ARP assay against oxidative agents. The ex-vivo infectivity of UNG overexpressed cells were determined in macrophages. Transcriptional level of LdUNG was determined in different condition of oxidative stress induced by ROS or antileishmanial drugs by qRT-PCR. **Results** We demonstrate the role of LdUNG in protection against DNA damaging agents and oxidative stress. Apart from this, we show that UNG is indirectly related to the mechanisms involving drug resistance in Leishmania. First, we demonstrated the presence of glycosylase activity in leishmanial lysate and then, in heterologously produced and purified recombinant LdUNG. We find that over-expression of enzyme in the parasite confers tolerance against oxidative stress by correcting DNA damage and surprisingly against antileishmanial drugs which are in use. Also, our studies revealed that stressed parasites or resistant isolates exhibit an up-regulated levels LdUNG transcription. Interestingly, overexpression of UNG enhances the ex-vivo macrophage infectivity of the parasite. **Conclusion** Our results emphasize the role of BER pathway in establishment of infection in human macrophage. Added to this, association of UNG expression with resistance against commonly used antileishmanial drugs opens a new window of opportunity to target LdUNG for development of combination therapy against leishmaniasis.

1789W

Generation of humanized CD4 knock-in mice using CRISPR/Cas9. *K.C. Chen[1], P.L. Chen[1,2,3,4].* 1) Graduate InsGraduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwantitute of Medical Genomics and Proteomics, National Taiwan Univ., Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan.

CD4, a glycoprotein expressed on the surface of helper T lymphocytes, is a critical molecule for immune regulation. CD4 is necessary to activate helper T cell by binding to class II major histocompatibility complex (MHC) molecules. Among the 4 domains, the Dl and D2 domains of CD4 are important regions to interact with class II MHC molecule. With the goal to study human immunity and autoimmunity using mouse as a model organism, we constructed a humanized CD4 knock-in mice which retains the promoter and regulatory elements of the endogenous mouse Cd4 gene and produced the chimeric CD4 molecule that contains all D1 and D2 domains of human CD4. The exons 3-5 of the endogenous mouse *Cd4* allele were replaced with the homologous human *CD4* gene sequence via CRISPR/Cas9 approach to make a humanized CD4 knock-in mouse strain. Through another project, we are also working on a humanized MHC class II knock-in mouse. Combination of the humanized chimeric CD4 mouse and humanized MHC class II knock-in mouse will benefit immunological modeling applications in the future.

1790T

The association of haploinsufficiency of ARID2 with Ras-MAPK signaling pathway. *M. Kang[1], B. Lee[1,2], J. Lee[1], G. Kim[2], I. Choi[2], J. Choi[2], E. Seo[2], H. Yoo[1,2].* 1) Asan Institute for Life Sciences, Seoul, South Korea; 2) Medical Genetics Center, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, Korea.

Somatic haploinsufficiency of *ARID2* has been reported in human cancer tissues including hepatic cellular carcinomas. Recently, its germline heterozygous loss-of-function mutations were identified in patients with neurodevelopmental disorders and Coffin-Siris like phenotype. Previous investigations on the tumor tissues with ARID2 haploinsufficiency suggested that its suppression is related to increased activity of Ras-MAPK pathway. In the current study, we described one patient with 12q12 microdeletion including *ARID2*. This patient had facial dysmorohism, short stature, cardiac defect, and intellectual disability. *ARID2* haploinsufficiency was confirmed by chromosome microarray, RNA and protein expression study. Transient knock-out study using sh-RNA-*ARID2* in Hela cells indicated that suppression of ARID2 is related with increased Ras-MAPK activity but no significant changes in the activity of AKT and mTOR signaling pathway was observed. In addition, transient KO study of two other genes, included in the deletion region, did not show any change of activity of Ras-MAPK, AKT and mTOR signaling pathway. ARID2 is a member of BAF (SWI/SNF-B) chromatin remodeling complex, and is expected to regulate transcriptional activation and repression of select genes by chromatin remodeling. The results of study indicates that haploinsufficiency of ARID is related with enhanced Ras-MAPK pathway. Further study is required for the association of its deficiency with dysmorphism and intellectual disability.

1791F

30 kDa fragment of beta-dystroglycan co-immunoprecipitates with Dp71 isoforms in PC12 cells. *C.O. Azotla Vilchis, J. Aragón Medrano, V. Ceja Orozco, R.M. Bermúdez Cruz, S.C. Montañez Ojeda.* Department of Genetics and Molecular Biology, CINVESTAV, Mexico city, Mexico.

Dystrophin Dp71 is a protein encoded by the *DMD* gene (MIM 300377) and its expression is relevant for nervous system function. Dp71 isoforms are generated by alternative RNA splicing, which are grouped according their carboxi terminal end in Dp71d, Dp71f and Dp71e. Each isoform has different subcellular localization; however, specific function of each isoform remains unknown. Beta-dystroglycan is part of the dystrophin-associated protein complex (DAPC). This complex allows multiple protein interactions and regulates intracellular localization and function of other proteins. A matrix metalloproteinase cleaves the extracellular domain of beta-dystroglycan and creates a 30 kDa fragment of this protein. Our group aim is to elucidate the function of Dp71 isoforms. For this, we are studying the protein complexes in which Dp71 isoforms are involved. Experimental strategy: PC12 cell stable clones overexpressing recombinant Myc-Dp71a, Myc-Dp71c and Myc-Dp40 proteins were generated. The cross-link with paraformaldehyde was standardized and total protein lysates of each clone were obtained. Cross-linked complexes were immunoprecipitated with Myc polyclonal antibody/protein G-sepharose and analyzed by blue native polyacrylamide gel electrophoresis (BN-PAGE). To visualize complexes containing recombinant Myc-Dp71 and beta- dystroglycan proteins, Western blot assays were carried out. Results: We found at least 3 complexes of different molecular weight containing specific Dp71 isoform in PC12 cells. Abundance of each complex is different between the different complexes. Beta-dystroglycan co-immunoprecipitated with Dp71/Dp40, being the 30 kDa beta-dystroglycan present in these complexes. These results demonstrate that 30 kDa beta-dystroglycan is one of the components of DAP-Dp71 complexes in PC12 cells, a well-established cell line model to study neuronal functions.

1792W

Analysis of copy number variation and association with facial shape in a large cohort of Bantu African children. *F. Yilmaz[1,2], M. Sorenson[3], D. Astling[2,4], H-C. Yu[2], J. Cole[2,5], S.A. Santorico[3,5,6], R.A. Spritz[2,5], A.E. Hendricks[3,5,6], T.H. Shaikh[2,5].* 1) Integrative and Systems Biology Program, University of Colorado Denver, Denver, CO; 2) School of Medicine, University of Colorado Denver, Aurora, CO; 3) Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO; 4) Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Aurora, CO; 5) Human Medical Genetics and Genomics Program, University of Colorado Denver, Aurora, CO; 6) Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO.

Recent genomewide association studies by the FaceBase Consortium have identified several genes associated with normal facial shape and appearance. These studies focused on analyzing SNPs in population based cohorts with normal facial characteristics. Images from 3D morphometric cameras were used to quantify facial distances and shapes. Copy number variations (CNVs), are a considerable source of genetic variation. While CNVs have been implicated in numerous genetic disorders, they are understudied in genetic association studies of normal traits including normal facial variation. We hypothesize that CNVs may play a role in the variability observed in normal facial appearance. To test this, we analyzed genotype and phenotype data from ~3700 Bantu Africans. CNV calls were generated by applying PennCNV, DNAcopy, and VanillaICE, to the raw intensity files obtained from the SNP array-based genotyping. After quality control to remove low-quality samples and CNVs, > 410,000 CNVs from 3463 individuals remained (~120 CNVs/individual). CNVs were further classified by size and probe content and annotated for overlap with known genes and CNVs. Although the vast majority of the CNVs were < 100 Kb in size, we identified 1482 CNVs that were ≥ 300 Kb, many of which overlapped regions previously implicated in genetic syndromes. CNVs were further classified based on the number of times they were observed in unrelated individuals. We observed ~500 CNVs in over 5% of the cohort, and ~2000 CNVs in 1-5% of the cohort, while the vast majority were seen in <1% of the cohort. We next examined the association between CNVs and 28 quantitative facial measurements, including 25 linear distances and 3 measures of overall face size, using a linear mixed effects model to control for cryptic relatedness and including known covariates (e.g., age, sex). We used two primary models; in model 1, deletions and duplications, have the same direction and magnitude of effect, while in model 2, deletions and duplications have opposite direction of effect. We identified several CNVs associated with facial measurements and many of our top hits overlap genes relevant to craniofacial development. For example, a CNV associated with centroid size (p-value = 5.92e-5, effect size = -4.44) overlaps *SLC2A3*, a gene expressed in craniofacial precursors. In addition to assessing the role of CNVs in facial shape and morphology, this study also provides a resource of CNVs generated from a large cohort of Africans.

1793T

***ARMC5* and PMAH: From human genetic defects to the *Armc5*+/- mouse.**

F.R. Faucz[1], A. Berthon[1], S. Espiard[2], L. Drougat[1], J. Bertherat[2,3], C.A. Stratakis[1]. 1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA; 2) Institut Cochin, INSERM U 1016, CNRS UMR8104, Université Paris Descartes; 3) Department of Endocrinology, Referral Center for Rare Adrenal Diseases, Assistance Publique Hôpitaux de Paris, Hôpital Cochin, 75014 Paris, France.

Bilateral adrenal hyperplasias (BAH) that lead to Cushing syndrome (CS) are linked to abnormalities of the protein kinase A (PKA) pathway. Almost half of all cortisol-producing adenomas harbor somatic mutations of the main PKA catalytic subunit *PRKACA* or Ca. In 2013, the first inactivating mutations in the Armadillo repeat-containing 5 (*ARMC5*) gene were identified in a form of BAH called primary macronodular adrenal hyperplasia (PMAH). Since, we have reported several new germline *ARMC5* mutations in patients with PMAH (*J Clin Endocrinol Metab. 2014;99:E1113* and *J Clin Endocrinol Metab. 2015;100:E926*) and additional ones in patients with primary hyperaldosteronism (*J Clin Endocrinol Metab. 2015;100:E900*). To better understand PMAH caused by Armc5 haploinsufficiency, we generated and characterized a new mouse model of *Armc5* deficiency. *Armc5* heterozygote mice (*Armc5*+/-) developed normally but at the age of 1 year, their corticosterone levels decreased, and this was associated with a decrease of PKA activity that was associated with a reduction of Ca expression at the RNA and protein levels. Similar decreases in PKA activity and Ca expression were also seen in tumors from patients with PMAH and *ARMC5* defects. Like in humans with PMAH, a subgroup of *Armc5*+/- mice developed hypercorticosteronemia at 18 months of age; interestingly, this was then associated with increases in both PKA activity and Ca expression. Adrenocortical tissue analysis from *Armc5*+/- mice at 18 months showed an abnormal activation of the Wnt/b-catenin signaling pathway in a subset of zona fasciculata cells. In conclusion, *Armc5* haploinsufficiency leads to age-dependent changes and CS in mice and this involves PKA, its catalytic subunit Ca, and the Wnt/b-catenin pathway. *F.R.F. and A.B. contributed equally to this work .

1794F

Establishment of primary cell lines from normal and abnormal human products of conception.

D. O'Day, T. Naluai-Cecchini, I. Glass. Department of Pediatrics, University of Washington, Seattle, WA.

Cell lines established from normal and abnormal human fetal donors can be a vital resource for molecular and cellular analyses of human development. For example, dermal fibroblasts are a mitotically active cell enabling observation of cell cycle dependent processes, such as ciliogenesis; they are validated starting materials for the generation of induced pluripotent stem cells (iPSCs); and they can be used to test for the presence of mosaicism. Here, we establish several primary cell lines from multiple fetal tissues including skin, brain, lung, spleen, liver, and kidney from both normal and abnormal human fetal donors. From 46 donor skin samples, we successfully established 38 different fibroblast cell lines derived from a wide range of congenital disorders such as Trisomy 21 and multiple congenital anomaly cases. For the 8 cases that failed to establish stable cell lines, 6 were exposed to unusually long intra-uterine fetal demise intervals (more than 24 hours) and 2 cases were lost as the result of technical failures. For 50 additional cases, we established astrocyte cell culture from cerebral cortex collected from donors 7 to 19 post-conception weeks (PCW); 16 of these cases have been successfully distributed to recipients of the laboratory for downstream applications. As a collaborative effort, we have also isolated 4 lung epithelial cell lines from 16 to 19 PCW cases as well as splenic stromal cell lines from 7 cases at 11 to 18 PCW for distribution to researchers for their specific projects. Cryogenically preserved cell lines generated from these otherwise difficult to obtain human fetal samples are a valuable resource for biomedical research. To facilitate research on normal and abnormal human development, the Laboratory of Developmental Biology will continue to establish, distribute, and store such cell lines for future use by biomedical researchers.

1795W

Genomic characterization of human induced pluripotent stem cells after CRISPR/Cas9 fluorescent tagging. *T.S. Grancharova, R.N. Gunawardane, Allen Institute for Cell Science.* Allen Institute for Cell Science, Seattle, WA.

The goal of the Allen Institute for Cell Science is to study and model cell behavior using image-derived data from human induced pluripotent stem cells (hiPSCs). We use the CRISPR/Cas9 system to fluorescently tag gene targets comprising the major signaling complexes and molecular machines in WTC hiPSCs. Because our goal is to model normal, non-cancer cell behavior, we are assessing the effects of gene editing and prolonged culturing on the transcriptomes and genomes of edited cell lines as part of our quality control protocol. To establish the baseline state of our unedited WTC line, we performed RNA-Seq on cell populations from passage numbers 8, 10, and 14 and found that their expression profiles are highly correlated indicating a relatively stable transcriptome in early passage unedited populations. Differential expression analysis comparing independently isolated clones from multiple edited lines (at passage ~25-35) to the unedited parental line revealed a set of ~50 genes that are down regulated (log2 fold change < -1; p<0.05) in all edited lines. This set of genes is enriched for gene ontology terms related to developmental processes suggesting that the editing/culturing process selects for cells in a more stem like state. In addition to characterizing the transcriptional profiles of edited cell lines, we looked for signatures of off-target Cas9 cutting using Sanger sequencing. For our initial assessment of off-target editing, we sequenced the top 8-10 predicted off-target locations for each crRNA used in generating ten clonal lines fluorescently tagged at different locations. Looking at 3-5 candidate clones for each line, we found no indels at the predicted cut sites indicating the absence of off-target cutting. We identified the baseline genomic variants in the unedited WTC line from 100X whole genome sequencing data and will use whole exome sequencing of each fluorescently tagged line to further assess Cas9 off-target editing and to identify other genomic aberrations that may arise during standard culturing and the clonal line generation process. All edited cell lines and their associated image and genomic data will be openly available to the non-profit scientific community.

1796T

Assessment of the impact of variants in constrained non-essential splice sites in fifty-two thousand type 2 diabetes cases and controls. *J.M. Mercader[1,2], J. Flannick[1,2,3], M. Udler[1,2], A. Mahajan[4], C. Fuchsberger[5], Sidi Zhang[6,7,8] for the AMP-T2D Consortium.* 1) Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 2) Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; 3) Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts, USA. University of Oxford, Oxford, OX3 7BN, UK; 4) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 5) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; 8) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts 02115, USA.

   Previous studies of genetic variation in type 2 diabetes (T2D) have uncovered more than 100 susceptibility loci for type 2 diabetes, most of them driven by common variants in regulatory regions. While low-frequency and rare coding variants have been recently shown to play a minor role, the potential for identification of rare coding variation associated with T2D may be enhanced by increasing the sample size and by improving our understanding of the functional impact of coding variation. Recent findings reported by Zhang et al. (Biorxiv, https://doi.org/10.1101/129312) identified specific nucleotide substitutions in non-essential splice sites near exon-intron boundaries that are on average more deleterious than missense variants and have higher likelihood of having an impact on splicing. To assess whether these non-essential splice-variants contribute to the susceptibility of T2D, we tested the association of this group of variants with T2D in a multi-ethnic study comprising whole exome sequences from 24,444 T2D cases and 24,502 controls. A total of 60,214 variants disrupted non-essential donor or acceptor sites, of which 11,697 were observed in at least 5 individuals. The most significant variant was located 5 base-pairs 5' to an exon-intron boundary of *WDR13* and was associated with 3.5 fold increased risk of developing T2D (MAF=0.0006, OR= 3.45, p=3.6x10[-6]). Multiple lines of evidence converge on a potential contribution of this gene to T2D risk. *WDR13* is a loss-of-function intolerant gene (pLI=0.94, ExAC), and lack of this gene in mice results in enhanced pancreatic beta cell proliferation, hyperinsulinemia and mild obesity. We also show that individuals carrying this variant have higher triglyceride levels. Replication in additional datasets and functional analyses will be needed to confirm the association with T2D and its impact on splicing of the *WDR13* gene. The second top variant was located 5 base-pairs 5' to an exon-intron boundary of the *HMGA2* gene, which despite not reaching experiment-wide significance (MAF=0.04, OR=1.26, p=5.5x10[-4]) is located in a region previously associated with T2D, and therefore suggesting an effector gene for this locus. In summary, analyzing constrained non-essential splice sites in the largest T2D exome sequencing project to date, we identified a novel large effect association with T2D, point to a possible effector transcript for a known T2D locus, and begin to demonstrate the impact of such type of variation in T2D.

1797F

Replication of newly identified type 2 diabetes candidate gene variants in Northwest Indian population groups. *V. Sharma[1], I. Sethi[1], I. Sharma[1], G. Singh[1,3], A. Mahajan[2], A. Angural[1], A.J.S. Bhanwer[3], M.K. Dhar[2], V. Singh[1], E. Rai[1], S. Sharma[1].* 1) Human Genetcs Research Group, Department of Biotechnology, Shri Mata Vaishno Devi University, Katra, Katra, India; 2) Department of Biotechnology, University of Jammu, 180006, India; 3) Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

   Recent large-scale common and rare variant genome-wide association studies (GWAS) have proven to be a powerful tool for the identification of new susceptible loci of T2D. Replication of those variants in different populations are of paramount importance, especially in Indian population which actually is conglomeration of different ethnic groups.

Aim
The aim of the study is to replicate newly identified T2D susceptibility loci, variants rs998451 and rs6723108 of *TMEM163*; rs7607980 of *COBLL1* and rs2296172 of *MACF1* in different ethnic groups of Northwest Indian population.

Method
We perform genotyping of the variants using Taqman allele discrimination assay using Real time Pcr**Results**Out of 4 SNPs examined in this study 3 did not show any significant association with T2D. The variant rs2296172 of gene MACF1 shows significant association with T2D with an expected odds ratio of 1.3 with confidence interval of 1.07-1.60 (p = 0.0009). We further evaluate variation rs2296172 of *MACF1* in endogamous Bania group specifically of Punjab region from the pooled population of Northwest India.

Discussion
Interestingly the results showed enhanced risk of T2D in Bania endogamous group than pooled population. These findings suggest the higher importance of studies within specific ethnic groups rather than conventional practice of pooling samples based on Geographical/Regional or Linguist affiliations like North or South Indian etc.

1798W

HLA imputation and allelic associations with type 1 diabetes in African Americans. *C.C. Robertson, S. Onengut-Gumuscu, W.M. Chen, S.S. Rich.* Center of Public Health Genomics, University of Virginia, Charlottesville, VA.

Background: Type 1 diabetes (T1D) is an autoimmune disease due to destruction of beta-cells in the pancreas. Risk of T1D is ~50% due to genetic factors, with the HLA region accounting for a majority of genetic risk. HLA allelic associations with T1D are robustly characterized in Europeans. Studies in African ancestry populations have been limited. **Methods:** We used the Illumina ImmunoChip and SNP2HLA software to impute classical HLA alleles in 1,036 T1D cases and 2,913 controls of African ancestry. Imputation used the Type 1 Diabetes Genetics Consortium (T1DGC) reference panel: 4,323 subjects of European, 251 African, and 622 with "other" ancestry. A lower bound of imputation accuracy was estimated using 50 previously HLA-typed subjects while excluding related subjects from the reference panel. Association with T1D was assessed for class I and II HLA loci and for classical alleles using logistic regression. Class II HLA haplotypes (*DRB1-DQA1-DQB1*) were inferred and haplotype associations were analyzed using logistic regression. Regression models were adjusted for population stratification using principal components. Likelihood ratio tests were used to assess association with T1D. **Results:** Estimated imputation accuracy (rho) of 4-digit classical alleles varied by locus. Lowest accuracy was seen at *HLA-B* and *HLA-DRB1* (rho=0.88). Accuracy was ≥ 0.93 at remaining loci, and highest accuracy was seen at *HLA-DQA1* (rho=0.99). All 8 class I and II HLA genes tested (*A, B, C, DRB1, DQA1, DQB1, DPA1, DPB1*) were significantly associated with T1D ($p<1\times10^{-7}$). The most significantly associated gene was *HLA-DQB1* ($p=2.2\times10^{-282}$). After adjusting for *HLA-DQB1*, four genes (*HLA-A, HLA-B, HLA-DRB1,* and *HLA-DQA1*) remained significant. The most significant classical allele was HLA-DQA1*03:01 (odds ratio (OR) = 5.5, $p=1.7\times10^{-144}$), and the most significant class II haplotype was HLA-DRB1*03:01-HLA-DQA1*05:01-HLA-DQB1*02:01 (OR = 3.9, $p=1.3\times10^{-73}$). While known European allelic and haplotype associations were recapitulated, African-specific associations were also identified, including the protective African-specific DR3 haplotype HLA-DRB1*03:02-HLA-DQA1*04:01-HLA-DQB1*04:02 (OR = 0.15, $p=8.6\times10^{-16}$). Evaluation of HLA associations with T1D in non-European populations may provide insights on the role of this most critically important locus on etiology of disease.

1799T

Genetic variability in energy expenditure and the risk of severe obesity. *A.C.P. Fonseca[1], B. Marchesini[1], V.M. Zembrzuski[1], J.R.I. Carneiro[2], J.F. Nogueira Neto[3], P.H. Cabello[1,4], G.M.K Cabello[1].* 1) Laboratory of Human Genetics, Oswaldo Cruz Institute, Rio de Janeiro, Rio de Janeiro, Brazil; 2) Clementino Fraga Filho University Hospital, Rio de Janeiro, Brazil; 3) Department of Pathology and Laboratory, Rio de Janeiro State University, Rio de Janeiro, Brazil; 4) Laboratory of Human Genetics, Grande Rio University, Rio de Janeiro, Brazil.

Obesity is defined as an increase in body fat mass which is enough to cause adverse health effects. The prevalence of obesity has increased rapidly, becoming a major public health problem worldwide. Such disease has a complex etiology, being clearly influenced by environmental and genetic factors. In this study, we aimed to investigate the difference in the molecular mechanism of individuals controlling energy expenditure which could disrupt the body homeostasis and may be a risk factor for obesity. A total of 391 adults subjects were selected, in which 189 were normal weight subjects (Body mass index, BMI: 22.8 [21.1; 23.9]) and 202 were severely obese individuals (BMI: 45.3 [39.7; 51.8]). Peripheral blood was collected for each patient, and we performed the biochemical (glucose, triglycerides, total cholesterol and fractions) and molecular analyzes. After genomic DNA was extracted, genotyping of *PPARGC1A* (rs8192678, rs3736265, rs2970847 and rs3755863) and *UCP1* genes (rs6536991 and rs12502572) were carried out by real time PCR (Taqman® assay). For the *FNDC5* gene, screening of exons 3, 4 and 5 as well as their intron-exon boundaries were performed using automatic sequencing. Clinical characteristic of the study population showed that anthropometric and biochemical data were significantly increased in the obese group. The exceptions were in HDL-cholesterol and height. Genotype and allele frequencies for all variants were obtained. Our results demonstrate a strong association between *PPARGC1A* rs2970847 and severe obesity susceptibility (χ^2=7.71; p=0.021). Allelic analysis showed that individuals carrying the rs2979847 (C) allele had a 1.91-fold increased risk of severe obesity (OR: 1.91 [1,2-3,03]; P = 0.003). The effect of *PPARGC1A* and *UCP1* polymorphisms on anthropometric and biochemical in the sample was analyzed by linear regression (adjusted by gender and age). *PPARGC1A* rs2970847 was associated with body weight (P = 0.039) and BMI (P = 0.030). Additionally, *UCP1* rs12502572 was associated with body weight (P=0.049). Furthermore, five rare variants were identified in *FNDC5* gene, which one is a novel missense mutation. The frequency of rs113173936 (AG) and rs72882318 (AG) were higher in severe obese group; however no statistic difference was found (P = 0.372 and P = 0.317, respectively). In conclusion, our study demonstrated that *PPARGC1A* and *UCP1* polymorphisms impact on severe obesity susceptibility.

1800F

Genome-wide meta-analysis of macronutrient intake identifies two novel loci: Cohorts for heart and aging research in genomic epidemiology. *J. Merino[1,2], H.S. Dashti[1,2], S.X. Li[3], A.E. Justice[4], M. Graff[4], C. Papoutsakis[5], C.E. Smith[6], G. Dedoussis[7], D.I. Chasman[2,8,9], T. Tanaka[10], A.Y. Chu[8,9] on behalf of the CHARGE Nutrition Working Group.* 1) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 2) Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA, USA; 3) MRC Epidemiology Unit, University of CambridgeCambridge, UK; 4) Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA; 5) Research International and Scientific Affairs, Academy of Nutrition and Dietetics Chicago, IL, USA; 6) Nutrition and Genomics Laboratory, Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, MA, USA; 7) Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece; 8) Division of Genetics, Brigham and Women's Hospital, Boston, MA, USA; 9) Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 10) National Institute on Aging, NIH, Baltimore, MD, USA.

Genome-wide association studies (GWAS) have facilitated the discovery of two relevant loci (*FGF21* and *FTO*) for macronutrient intake, an important modifiable risk factor for complex diseases with a significant heritable basis (h^2 ranging between 10 to 70%). Incomplete coverage of genetic variation and limited sample size in previous GWAS have precluded the discovery of additional loci. We conducted the largest GWAS to date of self-reported macronutrient intake in 91,112 participants from Cohorts for Heart and Aging Research in Genomic Epidemiology followed by replication in 32,545 participants and lookups in the UK Biobank (n=58,787). In addition, we applied complementary computational approaches to investigate mechanistic insights of loci associated with macronutrient intake. We identified 12 suggestively significant loci (P <10^{-6}) associated with intake of at least one macronutrient (expressed as percentage of total energy intake from carbohydrate, fat, and protein) in discovery meta-analysis. Four loci successfully replicated, revealing a novel common locus for carbohydrate intake [rs7619139 placed in a long-non coding RNA gene (*AC133680.1*), β=0.2, SE=0.05, $P_{1\text{-sided}}$=4×10^{-5}] and a novel low-frequency variant (MAF=0.01) for protein intake [rs77694286 near *DRAM1* (β =0.55, SE=0.19, $P_{1\text{-sided}}$=2.3×10^{-3})]. Both loci achieved genome-wide significance in the combined meta-analysis including results from discovery and replication cohorts (*AC133680.1*, $P_{combined}$=4.1×10^{-11}; *DRAM1*, $P_{combined}$=1.9×10^{-9}). Additionally, we confirmed previous findings between *FTO* and higher protein intake and *FGF21* with higher carbohydrate and lower protein and fat intake. All, but *DRAM1* locus, showed consistent directionality and effect-sizes in the UK Biobank lookups. We found a moderate concordant genetic correlation between protein intake and BMI (r=0.23, P=4×10^{-4}) indicating that genetic effects of higher protein intake are shared with higher BMI genome-wide. The novel variant in the *AC133680.1* gene is highly conserved across different species, involved in H1-neuronal progenitor cell differentiation (3.7 fold-change increase, P=1×10^{-3}) and primarily active in several brain regions, but not co-expressed with other quantitative loci linked mRNAs. Overall, the newly identified loci are likely to highlight relevant biological functions related to central nervous system control of food intake and body weight. Our findings reinforce current understanding of macronutrient intake biology.

1801W

Diabetes in cystic fibrosis and type 2 diabetes (T2D) have overlapping genetic risk architecture. *M. Atalar[1], B. Vecchio-Pagan[2], L.J. Strug[3], R.G. Pace[5], H. Corvol[4,7], J.M. Rommens[3], M.L. Drumm[6], M.R. Knowles[5], G.R. Cutting[1], S.M. Blackman[1].* 1) Johns Hopkins University School of Medicine, Baltimore, MD; 2) The Johns Hopkins University Applied Physics Laboratory, Laurel, MD; 3) The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Université Pierre et Marie Curie, Paris, France; 5) The University of North Carolina at Chapel Hill, Chapel Hill, NC; 6) Case Western Reserve University, Cleveland, OH; 7) Pediatric Pulmonology, Hôpital Trousseau, Paris, France.

Diabetes is a common age-dependent complication of cystic fibrosis (CF), affecting 40% of affected individuals by 30 years of age and 80% by 50 years of age. CF is an autosomal recessive disease caused by dysfunction of CFTR, a cAMP regulated chloride channel expressed in pancreas, intestine, lungs and sweat gland. Twin and sibling analysis indicated that variation in the age at onset of diabetes in CF is primarily determined by modifier genes. We report the results of genome-wide association and polygenic risk score analyses of 5364 unrelated subjects with CF derived from 4 study cohorts, 1307 of which had CF-related diabetes (CFRD). Association was tested using Cox proportional hazard regression within each cohort and fixed-effects meta-analysis on combined cohorts. Variants at two loci, *TCF7L2* (rs7903146; P=2.4E-11) and *SLC26A9* (rs4077468; P=8.29E-11), exceeded genome-wide significance. *TCF7L2* is the major genetic risk factor for type 2 diabetes (T2D) and *SLC26A9* encodes a chloride/bicarbonate exchanger that interacts with CFTR. The same *TCF7L2* SNV achieves the highest significance in CFRD and in T2D, and in the same direction. These findings provide critical replication of results derived from candidate studies (*TCF7L2*) and genome-wide association with a subset of subjects from the current study (*SLC26A9*). To further investigate genetic overlap between CFRD and T2D, polygenic risk scores were constructed using T2D-associated variants in the NHGRI-EBI Catalog of published GWASs. After pruning for MAF (≥5%) and LD (r²<0.8), genotypes were available for 155 T2D-associated SNVs. The T2D risk score was associated with CFRD (HR=1.02 per allele, 95% CI 1.02-1.03, P=6.0E-13, one-sided). A modified T2D risk score (removing 4 variants from previously reported candidate loci *TCF7L2*, *CDKAL1*, *CDKN2A/B*, and *IGF2BP2*), was also associated with CFRD (HR=1.02, P=5.0E-10, one-sided), indicating that additional T2D loci contribute to CFRD. A type 1 diabetes (T1D) risk score constructed with 30 SNVs (none typed in the *HLA* region) showed some evidence of association with CFRD (HR = 1.01, P=0.048, one-sided), suggesting that T1D risk alleles might also play a role in CFRD onset. These studies support overlap of disease mechanisms in T2D and CFRD. The high prevalence and early onset of diabetes in CF provide a unique opportunity to dissect the molecular mechanisms underlying CFRD and T2D. *Supported by CFF, NIH, Gilead Sciences, CF Canada, CIHR, Genome Canada.*

1802T

Impact of genetic variants identified in genome-wide association studies of diabetic retinopathy in Chinese patients with type 2 diabetes. *C.Y.Y. Cheung[1], C.H. Lee[1], K.H.M. Kwok[1], Y.C. Woo[1], M.M.A. Yuen[1], W.S. Chow[1], C.H.Y. Fong[1], R.L.C. Wong[1], A. Xu[1,2,3], P.C. Sham[4,5], K.S.L. Lam[1,2,3].* 1) Department of Medicine, The University of Hong Kong, Hong Kong; 2) Research Centre of Heart, Brain, Hormone and Healthy Aging, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 3) State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong; 4) Department of Psychiatry, the University of Hong Kong, Hong Kong; 5) Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong.

Background: Diabetic retinopathy (DR) is the most common microvascular complication of type 2 diabetes (T2DM). Novel DR-susceptibility loci have been identified from recent genome-wide association studies (GWAS) in various populations. If validated, these genetic markers may be used for risk profiling of DR in diabetic patients. This study aimed to examine the associations of these DR-associated single nucleotide polymorphisms (SNPs) with sight-threatening DR (STDR) in Chinese patients with T2DM. **Methods:** A total of 68 SNPs showing top association signals with DR ($P<5x10^{-4}$, $r^2<0.9$) in previous GWAS were genotyped in 3 groups of subjects: 1000 T2DM patients with STDR (STDR-cases), 2000 T2DM patients without retinopathy (non-STDR controls), and 1000 non-diabetic subjects recruited from the general population of Hong Kong (healthy-controls). STDR cases were defined as patients with either proliferative DR (PDR) or pre-PDR, or with clinically significant macular oedema. Multiple logistic regression models with adjustment for confounding factors were employed to examine for the independent association between the SNPs with STDR (STDR cases versus non-STDR controls; and STDR cases versus healthy controls). **Results:** When the STDR cases were compared with non-STDR controls, significant associations were observed at *IGSF21-KLHDC7A* rs3007729 (P=0.020; OR[95%CI]:0.86[0.76-0.98]), *SLC25A32* rs3098241 (P=0.046; OR[95%CI]:0.88[0.78-1.00]), and *COL5A1* rs7861012 (P=0.046; OR[95%CI]:0.88[0.78-1.00]), after adjustment for age, gender, duration of diabetes, hemoglobin A1c (HbA1c) and presence of hypertension. When the STDR cases were compared with healthy controls, significant association between *IGSF21-KLHDC7A* rs3007729 (P=0.021; OR[95%CI]:0.84[0.73-0.97]) was also detected. *BFSP2* rs1197310 (P=3.71x10⁻³; OR[95%CI]:1.23[1.07-1.41]), *CNTN5* rs10501943 (P=0.022; OR[95%CI]:1.66[1.08-2.57]), *FMN1* rs10519765 (P=0.037; OR[95%-CI]:0.80[0.64-0.99]), and *MAP3K7IP2* rs7772697 (P=0.044; OR[95%-CI]:0.84[0.70-0.99]) also demonstrated significant associations with STDR after adjustment for age, sex, HbA1c and presence of hypertension. None of these SNPs were associated with T2DM. **Conclusion:** We have successfully validated the significant and independent associations of several SNPs identified in previous GWAS with STDR in Chinese patients with T2DM. **Acknowledgement:** This study was supported by the Health and Medical Research Fund of the Food and Health Bureau, HKSAR (Project No. 03144016).

1803F

Chromatin accessibility landscapes in adipose tissue and preadipocytes at cardiometabolic trait loci. *K.W. Currin[1], M.E. Cannon[1], A. Safi[2], L. Song[2], Y. Wu[1], H.J. Perrin[1], M. Laakso[3], G.E. Crawford[2], K.L. Mohlke[1].* 1) University of North Carolina, NC.1.Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 2) 2.Department of Pediatrics, Division of Medical Genetics, Duke University, NC, USA; 3) 3.Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finl.

Epigenomic annotations can guide identification of regulatory variants at GWAS loci. We sought to functionally annotate cardiometabolic GWAS loci using adipose tissue chromatin accessibility and existing chromatin state data. We performed open chromatin profiling using ATAC-seq in the SGBS preadipocyte cell strain (2 replicates) and in frozen subcutaneous adipose tissue from 3 individuals from the METSIM study. We observed higher signal-to-noise and identified more ATAC-seq peaks in SGBS (~173,000-196,000) compared to adipose tissue (~36,000-58,000)(FDR<5%), which could reflect higher ATAC-seq efficiency in cells compared to frozen tissue. Compared to tissue-derived adipose nuclei promoter and enhancer chromatin states from the Epigenomic Roadmap, we observed more overlap with adipose tissue ATAC-seq peaks (85% of peak bases) than with SGBS ATAC-seq peaks (49% of peak bases). We tested for enrichment of GWAS loci for 11 cardiometabolic traits in our ATAC peaks and Roadmap adipose nuclei promoters and enhancers, controlling for gene proximity, allele frequency, and number of LD proxies. Loci for 6 traits, including waist-hip ratio adjusted for BMI, type 2 diabetes, and HDL, showed significant enrichment (Bonferroni P<4E-4) in both SGBS and adipose tissue peaks. Loci for 4 traits (triglycerides, LDL, cardiovascular endpoints, blood pressure) were significantly enriched in tissue peaks but not SGBS, suggesting that these loci may act through mature adipocytes or other adipose tissue cell types such as immune cells, vascular endothelial cells, or fibroblasts. Eight traits showed a stronger enrichment in tissue or SGBS peaks compared to the Roadmap adipose nuclei enhancers, suggesting that chromatin accessibility can pinpoint disease-relevant subsets of enhancers. We next compared our ATAC-seq peaks to variants at 110 cardiometabolic GWAS loci with coincident adipose eQTLs. Of 6,746 LD proxies (LD $r^2>0.8$) at these loci, 124 proxies at 52 loci overlapped both a tissue and SGBS peak, including 73 variants overlapping transcription factor motifs from JASPAR. Several of these motifs correspond to core adipose transcription factors, including PPARG, RXRA, ETS1, and RFX2. While a subset of cardiometabolic loci may act through non-adipocyte cells found in adipose tissue, our results suggest many loci may alter the function of preadipocyte active regulatory elements present in cell culture and primary tissue.

1804W

Multivariate genome wide association study uncouples "favourable" from "unfavourable" adiposity alleles. *Y. Ji, J. Tyrrell, S. Jones, R. Beaumont, M. Tuke, K. Ruth, R. Freathy, A. Murray, M. Weedon, A. Wood, T. Frayling, H. Yaghootkar.* Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK.

It is not clear if higher adiposity itself, or the adverse metabolic effects of higher adiposity, increases susceptibility to diseases such as colorectal cancer, depression and osteoarthritis. We aimed to identify genetic variants associated with "favourable adiposity" as a tool to test the effects of higher adiposity independently of adverse metabolic effects on non-metabolic outcomes. First, we identified alleles associated with higher body fat % ($P<5\times10^{-8}$) using 220,000 individuals from UK Biobank and published studies. Next, we performed a multivariate genome wide association study (GWAS) using published GWAS summary statistics of biomarkers of metabolic health: HDL-C, Triglycerides, fasting insulin, adiponectin, SHBG and a marker of liver fat. We defined "favourable adiposity" alleles as those which (i) had a multivariate $P<5\times10^{-4}$ (multiple test corrected) and (ii) were associated with the multivariate model stronger than body fat % alone. The latter criterion was set to uncouple "favourable" from "unfavourable" adiposity alleles. To validate the variants, we calculated the protective effect of "favourable adiposity" genetic score against cardiometabolic diseases. We performed Instrumental Variable analysis (IV) to estimate the causal effect of "favourable adiposity" on non-metabolic outcomes available in UK Biobank. We identified 85 independent genetic variants associated with 1.66-6.44% standard deviation (SD) per allele differences in body fat% ($P<5\times10^{-8}$). The multivariate model uncoupled "favourable adiposity" (e.g. the *PPARG* allele) from "unfavourable adiposity" alleles (e.g. *FTO* allele). We defined 14 "favourable adiposity" alleles; 10 previously published. In the genetic score analysis, each allele was associated with a 0.068 Kg/m² higher BMI (p = 2×10^{-62}) but lower risk of type 2 diabetes (0.96 OR; 2×10^{-21}), lower heart disease (0.98 OR; 6×10^{-9}) and lower systolic blood pressure (-0.156 mmHg; 1×10^{-7}). The IV analysis did not support causal association with depression, bone mineral density, bowel disease, cancer (any type) and osteoporosis (all P>0.3), but we could not exclude observational estimates of the associations. Our study supports the use of a multivariate analysis approach in identifying variants associated with higher body fat % but lower risk of cardiometabolic diseases. The identification of more "favourable adiposity" variants will provide additional power to test the role of higher adiposity without its adverse metabolic effects.

1805T

Whole exome sequencing and exome array genotyping in 3,943 Korean type 2 diabetes cases and controls. *S. Kwak[1], J. Chae[2], S. Lee[2], S.K. Choi[3], T. Park[3], K.A. Kim[4], J. Kim[2], K.S. Park[1].* 1) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea; 2) Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul, South Korea; 3) Department of Statistics, Seoul National University, Seoul, South Korea; 4) Biostatistics and Clinical Epidemiology Center, Samsung Medical Center, Seoul, Korea.

Type 2 diabetes (T2D) is a common complex disorder with strong genetic predisposition. Although different ethnic groups share most of the common variants of T2D, there might be rare or low frequency nonsynonymous variants that are specific to an ethnic group. In this study, we used whole exome sequencing to identify rare variants in 917 Korean T2D cases and controls (stage 1) and validated the findings in an independent set of 3,026 T2D cases and controls using customized exome genotyping array (stage 2). Whole exome capture was prepared using Agilent SureSelect version 4+UTR and sequencing was performed by Illumina HiSeq 2000. Nonsynonymous variants were incorporated into customized Axiom Biobank Plus Genotyping array. Genotype imputation was performed with 1,000 Genomes Project phase 3 reference. EPACTS software was used for association testing and METAL was used for meta-analysis of stage 1 and 2 results. We identified 728,838 variants using whole exome sequencing, which include 258,463 novel variants. A nonsynonymous variant in *PAX4* (rs2233580, R192H) was the only variant that was associated with T2D in genome-wide significance (OR = 1.81, $P = 6.36\times10^{-9}$). This variant was associated with lower age at diagnosis and decreased C-peptide level in T2D cases ($P < 0.05$). Other nonsynonymous variants that had suggestive association ($P < 1.0\times10^{-5}$) were located at *UVSSA*, *CRIPAK*, *GLP1R*, *TSPAN8*, *ARRDC2*, *KCNQ1*, *JMJD1C*, and *PLEKHG4B*. We also validated five known common variant loci (*MAEA*, *CDKAL1*, *PAX4*, *CDKN2A/2B*, and *KCNQ1*) to have genome-wide significant association in Koreans. Gene-wise analysis revealed that *SLC30A8* was most significantly associated with decreased risk of T2D ($P = 0.00010$) by Madsen and Browning method. We constructed a weighted genotype risk score using 32 variants that were confirmed in East Asians. T2D cases had significantly increased genotype risk score compared to controls (4.26±0.51 vs. 3.98±0.53, $P = 1.32\times10^{-57}$). The genotype risk score was significantly associated with lower age at diagnosis, lower body mass index, decreased C-peptide level, and decreased renal function in T2D cases ($P < 0.05$). It was also associated with increased HbA1c in controls ($P < 0.05$). In summary, we have identified several nonsynonymous variants suggestive to be associated with T2D in Koreans and validated previously confirmed common variants and gene-wise association using exome sequencing and exome array genotyping.

1806F

Identifying novel genetic variants associated with EHR-derived serum zinc in DiscovEHR study cohort. *R.P. Metpally, S. Krishnamurthy, D.T. Smelser, G.E. Breitwieser, D.C. Carey on behalf of the DiscovEHR collaboration.* Department of Molecular and Functional Genomics, GEISINGER CLINIC, DANVILLE, PA.

Background: Zinc (Zn^{2+}) is commonly measured in serum, and plays important roles in maintaining metabolism, cellular growth, wound healing, immune system and the breakdown of carbohydrates. Zn^{2+} is also needed for the senses of smell and taste. Zn^{2+} deficiency or overload can result in abnormal cellular function or damage. **Purpose:** Our goal was to evaluate the contribution of common genetic variation to normal physiologic variation in serum concentrations of the divalent metal. **Method:** We analyzed both whole exome sequencing (WES) and imputed Human Omni Express Exome bead-chip data from 2609 individuals with available outpatient serum Zn^{2+} values in the DiscovEHR cohort as described in Dewey et al 2016. We randomly split our data into two datasets to allow for discovery (70%) and replication (30%) by applying the two-step iterative resampling (TSIR) approach as described in Kang et al 2015. Mixed linear model based association analysis (MLMA) was performed on both the WES and imputed chip data using GCTA-v12.6 using median Zn^{2+} values. **Results:** A cut-off of discovering and replicating SNPs at least 20 times in 100 replications provided multiple variants associated with serum Zinc. Significant associations were found at two loci on 3q27.3 (most significant SNP rs7625980, P-value $9e^{-09}$; *HRG* gene) and 2p22.2 (most significant SNP rs11678807, P-value $1.18e^{-06}$; *VIT* nearest gene) from the imputed chip data. Coding variants from 3q27.3 locus were also identified from the WES analysis (most significant SNP rs1042445, P-value $2.3e^{-08}$; missense variant of *HRG* gene). The 3q27.3 locus contains multiple genes including the Histidine Rich Glycoprotein (*HRG*) gene which has been found to play a major role in transport and homeostasis of zinc and other metals. **Conclusion:** Zinc, an HRG ligand is essential in regulating the many processes of HRG protein such as regulating immune complex and pathogen clearance. HRG protein have potential prognostic/diagnostic value for cancer and the absence of the protein is also known to be associated with thrombophilia. This study provides important insights into genetic association with serum Zn^{2+} and identified genes that probably have important role in regulation of serum Zn^{2+} levels.

1807W

Evaluating tyrosine hydroxylase (*TH*) as a type 2 diabetes candidate gene in American Indians. *A. Nair, J. Sutherland, M. Traurig, S. Kobes, R. Hanson, C. Bogardus, L. Baier.* National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Phoenix, AZ.

We recently conducted a GWAS study in 7659 American Indians using a custom designed Affymetrix array. Among the top signals for type 2 diabetes (T2D) was rs10743152; this SNP had a significant sex interaction (P=0.0008) such that the T2D association was only significant in females ($P_{FEMALES}=4×10^{-7}$, OR=1.44). The nearest gene encodes tyrosine hydroxylase (TH) a rate limiting enzyme for the biosynthesis of catecholamines. Studies have shown a role of catecholamines, including dopamine, in the regulation of insulin secretion. Dopamine is synthesized in the beta-cells from circulating L-dopa and is co-secreted with insulin to provide a negative feedback loop for glucose stimulated insulin secretion in an autocrine manner. Based on the role of TH in dopamine synthesis, we hypothesized that *TH* may be the effector gene contributing to the T2D signal at rs10743152. Screening of the *TH* promoter identified two SNPs, rs10770140 (-581A/G) and rs10770141 (-824C/T) that are highly correlated (r^2=0.99) with our GWAS lead SNP rs10743152. Both SNPs have previously been shown to influence *TH* promoter activity in chromaffin cells, and rs10770141 has been shown to disrupt a SRY binding site. *SRY* is a male specific gene located on the Y chromosome, which could potentially explain the sex specific association seen with rs10743152. Direct genotyping of these 2 promoter SNPs identified an association with T2D only in females ($P_{T2D}<1×10^{-6}$, $P_{Sex Interaction}<0.001$). Based on the role of catecholamines in the regulation of blood pressure, we also analyzed the association of rs10770141 with hypertension and identified a significant association only in females ($P_{FEMALES}=0.0007$, OR=1.28). *TH* and *SRY* gene expression analyses using a human cDNA panel and pancreatic cell lines (NIT-1, NIT-2, α-TC6, 266-6 and AR42J) confirmed that *TH* is expressed in human pancreas and pancreatic cell lines, while *SRY* was detected in human pancreas and rat AR42J acinar cells. For functional studies, we cloned 1500 bps of the *TH* promoter, which contains 8 SNPs and forms 4 haplotypes (Hap 1-4), proximal to the luciferase gene in a pGL4.10 vector. All 4 haplotypes were examined in luciferase assays using NIT-1 beta-cells and a significantly higher luciferase activity was observed for Hap 1 and 3 compared to Hap 2 and 4. Hap 1 includes the diabetes risk alleles for rs10770141 and rs10770140. Further studies to identify the functional SNPs in Hap 1 and Hap 3 and to determine the effect of SRY are ongoing.

1808T

Glucose challenge metabolomics identifies C10- and C12-carnitines as possible contributors to insulin resistance. *C. Nowak[1], S. Trombley[1], S. Salihovic[1], C. Castillejo-Lopez[1], A. Ganna[2,3,4,5], C. Berne[6], N. Cook[1], C.D. Broeckling[7], J.E. Prenni[7], X. Shen[5], V. Giedraitis[8], J. Ärnlöv[9,10], L. Lind[11], J. Sundström[11], T. Fall[1], E. Ingelsson[1,12].* 1) Medical Sciences Department, Molecular Epidemiology, Uppsala University, Uppsala, SwedenDepartment of Medical Sciences and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, United States; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, United States; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, United States; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 6) Department of Medical Sciences, Clinical Diabetology and Metabolism, Uppsala University, Uppsala, Sweden; 7) Proteomics and Metabolomics Facility, Colorado State University, Fort Collins, CO, United States; 8) Department of Public Health and Caring Sciences, Geriatrics, Uppsala University, Uppsala, Sweden; 9) School of Health and Social Studies, Dalarna University, Falun, Sweden; 10) Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden; 11) Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden; 12) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, United States.

Background: Insulin resistance (IR) is mainly defined by a progressive impairment in insulin-stimulated uptake of circulating glucose. IR is an important risk factor for type 2 diabetes and cardiovascular disease. Metabolic challenges like the oral glucose tolerance test (OGTT) can reveal early pathogenic mechanisms not apparent in the fasting state. **Methods:** We studied 470 non-diabetic 71-year-old men in whom plasma samples obtained at 0 min, 30 min and 120 min during a 2-hour-OGTT were analyzed by untargeted metabolomics (ultraperformance liquid chromatography-time-of-flight mass spectrometry). IR was assessed by whole-body glucose uptake measured by hyperinsulinemic-euglycemic clamp. Our aim was to discover associations between IR and 192 metabolite trajectories during OGTT. **Results:** Nine trajectories were associated with IR. The most significant associations were observed for the medium-chain acylcarnitines C10 and C12. No associations between IR and OGTT trajectories were detected for short- or long-chain acylcarnitines. Concentrations of C10- and C12-carnitine decreased during OGTT in all men, but the decline was blunted in the most insulin resistant compared to more insulin sensitive persons. Acylcarnitines are intermediates of lipid metabolism and enable the transfer of activated fatty acids into mitochondria for beta-oxidation. Conflicting theories on their potential contribution to IR exist. To explore a possible causal role for medium-chain acylcarnitines in IR, we exposed murine 3T3-L1 adipocytes to different acylcarnitine species. We found evidence of blunted insulin-stimulated glucose uptake after exposure to C10- or C12-carnitine compared to other acylcarnitines and the control condition. **Conclusions:** These results suggest a possible causal role of C10- and C12-carnitine in the development of IR and highlight the importance of perturbation studies when dissecting complex polygenic traits.

1809F

A rare frameshift mutation in exon 9 of glucokinase regulator (GCKR) is associated with a severe progressive histological form of Nonalcoholic Fatty Liver Disease (NAFLD). *C.J. Pirola[1], T. Fernandez Gianotti[1], H. Dopazo[2], C. Rohr[2], G. Castaño[3], S. Sookoian[1].* 1) IDIM. University of Buenos Aires-CONICET, CABA, Buenos Aires, Argentina; 2) Biomedical Genomics and Evolution Laboratory. University of Buenos Aires-CONICET; 3) Liver Unit, Medicine and Surgery Department, Hospital Abel Zubizarreta, GCBA.

Background: Significant progress towards understanding the genetic component of NAFLD has been made, which uncovered common variants associated with small effects in the disease susceptibility. Exploration of plasma metabolome revealed that patients with NAFLD and Metabolic Syndrome show deregulated levels of amino acids, including enhanced glutamic acid/glutamine ratio. NAFLD is associated with changes in the expression of mitochondrial genes of the oxidative phosphorylation chain, as well as metabolic perturbations at the Krebs cycle (TCA)-level. We hypothesized that rare disease-associated variants in metabolic enzyme genes might be involved in NAFLD development. We performed an exploration of sequence variation in fourteen genomic loci that encode metabolic enzymes of the TCA-cycle, which included enzymes of the mitochondrial-matrix involved in glutamine and glutamate metabolism (*GLUD1*/*GLUD2*/*GLUL*/*GLS GLS2*), *ACO2* (second step of the TCA-cycle), *DLST* (2-oxoglutarate-dehydrogenase complex), isocitrate (*IDH1*/*IDH2*/*IDH3A*/*IDH3B*), malate and oxoglutarate dehydrogenases (*MDH1* and *OGDH*), and glucokinase regulator (*GCKR*). Methods: Sequencing of exons, exon-intron boundaries and 5' and 3'-UTR regions (62.1 Kb) was performed by next generation sequencing (> 60X coverage). Patients with NAFLD diagnosed by liver biopsy were included (*n* = 64); subjects without NAFLD (*n* =32) served as controls. Results: Deep sequencing revealed 132 single nucleotide polymorphisms (SNPs), including 17 novel SNPs. Clinical associations focused on nonsynonymous-SNPs showed a rare frameshift-mutation in exon 9 of *GCKR* (rs149847328 p.Arg227Ter, MAF<0.001) in a 48-years old patient with NAFLD, who had personal/family history of diabetes and hypertension in youth. Analysis of paired-liver biopsies at baseline and follow-up (5-years apart) showed progression of liver fibrosis from stage 0 to cirrhosis. Functional variant effect-prediction showed a high probability of deleterious effect (this variant introduces a stop-codon). Conclusions: The NAFLD phenotype could be modulated by rare variants, in particular its metabolic profile. *GCKR*, which is considered a susceptibility candidate gene for the Maturity-Onset Diabetes of the Young (MODY), might be involved in the pathogenesis of NAFLD. A search for this mutation might be recommended in NAFLD patients, particularly if they are young, and have family history of diabetes and hypertension.

1810W

Novel genetic determinants of diabetic kidney disease. *R.M. Salem[1], J.B. Cole[2,3,4], J.N. Todd[2,3,4], N. Sandholm[5,6], E.A. Valo[5,6], L.T. Hiraki[7,8], C.D. Liao[7,8], J. Ciao[7,8], M.G Pezzolesi[9,10], A. Smiles[10], S. McGurnaghan[11], P. McKeigue[11], S. Onengut-Gumuscu[12], W.M. Chen[12], C. Forsblom[5,6], M. Kretzler[13], K. Susztak[14], C. Godson[15], A.J. McKinight[16], A.P. Maxwell[16], H.M. Colhoun[11], A.S Krolewski[10], A.D. Paterson[7,8], P.H. Groop[5,6], S.S. Rich[12], J.N. Hirschhorn[2,4], J.C. Florez[2,3].* 1) University of California, San Diego, La Jolla, CA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Massachusetts General Hospital, Boston, MA; 4) Boston Children's Hospital, Boston, MA; 5) Helsinki University Hospital, Finland; 6) Folkhälsan Research Center, Finland; 7) University of Toronto, Canada; 8) SickKids, Toronto, Canada; 9) University of Utah, Salt Lake City, Utah; 10) Joslin Diabetes Center, Boston, MA; 11) The University of Edinburgh, Scotland; 12) University of Virginia, Charlottesville, VA; 13) University of Michigan, Ann Arbor, MI; 14) University of Pennsylvania, Philadelphia, PA; 15) University College Dublin, Ireland; 16) Queen's University Belfast, Belfast, United Kingdom.

Diabetic kidney disease (DKD) is a common and devastating complication affecting ~40% of those with type 1 diabetes (T1D). DKD is the leading cause of end stage renal disease (ESRD) and a significant independent risk factor for cardiovascular disease. No cure for DKD exists, although optimal glucose control may slow progression. Although there is strong evidence of a genetic component to DKD (heritability ~35% in persons with T1D), few DKD loci have been consistently identified. To enhance genetic discovery, we performed the largest genome-wide association study (GWAS) to date for DKD in 19,491 European ancestry participants with T1D, drawn from 20 individual cohorts genotyped on the Illumina HumanCoreExome BeadArray. We identified two novel loci attaining genome-wide significance, both in plausible candidate genes. SNP rs55703767, a common (MAF=21%) missense mutation in *COL4A3*, is associated with DKD defined by either macroalbuminuria or ESRD in T1D (OR = 0.80; P = 5.7 x 10^{-10}; P_{het} = 0.92). This variant is associated with both macroalbuminuria alone (OR = 0.79; P =1.6×10^{-7}; P_{het} = 0.94) and ESRD alone (OR = 0.81; P = 3.4×10^{-5}; P_{het} = 0.95), although not at genome`wide levels of significance. *COL4A3* encodes the alpha 3 subunit of Type IV collagen, a major structural component of basement membranes, particularly in the kidney. Pathogenic risk variants in *COL4A3* have been previously associated with thin basement membrane nephropathy, Alport Syndrome, and Goodpasture Syndrome. The second SNP, rs5872572, an infrequent variant (MAF = 4.3%), lies in a region near *EBF1*, a gene annotated as an enhancer in the kidney. This SNP is protective against macroalbuminuria (OR = 0.57, P = 1.1×10^{-8}; P_{het} = 0.18). The *EBF1* gene is an essential transcription factor in postnatal glomerular maturation, and the murine *Ebf1* knockout exhibits phenotypes similar to human DKD, including decreased glomerular filtration rate, albuminuria, and morphological kidney damage. These mice also have improper glomerular basement membrane differentiation with decreased *COL4A3* expression. Our work is the first to identify genome-wide significant loci associated with DKD or macroalbuminuria in T1D. Both genes implicated in DKD are key determinants of glomerular basement membrane function, suggesting that this aspect of kidney biology should be a focus of future research on the pathology, prevention, and treatment of DKD.

1811T

Diabetome: A comprehensive collection of diabetes phenotype and genotype data. *S. Shah[1], R.M. Anjana[2], S. Jebarani[2], V. Srinivasan[1], V. Radha[2], R. Praveen[1], B. Muthu Narayanan[1], R. Pradeepa[2], R. Unnikrishnan[2], J. Somasekar[1], S. Sumit[1], G. Rohit[1], S. Santhosh[1], V. Mohan[2].* 1) MedGenome Inc. Foster City, CA; 2) Madras Diabetes Research Foundation & Dr.Mohan's Diabetes Specialities Centre, Chennai India.

Diabetes is a chronic disorder of glucose metabolism, with complications including cardiovascular and end-stage renal disease, blindness, nontraumatic lower-extremity amputations, and premature mortality. It is a complex disease with many contributing genetic and environmental factors. Epidemiological and genetic studies, using large patient cohorts, with comprehensive phenotype, genotype and family pedigree information are needed to uncover novel mechanisms of diabetes complication and develop improved disease management and treatment strategies. Here we present the Diabetome Knowledgebase, a large database of diabetes-related phenotype and genotype data, that contains over 300,000 diverse patient records, including well-characterized clinical characteristics, biochemical results, family and treatment history, drug responses, genotype mapping, as well as secondary complications of diabetes, spanning over 25 years. In addition, whole genome or whole exome sequence data is also available for about 2500 patients with or without microvascular complications. The platform also provides analytical tools to facilitate data interpretation and analysis by conducting customized statistical analyses on large-scale patient data, to gain useful research insights and to identify patient cohorts for clinical trials and research studies. Findings from association studies using available information providing new insights into the biology, disease etiology, and a richer understanding of population-specific genetic and environmental risk factors, will be presented.

1812F

Measures of body composition and muscle fitness associate with nonalcoholic fatty liver disease (NAFLD). *E.K. Speliotes[1], A.P. Wright[2], V. Chen[2], Y. Chen[2], B.D. Halligan[2].* 1) Dept of Internal Med, Gastroenterology & Dept of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Dept of Internal Med. Gastroenterology, University of Michigan, Ann Arbor, Mi.

Background: Up to 25% of patients with NAFLD are not obese but may have abnormalities in body composition or muscle fitness that contribute to NAFLD. **Methods**: We investigated the effects of dual-energy X-ray absorptiometry (DXA) derived measures of body composition (appendicular lean mass (ALM)/height^2, central fat mass/height^2, and lower extremity fat mass/height^2), and grip and quadriceps strength measurements on computed tomography (CT) measured NAFLD in cross sectional data from 1306 individuals with all these measures in the population based Framingham Heart Study using logistic regression. We tested reported genetic variants that affect body composition and muscle fitness for effects on 7176 individuals with CT measured NAFLD. **Results**: 356 subjects (27%) had NAFLD based on imaging criteria. Subjects with NAFLD were more often male (64% vs. 50%), had diabetes (16.9% vs. 3.9%), hypertension (51.6% vs. 22.1%), and metabolic syndrome (52.9% vs. 15.2%) than patients without NAFLD; $p < 0.0001$ for all comparisons. Subjects with NAFLD had lower skeletal mass index (58.4 vs. 63.2 kg/kg), and greater indices of appendicular lean mass (8.4 vs. 7.5 kg/m^2), total lean mass (18.6 vs. 16.7 kg/m^2), total body fat (12.2 vs. 8.8 kg/m^2), lower extremity fat (3.6 vs. 2.9 kg/m^2), and central fat (2.2 vs. 1.9 kg/m^2); $p < 0.0001$ for all comparisons. Across both genders there was no difference in quadriceps or hand grip strength between patients with and without NAFLD; however, males with NAFLD had weaker hand grip strength (45.9 vs. 47.9 kg, $p = 0.004$) and females with NAFLD had weaker quadriceps strength (24.6 vs. 26.2 kg, $p = 0.04$). On multivariate logistic regression, central fat/height^2 (OR 2.22 95% CI(1.95-253)) and ALM/height^2 (OR 1.39 95% CI (1.15-1.63)) are positively associated with NAFLD, while lower extremity fat/height^2 (OR 0.59 95% CI (0.48-0.71)), quadriceps strength (OR 0.98 95% CI (0.97-0.99)), and hand grip strength (OR 0.97 95% CI (0.95-0.99)) are negatively associated with NAFLD. rs6689335 near LYPLAL1 that associates with waist-to-hip ratio at genome wide levels also associates with NAFLD (P < 0.05). **Conclusions**: Higher ALM was associated with NAFLD. Central fat is positively and lower extremity fat, quadriceps strength, and hand grip strength are negatively associated with NAFLD. Genetic influences on central fat distribution also associate with NAFLD.

1813W

A functional locus at 8q21.13 to FABP4 levels is modulated by BMI and kidney function: Meta-analysis of five GWAS. *S.W. van der Laan[1], J.S. Ngwa[2], A. Tönjes[3], P. Kovacs[3,4], M. Stumvoll[3], M. Fasshauer[3], J.F.C. Glatz[5], N.C. Onland-Moret[6], Y.T. van der Schouw[6], M.A. Kaakinen[7], G. Pasterkamp[8], F.W. Asselbergs[9], M.H. Chen[10], R.N. Lemaitre[11], I. Prokopenko[7], K.J. Mukamalan[12], P.I.W. de Bakker[6,13], R.S. Vasan[14], L. Djoussé[15,16], CHARGE Consortium Diabetes Working Group.* 1) Laboratory of Experimental Cardiology, University Medical Center Utrecht, Utrecht, the Netherlands; 2) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD, United States of America; 3) Department of Medicine, University of Leipzig, Leipzig, Germany; 4) Leipzig University Medical Center, IFB AdiposityDiseases, University of Leipzig, Leipzig, Germany; 5) Department of Genetics and Cell Biology, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, the Netherlands; 6) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands; 7) Department of Genomics of Common Disease, Imperial College London, London, United Kingdom; 8) Laboratory of Clinical Chemistry and Hematology, University Medical Center Utrecht, Utrecht, the Netherlands; 9) Department of Cardiology, University Medical Center Utrecht, Utrecht, the Netherlands; 10) Department of Neurology, Boston University, Boston MA, United States of America; 11) Department of Medicine, University of Washington, Seattle WA, United States of America; 12) Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston MA, United States of America; 13) Department of Medical Genetics, Center for Molecular Medicine, UMC Utrecht, Utrecht, the Netherlands; 14) Section of Preventive Medicine and Epidemiology, School of Medicine, Boston University, Boston MA, United States of America; 15) Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston MA, United States of America; 16) Department of Medicine, Harvard Medical School, Boston MA, United States of America.

Background Fatty acid binding protein (FABP4), estimated h^2=0.38, is produced by adipocytes and has been implicated in cardiovascular disease (CVD), type 2 diabetes (T2D), obesity, kidney function, and insulin resistance. A functional mutation in *FABP4* that reduces expression in adipocytes and plaques, has been associated with lower T2D risk and reduced visceral fat. Preclinical studies suggest that FABP4 inhibition may lower cardiometabolic disease risk. To further explore the genetic architecture of plasma FABP4 levels, we performed a GWAS meta-analysis and evaluated its genetic correlation to associated traits. **Methods** We meta-analyzed up to 8,270 individuals from 5 studies of plasma FABP4 measurements with 1000G-imputed data using a fixed-effects model. As FABP4 is strongly associated with BMI and estimated glomerular filtration rate (eGFR), we tested three models: 1) adjusted for age, sex, and principal components; 2) model 1 plus BMI; and 3) model 2 plus eGFR. To increase power of discovery, we performed gene-based analyses using MAGMA. To better understand the etiology of FABP4 levels, we used LD score regression (LDSC) to assess genetic correlation with BMI, kidney function and disease, glucose and insulin levels, CVD, and T2D. **Results** After correction for BMI, we identified two independent loci associated with plasma FABP4 near *FABP4* at 8q21.13. The novel lead variant was rs72684477[A] with beta(se)=-0.09(0.02), p=3.5x10^{-8}, EAF=0.18. The previously described functional mutation (rs77878271[G], beta(se)=-0.23(0.04), p=4.5x10^{-8}, EAF=0.03) was also independently associated with FABP4 levels. Model 3 yielded similar results. Gene-based analysis identified 5 genes associated with FABP4 levels (p<2.5x10^{-6}), including *FABP4*, *JADE1*, *SCLT1*, *STK36*, and *MOB2*. We applied LDSC revealing a correlation with BMI (r^2(se)=0.56(0.16), p=4.0x10^{-4}) and nominal correlations with CVD, T2D, and chronic kidney disease (p<0.03) in model 1. In model 2, the strongest correlation was with serum creatinine levels (p=5.5x10^{-3}) and nominally with CVD (p=0.03); no correlation was seen for model 3. **Conclusions** Two independent loci at 8q21.13 (one novel rs72684477) are associated with plasma FABP4 in models corrected for BMI and eGFR. LDSC suggests that FABP4 levels are partly modulated by loci also associated with BMI and kidney function. Given the estimated heritability of FABP4 plasma levels, including more samples in future meta-analyses of GWAS is likely to identify even more loci.

1814T

Systemic approach to understand human non-alcoholic fatty liver disease. *T. Yoo[1], Y. Lee[1], S.K. Joo[2], W. Kim[2], M. Choi[1].* 1) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, South Korea; 2) Department of Internal Medicine, Seoul National University Boramae Medical Center, Seoul, Republic of Korea.

Non-alcoholic fatty liver disease (NAFLD) features accumulation of fat in liver. NAFLD can be simply divided into 2 groups - non-alcoholic steatohepatitis (NASH) and simple steatosis (SS) - based on the degrees of liver damage and inflammation. However, the diagnostic criteria, molecular pathogenesis and prognosis mechanisms of the disease still need further investigation. To understand the molecular queues that differentiate pathogenesis and exacerbation of the disease, we collected ~100 human liver biopsy samples from Korean individuals and subjected them into the expression quantitative trait loci (eQTL) analysis through generating genome-wide gene expression and SNP genotyping data. The liver samples used in the analysis were divided into several groups according to their pathologic scores including fibrosis, steatosis and other inflammatory features. Among the 1,971 significantly differentially expressed genes (DEG), 64 DEGs were associated with significant cis-eQTL, and 3,197 significant trans-eQTL directed expression of 630 DEGs (P < 2.2e-16). Notably, distinct associations involving several trans-eQTL regulating immunoglobulin chains implicates genetic association between immune components and NAFLD pathophysiology. Through a series of conditional association tests, we are to identify the genotype-expression-phenotype relationships of variations in the immune loci and how they might induce the disease phenotype. This study will lead to the elucidation of molecular and functional immune mechanisms of human liver disease and potentially discovery of novel therapeutic targets of the disease.

1815F

Expanding the spectrum of type 2 diabetes risk alleles through a genome-wide association study imputed to the Haplotype Reference Consortium Panel. *D. Taliun[1], A. Mahajan[2] on behalf of the DIAMANTE, AMP-T2D, and DIAGRAM Consortia.* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

To identify and characterize genetic loci associated with type 2 diabetes (T2D), we conducted the largest meta-analysis of T2D genome-wide association studies (GWAS) to date in Europeans. We aggregated data on 50,160 T2D cases and 465,272 controls from 24 GWAS, each imputed up to HRC reference panel or study-specific reference panel of whole-genome sequences. We combined association summary statistics at >20M genetic variants across studies in a fixed-effects inverse-variance weighted meta-analysis. We performed approximate conditional analyses to identify distinct signals of T2D association within 500kb of each lead variant attaining genome-wide significance ($p<5 \times 10^{-8}$). We estimated the genetic correlations between T2D and a range of phenotypes with publically available GWAS summary statistics using LD score regression. Initial analyses revealed 105 loci at genome-wide significance, 38 of which did not overlap with established T2D susceptibility loci. The minor allele frequency (MAF) of lead variants at the 38 novel loci ranged from 2.7% to 46.3%; only the lead variant rs11820019 near *CCND1* gene had MAF<5% (MAF=2.7%, $p=4.8 \times 10^{-8}$, OR=1.17[1.11-1.23]). Conditional analyses revealed 44 additional (3 rare MAF<0.5%; 12 low-frequency MAF<5%) distinct association signals ($p<10^{-5}$) across the loci including 2 low-frequency variants at *TCF7L2* (rs182771291, MAF=0.4%, $p=1 \times 10^{-8}$, OR=1.64 [1.40-1.93] and rs142511108, MAF=1.1%, $p=3 \times 10^{-6}$, OR=1.24 [1.14-1.35]). Applying LD score regression, we confirmed the previously observed genetic correlations between T2D and obesity, fasting glucose, birth weight, lipids, and cardiovascular disease. In addition, we found novel negative genetic correlations with age of first birth ($p=1 \times 10^{-28}$), age at menarche ($p=3 \times 10^{-13}$), and father's age at death ($p=3 \times 10^{-7}$), and positive correlations with several metabolites including isoleucine ($p=3 \times 10^{-6}$) and glycoprotein acetyls ($p=3 \times 10^{-5}$). Our findings demonstrate that larger samples and reference panels continue to identify additional T2D variants, but that nearly all identified variants are common, suggesting the preponderant impact of common variants on T2D risk.

1816W

Profiling of the long non-coding RNA (lncRNA) *MALAT1* in the liver of patients with nonalcoholic fatty liver disease (NAFLD) shows association with an aggressive histological phenotype. *S. Sookoian[1], M.E. Garaycoechea[2], D.G. Rodriguez[2], C. Gazzi[3], G. Castano[4], D.M. Flichman[5], C.J. Pirola[6].* 1) Clinical and Molecular Hepatology, Institute of Medical Research (IDIM), School of Medicine, Buenos Aires University and National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina; 2) Hospital de Alta Complejidad en Red El Cruce, Buenos Aires, Argentina; 3) Department of Pathology, Institute of Medical Research A Lanari, School of Medicine-University of Buenos Aires, Buenos Aires Argentina; 4) Medicine and Surgery, Liver Unit, Hospital Abel Zubizarreta, Buenos Aires, Argentina; 5) Department of Virology, School of Pharmacy and Biochemistry, University of Buenos Aires and and National Scientific and Technical Research Council (CONICET), Buenos Aires, Buenos Aires, Argentina; 6) Molecular Genetics and Biology of Complex Diseases, Institute of Medical Research (IDIM), School of Medicine, University of Buenos Aires and National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina.

Background: LncRNAs orchestrate gene expression by modulating a plethora of molecular pathways. To identify lncRNAs involved in NAFLD severity we performed a multi-stage study that included: a) a systems biology screening-strategy, b) exploration of lncRNA-mRNAs interactions, and c) expression profiling of selected lnRNAs. Methods: Mining of biological interactions was performed by VisANT platform; LncRNA2Target dataset was used to predict lncRNAs-target genes interactions. Liver profiling of a candidate lncRNA was performed by qRT-PCR (n=81); *RPL19* was used for data normalization. Patients were selected from two different hospital-based settings, a cohort of morbid-obese subjects who underwent bariatric surgery (discovery-set) and a cross-sectional study of patients with NAFLD and MetSyn (replication-set). The histological phenotype was characterized by liver biopsy. Results: The NAFLD-interaction network yield a list of 494 genes; prioritization of mRNA-lncRNAs interactions showed ten candidate-lnRNAs potentially involved in gene expression regulation. Based on assessment of tissue-specific and co-expression patterns, we selected for further exploration the lncRNA *MALAT1* (metastasis-associated lung adenocarcinoma transcript 1), a~7kb single-exon transcript located in Chr 11q13 that is highly conserved across species.. We found that *MALAT1*, which is involved in regulation of cell-cycle and tumorigenesis, is constitutively expressed in the liver. However, liver abundance of *MALAT1* was significantly associated with NAFLD severity (p=0.00011). Specifically, liver-*MALAT1* expression was significantly up-regulated by 1.75-fold (p=0.029) and 5.81-fold (p=0.0038) in patients with NASH compared with simple steatosis (SS) (discovery and replication set, respectively, *ANCOVA* adjusted by age, HOMA and BMI). *MALAT1*-expression levels significantly correlate with the full spectrum of histological severity, including the degree of steatosis (R: 0.47 p=0.000007) and fibrosis (R: 0.45 p=0.00002). Exploration of liver *VEGFα*-mRNA, a proangiogenic factor driving tumor angiogenesis, which is a target of *MALAT1*, showed levels of both transcripts positively correlated (R:0.59, p=0.00002). Conclusion: This step-wise analysis of lncRNAs uncovered *MALAT1* as implicated in the molecular pathogenesis of NASH. Dysregulated expression of *MALAT1* seems to tip the transition from simple steatosis to a more aggressive phenotype. The AUROC for distinguishing NASH from SS was 0.84±0.07.

1817T

Multifactor dimensionality reduction (MDR) method to study association of type 2 diabetes mellitus with ENPP1 (K121Q), TCF7L2 (G>T) and GYS1 (A1>A2) gene variants in Punjabi population, India. *B. Doza[1], B. Barna[2], A.J.S. Bhanwer[2], K. Matharoo[2], M. Kaur[3].* 1) Aligarh Muslim University Murshidabad Centre, Murshidabad, West Bengal, India; 2) Guru Nanak Dev University, Amritsar, Punjab; 3) Government Medical College, Amritsar, Punjab.

 Introduction: The Multifactor Dimensionality Reduction (MDR) method detects and characterizes high order gene-gene interactions in case-control studies. The MDR method is model free and it does not assume any particular genetic model. Despite its high importance almost no study is available with this model approach to detect the status of T2DM Punjabi population. Hence, the present study tested the hypothesis that T2DM related gene variants (*ENPP1* (K121Q), *TCF7L2* (G>T) and *GYS1* (A1>A2) contribute to etiology of type 2diabetes mellitus in the Punjabi population or not. **Material and Methods:** It was a case control study in Punjabi population. A total of 500 subjects (250 T2DM patients and 250 controls) have been recruited from different clinical centers of Punjab and unrelated healthy controls were also selected from the same community and area. The entire data were analyzed with respect to three most important candidate gene polymorphisms such as *ENPP1* (K121Q), *TCF7L2* (G>T) and *GYS1* (A1/A2). **Results:** The present study found a strong association of *ENPP1* and *TCF7L2* variant with T2DM in Punjabi population but an insignificant association of the XbaI (A1>A2) polymorphism in *GYS1* with T2DM. The MDR analysis effectively found evidence for a significant gene-gene interaction among the three loci (*ENPP1*-K121Q, *TCF7L2*- G>T and *GYS1*-A1>A2) for pooled (male and female) T2DM patients and non-diabetic controls; two loci (*GYS1*-A1>A2 and *TCF7L2*- G>T) for males and two loci (*ENPP1*-K121Q and *GYS1*-A1>A2) for females among T2DM patients and non-diabetic controls. **Conclusion**: The present study showed MDR method to be effective in detecting multi-genetic interaction among different loci. It is suggested that gene-gene interaction is ubiquitous in determining the susceptibility of complex human diseases. Further studies on epistatic interactions are warranted to elucidate their possible underlying role in pathogenesis of T2DM.

1818F

A type 1 diabetes genetic risk score developed in Europeans discriminates between type 1 and type 2 diabetes in South Asian Indians in India.

J. Harrison[1], D.S.P Tallapragada[2], S. Sharp[1], K.A. Patel[1,3], A.T. Hattersley[1,3], G. Chandak[2], C.S. Yajnick[4], R. Oram[1,3], M.N. Weedon[1]. 1) Institute of biomedical and clinical science, University of Exeter Medical School, Exeter, Devon, United Kingdom; 2) CSIR-Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, 500 007, India; 3) National Institute for Health Research Exeter, Clinical Research Facility, Exeter, UK; 4) KEM Hospital, 489 Rasta Peth, Sardar Moodliar Road, Pune, 411011, India.

Many young adult patients with diabetes receive incorrect treatment because the type of diabetes they have is misdiagnosed and the problem is increasing with rising obesity rates. Misclassification is a particular problem in individuals of South Asian ancestry due to the high prevalence of early-onset Type 2 diabetes (T2D). Type 1 diabetes (T1D) patients require insulin treatment whereas patients with T2D or monogenic diabetes are often best treated with diet or tablets. We have previously shown that a T1D genetic risk score made up of 30 SNPs can be a useful tool for aiding the discrimination of T1D and T2D in Europeans. The aim of this study was to determine whether the T1D GRS discriminates T1D from T2D as well in South Asians as it does in white Europeans? We used two cohorts in this study: Indo-Europeans from Pune, India (clinically diagnosed) T1D (n=176), T2D (n=344), controls (n=171) and individuals of European ancestry (EUR) in UK Biobank (UKBB) T1D(n=251), T2D(n=6,357), controls(n=139,652). We genotyped 9 SNPs that capture the majority of the discriminatory ability of the T1D GRS. Comparison of individuals of European and South Asian ancestry showed that the background T1D GRS scores are significantly different between these two ethnic groups: (med(IQR)) controls Pune (0.65(0.62-0.71)) v UKBB EUR (0.63 (0.54-0.71)) p<0.0001, T2D cases UKBB EUR (0.63(0.55-0.72)) v Pune (0.64(0.60-0.70)) p=0.0016, T1D cases UKBB EUR (0.80(0.73-0.88)) v Pune (0.75(0.67-0.80)) p<0.0001. The T1D GRS was highly discriminative of T1D (AUC (95% CI)) 0.80 (0.75-0.84) in South Asians, but the ROC-AUC was significantly lower than in Europeans 0.86(0.84-0.88) p=0.0094. HLA variants alone were an independent predictor of T1D: 0.76 (0.72-0.81) but the T1D GRS performed better than an HLA score alone HLA v T1D GRS: 0.76(0.72-0.81) v 0.80(0.75-0.84), p=0.0001. The discriminative power of the HLA only score was significantly lower in the Pune cohort than in the European cohort Pune v EUR: 0.76(0.72-0.81) v 0.84(0.82-0.87), p=0.0026. We have shown a T1D GRS using SNPs defined in Europeans is still highly discriminative of T1D in South Asians despite genetic risk of T1D differing between the two populations. Refining the T1D-GRS for individuals of South Asian ancestry may increase the discriminative power. The T1D-GRS will be useful for classifying T1D in South Asian groups and help to ensure the correct treatment is given to patients diagnosed with diabetes.

1819W

Genetic factors influencing glycated hemoglobin, fasting glucose, and fasting insulin levels in the Population Architecture using Genomics and Epidemiology Study.

H.M. Highland[1], S. Bien[2], Y.M. Patel[3], M. Graff[1], A. Vishnu[4], B. Lin[5], R. Tao[6], D. Lin[5], K.L. Young[1], S. Liu[7,8], L. Phillips[9], Q. Qi[10], J.I. Rotter[11], L. Hindorff[12], S. Buyske[13,14], T.C. Matise[13], K.E. North[1], C. Kooperberg[2], J.S. Pankow[15], C. Haiman[3], R.J. Loos[4]. 1) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Keck School of Medicine, University of Southern California, Pasadena, CA; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 6) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN; 7) Department of Medicine, the Alpert Warren Medical School, Brown University, Providence, RI; 8) Brown Center for Global Cardiometabolic Health and Department of Epidemiology, Brown University, Providence, RI; 9) Atlanta Vanguard Clinical Center, Atlanta, GA; 10) Albert Einstein College of Med, Bronx, NY; 11) Institute for Translational Genomics and Population Sciences, LABiomed/Harbor-UCLA Medical Ctr, Torrance, CA; 12) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 13) Department of Genetics, Rutgers University, Piscataway, NJ; 14) Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ; 15) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN.

There are notable disparities in risk of type 2 diabetes (T2D) across ancestrally diverse populations. Knowledge of the genetic basis of related biomarkers for T2D-related traits, such as fasting glucose (FG), fasting insulin (FI) and glycated hemoglobin (HbA1c), is very limited in ancestrally diverse populations. In primarily European descent populations, nearly 100 loci have been identified for FG, FI, and HbA1c. Elevated levels of fasting glucose and fasting insulin precede the development T2D. To better understand the genetic basis of glucose dysregulation, the Population Architecture using Genetic Epidemiology (PAGE) Consortium genotyped 12,801 Hispanic/Latinos, 5,696 African Americans, 1,405 Native Hawaiians, 398 Native Americans, and 1,727 Asians without diabetes using the Multi-Ethnic Global Array (MEGA). This platform is enriched for coding variation and ancestrally diverse markers to improve the imputation of allelic dosages unmeasured variants across admixed populations. We performed imputation using 1000 Genomes Phase 3 reference panel, resulting in 40 million well-imputed variants. The association of these variants with FG, FI and HbA1c were tested using a GEE model that accounts for relatedness of individuals in our ancestry-combined samples. We identified a potentially novel association for HbA1c with variants near *KLHL29* (P = 4.0×10^{-11}). Other associations with HbA1c included the sickle cell anemia variant (rs334) in *HBB*, and an association with a *G6PD* variant known to cause X-linked hemolytic anemia (rs1050828, $P_{women} = 5.05 \times 10^{-62}$, $P_{men} = 5.08 \times 10^{-98}$), further highlighting the importance of both variation in hemoglobin genes and variants impacting red-blood cell lifespan when using HbA1c as a measure of glycemic control. Novel associations for FG include variants in *HDAC9* (rs118084662, P = 1.9×10^{-8}) and *DHRS7B (*rs540671238, P = 2.9×10^{-8}). We also replicated several known associations between FG and variants near *G6PC2, GCKR, MTNR1B,* and *FOXA2*. We observed a novel association between FI and intergenic variant rs116210206 (P = 7.8×10^{-9}) near the known *GRB14/COBLL1* locus. Further analyses including SNP by BMI interaction and gene-based tests of rare variants are underway. These additional tests will increase power to detect associations with ancestry-specific rare variants. Despite a modest sample size, we have identified novel associations with glycemic traits, in large part due to the diverse ancestral composition of the PAGE sample.

1820T

Association study of ENPP1 (K121Q), TCF7L2 (G>T), GYS1 (A1/A2) variants with type 2 diabetes mellitus (T2DM) in north Indian Punjabi population. *M. Kaur[1], B. Doza[2], B. Barna[3], A.J.S. Bhanwer[3], K. Matharoo[3].* 1) Government Medical College, Amritsar, Punjab, India; 2) Aligarh Muslim University Murshidabad Centre, Murshidabad, India; 3) Guru Nanak Dev University, Amritsar, Punjab, India.

Introduction: Over past decades genetic and genome wide association study (GWAS) on *ENPP1* (K121Q), *TCF7L2* (G>T), *GYS1* (A1/A2) have suggested a possible relationship between allele and genotype frequencies and pathogenesis of type 2 diabetes mellitus (T2DM). The purpose of this study was to confirm the reported association between these variants and increased risk of type 2 diabetes mellitus in north Indian Punjabi population. **Material and Methods**: A total of 500 participants (250 with T2DM and 250 healthy subjects) were recruited for this study. Genotyping was performed by PCR-RFLP method. Anthropometric and physiometric variables such as height, weight, body mass index (BMI), waist to hip ratio (WHR), waist circumference (WC), hip circumference (HC), biceps skinfold, triceps skinfold, SBP, DBP, MBP, pulse rate and pulse pressure were measured using standard protocol. The statistical analyses were performed to include chi-square, correlation, regression, principal component analysis and genetic association analysis. **Results**: The results revealed that anthropometric and clinical characteristics such as WHR, fasting and random glucose levels, SBP, DBP, pulse rate and pulse pressure have significant (p<0.001) differences between T2DM and control subjects. However, none of the anthropometric and clinical characteristics have found a significant difference with respect to their genotypic distributions for both of genes except DBP for *ENPP1* K121Q polymorphism. The association analysis of the present results showed K121Q variant of *ENPP1* and G/T variant of *TCF7L2* gene have a significant association with type 2 diabetes with dominant and recessive model of action, respectively, in north Indian Punjabi populations. Overall, the present study presented a strong association of *ENPP1* and *TCF7L2* variant with T2DM in north Indian Punjabi population and provided evidence of its key role in the pathopysiology of T2DM by influencing both insulin secretion and insulin resistance. However an insignificant association of the *GYS1* variant with T2DM was observed. The present investigation did not support a significant positive association of *ENPP1*, *TCF7L2* or *GYS1* variant with obesity and cardiovascular diseases in T2DM patients.

1821F

Pathway-informed genetic testing and analysis for type 2 diabetes. *C. Ma[1], M. Roth[1], K. Zhang[2], L. Shen[1].* 1) GB HealthWatch, San Diego, CA; 2) School of Medicine, University of California at San Diego, San Diego, CA.

Background: It remains a great challenge to predict genetic risk of complex diseases, such as type 2 diabetes (T2D), at the individual level. Twin studies show that genetics accounts for nearly 70% of disease susceptibility, however, with conventional statistical analysis, only about 10% of heritability can be explained with GWAS variants and an additional 2-3% when rare pathogenic variants are included. We developed a novel T2D risk model with a call accuracy reaching 80%. **Methods:** GBinsight T2D NGS panel analyzes coding and non-coding variants that are in or near genes within pathways known to affect T2D risk. 143 candidate genes are selected from the following T2D relevant pathways: Beta cell development and function Glucose sensor and insulin release Insulin signaling (organ resistance) Hypertriglyceridemia Obesity Low cholesterol Susceptibility to autoimmune diseases Choline, Betaine and Folate Methyl-cycle Pathways A novel pathway-based hierarchical structure model is used in risk score calculation. 1000 genomes data is used as population reference. 60 early-onset (diagnosed before age 60) T2D case samples from 4 ancestry groups: European (EUR), Hispanic (AMR), African American (AFR) and East Asian (EAS) and control samples are from 40 ancestry-matched volunteers without T2D over 70 years of age. Type 1 diabetes cases are excluded. **Results: 1.**GBinsight model provides a call accuracy of 0.83, 0.86, 0.77 and 0.73 in EUR, AFR, AMR and EAS samples respectively, with average true positive rate 0.78 and true negative rate 0.82. **2.** Each individual has a unique combination of risk and protective variants. Five known pathogenic variants in hypertriglyceridemia (LPL and APOC2), six rare damaging coding variants in MODY (BLK, PAX4 and PDX1) and eight in obesity genes (MC3R, SIM1, ADIPOQ, TUB, NROB2, CCKAR) are detected in case samples only. Common variants in APOA5, APOB, APOE, BHMT, HNF4A, LPL, MC4R, PCSK9, PSORS1C1 and UCP2 genes account for significant T2D risk. **3.** Combined effect of defects in multiple pathways account for risk of T2D at individual level. Case samples detected an average of 4.3 (95%CI 1.2-7.3) pathways scored as "high risk" (scored above 70% of reference population), and control samples detected an average of 1.9 (95%CI 0-4.3) pathways scored as high risk, two sample Wilcoxon-test P=7.5×10^{-14}. **Conclusion:** Genetic etiology of T2D is heterogeneous and complex. Pathway-based modeling significantly improves T2D risk assessment and may have clinical utility.

1822W

Fine-mapping fasting glucose and fasting insulin loci with whole genome sequence data from the Trans-Omics for Precision Medicine (TOPMed) Program. *A. Manning[1,2], J. Wessel[3], B. Hidalgo[4] on behalf of the Trans-Omics for Precision Medicine (TOPMed) Program Diabetes Working Group.* 1) Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA; 2) Department of Medicine, Harvard Medical School, Boston, MA; 3) Department of Epidemiology, Fairbanks School of Public Health, Indianapolis, IN; 4) Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL.

The majority of genetic variants significantly associated with diabetes-related glycemic traits reside in the non-coding genome, with many causal variants still unknown. Whole genome sequence association (WGSA) analysis allows us to fine-map known and novel loci across the non-coding genome without depending on imputation. Here, we present an initial WGSA of fasting insulin (FI) and fasting glucose (FG) levels in phase 1 (released in the fall of 2016) TOPMed data (N=7,121) with deep (>30x) sequence coverage in five cohorts, three European-ancestry (EA): Framingham Heart Study, N=3,209; Old Order Amish Studies, N=980, Cleveland Family Study, N=197, and two African-American (AA): Jackson Heart Study, N=2,487, Cleveland Family Study, N=248. For each cohort, we used linear mixed effects models (as implemented in EMMAX or MMAP software) adjusting for sex, age and BMI, with empirical kinship for relatedness and population structure. We restricted to variants with minor allele count > 5 in more than 1 cohort and meta-analyzed within and across ancestry with METAL. In EA+AA analysis, common (minor allele frequency [MAF]>0.05) variant associations ($P<5 \times 10^{-8}$) were identified for FG at known loci: *MTNR1B* (rs10830963, $P=2.5 \times 10^{-16}$; rs12792753, $P=1.4 \times 10^{-8}$), *GCK* (rs4607517, $P=1.16 \times 10^{-10}$, and 13 additional variants), and *G6PC2* (rs560887, $P=5.4 \times 10^{-10}$). At the *MTNR1B* locus, fine-mapping across ancestries reduced the significant FG associations from 46 (EA) to 2 (AA+EA). Novel FI associations were seen in rare (MAF<1%) variants: 18q12.1 (rs146884135, $P=2.1 \times 10^{-9}$), 4p12 (rs193168677, $P=1.2 \times 10^{-8}$), 1q25.3 (rs550837507, $P=3.8 \times 10^{-8}$), and *DCLK3* (rs573417731, $P=4.8 \times 10^{-8}$). These results, which are being extended with phase 2 TOPMed cohorts (expected release is in the summer of 2017; expected total N=27,830), highlight the value of multi-ancestry WGSA to refine known and discover novel associations for complex traits. Future analyses with combined phase 1 and phase 2 data will assess (1) enrichment of sub-significant associations using diabetes-specific functional annotations of the non-coding genome, and (2) the association of rare variants within these diabetes-specific functional annotations with fasting glucose and fasting insulin levels.

1823T

Glucose metabolism in pregnancy: Analysis of diabetes related genetic risk scores against glucose curve trajectories. *G.H. Moen[1,2], M. LeBlanc[3], C. Sommer[1], T. Lekva[4], E. Qvigstad[1], K.I. Birkeland[2,5], K.F. Frøslie[6].* 1) Endocrinology Morbid Obesity and Preventive Medicine, Oslo University Hospital, Oslo, Norway; 2) Faculty of Medicine, University of Oslo, Institute of Clinical Medicine, Oslo, Norway; 3) Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway; 4) Research Institute of Internal Medicine, Oslo University Hospital, Oslo, Norway; 5) Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway; 6) Norwegian National Advisory Unit on Women's Health, Oslo, Norway.

Little is known about the genetic risk factors for gestational diabetes (GDM), and few studies have been published, but observational studies have shown that high body mass index (BMI) is one of the many risk factors. Moreover, women with previously known GDM have a higher risk of developing type 2 diabetes (T2DM), and it has been suggested that similar underlying risk factors, including genetic markers, could be important. GDM is diagnosed in gestational week 24-28 by measuring blood glucose in the fasting state or after an oral glucose tolerance test (OGTT). Functional data analysis (FDA) of repeated measurements from OGTTs has found that more than 95% of the temporal variance in the glucose curves may be attributed to the glucose curve features "general level" and "timing of the peak", expressed by the first two functional principal components of OGTT curves. Genome-wide association studies (GWAS) associating a large number of SNPs to fasting blood glucose, 2-h post- OGTT glucose, T2DM, fasting insulin and BMI have previously been published. The aim of the present study was to examine whether the temporal variance in glucose curves could be explained by a genetic risk score for these traits using the publicly available GWAS data. The analyses were done on 529 healthy pregnant women of Norwegian background. The women underwent two OGTTs during pregnancy. The genetic risk scores were constructed using genome-wide significant SNPs of previously published GWAS. FDA was used to extract the general curve features, and the corresponding individual principal component (PC) scores from the OGTT data. We analyzed the correlation of weighted and unweighted genetic risk scores against the PC scores, the fasting glucose and 2h glucose at two time points in pregnancy, as well as BMI before pregnancy and in first trimester. We will discuss the results of the analysis and their potential implications in the understanding of glucose metabolism in pregnancy as well as the understanding and treatment of GDM.

1824F

Tissue specific isoform annotations in rare variant analysis. *A. Ndungu[1], M. van de Bunt[1,2], A. Mahajan[1], P.E MacDonald[4], A.L Gloyn[1,2,3], M.I McCarthy[1,2,3], T2D-GENES consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, United Kingdom; 3) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK; 4) Alberta Diabetes Institute, University of Alberta, Edmonton, Canada.

Due to ubiquitous alternative splicing in multi-exon genes, coding variants will often have more than one annotated consequence across multiple transcripts of the same gene. For such variants, the common practice is to assign the most severe consequence across all transcripts. However, the selected transcript might be rare, tissue-specific, or not expressed in tissues relevant to disease, leading to inaccurate annotation. To address this, we introduce an integrated 'omics approach in rare variant analysis that incorporates tissue-specific transcriptomic profiles to prioritise functional variants for gene-based analysis in type 2 diabetes (T2D). We use exon-level RNA-Seq data from human islets (118 cadaveric donors) to define the repertoire of expressed transcripts using a conservative exon-union based approach. Complete transcripts were defined as having all associated exons, (GENCODE.v19 transcripts), exceeding a minimum expression threshold (> 6 reads in ≥10 samples and TPM > 0.1 in ≥10). There were 94,569 transcripts in 17,642 genes with robust expression in islets. We explored whether applying an islet transcript filter when aggregating variants in gene based methods improved power to detect an association with T2D using whole exome sequence data across 12,940 individuals from 5 ancestries in the T2D-GENES consortium. For protein-truncating variants (PTVs), the commonly used approach had 69,956 PTVs across 14,415 genes while the islet-transcript filtered method had 39,776 PTVs in 9,392 genes. There were no genes reaching $p < 2.5 \times 10^{-6}$ in the conventional analysis and incorporating islet-transcript filters did not yield new signals. There were however notable differences with the two methods. For example, in the monogenic diabetes gene *HNF4A*, 3 previously identified PTVs in this gene were intronic in all islet-expressed transcripts due to an alternative promoter in islets. Variant prioritization based on transcript expression in disease relevant tissue could not only improve current methods of analysis, but also provide guided insight into disease mechanism. These analyses will be expanded to a much larger dataset of 52,000 individuals with exome sequence data providing increased power to detect rare variants associating with T2D. Furthermore, as full-length isoform reconstruction from short read data is challenging, incorporating long-read sequence data will be necessary for accurate estimation of isoform proportions in the human islet transcriptome.

1825W

Meta-genome-wide association study identifies multiple loci in the MHC region and a locus on chromosome 1 for serum C-peptide in type 1 diabetes. *D. Roshandel[1], R. Gubitosi-Klug[2], S.B. Bull[3], A.J. Canty[4], M.G. Pezzolesi[5], G.L. King[6,7], H.A. Keenan[6,7], J.K. Snell-Bergeon[8], D.M. Maahs[8,9], R. Klein[10], B.E.K. Klein[10], T.J. Orchard[11], T. Costacou[11], M.N. Weedon[12], R.A. Oram[12,13], A.D. Paterson[1,14], DCCT/EDIC Research Group.* 1) Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada; 2) University Hospitals Case Western Medical Center, Cleveland, OH, US; 3) Lunenfeld–Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada; 4) Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada; 5) Division of Nephrology and Hypertension, Diabetes and Metabolism Center, University of Utah, Salt Lake City, UT, US; 6) Research Division, Joslin Diabetes Center, Boston, MA, US; 7) Department of Medicine, Harvard Medical School, Boston, MA, US; 8) Barbara Davis Centre for Diabetes, University of Colorado School of Medicine, Aurora, CO, US; 9) Department of Paediatrics, Stanford School of Medicine, Stanford, CA, US; 10) Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, WI, US; 11) Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, US; 12) Institute for Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK; 13) National Institute for Health Research, Exeter Clinical Research Facility, Exeter, UK; 14) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

People with type 1 diabetes (T1D) vary in the progression of beta cell loss as measured by serum C-peptide. The factors beyond diabetes duration, e.g. genetic, that influence this heterogeneity are mostly unknown. Here, we aimed to identify genetic variants associated with C-peptide in T1D through a meta-genome-wide association study (meta-GWAS). We performed GWAS of stimulated, fasting or random C-peptide measured cross-sectionally in subjects from five T1D studies (n=3479): 1. Diabetes Control and Complications Trial (DCCT)(stimulated, n=1303) 2. Coronary Artery Calcification in Type 1 Diabetes (CACTI)(fasting, n=529) 3. Pittsburgh Epidemiology of Diabetes Complications (EDC)(fasting, n=150) 4. Joslin 50-Year Medalist Study (random, n=906) 5. Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) (random, n=591) Subsequently, the p-values across studies were combined through meta-GWAS taking into account direction of effect a study-specific sample-size weight. We identified a locus on chromosome 1, rs559047 (Chr1:238753916, T>A, minor allele frequency (MAF)=0.24-0.26), associated with lower C-peptide (p=4.13E-8) meeting the genome-wide significance threshold (p<5E-8). We also identified 3 independent loci with p-values close to genome-wide significance in the major histocompatibility complex (MHC) region where major known T1D risk loci are located. These included an indel, rs61211515 (Chr6:30100975, T/-, MAF=0.17-0.19), associated with lower C-peptide (p=1.51E-6); and rs3135002 (Chr6:32668439, C>A, MAF=0.02-0.06, p=7.19E-8) and rs9260151 (Chr6:29911030, C>T, MAF=0.07-0.10, p=8.43E-8) associated with higher C-peptide. Of these, rs61211515 was also associated with accelerated rate of C-peptide decline over time in a subset of individuals from DCCT (n=258) with repeated annual C-peptide measures (p=0.02). Conditional analyses suggested that all these loci are independent of known T1D risk loci in the MHC region. We also investigated association of all known T1D risk loci with C-peptide but only *IL27* (rs4788084, Chr16:28539848, risk allele (RA):C) and *INS* (rs689,Chr11:2182224, RA:T) showed nominal association with lower C-peptide. In conclusion, we identified a locus on chromosome 1 and multiple loci in the MHC, all distinct from T1D risk loci, associated with C-peptide. These associations need to be validated in independent populations. They could eventually provide insight into mechanisms of beta-cell loss and opportunities to preserve beta-cell function.

1826T

Replication of 93 T2D associated SNPs in Jat Sikhs, population of Punjab, India. *G. Singh[1,2], E. Rai[1], V. Singh[1], A.J.S. Bhanwer[2], R.N.K. Bamezai[3], S. Sharma[1].* 1) Department of Biotechnology, Shri Mata Vaishno Devi University, Katra, Jammu, India; 2) Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 3) National Centre of Applied Human Genetics, School of Life Sciences, Jawhar Lal Nehru University, New Delhi, India.

Type 2 Diabetes (T2D) is a complex, chronic illness in which both environmental and genetic factors interact and contribute to the etiology of the disease. Outcomes of the disease warrant continuous medical care with heterogeneous risk-reduction approaches which are not restricted to glycemic control only. In the present preliminary case-control study, a total of 430 subjects (204 Healthy Controls and 226 T2D Cases) were enrolled. Our focus was to study an endogamous community, Jat Sikhs from Punjab, India. In this association study, we included 93 SNPs from 56 different genes reported to be associated with T2D in literature. Anthropometric parameters like age, gender, BMI, SBP and DBP were also compared. Six SNPs showed significant association (p < 0.05) with T2D i.e. rs1044498, rs1801278, rs3813929, rs7961581, rs199925 and rs668514. This preliminary study indicated that genes ENPP1, IRS1, HTR2C, TSPAN8/LGR5, UCP2/UCP3 and CACNA1E are the respective most promising candidate genes associated with T2D in Jat Sikh population of Punjab. However, an extension of study in a larger sample size is required to make conclusive arguments.

1827F

Type 1 diabetes genetic risk score identifies neonatal diabetes patients with highest probability of mutations in Iranian population. *H. Yaghootkar[1], E. De Franco[1], S. Hosseini[2], M.R. Abbaszadegan[2], K. Colclough[1], S. Vakili[2], M. Ghahraman[2], A. Moleirinho[1], R. Maroufian[3], A. Haghighi[4], M. Razzaghy-Azar[3], K. Patel[1], F. Abbasi[3], M.M. Amoli[3], S. Ellard[1], R. Vakili[2], A.T. Hattersley[1].* 1) University of Exeter Medical School, Exeter, Devon, United Kingdom; 2) Mashhad University of Medical Science, Mashhad, Iran; 3) Tehran University of Medical Sciences, Tehran, Iran; 4) Harvard Medical School, Boston, USA.

Studies of white Europeans have indicated a type 1 diabetes genetic risk score test (T1D-GRS) as a tool to discriminate between T1D and other causes of diabetes. Iran is a Middle Eastern country with a high prevalence of diabetes (11.1%), obesity and consanguinity (37% of marriages). High rates of consanguinity mean undiscovered monogenic recessive forms of diabetes may exist in the population. We aimed to assess the utility of T1D-GRS in prioritizing patients for whole genome sequencing in the Iranian population. We studied patients with neonatal diabetes (diagnosed < 6 months) who were referred from Iran. We generated T1D-GRS using 10 common genetic variants associated with T1D and performed a comprehensive genetic testing including 6q24 methylation analysis and targeted next-generation sequencing of all known neonatal diabetes genes. Our cohort included 103 cases of neonatal diabetes including 50 females and 53 males from 62 consanguineous families (65%). Genetic testing identified the genetic cause of neonatal diabetes in 76 patients (74%). The common genetic causes were different from European patients. Mutations in *KCNJ11* and *ABCC8* account for 46% of NDM cases in Europeans while in our Iranian cohort, mutations in *EIF2AK3* (23%) and *PTF1A* (10%) are the most common causes of neonatal diabetes accounting for 33% of cases. We used T1D-GRS to divide the 26% of patients (n=27) with unknown diabetes etiology into 2 groups: probable T1D (patients with T1D-GRS > 50th T1D centile, n=10) and probable monogenic (patients with T1D-GRS ≤ 50th T1D centile, n=13). Probable T1D patients were diagnosed later (5 months vs. 2 months, P=0.001), had fewer syndromic features (11% vs. 43%, P<0.001), were less consanguineous (33% vs. 73%, P<0.001), were more males (70% vs. 51%, P= 0.006) and had higher birth weight (3150 g vs. 2400 g, P=0.01) than patients with monogenic neonatal diabetes. In contrast, the clinical characteristics of patients with probable monogenic neonatal diabetes were similar (P>0.05) to those of patients with genetically confirmed neonatal diabetes except that they were less consanguineous (46% vs. 73%, P<0.001). T1D-GRS can be used in the Iranian population to identify neonatal diabetes patients with the highest probability of a novel cause of monogenic diabetes. Our study suggests that 13% of patients in our cohort are likely to have a novel cause of neonatal diabetes. These patients have been prioritized for novel gene discovery by genome sequencing.

1828W

Trans-ethnic discovery of the genetic architecture of glycaemic control. *C. Langenberg on behalf of the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC).* University of Cambridge, Cambridge, United Kingdom.

Background/aims: We performed the first large-scale trans-ethnic discovery for fasting glucose (FG), fasting insulin (FI), glycated haemoglobin (HbA1c) and 2-hour glucose (2hG) to identify additional glycaemic trait-associated loci, evaluate heterogeneity across ancestries and conduct fine-mapping. **Materials and methods:** We ran ancestry-specific and trans-ethnic meta-analyses of 1000 genomes imputed data in up to 281,416 individuals without diagnosed diabetes from five ancestries (71% Europeans, 13% East Asian, 7% Hispanic, 6% African-American and 3% South Asian). Cohort-level association results were combined assuming a fixed-effects model in METAL, followed by trans-ethnic meta-analyses in MANTRA, where a $\log_{10}$ Bayes factor ($\log_{10}$BF) threshold of 6 was considered genome-wide significant. We used glycaemic and blood/iron-related summary statistics to classify the HbA1c signals as erythrocytic or glycaemic using a fuzzy clustering approach. GARFIELD and DEPICT (FDR<5%) were used to perform functional, tissue and pathway enrichment analyses and eQTLs in blood, islets and adipose were used to identify candidate genes. **Results:** We identified 102 trans-ethnic signals associated with FG, of which 48 did not overlap known FG-associated regions. The most significant novel FG signals, not previously known to be involved in diabetes, include those at or close to *NFX1* and *ZBTB38* genes ($\log_{10}$BF=12.42 and 11.79, respectively). We also identified 62 FI-associated trans-ethnic signals (43 novel), 130 HbA1c-associated trans-ethnic signals (57 novel) and 21 2hG-associated trans-ethnic signals (10 novel). The most significant novel signals for FI, HbA1c and 2hG, not previously known to be involved in diabetes, fall within or close to *BCL2*, *LRRC16A* and *CLEC14A* genes ($\log_{10}$BF = 11.66, 18.60 and 9.68, respectively). HbA1c signals were mostly classified as influencing HbA1c via erythrocytic pathways (~70%). Pathway analyses highlighted insulin signalling in HbA1c; insulin and MTOR signalling, maturity onset diabetes of the young, type 2 diabetes and circadian rhythm in FG; and cancer in FI, the latter being consistent with Mendelian Randomization analyses investigating the causal link between insulin resistance and cancer. **Conclusion:** This large international effort has identified over a hundred novel loci underlying glycaemic traits that pose new hypotheses about the biology and the genetic architecture of glucose and insulin-related traits.

1829T

Characterization of potential regulatory variants at the *SH2B1* body-mass index GWAS locus. *M.E. Cannon[1], K.W. Currin[1], S. Vadlamudi[1], A. Safi[2], L. Song[2], M. Laakso[3], G.E. Crawford[2], K.L. Mohlke[1].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Pediatrics, Division of Medical Genetics, Duke University, NC; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

Hundreds of loci have been identified in obesity GWAS, but the mechanisms at most loci are unknown. One body mass index (BMI) locus is located on chromosome 16 (P=3x10^{-25}, ß=0.031). eQTL associations have been observed for at least nine nearby genes; however, only two coincident eQTLs have been identified, for *SH2B1* (P=4.7x10^{-15}, ß=-0.39) and *ATXN2L* (P=2.5x10^{-11}, ß=-0.34) in adipose tissue. Rare *SH2B1* coding variants are also associated with obesity. *SH2B1* is an Src homology 2B family member involved in the regulation of leptin and insulin signaling. *Sh2b1* knockout mice exhibit hyperlipidemia, insulin resistance, and obesity. *SH2B1* has at least four isoforms that are expressed differentially across tissues. Neuron-specific restoration of *Sh2b1* reverses metabolic dysfunction in mice, *Sh2b1* over-expression in islets promotes beta cell growth and insulin secretion, *Sh2b1* regulates lipid metabolism in the liver, and it's role in adipose remains unclear. The mechanism of the GWAS and eQTL signals are not known. At this locus, 124 variants exhibit strong LD (r^2>0.8) with both the GWAS and eQTL lead variants. In ATAC-seq chromatin accessibility data in adipose tissue and SGBS preadipocyte cells, only 4 of the 124 variants overlapped ATAC-seq peaks. These 4 variants also overlapped regulatory histone marks of promoters and enhancers in adipose, brain, liver, and islets. At one variant, we observed a transcription factor motif for PU.1 (*SPI1*) and an allelic imbalance in ATAC-seq reads. rs7187776-G creates a stronger PU.1 motif and showed fewer reads (G=3, A=25) suggesting that the G-allele may alter PU.1 binding and/or chromatin accessibility. rs7187776 is located in the promoter of a long *SH2B1* isoform. In electrophoretic mobility shift assays, rs7187776-G showed increased protein binding compared to rs7187776-A using nuclear extract from SW872 cells or using purified PU.1 protein. In transcriptional reporter assays of a 456-bp element in SW872 adipocyte and SGBS preadipocyte cells, we observed extremely strong transcriptional activity (>200-fold compared to background); inconsistently, we observe no allelic differences between rs7187776 alleles. Taken together, rs7187776-G showed fewer ATAC-seq reads and stronger PU.1 binding and is associated with decreased expression level of *ATXN2L* and *SH2B1* and increased BMI. These data suggest that, in adipose tissue, a regulatory mechanism involving rs7187776 may contribute to the BMI GWAS and eQTL associations near *SH2B1*.

1830F

Analysis of whole exome and whole genome sequencing using family-based linkage suggests rare variants with large effects are relatively common in extended families. *N.D. Palmer[1], C. Gao[2], C. Lorenzo[3], J.M. Norris[4], J. Long[5,6], Y.I. Chen[7], J.I. Rotter[7], J.E. Curran[8], J. Blangero[8], L.E. Wagenknecht[9], C.D. Langefeld[10], D.W. Bowden[1].* 1) Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA; 2) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC, USA; 3) Department of Medicine, University of Texas Health Science Center, San Antonio, TX, USA; 4) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA; 5) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 6) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 7) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 8) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA; 9) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 10) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA.

Next generation sequencing has identified large numbers of rare variants, creating challenges for association analysis approaches, i.e. limited power, multiple comparisons, and increased false positives. Linkage analysis based on co-segregation of genetic markers with a phenotype through generations is a powerful complement. Data from 1205 Mexican Americans from 90 pedigrees in the Insulin Resistance Atherosclerosis Family Study were analyzed using two-point linkage and association on whole exome sequencing data (452,832 rare variants; MAF<5%). Cardiometabolic phenotypes were targeted (n=53; glucose homeostasis, adiposity, lipids, blood pressure, and biomarkers). Overlapping linkage and association signals were analyzed with family-specific linkage and fine-mapped using whole genome sequencing data. Correlated variants (r^2=1) on chr10 were strongly associated and linked with systolic blood pressure (SBP; rs201624882, rs142314590; P=1.3x10^{-13}, LOD=4.0, MAF=0.25%). rs142314590 is a synonymous variant located in *PRKG1*, a protein kinase family involved in cardiovascular homeostasis. Family-specific linkage identified a single family (LOD$_{Fam2014}$=4.1) with carriers (n=6, 182mmHg) having 1.6 times higher SBP than non-carriers (n=27, 113mmHg). In addition, correlated variants (r^2=1) were linked and associated with triglyceride levels (TG; rs189547099, chr4:157997598; P=6.3x10^{-8}, LOD=3.1, MAF=0.50%). Chr4:157997598 is an intronic variant in *GLRB* that modifies a *ZNF263* transcription factor consensus sequence. Family-specific linkage identified a single family (LOD$_{Fam1045}$=3.1) with carriers (n=13, 358mg/dl) having a 2.9 times higher TG than non-carriers (n=12, 124mg/dl). Significant linkage and association was also observed at chr11:35281477 with insulin secretion (AIR; LOD=3.1, P=2.50x10^{-7}, MAF=0.01%). chr11:35281477 is a 3'UTR variant in *SLC1A2*, a glutamate transporter implicated in insulin secretion. Family-specific linkage identified a single family (LOD$_{Fam2019}$=3.0) with carriers (n=6) having a 3.9 times higher AIR than non-carriers (n=10), conveying a protective effect. Overlapping linkage (LOD>2) and association (P<1x10^{-4}) signals were identified for 270 loci with 38 showing family-specific linkage of LOD>3. These results highlight the strength of family-based design for rare variant analysis, i.e. rare variants are enriched and frequent in families. The combination of association and linkage has the potential for large effect rare variant screening.

1831W

Meta-analysis in continental Africans and African Americans identifies PLCB3 as a novel locus of serum uric acid. *G. Chen, A. Doumatey, J. Zhou, A. Bentley, D. Shriner, A. Adeyemo, C. Rotimi, The AADM study investigators.* NHGRI/NIH, National Institutes of Health, Bethesda, MD.

Objects: Gout and hyperuricemia are disorders caused by impaired urate metabolism with an estimated 8.3 million people affected in the United States. The prevalence of hyperuricemia is highest among African ancestry individuals in the US and similar high prevalence have been observed among sub-Saharan Africans. Compared to European Americans, hyperuricemia has been associated with higher prevalence of cardiovascular and metabolic diseases including hypertension. In 2011, we identified novel GWAS loci for uric acid in African Americans. Here we describe our findings of the first large-scale GWAS in continental Africans. We meta-analyzed our findings with GWAS results from our African American cohort enrolled from Washington, DC as part of the Howard University Family Study (HUFS). Methods: A total of 4,126 individuals enrolled from Nigeria, Ghana and Kenya as part of the African American Diabetes Mellitus (AADM) study were genotyped on the MEGA high density GWAS array and imputed into the African Genome Resources Haplotype Reference Panel using the Sanger Imputation Service. The generalized linear mixed model association test (GMMAT) was used to perform association score test with cryptic relatedness random effect and covariates (sex, age, age2, BMI, diabetes status, and significant principal components). Imputation for the African American dataset used for replication and meta-analysis was also performed using the same Sanger reference panel. Results: A total of 111 SNPs reached genome-wide significance (p< 5.0 × 10-8); these SNPs were distributed across five loci [4p16.1: (SLC2A9, and WDR1), and 11q13.1: (SLC22A12, SF1, MAP4K2, and MEN1)] with the lowest p value of 4.40 × 10-17; The meta-analysis identified 235 SNPs that reaching genome-wide significance and distributed across six loci [4p16.1 (SLC2A9, and WDR1), and 11q13.1: (SLC22A12, SF1, MAP4K2, MEN1, and PLCB3)] with the lowest p value of 1.80 × 10-30. Conclusions: This first GWAS of serum uric acid in continental Africans confirms reported association between variants in SLC2A9, WDR1, SLC22A12, MAP4K2 and MEN1 and uric acid in other ancestral populations. Notably, the meta-analysis with African Americans identified a novel locus (PLCB3) that is associated with serum uric acid in African ancestry individuals thus potentially shedding new insight into the biology of this trait.

1832T

Transient genetic effects important for early growth programming. *O. Helgeland[1,2,4], M. Vaudel[1,3], L.F. Hernández[1], T. Zayats[5], J. Bacelis[6], J. Juoda-kis[7], H. de Haan[4], R. Myhre[4], I.H.G. Østensen[4], O.L. Holmen[8,9], B. Jacobsson[7,10], R.T. Lie[10,11], K. Hveem[8,9], A. Molven[12,13], G.P. Knudsen[10], P. Magnus[10], P.B. Julius-son[14,15], S. Johansson[1,3], P.R. Njølstad[1,2].* 1) K. G. Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 2) Department of Pediatrics, Haukeland University Hospital, Bergen, Norway; 3) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 4) Norwegian Institute of Public Health, Division of Health data and Digitalization, Department of Genetic Research and Bioinformatics, Oslo, Norway; 5) KG Jebsen Center for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway; 6) Department of Obstetrics and Gynecology, Sahlgrenska University Hospital Östra, Sweden; 7) Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Sweden; 8) K. G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway; 9) HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway; 10) Norwegian Institute of Public Health, Division of Health data and Digitalization, Oslo, Norway; 11) University of Bergen, Department of Global Public Health and Primary Care, Bergen, Norway; 12) K. G. Jebsen Center for Diabetes Research, Department of Clinical Medicine, University of Bergen, Bergen, Norway; 13) Department of Pathology, Haukeland University Hospital, Bergen, Norway; 14) Department of Clinical Science, University of Bergen, Bergen, Norway; 15) Morbid Obesity Centre, Vestfold Hospital Trust, Norway.

Background: Large meta-analyses have identified approximately 100 loci associated with adult BMI and 60 with birth weight. However, little is known about the genetic influences on infant and childhood BMI. We performed GWAS on longitudinal measurements of BMI from birth to age 8 years in a population-based child cohort. **Methods:** We examined 9540 children from the Norwegian Mother and Child Cohort Study (MoBa) with 12 measurements of height and weight at birth, at months 1, 3, 6, 8, 12 and 18, and at 2, 3, 5, 7 and 8 years. All children were of Norwegian ancestry with anthropometrics measured at hospitals and primary health care. Samples were genotyped using Illumina HumanCoreExome BeadChip. We tested variants for association with BMI SDSs using linear regression adjusting for sex. LD-score regression was used to estimate the genetic correlation between childhood BMI and other traits at each time point. **Results:** We found that variants in the LEPR/LEPROT locus were strongly associated with BMI as early as 3 months of age, peaking at 6-12 months (p=5.98E-15) and waning from age 3 with little effect at 8 years of age and no association with adult BMI. The associated markers were not in LD with markers in LEPR previously implicated in severe childhood obesity. Furthermore, we explored effects of established adult BMI SNPs on childhood growth. We identified increasing correlation from birth to 8 years (Pearson's r 0.24 to 0.68) with marked differences in timing and effect sizes between markers. Similarly, LD score regression showed that the pattern of genetic correlations with other traits change from BMI at early infancy to young childhood. **Discussion:** This is the first GWAS performed on BMI as early as 1 month of age with dense longitudinal BMI measurements from birth to 8 years of age. We identified variants in the LEPR/LEPROT region with strong transient effects on BMI in and around adiposity peak. Although LEPR has previously been associated with monogenic obesity and severe polygenic childhood obesity around 8 years of age, our variants are not in LD with previously implicated SNPs indicating a novel effect on weight gain in infants. The absence of association in large adult BMI meta-analyses illustrates a genetic effect specific to early weight gain providing insight into growth biology. Thus, our results are the first sign of transient genetic effects in BMI that might help decipher the complexity of childhood growth and development of obesity.

1833F

Genome-wide scan using Korea Biobank Array discovered that two rare missense variants on GPT gene were associated with liver enzyme level. *Y.J. Kim, S.M. Moon, M.Y. Hwang, B.J. Kim.* Division of Genome Research, Center for Genome Science, Korean National Institute of Health, Heung-deok-gu, Chungcheongbuk-do, South Korea.

Recently, genome-wide association studies (GWASs) with hundreds of thousands of samples have implicated unprecedented amount of loci including less common and rare functional variants associated with complex diseases and traits. Since 2014, with the goal of uncovering hidden genetics of complex diseases, we have been conducting the Korea Biobank Array (KBA) project to produce approximately 180,000 cohort-based genome information. KBA, a customized chip optimized for genome study in Korean population, contains about 830,000 markers including about 200K nonsynonymous variants derived from about 2,500 sequenced Korean samples. The KBA achieved higher imputation performance with imputation based genomic coverages of 95.8% and 73.65% for common (MAF >= 5%) and low-frequency variants (MAF 1-5%), respectively. In this study, we performed genome-wide association study for lipids and liver enzymes using about 7,000 genotyped samples with KBA and replicated the findings in about 6,000 samples of an independent cohort. In discovery study, all loci with common variants (P < 5×10^{-7}) were previously known. Among the known loci, we discovered 13 nonsynonymous variants (P < 10^{-4}, predicted to be damaging) including rare variants within *PCSK9* gene replicating the findings from Tang *et al.* 2015. Also, two novel rare missense variants (P < 5×10^{-7}) on *GPT* gene were discovered. In replication study, 15 variants were genotyped using taqman and tested for associations with lipids and liver enzymes. All functional variants were replicated at the level of P < 5×10^{-2}. Taken together, our study discovered numerous functional rare variants, possible novel therapeutic target, associated with lipids and liver enzymes.

1834W

Genome-wide association study of clinically-defined gout and subtypes identifies multiple susceptibility loci including transporter genes. *H. Matsuo[1], A. Nakayama[1], H. Nakaoka[2], K. Yamamoto[3], M. Sakiyama[1], A. Shaukat[4], Y. Toyoda[5], Y. Okada[6,7], Y. Kamatani[7], M. Nakatochi[8], T. Nakamura[9], T. Takada[5], K. Inoue[10], H. Nakashima[11], S. Shimizu[1], M. Kawaguchi[1], A. Hishida[12], K. Wakai[12], B. Stiburkova[13,14], L. Stamp[15], N. Dalbeth[16], T. Hosoya[17], I. Inoue[2], A. Takahashi[7,18], M. Kubo[19], H. Ooyama[20], T. Shimizu[21,22], K. Ichida[17,23], T. Merriman[4], N. Shinomiya[1], Eurogout Consortium, Japan Gout Consortium.* 1) Dept Integrative Physiol, National Defense Med College, Tokorozawa, Saitama, Japan; 2) Div Hum Genet, Dept Integrated Genet, National Inst Genet, Mishima, Shizuoka, Japan; 3) Dept Med Chem, Kurume Univ Sch Med, Kurume, Fukuoka, Japan; 4) Dept Biochem, Univ Otago, Dunedin, New Zealand; 5) Dept Pharmacy, Univ Tokyo Hosp, Tokyo, Japan; 6) Dept Statistical Genet, Osaka Univ Graduate Sch Med, Osaka, Japan; 7) Lab Stat Analysis, RIKEN Cent Integrative Med Sci, Yokohama, Kanagawa, Japan; 8) Statist Analysis Sect, Cent Adv Med Clin Res, Nagoya Univ Hosp, Nagoya, Aichi, Japan; 9) Lab Math, National Defense Med College, Tokorozawa, Saitama, Japan; 10) Dept Biopharmaceutics, Sch Pharmacy, Tokyo Univ Pharmacy Life Sc, Tokyo, Japan; 11) Dept Preventive Med Public Health, National Defense Med College, Tokorozawa, Saitama, Japan; 12) Dept Preventive Med, Nagoya Univ Grad Sch Med, Nagoya, Aichi, Japan; 13) First Faculty Med, Charles Univ Prague Gen Univ Hosp Prague, Inst Inherited Metabol Disorders, Prague, Czech Republic; 14) Inst Rheumatol, Prague, Czech Republic; 15) Dept Med, Univ Otago, Christchurch, New Zealand; 16) Dept Med, Univ Auckland, Grafton, Auckland, New Zealand; 17) Dept Pathophysiol Ther CKD, Jikei Univ Sch Med, Tokyo, Japan; 18) Omics Res Cent, National Cerebral Cardiovasc Cent, Suita, Osaka, Japan; 19) Lab Genotyping Develop, Cent Integrative Med Sci, RIKEN, Yokohama, Kanagawa, Japan; 20) Ryougoku East Gate Clin, Tokyo, Japan; 21) Midorigaoka Hosp, Takatsuki, Osaka, Japan; 22) Kyoto Industr Health Assoc, Kyoto, Japan; 23) Dept Pathophysiol, Tokyo Univ Pharmacy Life Sci, Tokyo, Japan.

Objective To identify novel gout loci including those that are subtype-specific, we performed a genome-wide association study (GWAS) of gout and its subtypes. **Methods** With 1396 clinically-ascertained gout cases and 1268 controls of Japanese males using a custom chip of 1961 single nucleotide polymorphisms (SNPs), putative causal association signals were replicated from a GWAS of 945 cases and 1213 controls. We also first conducted GWASs of gout subtypes. Replication with Caucasian and New Zealand Polynesian samples was done to further validate the loci identified in this study. **Results & Discussion** In addition to the five loci we reported previously, further susceptibility loci were identified at a genome-wide significance level (p<5.0×10[-8]): urate transporter genes (*SLC22A12* and *SLC17A1*) and *HIST1H2BF-HIST1H4E* for all gout cases, and *NIPAL1* and *FAM35A* for the renal underexcretion gout subtype. Whereas *NIPAL1* encodes a magnesium transporter, functional analysis did not detect urate transport via *NIPAL1*, suggesting its indirect association with urate handling. Localization analysis in the human kidney revealed expression of *NIPAL1* and *FAM35A* mainly in the distal tubules, which suggests the involvement of the distal nephron in urate handling in humans. Clinically-defined male gout cases and controls of Caucasian and Polynesian ancestries were also genotyped, and *FAM35A* was associated with gout in all cases. A meta-analysis of the three populations revealed *FAM35A* to be associated with gout at a genome-wide level of significance (Pmeta=3.58×10[-8]). Further findings after this GWAS will be presented at the meeting. Together with *ALDH2*, which is revealed by the fine mapping of *CUX2* region detected by the previous gout GWAS, our findings including novel gout risk loci provide further understanding of the molecular pathogenesis of gout and lead to a novel concept for the therapeutic target of gout/hyperuricaemia.

1835T

Genome-wide study suggests a parent-of-origin effect on birth weight at *ANK1-NKX6-3* type 2 diabetes locus. *R.N. Beaumont[1], O. Helgeland[2,3], R.B. Prasad[4], J.P. Bradfield[5], G. Zhang[6], F. Day[8], S. Johansson[2,3], S. Sebert[7], K. Ong[8], J.R.B. Perry[8], L. Muglia[6], S.F.A. Grant[5], P.R. Njolstad[2,3], Z. Kutalik[9,10], D.M. Evans[11,12], R.M. Freathy[1,11], Early Growth Genetics Consortium.* 1) University of Exeter, Exeter, United Kingdom; 2) University of Bergen, Bergen, Norway; 3) Haukeland University Hospital, Bergen, Norway; 4) University CRC, Skåne University Hospital Malmö, SE-205 02, Malmö, Sweden; 5) The Children's Hospital of Philadelphia, Philadelphia, USA; 6) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 7) University of Oulu, Oulu, Finland; 8) University of Cambridge, Cambridge, UK; 9) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 10) University Hospital (CHUV), Lausanne, Switzerland; 11) University of Bristol, Bristol, UK; 12) The University of Queensland Diamantina Institute, Brisbane, Australia.

Aims/hypothesis: Rare genomic anomalies at a number of known imprinted loci are associated with disorders affecting birth weight. We hypothesized that common genetic variants would also show parental origin-specific associations with birth weight. **Methods:** We analysed 28,836 parent-offspring trios or mother-child pairs from 11 studies to determine parental origin of alleles at 6.1M SNPs genome-wide (MAF >0.01; imputation quality >0.4; imputed to 1000 Genomes). We tested the associations between birth weight and [i] maternally-inherited alleles, and [ii] paternally-inherited alleles. In addition, for each SNP, we compared birth weight between A1$_m$A2$_f$ and A1$_f$A2$_m$ heterozygotes (test [iii]), where suffixes "m" and "f" represent, respectively, maternal and paternal inheritance of alleles A1 or A2. To complement this approach, we tested for differences in birth weight variance between heterozygotes and homozygotes at 12.2M variants (MAF>0.01; imputation quality >0.4) in 104,722 unrelated individuals from 13 additional, non-overlapping studies. **Results:** There were no loci showing genome-wide evidence of unequal parental effects (P<5x10[-8] in test [iii] and either test [i] or [ii]). However, focusing on 60 known birth weight loci, SNP rs13266210 at *ANK1-NKX6-3* (associated also with type 2 diabetes), showed some evidence for a paternal parent-of-origin effect (POE): each paternally-inherited risk allele was associated with a 0.04 SD (95%CI: 0.02, 0.06 SD) higher birth weight (P=9.8x10[-3]); P=6.0x10[-4] for difference between A1$_m$A2$_f$ and A1$_f$A2$_m$ heterozygotes (test [iii]); no evidence of association with the maternally-inherited allele (P=0.48). Additionally, among independent, unrelated individuals there was evidence of greater variance in heterozygotes than homozygotes for the same SNP (P=2.4x10[-3]). **Conclusions/interpretation:** Our analysis provides evidence for a possible POE on birth weight at the *ANK1-NKX6-3* locus already known to influence birth weight, although this requires replication in further samples. The upcoming addition of further data, including GWAS in ~200,000 unrelated individuals, will enhance our power to detect robust associations.

1836F

Meta-analysis in 93,701 East Asians identifies new loci associated with type 2 diabetes. *X. Sim[1], M. Horikoshi[2], C.N. Spracklen[3], Y.J. Kim[4], AGEN and DIAMANTE consortia.* 1) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 2) RIKEN, Centre for Integrative Medical Sciences, Japan; 3) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 4) Division of Structural and Functional Genomics, Center for Genome Science, Korean National Institute of Health, Osong, Chungchungbuk-do, South Korea.

To identify new T2D susceptibility loci in East Asians, we conducted genome-wide meta-analysis in 35,948 T2D cases and 57,753 controls from 21 studies based in China, Japan, Korea, Philippines, Singapore, Taiwan, and United States and imputed to 1000 Genomes reference panels. We performed association analyses with and without adjustment for body mass index (BMI), and combined study-specific association statistics using a fixed-effects, effective sample-size weighted meta-analysis approach assuming an additive genetic model. We analysed 6,957,373 variants, and identified 51 loci at genome-wide significance ($P<5\times10^{-8}$). Of the 51 loci, 42 mapped to known T2D loci. At these loci, only the lead variants at *TMEM154*, *SLC30A8*, *CDKN2A/B*, *TLE1*, *KCNQ1* were identical to previously reported lead variants. Among 9 novel loci, the strongest associated variant maps to an intron of *KSR2* (rs7305424, OR=1.08, $P=1.2\times10^{-9}$, $P_{het}=0.85$). *KSR2* regulates energy intake and expenditure, and rare variants in *KSR2* have been implicated in early-onset obesity and severe insulin resistance. Two of the new signals near *SIX2-SIX3* and *TMEM18* map to loci previously reported to be associated with T2D-related quantitative traits, fasting glucose (FG) and obesity, respectively. At *SIX2-SIX3,* the lead variant rs12712928 (OR=1.07, $P=1.3\times10^{-8}$, $P_{het}=0.20$) was in high linkage disequilibrium ($r^2=0.91$) with the East Asian FG-associated variant rs895636 ($P=6.9\times10^{-8}$, $P_{het}=0.42$), and reciprocal conditional analyses eliminated the signal at this locus. The allele associated with higher FG at rs895636 was associated with increased T2D risk. Associations at *TMEM18* attained genome-wide significance only in analyses without BMI adjustment (rs6719883, $P_{woBMI}=1.2\times10^{-8}$, $P_{BMI}=1.8\times10^{-5}$), suggesting mediation effects of BMI on T2D. Furthermore, conditional analysis on known obesity variant rs13021737 eliminated the association with rs6719883. Allele frequencies at variants for novel loci near *PTPN11* (rs77753011, OR=1.12, $P=9.9\times10^{-9}$, MAF=0.23) and *ZNF208* (rs142395395, OR=1.22, $P=5.0\times10^{-9}$, MAF=0.04) were higher in East Asians compared to the other ancestries (MAF<0.0014 in 1000G non-East Asians). Taken together, we replicated 42 known loci in East Asians and identified 9 new loci that highlight additional genes that may contribute to T2D pathophysiology.

1837W

Using genetics to understand the relationship between inflammation and cardiometabolic traits. *N.R. van Zuydam, A. Goel, C. Grace, TRIiC Consortium.* WTCHG, University of Oxford, Oxford, Oxfordshire, United Kingdom.

Chronic inflammation is associated with obesity, type 2 diabetes (T2D), cardiovascular disease (CAD) and their complications. It remains unclear whether chronic inflammation makes a causal contribution to the development of T2D and obesity, or is merely a consequence. If the latter, it is possible that chronic inflammation secondary to T2D and obesity plays a role in their complications. The patterns of genetic predisposition underlying these traits can be used to dissect these causal relationships. Genome-wide association statistics were assembled for 33 traits including (a) autoimmune diseases and inflammatory traits (e.g. psoriasis, rheumatoid arthritis [RA], CRP); (b) T2D, obesity/BMI and related traits; (c) CAD; and d) diabetic complications including chronic kidney disease (CKD). Polygenic scores (PGS) were generated and two-way Mendelian Randomisation analyses conducted to define causal relationships. A Bonferroni threshold of 8.1×10^{-4} was adopted which took account of the PGS comparisons made. Polygenic scores for autoimmune traits were not associated with T2D, obesity or CAD. The PGS for CRP was associated with increased risk of T2D (OR, 1.19 [1.09-1.30], $p=8.5\times10^{-5}$), BMI (Beta, 0.06[0.05-0.08] SD, $p=3.3\times10^{-14}$), WHR (Beta, 0.04[0.03[0.02-0.05] SD, $p=1.4\times10^{-3}$) but with decreased risk of CAD (OR=0.95[0.84-0.93], $p=2.6\times10^{-6}$). These associations were entirely attributable to variants near *HNF1A* and *APOC1* that are known to be pleiotropic and it is not clear whether the assocation is due to the effect of lipid traits rather than by CRP levels. Polygenic scores for metabolic traits showed some overlapping protection from psoriasis: T2D (OR, 0.85[0.79-0.92], $p=3.5\times10^{-5}$); BMI (0.47[0.37-0.59], $p=1.4\times10^{-10}$), WHR (0.26[0.14-0.48], $p=2.6\times10^{-5}$) and CAD (0.65[0.55-0.77], $p=9.3\times10^{-5}$). We found little evidence for a causal role for inflammation with respect to complications. The only significant association implicated the PGS for RA with CKD in subjects with T2D (OR, 0.90 [0.85-0.94], $p=1.4\times10^{-5}$). Overall, despite isolated signals of overlapping predisposition (some of which are likely to reflect pleiotropic signals), there was no consistent pattern supporting a causal relationship between chronic inflammation and cardiometabolic disease.

1838T

Phenome-wide association study of exome data from childhood obesity cohort reveals pleiotropic loci for 13 obesity-related traits. *S.B. Cho, H.B. Jang, H.J. Lee.* Korea National Institute of Health, Cheong-ju, South Korea.

Here, we performed phenome-wide association study using exome data with 13 obesity-related traits (BMI, waist to hip ratio, fat mass, fat percent, fat-free mass, SBP, DBP, fasting glucose, total-cholesterol, HDL-cholesterol, TG, ALT, and AST). We applied next generation sequencing of whole exome of 692 Korean adolescents who were recruited for childhood obesity cohort. We filtered loci having minor allele frequency < 0.01 and missing alleles > 85%, and 119,935 loci were used for analysis. Linear regression of each trait and exome was performed with age and sex covariates. The p values of each traits then were summed using Fisher's method. With the Bonferroni's multiple testing correction, 3,342 loci were significant (P < 0.05/119,935). The functional annotation of the loci indicated that the result contained a lot of metabolism-related loci that had been identified in the previous researches. We selected 10 loci and preformed validation study using independent cohort (n = 1,000). In the validation study, 7 loci showed significance in the phenome-wide association study with the 13 obesity-related traits. We believe that our results shows the possible genetic loci that are functionally related to the biology of obesity.

1839F

Molecular assessment of variants in inherited lipodystrophy genes: Prevalence and clinical impact in a large clinical care cohort. *C. Gonzaga-Jauregui[1], J. Altarejos[2], T.M. Teslovich[1], N.S. Abul-Husn[1], J. Gromada[2], A. Baras[1], O. Gottesman[1], F. Dewey[1], Geisinger-Regeneron DiscovEHR Collaboration.* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., Tarrytown, NY; 2) Regeneron Pharmaceuticals Inc., Tarrytown, NY.

Inherited Lipodystrophies are a group of genetic disorders characterized by loss of adipose tissue accompanied by metabolic dysregulation. They are broadly divided into Congenital Generalized Lipodystrophies (CGLs), which are mostly autosomal recessive disorders and Familial Partial Lipodystrophies (FPLs), which are inherited dominantly. Lipodystrophies are generally considered to be very rare genetic disorders with reported prevalences ranging from 1 in 10 million for CGLs, to 1 in 1 million for FPLs. We interrogated the electronic health record (EHR) information for ~1.6 million individuals in the Geisinger Health System for lipodystrophy/lipoatrophy diagnosis codes. We then performed genetic analyses of individuals with available genetic data through our Geisinger-Regeneron DiscovEHR collaboration in order to identify the likely causative variants of lipodystrophy in these patients. Of the 18 lipodystrophy-diagnosed patients with available genetic data, we identified potentially pathogenic variants in known lipodystrophy associated genes in 8 individuals. Of these, 4 individuals carry a pathogenic variant in *LMNA* (p. R482Q) associated with Dunnigan familial partial lipodystrophy. This variant is present in our clinical population at a frequency of 0.000115, with additional 10 individuals harboring a molecular finding for the condition but no documented diagnosis of lipodystrophy. A review of the EHR of these individuals showed that they have metabolic abnormalities consistent with lipodystrophy including diabetes, dyslipidemia, and pancreatitis. We then broadly surveyed the DiscovEHR cohort for pathogenic and likely pathogenic variants in lipodystrophy associated genes in order to evaluate the clinical and metabolic profiles of these variant carriers. We observed a substantial burden of metabolic dysregulation in these individuals. Overall, our data show that partial lipodystrophy is an underdiagnosed condition and that its prevalence, as defined molecularly, might be higher than previously reported. Additionally carriers of lipodystrophy-associated variants show metabolic abnormalities in the spectrum of lipodystrophy disease, however they are more often diagnosed as having type 2 diabetes and unspecified metabolic syndrome. Molecularly-guided stratification of patients with "common" disorders, such as diabetes and dyslipidemia, is an important step towards precision medicine and has the potential to lead to more effective therapies for patients.

1840W

Genetic association of irisin with obesity and metabolic syndrome. *H.B. Jang, H.J. Kim, H.J. Lee.* Korea National Institute of Health, Cheongju, South Korea.

Circulating irisin have been suggested to be involved in the pathogenesis of various complications of obesity, including metabolic syndrome (MetS) and Type 2 diabetes mellitus. We have performed exome-wide association studies to identify genetic variants that influence serum irisin levels and confer susceptibility to obesity and MetS in 677 Korean adolescents. After adjustment with age and sex, we identified 17, 9, and 4 single nucleotide polymorphisms (SNPs) as being significantly associated with irisin, obesity, and MetS ($P<5.0x10-5$), respectively. Of these, irisin-related 10 SNPs in seven genes (VPS25, EZH1, RAMP2, WNK4, CNTD1) contributed to the risk for obesity and MetS ($p<0.05$) and MetS-related genetic variants of MMP8 and MMP27 were associated with serum irisin levels and risk of the obesity ($p<0.05$). This study provides insight into the pleotropic effects of exome-wide association loci of irisin, obesity and MetS. Based on these results, we are in conducting replication study for independent Korean population.

1841T

Genotype determination: Analysis of *PNPLA3, GC*, and *LCP1* genes in nonalcoholic fatty liver disease in south of Iran. *S.S. Tabei[1], H. Faramarzi[2], M. Silawi[1].* 1) Medical Genetics, Shiraz University of Medical Sciences, Shiraz, Fars, Iran; 2) Community Medicine Department, Shiraz University of Medical Sciences, Shiraz, Fars, Iran.

Nonalcoholic fatty liver disease (NAFLD) is the most common liver disease and liver manifestations of metabolic syndrome. NAFLD is a complex trait which is influenced by polygenes, environmental factors, age and interaction between these factors and is specified by daily intake of <20 g alcohol, dysfunction of >5% hepatocells in lipid filtration, altered levels of hepatic transaminases and changed blood lipid profile. Obtained data by National Health and Nutrition Examination Survey during 1998 to 2008, represent that prevalence of chronic liver disease except NAFLD has remained stable. In the current study with regard to the growing number of NAFLD cases, we aimed to study the frequency and impacts of three different SNPs: rs738409 (*PNPLA3* gene), rs222054 (*GC* gene), and rs7324845 (*LCP1* gene) in NAFLD group in comparison with control group by RFLP-PCR approach (case group was comprised by 119 NAFLD patients and 118 individuals as control group). Additionally, demographic data were compared between two groups. We found significant association between rs738409 and fatty liver disease ($p<0.001$), but no association was seen between rs222054 and rs7324845 and NAFLD ($P>0.05$). In the next step raised results revealed that BMI, waist circumference, and cholesterol had the most significant association with NAFLD development ($p<0.01$, $p<0.002$, and $p<0.009$, respectively). Also, the incidence of NAFLD in females compared with males showed significant difference ($p<0.001$). rs738409 (C>G) which resides in *PNPLA3* gene and since 2008 is considered as a potential risk factor for development and progression of NAFLD. In the current study a significant relation between this variation and fatty liver disease was seen ($p<0.001$) and confirmed the previous results. The presence of rs738409 seems to be associated with obesity, insulin resistance and elevated levels of hepatic amino-transaminases, but the the last feature was not seen in the current study and needs further investigations.

1842F

Metabolomics screen in five metabolic tissues from healthy, prediabetic and type 2 diabetic subjects suggests new defects and points of gene environment interaction. *C. Wadelius[1], K. Diamanti[2], M. Cavalli[1], G. Pan[1], U. Riserus[3], S. Skrtic[4], J. Komorowski[2,5].* 1) Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala SE-75108, Sweden; 2) Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala SE-75124, Sweden; 3) Department of Public Health and Caring Sciences, Clinical Nutrition and Metabolism, Uppsala University, Uppsala SE-75185, Sweden; 4) AstraZeneca R&D, Mölndal, Sweden; 5) Institute of Computer Science, Polish Academy of Sciences, Warsaw 01-248, Poland.

Type 2 diabetes (T2D) is a complex disease associated with a large number of genetic variants and other risk factors identified in cross-sectional and longitudinal studies. The main metabolic organs suggested to contribute to disease are pancreatic islets, liver, muscle and fat with secondary effects in other tissues. However, the pathophysiological processes in the main tissues have not been elucidated. Many of the risk factors for diabetes identified in blood are secondary to events elsewhere. In order to search for possible primary disease events, we studied a unique material of tissues (pancreatic islets, liver, muscle, fat and serum) from healthy (16), prediabetic (12) and T2D (12) donors. A comprehensive metabolomics analysis was done on all samples and association to disease was determined based on fold change of 50% or p<0.05 between the groups, in ANOVA or two-tailed t-tests. In line with published data certain lysophosphatidylcholines (lysoPCs) were decreased in serum in T2D compared to healthy controls, while some bile acids were increased, thus validating the methodological approach. In addition, we found that several lysoPCs were decreased in liver, muscle and fat, indicating a widespread metabolic defect. LysoPCs are derived from components in cell membranes and other structures and act as signaling molecules. Furthermore, we found that several carnitines were increased in T2D tissues. Carnitines are conjugated to fatty acids and catabolized in mitochondria to create energy. LysoPCs derived from several polyunsaturated fatty acids (PUFAs) were decreased in different tissues. These PUFAs are synthesized by enzymes encoded by the polymorphic *FADS1/2* and *ELOVL2/5*, which are associated with a range of diseases and traits. We have recently shown that *FADS1* is regulated by rs174557 in an enhancer, which has several binding sites for SREBP1c and is activated by insulin. Analysis of the regulatory elements controlling *ELOVL2* is ongoing and suggests that other environmental signals act on this gene. In summary, our data suggests widespread metabolic defects in tissues of T2D patients that have not been identified before. Our analysis indicates possible involvement of key metabolic enzymes, which will be investigated further. It also underscores the value of studying disease processes in the relevant tissues and of combining genetic data with other –omics data sets.

1843W

Identification of I287S homozygous mutation in the MLX gene in an infant with non-alcoholic steatohepatitis: A case report. *Y. Watanabe[1,2], S. Watanabe[3], C. Sato[3], S. Ogasawara[4], J. Akiba[5], M. Kage[6], H. Yano[4], K. Fukui[1], T. Mizuochi[1], S. Yano[7], K. Hayasaka[8], Y. Yamashita[1], T. Kawaguchi[9], T. Torimura[9], KI. Yoshiura[3].* 1) Dept Pediatrics, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 2) Research Institute of Medical Mass Spectrometry, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 3) Dept of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 4) Dept of Pathology, Kurume University School of Medicine, Kurume, Japan; 5) Dept of Diagnostic Pathology, Kurume University Hospital, Kurume, Japan; 6) Kurume University Research Center for Innovative Cancer Therapy, Kurume, Japan; 7) Genetics Division, Dept of Pediatrics, University of Southern California, CA, USA; 8) Dept of Pediatrics, Yamagata University School of Medicine, Yamagata, Japan; 9) Division of Gastroenteorlogy, Dept of Medicine, Kurume University School of Medicine, Kurume, Japan.

Introduction: Nonalcoholic steatohepatitis (NASH) can be seen not only in adults but also in infants. In infants, various genetic disorders are thought to be involved in the development of NASH although it is often difficult to identify the etiology. A high concentration of intracellular glucose activates lipid synthesis through up-regulation of the lipogenic enzymes including acetyl-CoA carboxylase and fatty acid synthase. Activities of the lipogenic enzymes are known to be regulated by several transcription factors including a Max-like factor X (Mlx) in mice. Mlx contains the basic helix-loop-helix leucine zipper and the dimerization and cytoplasmic localization domain (DCD). Mlx plays an important role in a glucose-mediated transcription of the lipogenic enzymes in the liver through forming a heterodimer with carbohydrate response element-binding protein (ChREBP) and Mondo-A. Hyperglycemia causes nuclear localization of Mlx-ChREBP and Mlx-Mondo-A heterodimers. The former binds the carbohydrate response element leading to activate transcription of the lipogenic enzyme genes and the latter activates transcription of the genes coding the glycolytic enzymes. We present, for the first time, a patient with infantile onset of NASH with a homozygous mutation in the MLX gene. **Case Report:** A 6-month-old Japanese male, a product of consanguineous parents (half siblings through their mother), was presented with developmental delay, hypotonia, massive hepatomegaly, hypoglycemia, ketosis, and lactic acidosis. Liver biopsy at age 6 months showed macrovesicular steatosis and the pathohistological diagnosis of NASHwas made. Whole exome sequencing study identified 12 homozygous mutations on the autosomes in the detected 27 runs of homozygosity. Among the identified 12 genes, *MLX* is the only one involved in fatty acid metabolism, suggesting involvement of the homozygous p.I287S mutation in the DCD of *MLX* in the development of NASH. Immunohistochemical studies using anti-MLX antibody clearly showed nuclear hyperactive signals of *MLX* in our case but not in an age matched control. Nuclear expression of *MLX* was seen in the hepatocytes of adult patients with NASH. **Discussion:** We demonstrated the first case of infantile NASH with a p.I287S homozygous mutation of *MLX* with the increased expression of MLX in the hepatocytes suggesting involvement of *MLX* in the development of infantile onset NASH. More cases are needed to confirm this hypothesis.

1844T

Causality links between gut microbiome and glucose/insulin metabolism and type 2 diabetes. *S. Sanna[1], U. Võsa[1], A. Vich Vila[1,2], M.J. Bonder[1], A. Kurilshikov[1], F. Imhann[1,2], R. Weersma[2], L. Franke[1], A. Mahajan[3], N. Van Zuydam[3], M. Costa[3], M. McCarthy[3], A. Zhernakova[1], J. Fu[4], C. Wijmenga[1].* 1) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands; 2) University of Groningen, University Medical Center Groningen, Department of Gastroenterology and Hepatology, The Netherlands; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 4) University of Groningen, University Medical Center Groningen, Department of Pediatrics, The Netherlands.

Accumulating evidence suggests that the gut microbiome influences the development of obesity and metabolic diseases. Studies in mice have proposed a causal effect of gut microbiome on the development of obesity and type 2 diabetes (T2D). Due to the complex host-microbe interaction, we hypothesized that a bi-directional causal effect exists between gut microbiome, T2D and T2D intermediary phenotypes. We investigated the effects of impairments in glucose/insulin metabolism on gut microbiome in the LifeLinesDeep population cohort (952 normo-glycemic individuals with metagenomics and genomic profiles) using a Mendelian Randomization (MR) approach. We extracted summary statistics from 32 published genome-wide association studies on obesity, fasting insulin and glucose, insulin secretion, proinsulin and T2D, and constructed 52 polygenic risk scores (PRS) using LD-pruned SNPs at two levels of significance (5x10-8 [N=20] and 1x10-5 [N=32]). Of the 273 taxonomies and 523 bacterial pathways assessed, 32 and 22, respectively, were significantly correlated with at least one PRS (FDR<0.1), indicating which specific microbiome composition and function is affected by the host's innate ability to maintain a correct insulin/glucose metabolism. For example, several species of *Actinobacteria* and *Archaea*, such as *Methanobrevibacter smithii* which are important for the digestion of polysaccharides, were less abundant in individuals genetically predisposed to secrete high levels of insulin after glucose intake, and more abundant in those volunteers at high genetic risk for T2D. Individuals at high-risk for T2D showed also a reduced abundance of bacteria involved in degradation of carbohydrates. Our results highlight a causal relation of insulin/glucose metabolism in modulating the microbiome composition and function, bringing new insights into host-microbe interaction.

1845F

Zebrafish larvae as a model system for high-throughput, image-based screens in insulin resistance and diabetes. *A. Emmanouilidou[1,2], P. Ranefall[2,3], M. Bandaru[1,2], A. Allalou[2,3], E. Ingelsson[4], M. den Hoed[1,2].* 1) Department of Immunology, Genetics and Pathology, Uppsala Universitet, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala University; 3) Department of Information Technology, Division of Visual Information and Interaction, Uppsala University, Sweden; 4) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA.

Background Genome-wide association studies identified >100 loci that are robustly associated with the risk of insulin resistance and diabetes. The causal genes and mechanisms remain unidentified for most loci. We aim to examine if zebrafish larvae can serve as a model system for image-based genetic screens in insulin resistance and diabetes. **Methods** In a dietary screen, ~500 larvae with transgenically expressed fluorescent labels on β-cells and/or hepatocytes were overfed on a normal or cholesterol-supplemented diet in the presence or absence of 3% glucose in the medium from 5 to 10 days post-fertilization. In a separate effort, ~800 larvae were metabolically challenged with or without concomitant treatment with rosiglitazone or metformin. At 10dpf, larvae were soaked in a lipid staining dye, followed by optical sectioning of the pancreatic islet, liver and subcutaneous lipid stores using an automated positioning system and fluorescence microscope. ß-cell volume and subcutaneous and liver fat accumulation were quantified objectively using custom-written scripts. Quantitative outcomes were inverse normally transformed and data were analyzed using linear (beta±SE) or logistic (OR [95% CI]) regression, adjusting for time of imaging and batch. **Results** Continuous exposure to 3% glucose showed a positive main effect (0.38±0.17 SD) and a positive interaction with cholesterol supplementation (0.47±0.23 SD) for β-cell volume. Glucose exposure also increased the odds of subcutaneous (2.24 [1.28-3.90]) and hepatic fat accumulation(2.44 [1.20-4.96]). Dietary cholesterol supplementation increased the odds of hepatic fat accumulation (2.15 [1.10-4.22]) but reduced the odds of subcutaneous fat accumulation (0.40 [0.23-0.71]). Treatment with rosiglitazone reduced hepatic fat accumulation by -0.82±0.10 SD and increased the odds of subcutaneous fat accumulation (2.16 [1.07-4.37]). Treatment with metformin had the opposite effect (0.33±0.09 SD; 0.38 [0.16-0.92]). Neither rosiglitazone nor metformin affected ß-cell volume (*P*>0.3). **Conclusion** Challenging zebrafish larvae metabolically for five days induces an insulin resistant state that can to some extent be prevented by treatment with rosiglitazone. Hence, zebrafish larvae are a promising model system for genetic screens in insulin resistance and diabetes. Multiplex CRISPR-Cas9 mutant models for proof-of-concept genes are currently being screened to further validate the model system.

1846W

New insights into the role of genetic variation within *FGF21* in the pathogenesis of obesity. *E. Aerts[1], E. Geets[1], A. Verrijken[2], G. Massa[3], K. Van Hoorenbeeck[4], S. Verhulst[4], L. Van Gaal[2], W. Van Hul[1].* 1) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) Department of Endocrinology, Diabetology and Metabolic Diseases, Antwerp University Hospital, Antwerp, Belgium; 3) Department of Pediatrics, Jessa Hospital, Hasselt, Belgium; 4) Department of Pediatrics, Antwerp University Hospital, Antwerp, Belgium.

Objective Brown adipose tissue is a key site of heat production that has the potential to increase energy expenditure. After the discovery of active brown fat and inducible "brown-like" or beige adipocytes in human adults, browning of white adipose tissue has been considered an attractive obesity target. The fibroblast growth factor 21 (FGF21) is one of the most essential inducers of beige adipocytes and so dysregulation of this endocrine factor may be implicated in energy regulation. Therefore, we hypothesized that pathogenic mutations within *FGF21* might be involved in the pathogenesis of obesity. **Methods** We screened 406 severely overweight or obese children and adolescents for mutations in the coding region of the *FGF21* gene with high-resolution melting curve analysis. Direct sequencing was performed for samples with melting patterns deviating from wild-type. To determine the effect of the mutations on the cAMP and NFAT signaling activity of FGF21, luciferase reporter assays were performed. **Results** Our mutation analysis resulted in the identification of two non-synonymous coding variants in two unrelated obese individuals: p.R47Q and p.L142V. For p.L142V we were able to demonstrate a significant decrease of the cAMP and NFAT signaling activity of FGF21. **Conclusion** With this study, we have demonstrated the presence of a rare variant in a severely obese 6-year-old girl (BMI z-score 3.01) and confirmed its pathogenic effect by performing a luciferase reporter assay. In conclusion, these data confirm the importance of genetic variation within the *FGF21* gene in energy regulation.

1847T

Genetic evidence that early carbohydrate-stimulated insulin secretion affects accumulation and distribution of adiposity. *C.M. Astley[1,2,3], S.W. Kim[3], J.N. Todd[1,2,3,4], R. Salem[5], R. Fine[3,2], C.B. Ebbeling[1,2], D.S. Ludwig[1,2], J.C. Florez[2,3,4], J.N. Hirschhorn[1,2,3].* 1) Boston Children's Hospital, Boston, MA; 2) Broad Institute, Cambridge, MA; 3) Harvard Medical School, Boston, MA; 4) Massachusetts General Hospital, Boston, MA; 5) University of California, San Diego, CA.

According to the Carbohydrate-Insulin Model of obesity, a large carbohydrate load exaggerates early postprandial insulin secretion, thereby altering substrate partitioning toward fat deposition and promoting weight gain. This hypothesis is supported by mechanistic studies of metabolic fuels, observational studies, and dietary intervention clinical trials, but not all studies show consistent findings. Genetic predictors of insulin level at 30 minutes following an oral glucose load (insulin-30) can be used as instrumental variables in Mendelian randomization to help further test causality for the association between genetically-determined insulin-30 and obesity and weight gain and to estimate the effect of insulin-30 on anthropometric traits. These genetic-based methodologies are less susceptible to confounding bias due to the randomization of insulin-30 increasing alleles at meiosis. We sought to test the Carbohydrate-Insulin Model using genetic risk score and Mendelian randomization methods, applied to large data sets from MAGIC and GIANT Consortia, and UK Biobank. We assessed effects on both measures of obesity (waist-to-hip ratio, WHR and body mass index, BMI) as well as interval change in BMI. Genetic risk scores for insulin-30 and anthropometric traits were validated and tested in data sets that are independent from the discovery genome-wide association studies. The genetic risk score for insulin-30 predicted interval gain in BMI (p-value < 0.05) better than the genetic risk score for BMI (p-value = 0.61). Mendelian randomization using genetic instruments for the exposure of insulin-30 support causal relationships of increased insulin-30 with increased BMI, waist-to-hip ratio (WHR), and interval gain in BMI (p-values < 0.05). One of the insulin-30 genetic instruments (near *GIPR*) may be mechanistically linked to enteral carbohydrate stimulated insulin secretion, and is associated with a more metabolically unhealthy phenotype, as assessed by WHR. In conclusion, these results provide further support of the Carbohydrate-Insulin Model of obesity suggesting that early insulin secretion after a carbohydrate load, likely in combination with a higher glycemic load diet, contributes to increased adiposity, central adiposity, and continued weight gain over time.

1848F

Exome sequencing in African American children with early-onset obesity reveals new insights. *A. Chesi[1], S.E. McCormack[2,3], V.V. Thaker[4,5], H. Hakonarson[3,6], J.N. Hirschhorn[4], S.F.A. Grant[1,2,3,6,7,8].* 1) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Diabetes and Endocrinology, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Division of Endocrinology, Boston Children's Hospital and Harvard Medical School, Boston, MA; 5) Division of Molecular Genetics, Columbia University Medical Center, New York, NY; 6) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 7) Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 8) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

African American (AA) youth have an excess burden of obesity compared to children of European ancestry (19.5% vs 14.7%). Given that genetics of obesity have been relatively understudied in populations of African descent, this ethnicity offers an opportunity to uncover of additional causal variants over what has been previously reported. We prioritized for whole exome sequencing AA children (n=181) with both early-onset obesity (<6 yrs) and the most rapid rise in BMI to allow for exploration of rare coding variants typically undetected by GWAS. 'Super Imposition by Translation And Rotation' (SITAR) was leveraged to select cases by fitting BMI curves using a shape-invariant model with a regression B-spline mean curve and subject-specific random effects on both measurement and age scales. Children with known genetic etiologies and/or underlying predisposing conditions were excluded. To reduce the analytical search space and to identify variants more likely to contribute to the trait, we employed two approaches: 1) leveraging a panel of 32 genes with known association to monogenic forms of obesity, 2) examining all genes (n=166) harbored within corresponding topologically-associating domains (TADs) to established BMI/obesity GWAS signals. We focused our analysis on variants predicted to confer a severe (stop-gain, splice-site donor/acceptor, frameshift indel) or moderate (missense, non-frameshift indel) effect on their transcripts, and a MAF<1% in the African population available in ExAC. In known monogenic obesity genes, we identified 8 severe heterozygous variants (4 novel, plus notably a stop-gain R143X in *UCP3* annotated as pathogenic in ClinVar for severe obesity and T2D and a frameshift in *BBS10* annotated as likely pathogenic for recessive Bardet Biedl syndrome) and 63 moderate variants (20 novel, plus the following knowns: *POMC* R236G associated with early-onset obesity, *MC4R* F202L annotated as of uncertain significance for obesity susceptibility and *VPS13B* V3941VD annotated as probably pathogenic for recessive Cohen syndrome). Going TAD-wise, we identified an additional 34 severe variants (9 novel, including a R734X stop-gain in *RPGRIP1L*, a neighboring gene of *FTO*) and 173 moderate variants (34 novel). As such we uncovered rare genetic variants in AA children with early-onset and persistent obesity, a phenotype more likely to have a genetic etiology. These findings warrant further investigation in larger cohorts and assessment in functional studies.

1849W

Identifying subject-specific regulatory networks of diet-induced weight loss. *D.C. Croteau-Chonka[1], K. Glass[1], S.R. Smith[2], B.J. Stubbs[1], N.M. Laranjo[1], F.M. Sacks[1,3], G.A. Bray[4], L. Qi[1,5], V.J. Carey[1], B.A. Raby[1].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Translational Institute, Florida Hospital, Orlando, FL; 3) Department of Nutrition, Harvard School of Public Health, Boston, MA; 4) Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, LA; 5) Obesity Research Center, Tulane University, New Orleans, LA.

Our aim was to characterize the effects of weight loss on adipose tissue biology. We hypothesized that weight loss induces specific structural changes to the transcriptional regulatory network and that the extent of change is correlated with the amount of weight loss. Preventing Overweight Using Novel Dietary Strategies (POUNDS Lost) was a randomized clinical trial testing the effect of dietary interventions on weight loss. We studied 101 POUNDS Lost subjects (mean age = 52.4 years, standard deviation (SD) = 7.9) who lost an average of 8.3 kg (SD = 4.6) over six months. Paired subcutaneous adipose tissue (SAT) samples collected at baseline and six months later for each subject were assayed for 48,803 gene expression probes. We used LIONESS, or Linear Interpolation to Obtain Network Estimates for Single Samples, an extension of the Passing Attributes between Networks for Data Assimilation (PANDA) method, to construct subject-specific regulatory networks before and after weight loss. Overall, there was a trend towards more within-subject network remodeling in SAT of subjects who lost more weight ($r = 0.18$, $P = 0.072$). Subjects who lost less weight tended to have pre- and post-dieting networks that were more similar to one another, while those individuals who lost more weight tended to have more network alterations. Among 462,340 transcription factor (TF)-gene pairs in the aggregate network, sets of 89 and 119 genes demonstrated significantly increased and decreased targeting, respectively, by specific TFs in relation to amount of weight loss ($P < 0.05$). The 89 increasingly targeted genes were enriched for targets of three TFs (adjusted $P < 0.05$): LKLF, SLUG, and PPARA. PPARA, or peroxisome proliferator activated receptor alpha, plays a role in the management of energy stores during fasting. LKLF, or Krüppel-like factor 2, is highly expressed in fat and is a negative regulator of adipocyte differentiation, acting to inhibit PPAR gamma (PPARG). SLUG, or snail family transcriptional repressor 2, also mediates adipogenesis via PPARG. Diet-induced weight loss alters gene expression in adipose tissue, reflecting perturbations of regulatory networks representing coherent and specific biological pathways. Network analyses will help further characterize specific regulatory relationships among genes responding to weight loss at aggregate and individual scales.

1850T

Copy number variation and mutation analysis indicate a possible interesting role of *POU3F2* in the Prader Willi *like* phenotype. *E. Geets[1], E. Aerts[1], A. Verrijken[2,3], G. Massa[4], K. Van Hoorenbeeck[5], S. Verhulst[5], L. Van Gaal[2,3], W. Van Hul[1].* 1) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) Department of Endocrinology, Diabetology and Metabolic Diseases, Antwerp University Hospital, Antwerp, Belgium; 3) Laboratory of Experimental Medicine and Paediatrics, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; 4) Department of Paediatrics, Jessa Hospital, Hasselt, Belgium; 5) Department of Paediatrics, Antwerp University Hospital, Antwerp, Belgium.

   Background: Prader-Willi syndrome (PWS), caused by a paternal defect on 15q11.2 – q13, is the most common form of syndromic obesity. Once a clinical diagnosis of PWS cannot be confirmed molecularly with the use of MS-MLPA or additional techniques, patients are classified as Prader Willi *like* (PWL). Deletions at 6q14.1 – q16.3 encompassing *SIM1* were reported in some individuals with a PWL phenotype. More recently, Kasher and colleagues also identified small deletions encompassing *POU3F2*, and not *SIM1*, in ten individuals from six families with developmental delay, intellectual disability, neonatal hypotonia, susceptibility to obesity and hyperphagia, characteristics resembling the PWL phenotype. In addition, with the use of morpholino and mutant zebrafish models, they showed that the gene lies downstream of *SIM1* in the leptin-melanocortin signaling pathway which indicates that *POU3F2* also has a potentially interesting role in the obesity phenotype of PWL patients. Consequently, we decided to perform copy number variation (CNV) analysis of the *POU3F2* region in a PWL cohort followed by mutation analysis of the gene in both a PWL cohort and obese children and adolescents. **Methods:** A genome-wide microarray analysis was performed in a group of 109 PWL patients. Next, we screened 95 PWL patients and 459 obese children and adolescents for mutations in *POU3F2* using high-resolution melting curve analysis and Sanger sequencing. **Results:** CNV analysis of 109 PWL patients did not show any structural variation encompassing the *POU3F2* gene. The mutation screening resulted in the identification of three interesting nonsynonymous variants: a previously unreported variant p.H255R (c.764A>G) in one PWL patients and two variants, p.G85R (c.253G>A) and p.G352V (c.1055G>T), in two obese adolescents. The latter was also not previously described in public sequencing databases. Additionally, *in silico* analysis showed a probably damaging effect of all three variants on the protein structure of POU3F2. **Conclusions:** In contrast to Kasher and colleagues, no gene harboring deletions were identified in the *POU3F2* region in our PWL cohort. Secondly, taking into account their absence or very low minor allele frequencies in public sequencing databases and the results of *in silico* prediction programs, further functional analysis of p.G85R, p.H255R and p.G352V is useful. This would provide further support for a possible role of *POU3F2* in the pathogenesis of the PWL phenotype.

1851F

A dinucleotide deletion in a putative miRNA target site in long-chain fatty acid elongase Elovl6 associates with higher thermogenesis and lower body mass index in Pima Indians. *P. Kumar, P. Piaggi, M. Traurig, L.J Baier.* Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ.

   To determine whether miRNAs have a role in influencing obesity among Pima Indians, variation in miRNAs and predicted miRNA binding sites was identified in whole-genome sequence data from 335 Pima Indians. A dinucleotide deletion (CT; rs200193952) in the 3'-UTR of the *ELOVL6* gene was identified that may disrupt a predicted binding site for miR-204/211. *Elovl6* encodes the long-chain fatty acid elongase that catalyzes the conversion of C16 saturated and monounsaturated fatty acids to C18 fatty acids. *Elovl6* also has an important role in thermogenic action of brown adipose tissue (BAT). Loss-of-function of *Elovl6* causes reduced expression of mitochondrial electron transport chain components that lower the BAT thermogenic capacity. *Elovl6* deficient mice gain weight with increased subcutaneous white adipose tissue and impaired carbohydrate metabolism when beige fat was disabled by thermoneutrality or aging. The CT deletion has a frequency of 5.1% in Pima Indians, while among the populations reported in the 1000 Genomes database, this variant has a frequency of less than or equal to 4.1%. Association analysis of rs200193952 in Pima Indians who had been longitudinally studied identified a modest association with maximum BMI in adulthood ($p=0.009$, beta=-0.0287 BMI log_e units, N =6714), and maximum BMI Z-score in childhood ($p=0.005$, beta=-0.1611, N =5256), where the deletion predicted leanness. We hypothesize that the dinucleotide deletion within the miR-204/211 target site would lead to higher Elovl6 protein levels, which would increase the thermogenic capacity resulting in a lean phenotype. Therefore, we analyzed energy expenditure (EE) data as continuously measured over 24 hours inside a whole-room indirect calorimeter during energy balance and weight stability. The 24-hour EE trajectory of deletion carriers showed a higher post-prandial thermogenesis despite similar sleeping EE. As a result, carriers had higher "awake and fed" thermogenesis (AFT), calculated as energy expenditure of the non-active fed condition above the minimum metabolic requirement during sleep ($p=0.008$, beta =94.2 kcal/day, N =382). In summary, we have identified variation in a regulatory miRNA target site in *ELOVL6* that associates with lower BMI in Pima Indians and may regulate Elovl6 levels ultimately increasing daily thermogenesis.

1852W

Low serum insulin-like growth factor-II levels correlate with high body mass index in older American Indian adults. *Y. Muller, D. Mahkee, P. Piaggi, W. Knowler, C. Bogardus, R. Hanson, L. Baier. Phoenix Epidemiology and Clinical Research Branch, National Institutes of Health, Phoenix, AZ.*

The hormone insulin-like growth factor-II (IGF-II) regulates metabolism and growth. Data from experimental and gene-association studies have shown that IGF-II may influence body weight or Body Mass Index (BMI). In mouse models, there is an inverse relationship between IGF-II levels and body weight, where overexpression of IGF-II leads to a significant reduction in fat mass and lipid content. In humans, however, both positive and inverse relationships have been reported between circulating IGF-II levels and weight gain or obesity. Therefore, in the current study we assessed the relationship between serum IGF-II levels and BMI in a population based sample of American Indians. Serum samples for measurement of IGF-II levels were available on 713 subjects who were part of a longitudinal study of the etiology of type 2 diabetes among the Gila River Indian Community in Arizona. Blood samples were drawn during an exam when BMI was measured by a medical professional and the subject was determined to be non-diabetic. Serum IGF-II was measured using the method of Enzyme-linked Immunosorbent Assay (ELISA). Cross-sectional analysis showed that serum IGF-II levels did not differ by gender (mean IGF-II=444.3 ng/ml in 285 males; mean IGF-II=449.0 ng/ml in 428 females; P=0.77); however, serum IGF-II levels had a strong inverse correlation with age (r=-0.39, P=2.6x10^{-29}). Among all subjects, lower IGF-II levels correlated with higher BMI (r=-0.17, P=8.2x10^{-6}, adjusted for age, sex and serum storage time). When stratified by age, the strongest correlation was observed in subjects ≥30 years (r=-0.29, P=5.2x10^{-7}, N=283), a modest correlation was observed in subjects 20-29 years (r=-0.15, P=0.016, N=267) and no correlation was observed in subjects 15-20 years (r=0.08, P=0.21, N=192). Among the 713 subjects, 231 subsequently developed type 2 diabetes. In a Cox proportional hazards model, IGF-II levels did not significantly predict subsequent development of diabetes (HRR=0.98, P=0.80, adjusted for age, sex, serum storage time and BMI). In summary, our data from a population based sample of American Indians indicate an inverse relationship between serum IGF-II levels and BMI, where the correlation is stronger at older ages.

1853T

Contrasting the genetic architecture of human thinness and severe obesity. *F. Riveros Mckay Aguilera[1], V. Mistry[2], R. Bounds[2], A.E. Hendricks[1,3], J.M Keogh[2], H. Thomas[2], E. Henning[2], S. O'Rahilly[2], E. Zeggini[1], E. Wheeler[1], I. Barroso[1,2], I.S. Farooqi[2], Understanding Society Scientific Group.* 1) Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom; 3) Department of Mathematical and Statistical Sciences, University of Colorado-Denver, Denver, CO 80204, USA.

Obesity is a highly heritable trait and one of the leading causes of preventable death worldwide. In recent years, the prevalence of the condition has been steadily increasing making it a major global public health priority. In Western countries this rise in prevalence has been linked to environments facilitating poor-quality food choices and sedentary behaviour. Genetic association studies so far have been mainly focused on identifying genes and pathways associated with either body mass index or obesity. From a drug development perspective though, identifying biological pathways associated with lifelong healthy thinness within an "obesogenic" environment can lead to novel plausible therapeutic targets. To this end, we characterised the heritability of persistent human healthy thinness in 1471 individuals, and contrasted it with that of early onset severe childhood obesity (N=1456). Our analysis revealed that persistent leanness is a heritable trait (h^2=28.07%) comparable to childhood obesity (h^2=32.33%). Persistent thinness has a negative genetic correlation with adult BMI (r^2=-0.63, 95% CI [-0.44,-0.82], p=3.54x10^{-11}) while childhood obesity has a positive and stronger genetic correlation (r^2=0.86, 95% CI [0.74, 0.98], p=1.86x10^{-43}). We also found that the loci influencing these traits at the tails of the BMI distribution are not complete mirror images of each other. Instead some loci, like *CADM2*, have a stronger effect on the lower end of the BMI distribution, supporting the hypothesis of different mechanisms, or gene-environment interactions in the two extremes of the distribution. Finally, using a genome-wide association approach with imputation to UK10K + 1000 genomes (genome-wide significance threshold P=5x10^{-8}), we highlight a novel variant associated with persistent thinness near *PI15* (rs12362703, p=1.52x10^{-8}) and a new locus associated with obesity (*FAM150B*, OR=1.72, p=5.68x10^{-9}). Altogether, our results show that studying persistent thinness can be a powerful approach to further understand the genetic architecture of body weight regulation.

1854F

The role of genetic and self-identified ancestry in determining obesity among African and Hispanic Americans. *A. Vishnu[1,2], G. Belbin[1], E. Kenny[1], E. Bottinger[1], R.J.F. Loos[1,2].* 1) The Charles Bronfman Institute for Personalized Me, Icahn School of Medicine at Mount Sinai, New York, NY; 2) The Genetics of Obesity and Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY.

Individuals of African and Hispanic ancestry are at higher risk of obesity than individuals of European ancestry in the United States (US). This may be due to differences in environmental factors, such as cultural and social practices, but may also in part be genetically driven. Here, we examined to what extent African ancestry is associated with body-mass index (BMI) among 6,368 African-American (AA) and 7,569 Hispanic American (HA) participants in the New York City based BioMe™ biobank. We quantified global genetic ancestry using ADMIXTURE, specifying two putative ancestral populations (k=2) to distinguish African from non-African ancestry. We excluded participants with significant Native American ancestry i.e. ≥30% (determined with k=3). We used linear regression to examine association between percentage African ancestry and BMI among AAs and HAs. Covariates examined were age, sex, self-reported ancestry (HA vs. AA), US vs. non-US born status, and their interactions. Stepwise regression procedures were used to select variables which significantly ($P<0.05$) contributed. Participants exhibited varying levels of admixture, resulting in a wide spectrum of African ancestry present in both self-reported AAs (25th, 50th, 75th, 90th pcts: 79%, 86%, 92%, 96%, respectively) and HAs (15%, 26%, 41%, 60%, respectively). Among AAs and HAs, 18% and 60% of participants are non-US born, respectively. As analysis showed a strong interaction between ancestry and sex, we report results for men and women separately. Among women, genetic and self-reported ancestry, age and country of birth were significantly associated with BMI. However, the association differed between US-born and non-US-born women (P-interaction=0.004). For every 10% higher genetic African ancestry, BMI increases by 0.40 kg/m^2 (P=1.7E-07; ~1.09kg for a 1.65m tall woman) and 0.19 kg/m^2 (P=3.1E-04; ~0.51kg) among US-born and non-US born women, respectively. Independent of genetic ancestry, BMI in women who self-report as AA was 0.79 kg/m^2 (P=1.2E-02) lower than self-reported HAs. In contrast, among men self-reported HA status (+0.96 kg/m^2, ~2.94 kg for a 1.75m tall man) and being US-born (+1.37 kg/m^2, ~4.21kg) influenced BMI, while no significant association was detected for genetic ancestry (P=0.2). Besides the contribution of environmental factors, genetic ancestry contributes to variation in BMI among HA and AA women. However, among men, only country of origin and self-reported ancestry play a role.

1855W

Genetic diversity and functional genomic mapping in an Emirati population with type 2 diabetes. *K.S. Elliott[1], H. Daggag[2], G.B. Busby[1], L.J. Petherbridge[1], E. Lau[1], M. Butler[3], A. Yavari[3], M.T. Barakat[2], H. Ashrafian[3], J.C. Knight[1].* 1) Wellcome Trust Centre for Human Genetics, Univeristy of Oxford, UK; 2) Imperial College of London Diabetes Centre, Abu Dhabi, UAE; 3) Experimental Therapeutics, Radcliffe Department of Medicine, University of Oxford, UK.

Type 2 diabetes (T2D) is a leading global health problem, particularly acute in the Middle East where prevalence in Arab countries is 25.5% compared to 9.3% in the US. Unique genetic architecture due to marriage culture can provide a powerful resource to uncover novel genetic factors previously unidentified in large Caucasian cohorts. To explore this, over 700 individuals of Emirati descent (comprising confirmed and undiagnosed T2D) have been recruited through the ICLDC in Abu Dhabi. Genotyping used Illumina's Multi Global Ethnic SNP array was followed by principle components analysis (PCA) with 1000 Genomes phase 3 data. Overlap with West Asian and African populations is consistent with geography and human population history. Further PCA centred on the Abu Dhabi Emirati cohort has shown that many individuals with the same family name cluster together indicating a shared ancestry. More detailed analysis using Finestructure revealed distinct groups which share more ancestry with each other than other individuals within the cohort which also correlates with family name. T2D and obesity related clinical records have allowed GWAS to search for disease risk alleles. Preliminary analysis has revealed an excess of p<10^{-4} signals in known BMI genes (p=2.3x10^{-4}) and one of the top signals in T2D is a SNP in the *CDKAL1* gene (p=4.6x10^{-6},OR=1.2) independent of the previously reported GWAS hit in Europeans. The presence of this variant in one of the largest effect size T2D genes increases confidence in positive replication. Such an association in a relatively small sample size infers that the genetic component of metabolic disease in the Emirati population may be greater compared to Caucasian populations, where very large sample sizes have been required to detect equivalent signals. This gives optimism for discovery of further novel T2D and obesity variants as analysis continues. Verification of these results is being sought in appropriate related populations. Using publicly available data we will extend the analysis to include other geographically neighbouring populations from the Middle East, North Africa, Caucasus and Western Asia also providing further insights into the genetic architecture and evolution of the Emirati population. This novel population makes a powerful and innovative dataset for identifying T2D and obesity variants specific to the Emirati population which could be used in the future to direct appropriate healthcare and treatment.

1856T

Effects of long noncoding RNA regulation of gene expression on type 2 diabetes. *A.J. Payne[1], M. van de Bunt[1,2], C.M. Lindgren[1,3,4], M.I. McCarthy[1,2,5], The GTEx Consortium.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Oxford, United Kingdom; 3) The Big Data Institute, University of Oxford, Oxford, United Kingdom; 4) Broad Institute of Harvard and MIT, Cambridge, MA, United States; 5) NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

Long noncoding RNAs (lncRNAs) account for a large proportion of human transcripts and have been implicated in cellular function which when disrupted can contribute to disease. We assessed lncRNA associations with type 2 diabetes (T2D) and relevant processes using data from the GTEx Project. Using 48 tissues from up to 565 individuals in GTEx v7 (22% with T2D), analysis identified 355 genes with significant (FDR of 0.01) differential expression (DE) between T2D cases and controls, 344 of them coding genes, and 11 lncRNAs. To assess the potential of lncRNAs as regulators in T2D-associated processes, co-expression modules were constructed using all expressed genes in each tissue. Enrichment for T2D-associated lncRNAs was assessed in each module by comparing the Fisher score of DE p-values of its lncRNAs to a distribution of Fisher scores from random lncRNAs. Of the 5580 total modules, 131 were enriched (FDR of 0.01). Two-sample Mendelian Randomisation (MR) using genetic instruments associated with each module eigengene found no causal effects of the 131 T2D-lncRNA-enriched module eigengenes on any of 8 diabetes-related GWAS phenotypes, including T2D, fasting insulin, and fasting glucose. One module of 54 co-expressed genes (from pancreas) was enriched for the GO terms 'insulin secretion' and 'regulation of insulin secretion': no other modules over-represented terms containing 'insulin' or 'glucose'. This module's eigengene was significantly associated with T2D risk (logistic regression $p=6.75\times10^{-5}$). The module contained 5 coding genes with significant T2D DE (*PPP1R1A*, *IAPP*, *G6PC2*, *HADH*, *HHATL*), and a lncRNA with nominal DE (*CTB-12O2.1*; $p_{adjusted}=0.012$). T2D-risk was associated with lower expression of each. *CTB-12O2.1*, a long intergenic noncoding RNA, has no known links to T2D or glycaemic traits. Two-stage least squares regression MR using elastic-net-derived genetic instruments found that *CTB-12O2.1* expression affected the module's eigengene (p=0.0013), suggesting a regulatory function of *CTB-12O2.1* amongst the module's genes. This study identified a cluster of genes co-expressed in whole human pancreas that is enriched for lncRNA T2D DE, is over-representative of genes involved in insulin secretion, and has aggregate expression associated with T2D risk. LncRNA *CTB-12O2.1* has a potential regulatory effect on expression of this module, making it a candidate regulator of a T2D-associated process in humans.

1857F

Identification of eQTLs affecting expression levels in both adipose and skeletal muscle tissues. *W.-C. Hsueh, S. Kobes, R.L. Hanson.* NIDDK, NIH, Phoenix, AZ.

Transcriptomic studies in tissues related to glucose metabolism may yield insights into genetic causes of diabetes. By analyses of empirical identity by descent (IBD) sharing between individuals in a founder population, we previously showed significant heritability in many of 16,820 transcripts measured in both skeletal muscle (n=220) and subcutaneous abdominal adipose tissues (n=209). We now extend this to examine the role of expression quantitative trait loci (eQTL) across tissues. Transcript levels were measured on the Affymetrix Human Exon 1.0 ST Array and genotypes on the Illumina HumanCore-Exome-24 BeadChip. Extent of both genome-wide and gene-specific (200kb up- and down-stream of its coding region) IBD between all pairs of subjects was estimated based on genotypes for ~217,000 SNPs with Beagle. We estimated narrow-sense heritability (h^2) of each transcript affecting levels in both tissues, and partitioned out the contributions of cis- (h^2_{cis}) and trans-elements (h^2_{trans}), adjusting for age, sex and an ancestry estimate. With correction for multiple comparisons by the false discovery rate (FDR), we identified 470 significant ($p_{FDR}<0.05$) *cis*-eQTL for muscle tissue and 196 *cis*-eQTL for adipose. *Cis*-eQTL identified in one tissue were significantly more likely to also be *cis*-eQTL in the other (Yule's Q=0.89, $p=3.9\times10^{-119}$). Transcript level measures in both muscle and adipose tissues were available from 103 individuals. Levels of 1,028 transcripts (6%) were nominally (p<0.05) correlated across tissues. Adipose *cis*-eQTL were 6.4 times more likely to correlate across tissues than the usual transcript, while muscle *cis*-eQTL were 4.6 times more likely (p<0.0001 for both). For 53 transcripts that were significantly correlated across tissues ($p_{FDR}<0.05$), we conducted bivariate analysis to partition the correlation into *cis*-genetic, *trans*-genetic and environmental effects. The majority (47 or 89%) of these transcripts shared significant *cis*-genetic determinants, while 13% shared significant t*rans*-effects, and non-genetic factors were significant for 45%. On average, 46% of the cross-tissue correlation was due to *cis*-eQTL, 17% due to *trans* effects and 37% to non-genetic effects. These analyses suggest that *cis*-eQTL are major genetic determinants of cross-tissue correlation in transcript levels, and that lessons learned by identification of *cis*-eQTL may often generalize across tissues.

1858W

Evaluating the contribution of alternative splicing in the liver to variation in lipid levels. *K.A.B. Gawronski[1,2], M. Caliskan[2], Y. Park[2], S. Rao[2], K.M. Olthoff[7], A. Shaked[7], B.E. Engelhardt[8], D.J. Rader[2,5,6], B.F. Voight[2,3,4], C.D. Brown[2].* 1) Genetics and Gene Regulation Program, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 4) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 5) Division of Translational Medicine and Human Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 6) Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 7) Department of Surgery, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 8) Department of Computer Science, Princeton University, Princeton, NJ.

Evaluating the contribution of alternative splicing in the liver to variation in lipid levels

Lipid levels are heritable traits associated with cardiovascular disease risk, and previous genome-wide association studies (GWAS) have identified more than 150 loci associated with these traits. However, the underlying genetic mechanism(s) for the majority of lipid loci are not well understood. While intense effort has been devoted to the identification of genetic variants that causally affect complex traits via gene expression (eQTLs), recent research indicates that changes in the abundance of alternatively spliced transcripts may be another contributing mechanism (Zhang et al., 2015). Consequently, identifying genetic loci that associate with alternative splicing (*i.e.*, sQTLs) in primary liver tissue - an organ highly relevant for lipid traits - would be ideal, but has not been described to date.

To address this gap, we present results from an sQTL discovery effort using data from more than 300 primary liver and stem cell-derived hepatocyte samples. Gene ontology analysis confirms that the alternatively-spliced genes associated with identified sQTLs are enriched for liver-specific processes including alcohol and lipid metabolism. To evaluate the contribution of our sQTLs to variation associated with lipid levels, we present the results of co-localization analysis using GWAS data generated by the Global Lipid Consortium (Global Lipids Genetics Consortium et al., 2013). We demonstrate that sQTLs co-localize with lipid loci previously unexplainable using eQTL data alone. Finally, given that liver-dependent gene regulation has been shown to impact lipid levels, we compare the enrichment of presumed tissue-restricted sQTLs in GWAS lipid loci to the enrichment of multi-tissue sQTLs (identified using 44 tissues from GTEx) in GWAS lipid loci. Given that relatively few studies have evaluated the relationship between alternative splicing and complex traits, our efforts provide an important contribution to understanding genetic mechanisms influencing both cardiovascular disease risk and complex traits in general.

1859T

Meta-analysis of >150 genome-wide studies for association with blood lipid levels. *S.E. Graham[1], G. Hindy[2,3], W. Zhou[4], G. Abecasis[5], G. Peloso[3,6,7], S. Kathiresan[3,6,8,9], C.J. Willer[1,4,10], Global Lipids Genetics Consortium.* 1) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI; 2) Department of Clinical Sciences, Lund University, Malmö, Sweden; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 5) Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI; 6) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 8) Department of Medicine, Harvard Medical School, Boston, MA; 9) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 10) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Cardiovascular disease is a predominant cause of death worldwide and is mediated by circulating levels of blood lipids. Approximately half of all cases of cardiovascular disease are attributed to inherited genetic risk factors, yet only ~10-12% of genetic variation is thought to be explained by previously identified variants. To further understand the genetic contributions to cardiovascular disease we are performing a meta-analysis of over 150 studies with an estimated final sample size of >500,000 individuals collected through the Global Lipids Genetics Consortium (GLGC). All samples were genotyped with a genome-wide array and were imputed using 1000 Genomes Phase 3 with European samples additionally imputed using the Haplotype Reference Consortium. We tested for association between single variants and indels with 5 lipid phenotypes: total cholesterol, HDL-cholesterol, LDL-cholesterol, non-HDL cholesterol, and triglycerides. We collected covariance matrices from all cohorts to allow for conditional analysis and rare variant burden testing. We will present our findings aimed at discovering novel lipid loci and uncovering biological mechanisms for association at these loci by integrating with epigenetics data. We will also determine the causal role of these lipid phenotypes in cardiovascular disease by performing multivariate Mendelian randomization studies and examining lipid genetic risk score associations with diseases in the UK biobank.

1860F

Population and medical genetics of the Kibbutzim Family Study. *S. Carmi[1], E. Granot-Hershkovitz[1], D. Karasik[2,3], I. Peter[4], Y. Friedlander[1], H. Hochner[1].* 1) Public Health, The Hebrew University of Jerusalem, Jerusalem, Israel; 2) Medicine, Bar Ilan University, Safed, Israel; 3) Institute for Aging Research, Hebrew SeniorLife, Boston, MA; 4) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Background: The Kibbutzim Family Study (KFS) tracks large families living in long-term social stability and homogeneous environment in Northern Israel. Extensive information on anthropometric and cardio-metabolic phenotypes was collected on 1032 participants during two visits, eight years apart. We genotyped 901 individuals on an exome-chip array (500k SNPs), making the KFS one of the largest genotyped family-based cohorts in Israel. Here, we report results for the population-genetic characteristics of the KFS as well as for genotype-phenotype associations. Population genetics: The KFS participants are mostly self-reported Ashkenazi Jewish (AJ; 85%), with the remaining reporting various other Jewish ancestries. Principal Component Analysis of the KFS, jointly with Jewish and world-wide samples, confirmed self-reported ancestries and allowed precise assignment of ethnic origins to the remaining individuals. Interestingly, we observed, for the first time, a weak genetic distinction between AJ originating from Eastern and Western Europe. Using family-based phasing and shared haplotypes analysis, we confirmed claims of a recent severe AJ bottleneck (point estimates: effective size ≈450 individuals, 23 generations ago). Complex traits: We imputed the KFS using 1000 Genomes data and 128 AJ whole-genomes, reaching 6.9M variants with imputation quality >0.9 and minor allele frequency >1%. We used linear mixed models to conduct a genome-wide association study of 17 cardio-metabolic risk phenotypes. Our strongest associations, which replicated previously discovered regions, were for Lipoprotein (a) (rs56393506, upstream of *LPA*; $P=9 \cdot 10^{-17}$), weight (chr.8p23.1; $P=3.6 \cdot 10^{-8}$) and low-density lipoprotein (chr.16q22.1-2; $P=8.8 \cdot 10^{-7}$). Effect sizes were largely correlated between baseline values and the differences across visits, suggesting shared genetic architecture. Finally, we observed, for some phenotypes, an enrichment of association signals at alleles with an elevated frequency in the KFS compared to Europeans, in line with the expected value of the AJ founder population for trait mapping. Summary: The KFS is a valuable resource for the study of the population genetics of Israel as well as of cardio-metabolic traits under a homogeneous environment.

1861W

Regulatory activity and deletion of rs3780181 suggests a molecular mechanism at the *VLDLR* lipid GWAS locus. *J. Davis, T. Roman, S. Vadlamudi, M. Zeynalzadeh, A. Iyengar, K. Mohlke.* Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Blood lipids, including total cholesterol (TC) and low-density lipoprotein cholesterol (LDL-C), are heritable risk factors for cardiovascular disease. The molecular mechanisms underlying the majority of blood lipid-associated GWAS signals remain elusive. One association signal is located in intron 3 of *VLDLR*: rs3780181-A is associated ($P \le 2 \times 10^{-9}$) with increased TC and LDL-C. rs3780181 is also associated with the expression level of *VLDLR* in liver samples. To investigate the variants and genes underlying the association signal, we evaluated allelic differences in functional regulatory assays and used genome editing to delete the putative regulatory element. We tested a 222-bp DNA element surrounding rs3780181 for enhancer activity using transcriptional reporter assays in human HepG2 hepatocytes, THP-1 monocytes, and SGBS preadipocytes. We generated 3-4 independent clones for each allele in both forward and reverse orientations with respect to a minimal promoter. We observed the highest enhancer activity in THP-1 cells, 5-fold more than an empty vector (EV) control. In HepG2 and THP-1 cells, the risk allele rs3780181-A had decreased enhancer activity than the non-risk G-allele element in the both orientations. The allelic difference in enhancer activity was significant ($P \le 0.009$) in both orientations in HepG2 cells, and significant ($P=0.02$) in the forward orientation in THP-1 cells. In electrophoretic mobility shift assays using nuclear extracts of THP-1 and HepG2 cells, the risk allele rs3780181-A showed stronger specific protein binding than the G-allele. To determine if the enhancer containing rs3780181 affects expression of nearby genes *in vivo*, we used CRISPR-cas9 to engineer a deletion of the DNA surrounding rs3780181 in HEK293T kidney cells, and we measured the effect of the homozygous deletions on expression of nearby genes: *VLDLR*, a lipoprotein receptor, and the adjacent gene, *SMARCA2*, a chromatin-remodeler. All expression levels were normalized to the housekeeping gene *ACTB*. The enhancer deletion cell lines (n=8) showed 1.2-fold significantly ($P<0.04$) decreased *VLDLR* and *SMARCA2* expression when compared to the wild-type (n=3) and mock-edit control cell lines (n=6). Together, these results suggest that altered binding of a transcriptional repressor to the A-allele of rs3780181 decreases activity of an enhancer of *VLDLR* and/or *SMARCA2*, leading to increased TC and LDL-C.

1862T

Heritability and genetic correlation of 25 complex traits in Taiwanese population. *C. Lin[1,2], M. Su[1,2], C. Hsiung[1,2], C. Shen[1,2].* 1) Taiwan Biobank, Academia Sinica, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Recent studies have shown that the heritability variation across a wide range of multiple traits, which is always specific to a particular population. Taiwan Biobank (TWB) provides a unique opportunity to evaluate the heritability of multiple traits in Han Chinese population. We estimated the proportion of phenotypic variance explained by the common variants for 25 complex traits derived from the TWB. The genetic correlation between pairs of traits in Taiwanese population were estimated by Genome-wide Complex Trait Analysis (GCTA). More than 16,000 samples were genome-wide genotyped using the Affymetrix Axiom TWB array, which was specifically designed for Taiwanese population. After data quality control, genotype data of 598,060 single nucleotide polymorphisms (SNPs) in 13,231 unrelated individuals were analyzed. We found the highest heritability (0.47) estimated by common SNPs is body height among 25 complex traits. Based on the phenotypic correlation and genetic correlation between 25 traits, triglyceride and uric acid show a relatively high correlation with multiple traits such as systolic blood pressure (SBP), Hemoglobin A1c (HbA1c), and high-density lipoprotein (HDL). It is indicated that triglyceride and uric acid may play an important role in most of common disease such as hypertension, type 2 diabetes, and hyperlipidemia. We also noticed that increased triglyceride or increased uric acid might causally increases SBP and HbA1c and decreases HDL. Our study describes the phenome-wide heritability and genetic correlation in Taiwanese population, and reveals that triglyceride and uric acid may play an important role in the causal relationship among type 2 diabetes, hypertension, and hyperlipidemia. The purpose of these results may demonstrate a good model of potential of TWB how to make good use of TWB information to estimate heritability of common diseases.

1863F

Genome-wide association study of HDL efflux phenotypes in 5,143 French Canadians. *K.S. Lo[1], C. Low-Kam[1], D. Rhainds[1], M.-P. Dubé[1,2], G. Lettre[1,2], J.-C. Tardif[1,2].* 1) Montreal Heart Institute, Montreal, Quebec, Canada; 2) Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada.

High-density lipoproteins (HDL) are the centerpiece of reverse cholesterol transport by which cholesterol is removed from peripheral cells and transported to the liver for excretion. Despite this potent anti-atherosclerosic role, neutral clinical trials and Mendelian randomization studies have shed doubt on the "good" cholesterol hypothesis (Barter et al. 2007; Schwartz et al. 2012; Voight et al. 2012). These results have shifted the interest of the community from merely increasing HDL-cholesterol (HDL-C) concentration to understanding HDL composition and functions. Then, recent studies have shown that cholesterol efflux of serum HDL (the capacity of HDL particles to accept cholesterol from macrophages) is an independent risk predictor of coronary artery disease (CAD) (Khera et al. 2011; Rohatgi et al. 2014). Here, we attempted to identify the genetic determinants of HDL efflux in the French Canadian population. In a subset of 2,002 participants (~50% with myocardial infarction (MI)), we obtained HDL-related measures (HDL-C, apoA-I and LCAT concentrations), nuclear magnetic resonance spectroscopy data, and HDL efflux using 3 different cellular assays (J774, BHK, HepG2). We observed significant correlations between HDL-C concentration and triglyceride levels (r=-0.47), large HDL particle concentration (r=0.87), HDL particle size (r=0.7), and HDL efflux (J774 basal efflux, r=0.65; J774 stimulated, r=0.4). HDL efflux was independently associated with reduced MI risk after adjustment for traditional risk factors (P=6x10^{-11} for J774 basal efflux). We measured HDL efflux in an additional 3,397 participants. We combined low-pass whole-genome sequencing and genome-wide genotyping followed by imputation to test the association between HDL efflux phenotypes and genotypes at >31M variants. The strongest hits included loci previously associated with HDL-C concentration (*e.g. CETP, LIPC, LPL, APOA1*). We repeated the association analyses, correcting for HDL-C and triglyceride levels. We saw no enrichment of association results for HDL efflux when focusing on variants previously associated with CAD. Also, there was no enrichment of association signals when considering regulatory sequences from the liver or hepatocytes. Overall, our initial results suggest that HDL effluxes are typical human complex traits controlled by variants of weak effect sizes. This highlights the importance of increasing sample size to define the genetic determinants of this promising CAD biomarker.

1864W

Genetic effect assessment of functional variants on blood lipid traits by exome-wide association study. *S. Moon, S. Han, M.Y. Hwang, K.H. Yoon, Y.J. Kim, B-J. Kim.* Division of Genome Research, Korea National Intsitute of Health, Cheongju-si, Chungcheongbuk-do, South Korea.

Dyslipidemia of high density lipoprotein cholesterol (HDLc), low density lipoprotein cholesterol (LDLc), triglyceride (TG) and total cholesterol (TC) involved in the lipid metabolism has a crucial role in the pathogenesis of metabolic syndrome emerging as a major public health issue. Although large-scale genome-wide association identified many genetic loci associated with the metabolic syndrome and related traits, there is a few studies on the effect of coding polymorphisms. Here, we evaluated the role of exonic variants related the lipid traits (HDLc, LDLc, TG, and TC) to discover exonic variants associated with the development of metabolic syndrome. We performed exome-wide analysis using about 14,000 individuals (7,442, 3,407 and 3,037 in discovery, replication 1 and 2, respectively) from Korean Genome and Epidemiology Study genotyped with the Illumina HumanExome Beadchip (hereafter referred to as the ExomeChip). Using the Exome Aggregation Consortium browser, we examined genetic effects of all functional variants including missense, loss-of-function in the ExomeChip. Consequently, a total of 3,474 functional variants of 24 genes which have at least a functional variant associated with lipid traits. From the results, we confirmed that 45 common variants listed in GWAS catalog are replicated in our study regardless of ethnic difference. Moreover, we assessed genetic effects of 3,474 functional variants. Of these, 138 variants were significantly associated with lipid traits ($P < 5 \times 10^{-6}$ in discovery, $P < 5 \times 10^{-2}$ in each replication stage). Of 138 associated variants, 53 variants showed a pleiotropic effect in at least two traits. Some of variants were significant only in Asian population. Our results provide researchers with the useful evidence to understand causal variants and risk factors associated with lipid-related traits.

1865T

A multi-trait genetic association approach to identify genetic loci not identified before in single-trait GWAS of lipid traits. *M. Preuss[1,2], G. Nadkarni[1], R. Loos[1,2,3].* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai; 2) The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai; 3) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai.

Obesity, combined with a poor lipid profile, increases the risk of cardiovascular disease substantially. Large-scale genome-wide association studies (GWAS) have identified more than 360 genetic loci associated with body mass index (BMI), and circulating triglyceride, HDL- and LDL-cholesterol (HDL-C, LDL-C) levels. These GWASs have mainly focused on one single trait at a time, and the loci identified may not fully represent the biology that underlies these correlated traits. Here, we performed a multi-trait genome-wide association analysis that considers those four traits simultaneously in one analysis. We speculate that such multi-trait approach will identify genetic loci that highlight a more comprehensive biology to further elucidate the link to cardiovascular disease. We used mtSet (Casale FP et al. 2015), a multi-trait LMM set test that enables joint analysis across multiple correlated traits and multiple sets of variants while accounting for population structures and relatedness. MtSet uses as a sliding window (of 30kb) approach, resulting in ~174,000 windows and a Bonferroni corrected genome-wide significance (GWS) of $P = 2.9 \times 10^{-7}$. Each trait was adjusted for age, and residuals were log- and rank-transformed to normality in men and women separately. Phenotype data was extracted from electronic health records (EHRs) for 1,560 African ancestry (AA) and 1,982 Hispanic ancestry (HA) participants of the Mount Sinai Bio*Me* BioBank for whom 1000 Genomes (Phase 3) imputed data was available. No loci reached genome-wide significance, but six reached suggestive significance ($P<10^{-5}$), covering *NAALADL2* ($P=1.5\times10^{-6}$), *KIAA1324L* ($P=8.5\times10^{-6}$) and a CNV at 5q14 ($P=1.2\times10^{-6}$) in individuals of African ancestry AA, and *LMCD1* ($P=8.7\times10^{-7}$), *a lincRNA at* 4q28 ($P=1.8\times10^{-6}$), and *ZNF431* ($P=2.9\times10^{-6}$) in individuals of Hispanic Ancestry. Even though our study has still insufficient power, results are promising despite the small sample size. We are now increasing the sample size to confirm our observations and to subsequently narrow-down the loci and interpret their biological relevance. None of the suggestive loci have been previously implicated in obesity and lipid-related traits. We believe that a multi-trait genome-wide analysis provides and alternative approach to further and more comprehensively elucidate the broader biology that underlies correlated traits and diseases.

1866F

Low LDL cholesterol concentrations are associated with increased risk of type 2 diabetes. *Q. Feng[1], W.Q. Wei[2], C.P. Chung[3], R.T. Levinson[4], A.C. Sundermann[5], J.D. Mosley[1], L. Bastarache[2], J.F. Ferguson[6], N.J. Cox[4], D.M. Roden[1], J.C. Denny[2], M.F. Linton[4,5], D.R.V. Edwards[4,5], C.M. Stein[1,6].* 1) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 3) Division of Rheumatology, Vanderbilt University Medical Center, Nashville, TN; 4) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 5) Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University, Nashville, TN; 6) Division of Cardiovascular Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.

Background Observations from Mendelian randomization studies and statin clinical trials suggest that low LDL-C levels may be associated with increased risk of type 2 diabetes (T2DM). **Methods** We used de-identified electronic health records (EHRs) at Vanderbilt University Medical Center to compare the risk of T2DM among individuals with very low (≤60 mg/dL, N=8,943) and normal (90-130 mg/dL, N=71,343) LDL-C levels. LDL-C measurements associated with statin use, hospitalization, or a serum albumin level < 3 g/dl were excluded. We used a two-phase approach; in 1/3 of the sample (discovery) we used T2DM phenome-wide association (PheWAS) codes, and in the remaining 2/3 (validation) we identified T2DM cases and controls using a validated algorithm. The prevalence of T2DM in the low and normal LDL-C groups was compared using logistic regression with adjustment for age, race, sex, body-mass index (BMI), HDL-C, triglycerides, and duration of care. Sensitivity analyses included exclusion of subjects with a previous medical condition representing a possible indication for statin therapy (diabetes or myocardial infarction) and those with advanced renal disease and organ transplantation. Secondary analyses included stratification of sex, race, BMI, and LDL-C level. **Results** In the discovery cohort, PheWAS code 250.2 (T2DM) was significantly more frequent in the low LDL-C group (p= 1 × 10⁻³³). In the validation cohort, the risk of T2DM was increased in the low compared to normal LDL-C group (OR 2.4, 95% CI 2.1-2.8; P<2 × 10⁻¹⁶). The findings were near identical in sensitivity analyses. The association between low LDL-C levels and T2DM was significant in males (OR 2.4, 95% CI 2.0-3.0) and females (OR 1.7, 95% CI 1.4-2.1), and in normal weight (OR 2.2, 95% CI 1.6-3.0), overweight (OR 2.2, 95% CI 1.7-2.8) and obese (OR 2.0, 95% CI 1.7-2.4) individuals, and in individuals with LDL-C < 40 mg/dL (OR 2.3, 95% CI 1.7-3.1) and LDL-C 40-60 mg/dL (OR 2.0, 95% CI 1.7-2.3); P-values all <10⁻⁶. The association was significant in whites (OR 2.7, 95% CI 2.3-3.2; P<2 × 10⁻¹⁶) but not blacks (OR 1.1, 95% CI 0.8-1.5; p=0.56). **Conclusions** Low LDL-C concentrations occurring in the absence of statin treatment were significantly associated with T2DM risk in a large EHR population; this increased risk was present in both sexes and all BMI categories, and in whites but not blacks. The findings suggest special attention be paid to T2DM risk with therapies aiming to markedly lower LDL-C.

1867W

Whole exome sequencing identifies coding variants associated with NMR-based lipid phenotypes in a large cardiovascular cohort. *S. Giamberardino[1], R. McGarrah[1], L.C. Kwee[1], J. Overton[2], J. Reid[2], S.G. Gregory[1], A. Baras[2], C.B. Newgard[1,3], W.E. Kraus[1], S.H. Shah[1].* 1) Duke Molecular Physiology Institute, Duke University, Durham, NC; 2) Regeneron Genetics Center, Tarrytown, NY; 3) Sarah W. Stedman Nutrition and Metabolism Center, Duke University, Durham, NC.

Although high density lipoproteins (HDL) are known colloquially as "good cholesterol," with lower HDL levels associated with higher cardiovascular (CV) risk, increasing evidence suggests that more granular HDL determinations around HDL particle size are better predictors of risk. Studies have shown that higher levels of small HDL particles exhibit strong CV protective effects. And while genetic studies of lipids have identified novel therapeutic targets, they have not focused on these more refined evaluations which may provide better signals. Utilizing the CATHeterization GENetics (CATHGEN) cohort, a cohort of patients who received cardiac catheterization, we measured HDL subclasses using NMR in frozen, fasting plasma. Whole exome sequencing was performed and resulted in 8,557 individuals for analysis. Given prior studies showing the strongest association between the molar sum of concentrations of small and medium HDL subclasses (HMSP) with CV risk, analyses were focused on this lipid phenotype. After quality control measures were applied, single variant association testing was performed using linear regression models adjusting for within-ancestry population stratification, batch, age and sex, stratified by ancestral groups. With a MAF threshold of 0.01, stratified analysis datasets contained 5,882 individuals and 170,081 autosomal variants for the European ancestral group and contained 1,671 individuals and 294,165 autosomal variants for the African ancestral group. Results from these stratified tests were then combined and meta-analyzed. Five exomic variants met genome-wide significance (p≤5×10⁻⁸) for association with HMSP in the meta-analysis and localized to three genomic regions: chr2:27.3-27.5Mbp (rs1260326 and rs4665972 in *GCKR* and *EIF2B4*, min. p=1x10⁻¹⁰), chr19: 44.9Mbp (rs7412 in *APOE*, p=4x10⁻⁰⁸) and chr20: 45.9-46.0Mbp (rs3918251 and rs7679 in *SLC12A5-AS1* and *PCIF1*, min. p=1x10⁻⁸), with a consistent direction of effect across ancestral groups. Using a large CV cohort, we have identified genetic loci strongly associated with the HDL subclass biomarker HMSP, a more granular marker of HDL cholesterol and a stronger predictor of CV risk than conventional HDL. These loci highlight potential novel lipid loci mediating inter-individual variation in HDL cholesterol phenotypes but also identify a known lipid and CV locus (*APOE*).

1868T

Novel genetic variants associated with lipid levels in a multi-ethnic population from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *Y. Hu[1], M. Graff[2], K. Nishimura[1], J. Haessler[1], S. Buyske[3], C. Haiman[4], L. Le Marchand[5], L.W. Martin[6], L. Wilkens[5], N. Zubair[1], S. Bien[1], J.L. Ambite[7], R. Do[8], Y. Lu[8,9], M. Verbanck[8], R. Tao[9], D. Stram[4], R.J.F. Loos[8], T. Matise[10], C. Kooperberg[1], K.E. North[2], I. Cheng[11], U. Peter[1].* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 3) Department of Statistics & Biostatistics, Rutgers University, New Brunswick, NJ; 4) Keck School of Medicine, University of Southern California, Los Angeles, CA; 5) University of Hawaii Cancer Center, Honolulu, HI; 6) School of Medicine & Health Sciences, George Washington University, Washington, DC; 7) Information Sciences Institute, University of Southern California, Los Angeles, CA; 8) Icahn School of Medicine at Mount Sinai, New York, NY; 9) Vanderbilt University Medical Center, Nashville, TN; 10) Department of Genetics, Rutgers University, New Brunswick, NJ; 11) Cancer Prevention Institute of California, Fremont, CA.

Lipid levels are heritable polygenic traits associated with cardio-metabolic outcomes. Extensive genome-wide association studies (GWAS) have been conducted in European-ancestry populations, identifying >150 loci. However, few studies included African Americans (AAs) or Hispanics, who are characterized by distinct genetic backgrounds and lipid profiles. In this study, we performed a GWAS for high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglycerides (TG) and total cholesterol (TC) in a multi-ethnic population from the PAGE Study, consisting of 10,137 AAs, 17,856 Hispanics, 2,387 Asians, 1,915 Hawaiians, 604 Native Americans and 360 other individuals. Participants were genotyped using the Multi-Ethnic Genotyping Array, followed by imputation using all 1000 Genomes Phase 3 data. Lipid levels were adjusted for medication by adding a constant, and TG values were log-transformed. All participants were pooled for association testing, adjusted for age, sex, study, self-identified ethnicity, center, household membership and the first 10 principal components (PCs). Stratified analyses within each ethnic group, conditional analyses adjusting for previously defined loci and interaction analyses between SNPs and PCs were also performed. We identified two novel loci for HDL (*8q22*-rs11782435 and *14q22*-rs75405126, *P*=2.94E-9 and 2.07E-8, respectively), one novel loci for TG (*SLC4A5*-rs186278890, *P*=3.34E-10) and one novel loci for TC (*CREB3L2*-rs73729087, *P*=1.52E-8), with consistent directions and comparable effect sizes across ethnic groups (*P*$_{heterogeneity}$>0.05). Multiple novel independent signals were identified in reported loci (*CETP, LCAT, LIPG, SCARB1, TOM1, LDLR, HRP, LPL, APOE* and *SAMM50*), and substantial refinement was observed at *APOA5* for TG, reducing the number of SNPs in high LD (r^2>0.6) with the top hit in AAs by 100%. Of the 40 reported loci replicated in our study (*P*≤5E-8), 18 showed significant heterogeneity (*P*<0.05). In addition, four AA-specific novel loci (*MTA1*-rs60255124, *CDH3*-rs564672295 and *STEAP1*-rs373140531 for HDL and *MTHFD2*-rs186278890 for TG, *P*≤3.52E-8) and two Hispanic-specific novel loci (*7q22*-rs552830302 for HDL and *RNF44*-rs188713108 for TC, *P*≤4.30E-8) were identified, showing disparate allele frequencies between ancestral populations. These results highlight the importance of conducting GWAS in diverse populations, in order to provide a more complete portrait of the genetics basis underlying lipid traits.

1869F

Genetic analysis of lipids in >300,000 participants in the Million Veteran Program. *D. Klarin[1], S. Damrauer[2], K. Cho[3], S. Duvall[4], G. Peloso[5], K.M. Chang[2], J. Huang[3], J. Lynch[4], Y.L. Ho[3], D. Liu[6], D. Saleheen[2], S. Pyarajan[3], C. Willer[7], Y. Sun[8], G. Abecasis[9], J. Concato[10], J.M. Gaziano[3], C.J. O'Donnell[3], D.J. Rader[2], S. Kathiresan[1], P.S. Tsao[11], P.W.F. Wilson[8], T.L. Assimes[11] on behalf of the VA Million Veteran Program.* 1) Massachusetts General Hospital, Boston, MA; 2) Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) VA Boston Healthcare System, Boston, MA; 4) University of Utah School of Medicine, Salt Lake City, UT; 5) Boston University School of Public Health, Boston, MA; 6) Penn State University, Hershey, Pennsylvania; 7) University of Michigan Medical School, Ann Arbor, MI; 8) Emory University School of Medicine, Atlanta, GA; 9) University of Michigan School of Public Health, Ann Arbor, MI; 10) VA Connecticut Healthcare, West-Haven, CT; 11) Stanford University School of Medicine, Stanford, CA.

Background: Large-scale biobanks that integrate biosamples with clinical phenotypes from an electronic health record (EHR) may catalyze better understanding of disease etiology, prevention, and treatment. The Million Veteran Program (MVP) is a national, voluntary research program to study how genes affect health in US veterans. Circulating plasma concentrations of high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglycerides, and total cholesterol are known risk factors for cardiovascular disease, and represent the main therapeutic targets for disease prevention and treatment. **Methods:** For 353,323 participants genotyped using a custom MVP array, we extracted clinical phenotypes from the EHR spanning an average of 10 years per participant. We tested the association for up to 32 million genotyped and imputed DNA sequence variants for association with plasma HDL, LDL, triglycerides, and total cholesterol, and combined association results with published literature via meta-analysis. **Results:** Of 353,323 genotyped participants, 302,271 had lipid data available for analysis. 72% were of European ancestry, 20% were African American, and 8.3% were Hispanic. Genome-wide association testing in MVP and meta-analysis with published literature (combined N>600,000) yielded 129 novel associations at genome-wide level of significance. For example, carriers of a rare missense mutation in Perilipin-1 (*PLIN1* p. L90P, minor allele frequency = 0.08% in MVP) demonstrated a 2.65 mg/dL (0.233 standard deviation) higher plasma HDL after testing in 596,475 individuals. Perilipin-1 is required for lipid droplet formation, triglyceride storage, and free fatty acid metabolism. Frameshift loss of function mutations in Perilipin-1 have been reported to result in severe lipodystrophy. **Conclusions:** In a large-scale biobank leveraging the electronic health records of US veterans, we have identified 129 new genetic signals for blood lipid levels.

1870W

Genome-wide association study of anthropometric, cardiovascular, and lipid biomarkers in an ethnically diverse cohort of sub-Saharan Africans.

M. Hansen[1], A. Ranciaro[1], S. Fan[1], Y. Lo[1], S. Mpoloka[3], G. Mokone[4], M. Bolaane[5], T. Nyambo[6], S. Omar[7], D. Meskel[8], G. Belay[8], S. Thompson[9], J. Hirbo[10], W. Beggs[1], M. Yeager[11,12], S. Chanock[12], S. Tishkoff[1,2]. 1) Genetics, University of Pennsylvania, Philadelphia, PA; 2) Biology, University of Pennsylvania, Philadelphia, PA; 3) Department of Biological Sciences, University of Botswana, Gaborone, Botswana; 4) Department of Biomedical Sciences, University of Botswana School of Medicine, Gaborone, Botswana; 5) History Department, University of Botswana, Gaborone, Botswana; 6) Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; 7) Kenya Medical Research Institute, Center for Biotechnology Research and Development, 54840-00200 Nairobi, Kenya; 8) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 9) Global Alliance for Chronic Diseases, UCL Institute for Global Health, London, United Kingdom; 10) Genetics Institute, School of Medicine, Vanderbilt University, Nashville, TN; 11) Cancer Genomics Research Laboratory, NCI-DCEG,SAIC-Frederick Inc, Frederick, MD; 12) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD.

   Sub-Saharan Africa is home to a diversity of cultures and ethnic groups whose complex history and regional adaptations have led to high levels of genetic and phenotypic variation across the continent. Sub-Saharan populations typically have low linkage disequilibrium as they did not undergo an out-of-Africa population bottleneck, allowing for fine-mapping of trait-associated loci. We present genome-wide association results across a cohort of up to ~2,000 individuals from ethnically diverse rural Central, Eastern, and Southern African populations encompassing a variety of lifestyles (hunter-gatherers, pastoralists, and agriculturalists). Genotypes are derived from array data augmented by imputation based on whole genome sequence data. The set of quantitative traits investigated include anthropometric measurements (including height, BMI, percent body fat), cardiovascular traits (including pulse, blood pressure), and circulating lipid biomarkers (including cholesterol, HDL, LDL). For many traits we find large variation among populations and ethnicities, including height, blood pressure, LDL, and BMI. Estimates of heritability, ethnic and regional variation in trait associations, and comparison to non-African associations are discussed. The association results are a unique resource for understanding the genetic architecture of quantitative traits relevant to human health in a diverse set of African populations under-represented in human genetic studies.   NIH grant 1R01DK104339-01 and 1R01GM113657-01 to SAT. This work was supported in part by a pilot-project funded by the Center of Excellence in Environmental Toxicology at the University of Pennsylvania (NIH P30-ES013508) and by NIEHS Training Grant (T32-ES019851) to MH.

1871T

_TM6SF2_ rs58542926 impacts lipid processing in liver and small intestine.

N.A. Zaghloul[1,2], E.A. O'Hare[2], R. Yang[2], L. Yerges-Armstrong[1,2], U. Sreenivasan[2], R. McFarland[2], C.C. Leitch[2], M. Wilson[3], S. Narina[2], A. Gorden[2], K. Ryan[1,2], A.R. Shuldiner[1,2], S.A. Farber[3], G.C. Wood[4], C.D. Still[4], G.S. Gerhard[4], J.D. Robishaw[4], C. Sztalryd[2,5]. 1) Program in Personalized & Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School Medicine, Baltimore, MD; 3) Carnegie Institution for Science, Department of Embryology, Baltimore, MD; 4) Geisinger Clinic, Geisinger Obesity Research Institute, Danville, PA; 5) Baltimore VA Medical Center, VA Research Service, Geriatric Research, Education and Clincial Center, VA Maryland Health System, Baltimore, MD.

   The transmembrane 6 superfamily member 2 (_TM6SF2_) loss-of-function variant, rs58542926, is a genetic risk factor for nonalcoholic fatty liver disease and progression to fibrosis, but is also associated with lower levels of hepatic-derived triglyceride-rich lipoproteins. _TM6SF2_ is expressed predominately in liver and small intestine, sites for triglyceride rich lipoprotein biogenesis and export. In light of this, we explored whether TM6SF2 may exhibit analogous effects on both liver and intestine lipid homeostasis. To test this, we genotyped rs58542926 in 983 bariatric surgery patients from the Geisinger Medical Center for Nutrition and Weight Management, Geisinger Health System (GHS) in PA and from 3,556 study participants enrolled in the Amish Complex Disease Research Program (ACDRP). Although these two cohorts have different metabolic profiles, carriers in both cohorts had improved fasting lipid profiles. Importantly, following a high fat challenge, carriers in the ACDRP cohort exhibited significantly lower postprandial serum triglycerides suggestive of a role for TM6SF2 in the small intestine. To gain further insight into this putative role, effects of TM6SF2 deficiency were studied in a zebrafish model and in cultured human Caco-2 enterocytes. In both systems TM6SF2-deficiency resulted in defects in small intestine metabolism in response to dietary lipids including significantly increased lipid accumulation, decreased lipid clearance and increased endoplasmic reticulum (ER) stress. Consistent with this, we also observed increased hepatic ER stress in human carriers of the variant. These data strongly support a role of TM6SF2 in regulation of systemic lipids potentially through a role for TM6SF2 in intracellular lipid metabolism in liver and intestine. .

1872F

Gender differences in genetics of body composition and obesity traits after an intensive exercise intervention. *A.I. Vazquez[1], F. Dong[2], H. Kim[1], Y.L. Bernal-Rubio[1], J.R. Fernandez[3], M.S. Bray[2].* 1) Epidemiology and Biostatistics, Michigan State University, East Lansing, MI; 2) Department of Nutritional Sciences, University of Texas, Austin, TX, USA; 3) Department of Nutrition Sciences, University of Alabama at Birmingham, Birmingham, AL, USA.

The genetic variation on body composition measures has been shown to differ across gender and race/ethnicity; however, little is known about how genes differentially influence body composition changes, particularly in the context of potential hormonal differences associated with gender. In the Training Interventions and Genetics of Exercise Response (TIGER) study genetic variation was examined in sedentary adults (18-35 y) who participated in a controlled aerobic exercise 15-week protocol. 1012 non-Hispanic white, 700 African American, and 332 Hispanic participants were genotyped using the MetaboChip (Illumina, Inc.). Measures of body weight, BMI, waist/hip ratio (WHR), lean mass (Lean-M), fat mass (Fat-M), bone mass (Bone-M) and percent fat (%Fat) were obtained pre- and post-intervention. We performed genome wide association studies (GWAS) for baseline traits, and change between pre and post-intervention. GWAS were done in within races and in the PC-adjusted multiethnic group. Genomic heritability (GH; inter-individual variation explained by SNPs) for baseline and change measures were estimated through Bayesian analysis. Weight, BMI, waist, hip, Fat-M and %Fat were reduced with the intervention, while there was an increase in Lean-M and Bone-M. Body changes after an exercise intervention have a large genetic component, with GH ranging between 0.2 and 0.4. GH of change measures was variable across genders. Females (F) had the highest GH for change in WHR (mean [95% Confidence Region]: 0.56 [0.33;0.79] for F vs 0.28 [0.07;0.52] for males [M]), an indicator of fat topography. Males had the highest GH for change in weight (0.37 [0.12;0.64] for M vs 0.17 [0.05; 0.3] for F), BMI (0.44 [0.18;0.72] for M vs 0.25 [0.08;0.44] for F), fat-M (0.45 [0.16;0.75] for M vs 0.19 [0.06;0.34] for F), lean-M (0.29 [0.08;0.53] for M vs 0.19 [0.06;0.34] for F), and Bone-M (0.38 [0.11;0.67] for M vs 0.21 [0.06;0.38] for F). Among the top genes and pathways represented by the significant genes in gender-specific analyses were the FTO pathway in females, and the Nuclear Receptors in Lipid Metabolism pathway in males. Our results establish the existence of gender-specific differences associated with change in weight and body composition and demonstrate that changes in body size/mass traits between gender differ in inter-individual variation and pathways, supporting the importance of examining genetic effects specific to genders.

1873W

Evaluation of loss-of-function mutation in *PCSK9* gene in large nationwide health registry based PheWas study in Finland. *M. Alanne-Kinnunen[1], H. Mattsson[1,2], E. Kilpeläinen[1], P. Häppölä[1], P. Palta[1], A. Havulinna[1,2], M. Männikkö[5], V. Salomaa[2], P. Jousilahti[2], S. Koskinen[2], J. Eriksson[6], O. Raitakari[7], T. Lehtimäki[8], H. Laivuori[1], M. Perola[2], S. Ripatti[1], M. Daly[3,4], A. Palotie[1,3,4].* 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki. P.O.Box 20 FI-00014 University of Helsinki, Finland; 2) National Institute for Health and Welfare. P.O.Box 30 FI-00271 Helsinki, Finland; 3) Broad Institute. 75 Ames Street, Cambridge MA 02142, USA; 4) Department of Medicine, Massachusetts General Hospital. 185 Cambridge St, CPZN-6818, Boston, MA 02114, USA; 5) Faculty of Medicine, University of Oulu. P.O. Box 5000 FI-90014 University of Oulu, Finland; 6) Faculty of Medicine, University of Helsinki. PO Box 20 FI-00014 University of Helsinki, Finland; 7) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku. Kiinamyllynkatu 10 FI-20520 Turku, Finland; 8) Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Life Sciences, University of Tampere. P.O.Box 100 FI-33014 Tampere, Finland.

Proprotein convertase subtilisin/kexin type 9 (*PCSK9*) is a secreted protease that has profound effects on plasma levels of LDL cholesterol (LDL-C) by mediating degradation of LDL receptor. Loss-of-function (LoF) mutations have been shown to associate with reduced risk for cardiovascular events. PCSK9 inhibitors have been developed and have been shown to effectively and safely induce more favorable lipid profile and to reduce the rate of cardiovascular events. To study weather carriers of the LoF variant had any adverse or beneficial health outcomes we analyzed life-long national health registry data in 44500 Finnish individuals. We examined the health data of 3340 carriers of the *PCSK9* R46L (rs11591147) LoF variant (3270 heterozygotes and 70 homozygotes) and compared them to 41160 non-carriers. Our large harmonized phenome-wide association study data set comprises of 97 different quantitative traits, 440 disease endpoints and 174 drug categories derived from nationwide registries that include follow-up data from hospitalizations since 1969, cancers (1952), specialty clinic visits (1998), drug purchases (1995), and drug reimbursements (since 1964). By utilizing drug purchase data, we were able to analyze the usage of drugs from several categories (eg. cardiovascular system, metabolism, and systemic hormonal preparations). As expected, the variant carriers have lower levels of LDL-C (beta=-0.4 mmol/l, $p=9\times10^{-167}$), total cholesterol (beta=-0.38 mmol/l, $p=4\times10^{-104}$), apolipoprotein B (ApoB) (beta=-0.08 g/l, $p=1\times10^{-74}$), ApoB to ApoA1 ratio (beta=-0.07, $p=6\times10^{-65}$) and higher levels of HDL cholesterol (beta=0.02 mmol/l, $p=6\times10^{-6}$). Underscoring the validity of the Finnish registry data, variant rs11591147 carriers also had reduced risk of hyperlipidemia (OR=0.56 (0.44-0.7), $p=2\times10^{-9}$), myocardial infarction (OR=0.68 (0.55-0.82), $p=6\times10^{-5}$) and they were significantly less likely to have ever purchased lipid modifying agents (ATC C10A) (OR=0.57 (0.52-0.63), $p<2\times10^{-17}$). Most importantly, across a diverse set of disease endpoints covering neurodegenerative, gastrointestinal, cardiovascular, respiratory, cancer, allergy and immune mediated diseases, we did not observe any significantly increased risk for any diseases or increased use of prescribed medicines. Our study demonstrates the benefit of large population cohorts, biobank samples and decades of centrally aggregated health registry data in evaluating potential comorbidities or adverse effects of drug targets.

1874T

Genetic effects of familial hypercholesterolemia variants on LDL cholesterol levels among multi-ethnic veterans: The Million Veteran Program Study. *Y.V. Sun[1,2], S.M. Damrauer[3], Q. Hui[1], T.L. Assimes[4], Y. Ho[5], P. Natarajan[6], D. Klarin[7], J. Huang[5], J. Lynch[8,9], J.P. Honerlaw[5], K. Cho[5,10], D.J. Rader[11], C.J. O'Donnell[5,12], P.S. Tsao[4], P.W. Wilson[13] on behalf of the VA Million Veteran Program.* 1) Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA; 2) Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA; 3) Corporal Michael Crescenz VA Medical Center, Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) VA Palo Alto Health Care System, Department of Medicine, Stanford University School of Medicine, Stanford, CA; 5) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA; 6) Center for Genomic Medicine and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA. Department of Medicine, Harvard Medical School, Program in Medical and Population Genetics, Broad Institute of Harvard & MIT, Cambridge, MA; 7) Massachusetts General Hospital, Boston, MA, Broad Institute of Harvard & MIT, Cambridge, MA; 8) University of Massachusetts College of Nursing & Health Sciences, Boston, MA; 9) Department of Veterans Affairs Salt Lake City Health Care System, Salt Lake City, UT; 10) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 11) Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA; 12) Harvard Medical School, Boston, MA; 13) Atlanta VA Medical Center and Emory Clinical Cardiovascular Research Institute, Atlanta, GA.

    Background: Familial hypercholesterolemia (FH), characterized by inherited high levels of low-density lipoprotein cholesterol (LDL-C) and consequently premature coronary heart disease, is caused by many rare variants in a few genes, including *LDLR, APOB* and *PCSK9*. Hundreds of FH variants have been documented, but the impact of these variants at the population level is poorly studied due to their low frequencies. **Design and Methods**: Using genome-wide genotypes from a customized Affymetrix Axiom® array from 331,107 ethnically diverse participants of the Million Veteran Program (MVP), we identified FH variants based on clinical significance categories ("pathogenic" or "likely pathogenic") from ClinVar. We then determined the collective and individual impact of each of these variants on untransformed maximum available laboratory measures of LDL-C over up to 15 years (maxLDL). **Results**: Of the 62 "pathogenic" or "likely pathogenic" SNPs archived in ClinVar and genotyped in the MVP, we observed 58 variants from *LDLR* and *APOB* at least once. 1,616 individuals in total had at least one of these variants, resulting in an estimated carrier frequency of 1: 205 among the MVP participants, and 1:84 (1.2%) in those with maxLDL > 190 mg/dL. On average, FH variants in *APOB*, *LDLR* or both genes associate with greater maxLDL by 23.6 mg/dL ($p=1.1\times10^{-54}$), 18.3 mg/dL ($p=2.4\times10^{-52}$), and 20.4 mg/dL ($p=1.5\times10^{-103}$), respectively. We then examined the genetic effects of 12 rare FH variants, each with more than 30 observed carriers (carrier frequency between 0.01% and 0.56%), and confirmed that carriers of seven variants had higher mean levels of maxLDL (6-80 mg/dL) than non-carriers. Phenotypic effects for European Americans and African Americans (AAs) were largely similar, but several variants, including rs201573863 and rs151207122, were predominantly observed in AAs. **Conclusion**: Putatively pathogenic FH variants have a wide spectrum of phenotypic effects on LDL-C, and have various frequencies in the population - some are race-specific. More robust evidence of genotype-phenotype associations of FH variants is needed to accurately infer at-risk individuals from genotype-based screening to aid in decisions regarding the intensity of LDL-lowering.

1875F

Lipid loading in human liver cells induces differential expression of 88 genes. *M. Alvarez[1], K. Garske[1], J. Benhammou[1], Y. Bhagat[1], P. Pajukanta[1,2,3].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, USA; 3) Molecular Biology Institute at UCLA, Los Angeles, USA.

    Fatty acids are building blocks of triglycerides (TG), associated with obesity and metabolic disorders, such as type 2 diabetes and non-alcoholic fatty liver disease. The liver plays a key role in metabolizing free fatty acids (FFAs), and this process is highly regulated to ensure an appropriate energy balance. In insulin resistance and metabolic diseases, FFAs are continually released from adipose tissue into the blood, inducing the liver to ectopically accumulate lipids by taking in fatty acids. It is therefore important to understand the human hepatocyte transcriptomic response to lipid uptake. To this end, we explored the effect of uptake of the saturated fatty acid, palmitic acid (PA), on global mRNA transcription in the human liver cell line, HepG2. We incubated HepG2 cells with PA conjugated to BSA, in serum-depleted media for 24 hours. We also incubated HepG2 cells with BSA as a vehicle-only control in serum-depleted media. We prepared the RNA-seq libraries using the Illumina TruSeq Stranded mRNA library prep kit. Libraries were sequenced at a depth of 25-30 M paired-end reads on an Illumina HiSeq 4000 platform. Each group was conducted and sequenced in triplicates. We mapped the reads using the STAR 2-pass protocol, and counted against the Gencode version 26 annotation using HTSeq. Differentially expressed (DE) genes were identified using the edgeR pipeline, employing FDR<0.05. Of the 11,408 expressed genes, 9 were significantly down-regulated while 79 genes were up-regulated in response to lipid loading. After running DAVID pathway analysis, the most enriched pathways passing the multiple testing correction were glycolysis/gluconeogenesis, steroid biosynthesis, and pentose phosphate pathways. Sixteen genes encoding mitochondrial respiratory chain enzymes were also up-regulated. Noteworthy, we identified several DE genes with no previous implication in fatty acid metabolism, including 3 metallothionein genes and 3 ribosomal genes. Next, we investigated how much the +/- 500 kb local *cis* regions of these genes explain of heritability of TG levels using LD Score. We found that these *cis* regions explained 3.3% of the TG heritability, a significant 4.3-fold enrichment (p=0.019). Overall, the 88 DE genes induced by lipid loading help elucidate the cellular responses to environmental dietary cues in the human liver, and provided 88 candidate genes of which 39 are novel for exploring gene-environment interactions in metabolic diseases.

1876W

Human liver transcriptomes reveal potential new cholesterol genes under tight co-regulation with statin-targeted cholesterol synthesis pathway genes. *A. Ko[1,2], D. Kaminska[1,3], E. Nikkola[1], J. Pihlajamäki[3,4], P. Pajukanta[1,2,5].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Molecular Biology Institute at UCLA, Los Angeles, CA; 3) Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland; 4) Clinical Nutrition and Obesity Center, Kuopio University Hospital, Kuopio, Finland; 5) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA.

Obesity may perturb cholesterol synthesis and other key liver functions by inducing ectopic fat deposition to the liver, which can cascade to non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH), cirrhosis, and cancer. To identify genes affected in liver due to obesity, we performed RNA-sequencing (RNA-seq) on liver biopsies and cataloged metabolic and histological traits from 259 obese bariatric surgery patients. Using 15,670 expressed genes, we carried out weighted co-expression network analysis to identify co-regulated genes critical to liver's main functions as well as differential expression (DE) analysis between healthy and NASH livers (n=69 and 43) to uncover disease-perturbed genes. Overall, 19 signed co-expression modules were identified of which 11 are significantly associated (FDR<0.05) with fatty liver, fibrosis, an aggregated NAFLD phenotype, and lipid traits. Notably, a 75-gene module that positively correlates with statin usage (FDR=2.0×10^{-15}) displays enrichment for cholesterol biosynthesis pathway (FDR=9.68×10^{-14}) and protein-protein interactions (P<0.00001), implying statin's strong effect on cholesterol synthesis pathway at both mRNA and protein levels. Importantly, the module is highly preserved in an independent liver RNA-seq cohort from GTEx (n=96, Z-score>10), suggesting their consistent liver mRNA co-regulation. The module comprises the statin target gene, *HMGCR,* and 19 other known cholesterol genes (e.g. *SREBF2, LDLR* and *PSCK9*) in high correlation with new candidates for cholesterol regulation (e.g. *DERL3, SLC29A2, C14ORF1, and DNHD1*). The flanking (±500kb) variants of the 67 module genes that are individually associated with statin usage (P<6.67×10^{-4}) show high enrichment for LDL-cholesterol (LDL-C) GWAS signal (P=0.0012) and account for 5% of LDL-C heritability using LD Score, indicating their genetic effect on LDL-C. Based on the DE analysis, 13 statin-module genes are significantly up-regulated in NASH subjects, including key members from cholesterol biosynthesis and fatty acid pathways (e.g. *SQLE*, *MVK*, *FADS1, and ACLY*), suggesting possible disruption of their co-regulation with other statin-module genes by NASH. As statins can induce adverse side effects, such as liver damage or type 2 diabetes, the novel statin-associated genes we identified by integrating phenotype, liver histology, and transcriptomic data provide potential targets for new therapeutic agents of hypercholesterolemia.

1877T

Genetic regulation of adipose tissue transcript expression is involved in modulating serum triglyceride and HDL-cholesterol. *S.K. Das[1], S.P. Sajuthi[2], N.K. Sharma[1], M.E. Comeau[2], J.W. Chou[2], D.W. Bowden[3], B.I. Freedman[4], J.S. Parks[5], C.D. Langefeld[2].* 1) Internal Medicine-Endocrinology, Wake Forest School of Medicine, Winston-Salem, NC; 2) Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 3) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Internal Medicine, Section on Molecular Medicine, Wake Forest School of Medicine, Winston-Salem, NC.

Dyslipidemia is a major contributor to the increased cardiovascular disease and mortality associated with obesity and type 2 diabetes. We hypothesized that variation in expression of adipose tissue transcripts is associated with serum lipid concentrations in African Americans (AAs), and common genetic variants regulate expression levels of these transcripts. Fasting serum lipid levels, genome-wide transcript expression profiles of subcutaneous adipose tissue, and genome-wide SNP genotypes were analyzed in a cohort of non-diabetic AAs (N=250). Serum triglyceride (TRIG) and high density lipoprotein-cholesterol (HDL-C) levels were associated with expression level of 1458 and 2089 adipose tissue transcripts, respectively, but poorly associated with total cholesterol or LDL-C levels. Serum HDL-C-associated transcripts were enriched for salient biological pathways, including branched-chain amino acid degradation, and oxidative phosphorylation. Genes in immuno-inflammatory pathways were activated among individuals with higher serum TRIG levels. We identified significant cis-regulatory SNPs (cis-eSNPs) for 449 serum lipid-associated transcripts in adipose tissue. The cis-eSNPs of 12 genes were associated with serum lipid level in genome wide association studies in Global Lipids Genetics Consortium (GLGC) cohorts. Allelic effect direction of cis-eSNPs on expression of *MARCH2*, *BEST1* and *TMEM258* matched with effect direction of these SNP alleles on serum TRIG or HDL-C levels in GLGC cohorts. These data suggests that expressions of serum lipid-associated transcripts in adipose tissue are dependent on common cis-eSNPs in African Americans. Thus, genetically-mediated transcriptional regulation in adipose tissue may play a role in reducing HDL-C and increasing TRIG in serum.

1878F

The X-factor of complex disease: Development of methods and software for analysis of the X chromosome in GWAS and RVAS reveals X-autosomal gene-gene interactions and X-linked associations underlying lipid levels and their sexual dimorphism. *A. Keinan[1,2,3,4], S. Carmi[5], F. Gao[1], L. Lo[1], L. Ma[6], R. Ma[1], Y. Qiu[1], A. Slavney[1], K. Ye[1].* 1) Department of Biological Statistics & Computational Biology, Cornell University, Ithaca, NY; 2) Cornell Center for Comparative and Population Genomics, Cornell University, Ithaca, NY; 3) Center for Vertebrate Genomics, Cornell University, Ithaca, NY; 4) Center for Enervating Neuroimmune Disease, Cornell University, Ithaca, NY; 5) Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; 6) Department of Animal and Avian Sciences, University of Maryland, College Park, MD.

   A unique role of ChrX in complex disease and quantitative traits is expected from its role in model organisms, human Mendelian disorders, and sexual dimorphism. However, its role remains poorly tested as GWAS largely ignored or incorrectly analyzed it. A similar picture emerges for sequence-based association studies (lit. review). For several years, we developed methods and software for X that span all GWAS stages (XWAS), including QC, genotype calling, stratification correction, and statistical tests that both account for and exploit X-inactivation, which is becoming commonly used. In a first large-scale XWAS, we analyzed 27 GWAS datasets of different autoimmune diseases and reported 22 X-linked replicated associations, most with sex-specific effect size and related functions. However, we were puzzled by association of a locus in *COSMC* with 2 IBDs. Recently, this result led to a functional mouse study, revealing *COSMC* loss leads to IBD-like microbiota disruption and results in gut inflammation, mostly in males, and to extensive mechanistic insight. This example stresses the importance of studying X's role in complex diseases as a stepping stone toward sex-specific diagnosis and treatment. I will present results from the following efforts, with only brief examples below: 1) Beyond sex-specific X-linked effects, we test sex-specific X-linked modulation of previously associated autosomal loci, using lipid levels as initial testbed, partly due to our experience with these in GWAS, pharmacogenomics, and nutrigenomics; 2) Large-scale XWAS meta-analysis of lipid levels based on >50 datasets; 3) XWAS of several psychiatric disorders since they are highly sexually dimorphic and X genes play key roles in neurological development; 4) We develop methods for sequence-based association studies, including X-tailored variant and genotype calling, as part of and in coordination with GATK, and optimizing statistics for testing a region or a set of X-linked rare variants; 5) We broadly apply these newly-developed methods, initially to lipid levels. 1) We discovered 4 such X-Auto interactions underlying HDL levels alone. 2) One association with both diastolic and systolic BP is within *GRIA3*, which product interacts with a β2-adrenergic receptor encoded by *ADRB2*, which has been implicated in BP response to inhibitors 5) In RVAS of UK10K, we optimized SKAT's weight function based on theoretical predictions to be steeper for X, which greatly improved power (but not for autosomes).

1879W

Common and rare genetic variants for asthma, hay fever and eczema. *W. Ek, T. Karlsson, M. Rask-Andersson, A. Johansson.* Department of Immunology, Genetic, Science for Life, Uppsala University, Uppsala, Uppsala, Sweden.

   Purpose The heritability for asthma has been estimated to 82 % and for hay fever and eczema to 60-80% [1]. Previous genome-wide association (GWA) studies have reported a number of common variants explaining part of this heritability. This study, aims to investigate the effect of common as well as rare genetic variants behind asthma, hay fever and eczema using the UK Biobank (UKBB) including more than 500,000 participants with ~70 million densely imputed genetic variants. **Methods** We performed GWA studies for asthma and for the combined diagnose hay fever/eczema, using 116,137 non-related Caucasian participants with self-reported diagnosis. We used logistic regression models, adjusted for a number of covariates and performed conditional analysis to explore secondary signals. We further estimated the SNP heritability for each disease using GCTA, and investigated the shared genetic contribution to the diseases. . **Summary of results** The SNP heritability was estimated to 35 % for asthma and 30 % for hay fever /eczema. Our study resulted in 30 loci for asthma and 29 loci for hay fever/eczema, passing a genome wide significance level ($P<5 \times 10^{-8}$). We confirmed many previously reported loci and report a number of novel loci for each disease including *RTEL1, TRIM26* and *RAD51L* for asthma. *RTEL1* is biologically interesting due to its involvement in lung function. For hay fever/eczema, examples of novel loci were *ZBTB38, EVI5, NR* and *LCE1E*. Interestingly, some previously reported loci seem to be more than one independent loci, For example, both the epidermal differentiation complex harboring the *FLG* gene and the MHC region, contains more than one independent loci associated to atopic diseases. Interestingly, many genes seem to overlap between the diseases, while others seem to be disease specific.   1 *Allergy Eur J Allergy Clin Immunol* 2016.

1880T

Multi-omic approaches to identifying clinical biomarkers of asthma exacerbations in African Americans. *H. Gui[1], S. Xiao[1], A.M. Levin[1], M. Yang[1], J.J. Yang[2], D. Rynkowski[1], E.G. Burchard[3], L.K. Williams[1].* 1) Henry Ford Health System, Detroit, US; 2) University of Michigan, Ann Arbor, US; 3) University of California San Francisco, San Francisco, USA.

Severe asthma exacerbations, which result in rescue medication use, emergency department visits, and hospitalizations, contribute to significant disease-related morbidity and account for a large proportion of asthma-related expenditures. Until now there are few measures that can be used clinically to predict impending events and thereby guide treatment. Multi-omic approaches many provide complementary information when attempting to identify exacerbation biomarkers. The Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-ethnicity (SAPPHIRE) is a longitudinal asthma study; study participants had detailed electronic medical records data on severe exacerbations (oral steroid use, ER visits, or hospitalizations). In total 875 participants had genome-wide SNP array data (518341 common variants), 199 had RNA-sequence expression data (15,022 protein-coding genes), and 185 had both. Cox proportional hazard regression models were used to test the influence of single nucleotide polymorphism (SNP) genotypes (additive genetic model) and gene expression levels (normalized read counts) on time-to-exacerbation. Models accounted for potential confounders. Top hits from both data sets were used to search for cis-eQTL within 2Mb of top (SNP) associations. The top SNP association, rs16944353, showed a hazard ratio (HR) of 1.55 (P=1.73E-07) but did not reach the pre-defined Bonferroni threshold (P<5.00E-08) in the genome-wide association study (GWAS). The top association from the RNASeq analysis was for *AHR*, which showed a HR=3.56 (P=1.39E-05, FDR < 0.25) for time-to-exacerbation, but this did not reach the pre-defined threshold of P<5.74E-06. There were no overlapping genes in the top 10 signals from both analyses. In addition, none of the top 10 GWAS SNP associations were eQTL loci for the top genes identified in the transcriptomic analysis. In conclusion, we discovered one GWAS SNP and one gene transcript that were of borderline significance in their relationship to asthma exacerbations. Multi-omic approaches may identify different, and potentially complementary, predictors of disease outcomes. Future analyses with additional samples are needed to validate these initial findings. .

1881F

Shared genetic etiology and ancestry variations between asthma and major complex diseases. *T. Mersha[1], Y. Gauta[1], S. Ghandikota[1], D. Lin[2].* 1) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Karlstad University, Karlstad, Sweden.

Although epidemiological and clinical evidence supports the observation that subjects with asthma often have other complex diseases such as autoimmune, infectious, and metabolic disorders, little is known about genetic architecture for such comorbidities. The objectives of this study were twofold: a) determine the shared genetic risk factors and biological pathways between asthma and other complex diseases in the context of GWAS data; b) analyze racial ancestry variations among the shared GWAS risk variants. W explore over 20,000 unique GWAS SNPs associated with 1,480 diseases or traits to extract shared risk variants among asthma and other disease ontologies (i.e., asthma, metabolic, autoimmune, cardiovascular, inflammation, infection, mental and cancer). We calculated Jaccard similarity index to determine relatedness and pathways to identify shared functional enrichments at the gene and pathway level. We found more shared genetic loci than expected by chance alone (significant hypergeometric p-value for asthma-metabolic (<3x10^{-15}), asthma-autoimmune (<2x10^{-29}), asthma-cardiovascular (<1.1x10^{-7}), asthma-inflammation (<1.9x10^{-22}), asthma-infection (<5x10^{-7}), asthma-mental (<6x10^{-10}) and asthma-cancer (<4x10^{-16}). The Jaccard index for asthma-complex disease etiology pairs ranged 7%-20%. Significantly over-represented pathways include the T Helper cell differentiation (fold enrichment = 13.7, p-value = 2.1 x 10^{-14}). Ancestry variation in allele frequency was found for multiple diseases particularly for infection, metabolic and autoimmune-related risk variants. Our results show that the degree of interconnection between asthma and specific complex disease genetic loci couldn't be explained by random chance, suggesting a shared molecular mechanisms linking asthma with the rest of complex diseases. An improved understanding of these pathways in shared genomic regions could provide valuable insights into the causal pleiotropic effects that may contribute to comorbidity between asthma and other diseases. Future research is warranted to explore further to identify underlying shared mechanism and novel therapeutic targets associated with the genetic pathways. .

1882W

Mapping human airway smooth muscle cell transcriptional and epigenetic responses to asthma-promoting cytokines reveals enrichments for asthma-associated SNPs. *E.E. Thompson[1], Q. Dang[2], B. Mitchell-Handley[1], K. Rajendran[2], S. Ram-Mohan[2], J. Solway[3], R. Krishnan[2], C. Ober[1]. 1) Department of Human Genetics, The University of Chicago, Chicago, IL; 2) Center for Vascular Biology Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; 3) Department of Medicine, The University of Chicago, Chicago, IL.*

Airway hyperresponsiveness (AHR) to inhaled stimulants is a key feature of asthma. AHR has a strong genetic component; however, the genetic architecture of AHR is largely unknown. We used cell models of AHR to discover functional single nucleotide polymorphisms (SNPs) that have direct effects on transcriptional and epigenetic responses to two asthma and AHR-promoting cytokines, IL-13 and/or IL-17, in cultured airway smooth muscle cells (ASMCs) from 71 individuals. We hypothesized that some of these SNPs will be associated with asthma and, therefore, enriched for small p-values among SNPs included in asthma GWAS. Expression and DNA methylation data were obtained using Illumina Human HT-12 v4 arrays and Infinium MethylationEPIC BeadChip, respectively; samples were genotyped with the Illumina OmniExpress BeadChip. We conducted expression quantitative trail locus (e) QTL and methylation (me)QTL mapping using Matrix eQTL in cytokine-treated samples. Genes regulated by eQTLs in IL-13-exposed ASMCs were enriched for T cell activation involved in immune responses (GO:0002286; p=0.0072), and genes regulated by eQTLs in IL-17-exposed ASMCs were enriched for microtubule organization (GO:0031122; p=0.0023). We then asked whether SNPs that were QTLs at a FDR 5% were more likely to have small p-values (p<0.05) in asthma GWAS. Indeed, eQTLs (SNPs ±250Kb of the transcription start site of each gene) in IL-13, IL-17 and IL-13+IL-17-exposed ASMCs were enriched for asthma-associated GWAS SNPs (p=1.48x10^{-5}, 7.91x10^{-8}, and 4.63x10^{-8}, respectively); meQTLs (SNPs ±5kb of a CpG site) in IL-17 and IL-13+IL-17-exposed ASMCs, but not in IL-13-exposed ASMCs, were enriched for asthma-associated GWAS SNPs (p=0.00042, 0.011 and 0.43, respectively). Overall, 249 asthma-associated GWAS SNPs were associated with gene expression (eQTLs: mean of 15 SNPs, 7 genes per exposure) or with methylation (meQTLs: mean of 88 SNPs, 76 CpGs per exposure) levels in ASMCs. These results indicate that IL-13 and IL-17-exposed cultured ASMCs serve as a model for elucidating the function of asthma-associated SNPs identified in GWAS, and for identifying additional candidates that did not meet stringent criteria for genome-wide significance in GWAS. Moreover, our data suggest that IL-13- and IL-17-mediated arms of asthma pathogenesis and AHR may be influenced by genetic variation in different sets of genes. This work was supported by R01 HL113395. .

1883T

Unraveling the genetic architecture of generalized vitiligo in a homogeneous, isolated Romanian village. *G. Andersen[1], S. Birlea[1,2], S. Santorico[1,3], R. Spritz[1,4]. 1) Human Medical Genetics and Genomics, University of Colorado, Anschutz Medical Campus, Aurora, CO; 2) Department of Dermatology, University of Colorado School of Medicine, Aurora, CO, USA; 3) Mathematical and Statistical Sciences, University of Colorado, Denver, CO, USA; 4) Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO, USA.*

Vitiligo is a genetically complex disease in which patches of depigmented skin result from autoimmune destruction of melanocytes. By genome-wide association studies (GWAS) of vitiligo in individuals of European ancestry (EUR), we identified 48 vitiligo susceptibility loci. While proteins encoded at these loci highlight a framework for melanocyte-directed autoimmunity, the overall picture of vitiligo pathobiology remains incomplete. Here, we report analyses of an isolated, inbred Romanian village with vitiligo prevalence ~8 times higher than in the general EUR population, suggesting an enrichment of genetic and/or environmental risk factors. Due to the small number of ancestral founder genomes (~23), villagers are expected to have reduced causal allele heterogeneity and perhaps fewer causal alleles overall, thus constituting a less genetically complex subset of the general EUR population. We have genotyped 50 cases and 120 unaffected villagers for 591,361 SNPs genomewide, and we have used these genotypes to compare genetic architectures under three alternative models of vitiligo pathogenesis: (1) a polygenic model with unusually high prevalence of common EUR risk variants detected by GWAS; (2) a mono- or oligogenic model in which a highly-penetrant variant rare in the general EUR population explains most vitiligo risk; or (3) both. Specialized association analyses of the 48 known susceptibility loci are being used to assess common, genetic risk variants under hypothesis 1, and genome-wide linkage analysis is being used to identify variants consistent with hypothesis 2. Our combined approach of association and linkage within this remarkable isolated village provides a unique opportunity to ascertain the genetic architecture of vitiligo in a relatively simple, homogeneous population. Implications of this study will be to highlight key pathobiological pathways, either by identifying a combination of common risk variants that confer particularly high risk, or by identifying a novel vitiligo susceptibility locus that confers unusually high disease risk.

1884F

Polygenic risk score predicts development of HCV-associated mixed cryoglobulinemia and response to interferon-free therapy. *M. Artemova[1], N. Mukhin[1,2].* 1) Fundamental Medicine, Moscow State University, Moscow, Russian Federation; 2) Tareev Clinic for Nephrology, Internal and Occupational Diseases, Moscow, Russian Federation.

Mixed cryoglobulinemia (MC) can be detected in 40-60% of patients chronically infected with hepatitis C virus (HCV). Overt MC vasculitis (MCV) appears in a minority (5-10%) of patients and includes a spectrum of symptoms from mild to life-threatening. Despite the success of new direct acting antivirals, several recent studies demonstrate persistent MC or MCV following virus eradication after interferon-free treatment (PMID: 27943361; 19571811). Recent GWAS confirmed importance of the host genetic background as a predisposing factor in MCV development (PMID: 25030430). Here, we present a method for quick evaluation of genetic predisposition to MC among HCV patients, using polygenic risk score (PRS). We focused on cases with only HCV and no other infections or autoimmune diseases to ensure cohort is genetically-enriched. Study included 54 affected samples diagnosed with HCV-associated MC (31 MCV and 23 asymptomatic MC) and 26 well-matched controls with long duration of HCV infection without notable production of cryoglobulins. Using qPCR, we screened the two independent SNPs most associated with MCV: rs2071286 and rs9461776. PRS was calculated as a sum of alternative alleles for these two SNPs weighted by their logistic regression coefficients. Individuals with asymptomatic MC and MCV had similar PRS (1.1 and 0.816, respectively); thus, we consolidated these cohorts for further analysis. Samples with MCV and asymptomatic MC on average had a 3.2-fold increase in risk (p=3.4×10^{-5}) compared with controls. Analysis of the clinical data demonstrated that immunological response to the interferon-free treatment depends on interaction of HCV duration and inherited risk (PRS), p=0.033 and on the rate of rheumatoid factor production (p=0.03). Patients with long duration of HCV, high PRS and rheumatoid factor kept cryoglobulins production for at least 11 weeks longer than controls.

1885W

Identification of several genes modifying multiple sclerosis risk conferred by tobacco smoke: A case-only analysis. *F.B.S. Briggs.* Neuroimmunological Disorders Gene-Environment Epidemiology Lab, Department of Population and Quantitative Health Sciences, Case Western Reserve University, School of Medicine, Cleveland, OH.

Background: Multiple sclerosis (MS) is an idiopathic neuro-inflammatory disease with prominent genetic and environmental risk components. It is likely specific genetic variants contribute to MS risk for a given environmental condition. Thus, exploring genetic heterogeneity in the context of environmental risks presents a unique opportunity to further characterize the heritable component of MS. Tobacco smoke (TS) is an established risk factor, and prior work demonstrated *NAT1* conferred MS risk in the presence of TS exposure. Thus, a gene-centric, multi-stage analysis was conducted to identify novel genic variants modifying MS risk due to TS. **Methods:** Genetic and survey data for 1,056 European Americans cases (onset age ≥18 yrs) from the Accelerated Cure Project for MS were available. DNA samples were genotyped on an Illumina HumanExome BeadChip with custom content (~270,000 autosomal SNPs). The study population was divided into a discovery (N=795) and replication (N=261) set. Detailed TS history was available and cases were classified as a smoker if they were an active smoker within 5 years prior onset. Power calculations were conducted for the discovery data assuming a 2-sided α and TS prevalence of 33% for a range allele frequencies and effect sizes. A total of 42,346 variants (MAF≥5%) in 11,500 genes were interrogated in the discovery data with smoking status as the outcome. Models were adjusted for gender, onset age, birth year, education, disease subtype, and ancestry. Discovery associations (2-sided p<0.05) for which there were sufficient power (>70%) were pursued in the replication data where a 1-sided p<0.05 determined significance. The discovery and replication data were combined to determine final associations. **Results:** 1,386 discovery associations were pursued in the smaller replication data. In the replication data, there were 63 SNPs in 40 genes with significant associations. In the combined data, the most significant associations were for missense SNPs in *DNAH9* (odds ratio [OR]=1.5, p=9×10^{-5}) and *ADRB2* (OR=1.5, p=1×10^{-4}). The strongest effects were observed for an *OR2L8* missense SNP (OR=0.43, p=8×10^{-4}) and a *TLR4* intronic SNP (OR=0.45, p=8×10^{-4}). TS alters *DNAH9* and *TLR4* expression in airway epithelium, and *DNAH9*xTS contributes to bronchial hyperresponsiveness and *TLR4*xTS contributes to the accrual of neutrophils in airways of smokers. Thus, the results suggest genetic variation in physiological response to TS contributes to MS risk. .

1886T

Admixture mapping of 13,569 individuals provides evidence for increased European origin of the major histocompatibility complex class i region in multiple sclerosis. *C. Chi[1,2], X. Shao[1], B. Rhead[1,2], E. Gonzales[3], J.B. Smith[3], C. Schaefer[4], E. Waubant[5], L.F. Barcellos[1], A.M. Langer-Gould[3,6], J. Graves[5], A. Waldman[7], T. Lotze[8], T. Schreiner[9], B. Weinstock-Guttman[11], G. Aaen[12], J.M. Tillema[13], J. Ness[14], M. Candee[15], L. Krupp[16], M. Gorman[17], L. Benson[17], T. Chitnis[18], S. Mar[19], A. Belman[20], C. Casper[15], J. Rose[15], M. Moodley[21], M. Rensel[21], M. Rodriguez[13], B. Greenberg[10].* 1) Genetic Epidemiology and Genomics Laboratory, UC Berkeley, Berkeley, CA; 2) Computational Biology Graduate Group, UC Berkeley, Berkeley, CA; 3) Kaiser Permanente Southern California, Department of Research & Evaluation, Los Angeles, CA; 4) Kaiser Permanente Division of Research, Oakland, CA; 5) UC San Francisco, Department of Neurology, San Francisco, CA; 6) Southern California Permanente Medical Group, Los Angeles Medical Center, Neurology Department, Los Angeles, CA; 7) Children's Hospital of Philadelphia, Philadelphia, PA; 8) Texas Children's Hospital, Houston, TX; 9) University of Colorado School of Medicine, Aurora, CO; 10) University of Texas Southwestern, Dallas, TX; 11) SUNY Buffalo, Buffalo, NY; 12) Loma Linda University, Loma Linda, CA; 13) Mayo Clinic, Rochester, MN; 14) University of Alabama at Birmingham, Birmingham, AL; 15) University of Utah, Salt Lake City, UT; 16) SUNY Stony Brook, Stony Brook, NY; 17) Boston Children's Hospital, Boston, MA; 18) Brigham and Women's Hospital, Boston, MA; 19) Washington University St. Louis, St. Louis, MO; 20) NYU Langone Medical Center, New York, NY; 21) Cleveland Clinic, Cleveland, OH.

Multiple sclerosis (MS [MIM 126200]) is a demyelinating autoimmune disease with genetic and environmental risk factors. To date, over 2.5 million individuals worldwide have been diagnosed with MS. Disease prevalence is highest in North America and northern Europe at ~100 cases per 100,000 individuals. MS is a multifactorial disease, and a majority of established genetic risk factors have been discovered in white populations. However, it is not known to what extent these genetic risk factors originated from European ancestral populations. We hypothesize that the genetic risk of MS is largely driven by European ancestry, contributing to the observed prevalence disparity among different populations. We performed admixture mapping and tested more than 200 established major histocompatibility complex (MHC) and non-MHC genetic risk factors for increased European ancestry in cases compared to controls. We collected genotype data from 1,589 admixed cases and 11,980 admixed controls representing the black, Hispanic, and Asian populations recruited through Kaiser Permanente Northern and Southern California, the U.S. Pediatric MS Network, Genetic Epidemiology Research on Aging cohort, UC San Francisco and the International MS Genetics Consortium. Using fastStructure, we estimated the average global admixture proportions of black (74% African, 26% European), Hispanic (6% African, 52% Native American, 42% European), and Asian (92% East Asian, 8% European) study subjects. We inferred local ancestry from genotype data for each subject using RFMix to map the admixture proportions across the genome. For each admixed population, we performed hypothesis testing with a non-parametric test statistic (Pritchard and Montana, 2004) that adjusts for global admixture differences between cases and controls. We found no evidence supporting increased European ancestry for the non-MHC genetic risk loci in MS cases. However, we found evidence for increased European ancestry in regions within the MHC in black MS cases. The difference in European ancestry within the Class I MHC region is 1.8% greater than the difference in genome average between cases and controls (p < 0.05). Our findings support the hypothesis that the non-MHC genetic risk factors are not exclusively European but that some MHC risk alleles may originate from European ancestry.

1887F

Immune-phenotypes among patients with systemic lupus erythematosus and their association with *HLA-DRB1* alleles. *L.M. Diaz-Gallo[1], E. Lundström[1], V. Oke[1], K. Elvin[2], YL. Wu[3], J. Gustafsson[1], A. Jönsen[4], D. Leonard[5], A. Zickert[1], G. Nordmark[5], A.A. Bengtsson[4], L. Rönnblom[5], E. Gunnarsson[1], C.Y. Yu[3], E. Svenungsson[1], L. Padyukov[1].* 1) Department Medicine Solna, Rheumatology Unit, Karolinska Institutet, Stockholm, Stockholm, Sweden; 2) Department of Clinical Immunology and Transfusion Medicine, Unit of Clinical Immunology, Karolinska Institutet/Karolinska University Hospital, Stockholm, Sweden; 3) The Research Institute at Nationwide Children's Hospital, Columbus, Ohio 43205, USA; 4) Department of Clinical Sciences, Section of Rheumatology, Lund University, Skåne University Hospital, Lund, Sweden; 5) Department of Medical Sciences, Section of Rheumatology, Uppsala University, Uppsala, Sweden.

Systemic lupus erythematosus (SLE) is a remarkably heterogeneous autoimmune disease, with a broad display of autoantibodies and there is a need to define SLE sub-phenotypes. Therefore, we set out to investigate whether there are phenotypic clusters in SLE, based on autoantibody profile, and to investigate their association with one of the most important genetic risk factors for SLE, *HLA-DRB1* alleles. We evaluated a total of 908 SLE patients of European Caucasian origin, and 3654 age- gender- and ethnicity-matched healthy controls (HC). Blood samples from the patients were used to determine 13 autoantibodies: dsDNA, nucleosomes, ribosomal P, RNP68, RNPA, Sm, Sm/RNP, SSA52, SSA60, SSB, aCL-IgG/IgM and ab$_2$GP1. *HLA*-DRB1 typing was performed by sequence-specific primer polymerase chain reaction assay. Cluster analysis was done using Gower distance matrix, followed by partition around medoids cluster calculation and Silhouette metric for number of clusters validation. Chi-square test, odds ratios and 95% confidence intervals were calculated for the association tests. Our results demonstrate that there are differential associations between the four defined clusters and specific *HLA-DRB1* alleles. A cluster where the anti-SSA52/60/SSB are dominating was specifically and strongly associated with *HLA-DRB1*03* when compared with HC (OR=4.1 95%CI=3.4-4.9, PFDR=6.4E-56) and other clusters (OR=2.9 95%CI=3.3-3.6, PFDR=1.1E-19). A second cluster, dominated by anti-SmRNP/Sm/DNA/RNPA/RNP68/nucleosome, was specifically associated with *HLA-DRB1*15* when compared with HC (OR=1.7 95%CI=1.6-2.1, PFDR=5.7E-6) and other clusters (OR=1.5 95%CI=1.1-1.9, PFDR=0.01), suggesting that the overall association of these alleles with SLE comes from these autoantibodies clusters. The third cluster, dominated by anti-B2GP1/aCL-IgG/IgM, was associated with *HLA-DRB1*04* when compared with other clusters (OR=1.8 95%CI=1.4-2.4, PFDR=2E-4). No significant association was detected between the fourth cluster, negative for the 13 tested autoantibodies, and *HLA-DRB1* alleles. Previous reports and empirical observations from the clinic support that SLE subgroups, based on autoantibody profile, are associated with specific clinical manifestations, damage accrual and long term prognosis. Our approach confirms these observations and frame them into the genetic background. These results could be used in the definition of predictive biomarkers and inclusion criteria for clinical trials.

1888W

GWAS polygenic model approach applied to primary biliary cholangitis (PBC) in a Japanese population. *O. Gervais[1], K. Ueno[1], K. Kojima[1], Y. Aiba[3], M. Kawashima[2], Y. Kawai[1], Y. Hitomi[2], K. Tokunaga[2], M. Nakamura[3,4], M. Nagasaki[1].* 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 2) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 3) Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center, Nagasaki, Japan; 4) Department of Hepatology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan.

Primary Biliary Cholangitis (PBC) is an autoimmune liver disease which affects predominantly women over the age of 40 and is typically associated with anti-mitochondrial antibodies (AMA). Although the etiology of PBC is still unknown, genetic predisposition is well established, as exemplified by studies which have shown familial clustering and higher concordance in monozygotic twins than in dizygotic twins. A number of GWAS analyses have been conducted to elucidate the role of genetic variation in the development of PBC, and several susceptibility genes associated with PBC have been discovered, notably in the highly polymorphic HLA region. At the same time, the majority of the GWAS studies on PBC published so far solely focused on individuals of European descent. The first GWAS for PBC in the Japanese population revealed potential ethnic differences, for instance with regard to the *IL12A* locus, which was not found to be significantly associated with PBC, unlike previous research conducted in a North American cohort. This finding suggests a need for additional research to clarify the regional genetic characteristics of PBC. In this study, polygenic model analysis was used to examine data collected from Japanese individuals (>1300 PBC cases and >1500 healthy controls) genotyped for approximately 600,000 SNPs using the Affymetrix Axiom ASI Array, and to estimate the genetic parameters (heritability, etc.) associated with PBC in this population.

1889T

Lupus-associated functional polymorphism in *PNP* causes cell cycle abnormalities and interferon pathway activation in human immune cells. *Y. Ghodke-Puranik[1], J.M. Dorschner[1], D.M. Vsetecka[1], S. Amin[2], A. Makol[2], F. Ernste[2], T. Osborn[2], K. Moder[2], V. Chowdhary[2], E. Eliopoulos[3], M. Zervou[4], G.N. Goulielmos[3], M.A. Jensen[1], T.B. Niewold[1].* 1) Dept of Immunology and Division of Rheumatology, Mayo Clinic, Rochester, MN; 2) Division of Rheumatology, Mayo Clinic, Rochester, MN; 3) Department of Agricultural Biotechnology, Agricultural University of Athens; 4) Laboratory of Molecular Medicine and Human Genetics, Department of Internal Medicine, Medical School of Crete, Heraklion, Greece.

Background: Systemic lupus erythematosus (SLE) is frequently characterized by type I interferon (IFN) pathway activation. We previously found a missense SNP in the *purine nucleoside phosphorylase* (*PNP*) gene was associated with high IFN in SLE (rs1049564). *PNP* is a key enzyme in purine metabolism. In this study, we perform functional follow-up of this polymorphism in human cells. **Methods:** Type I IFN was measured in patient sera using a reporter cell assay. The *PNP* variant was modeled structurally using Pymol software. *PNP* mRNA and protein levels, and type I IFN-induced gene expression were measured in lymphoblastoid cell lines with known *PNP* rs1049564 genotypes. Cell cycle was assayed using flow cytometry. **Results:** Structural modeling indicated no major disruption in folding related to rs1049564. Human lymphoblastoid cells homozygous for rs1049564 TT had decreased *PNP* mRNA and protein levels. Rs1049564 TT cells had reduced PNP enzymatic activity even when the amount of PNP was controlled. TT cells had a ~2-fold increase in S-phase block as compared to homozygous CC cells. The S-phase block could be pharmacologically reversed with hypoxanthine and adenosine, supporting relative PNP deficiency as the cause of the S-phase block. Type I IFN-induced transcripts were increased in a dose-response fashion related to the rs1049564 T allele both at baseline and after type I IFN stimulation. **Conclusions:** The rs1049564 *PNP* T allele is a loss-of-function variant, inducing S-phase block and IFN pathway activation in lymphocytes. The S-phase block can be rescued in our in vitro experiments, suggesting a potential for personalized therapeutics.

1890F

Association of protein tyrosine phosphatase non-receptor N22 gene functional variant R620W with systemic lupus erythematosus patients from Kuwait. *M.Z. Haider[1], A.M. Al-Awadhi[2,3], J. Sukumaran[1], S. Balakrishnan[1].* 1) Pediatrics Department, Faculty of Medicine, Kuwait University, Jabriya, Kuwait; 2) Department of Medicine, Faculty of Medicine, Kuwait University, Jabriya, Kuwait; 3) Al-Amiri Hospital, Kuwait.

Systemic lupus erythematosus (SLE) is a common, complex autoimmune disease with multiple-organ involvement, characterized by the production of pathogenic autoantibodies directed against cytoplasmic and nuclear cellular components. SLE is thought to results from multifactorial etiology, involving hormonal factors, environmental triggers and genetic susceptibility. The aim of this study was to investigate the role of protein tyrosine phosphatase non-receptor type 22 (*PTPN22*) gene functional variant R620W in genetic susceptibility of the SLE in Kuwaiti Arabs. The study included 134 SLE patients and 214 healthy controls from Kuwait. The genotypes of *PTPN22* gene functional variant R620W were determined by PCR-RFLP and confirmed by DNA sequencing and the genotype and allele frequencies were compared between SLE patients and controls. A relatively high frequency of the variant 620W (T-allele) of the *PTPN22* gene was detected in SLE patients from Kuwait. 35.7% of the SLE patients had at least one variant allele (T-allele) compared to 15.9% in the controls. A statistically significant difference was detected in the frequency of variant genotypes, TT and CT between SLE patients and the controls (p <0.0001). No association was detected between the *PTPN22* gene variant and the Raynaud's phenomenon, renal involvement and severity of the SLE. The frequency of *PTPN22* gene functional variant R620W reported in this study is amongst the highest compared to other world populations. A high prevalence of this variant in SLE patients in comparison to controls suggests its significant contribution in conferring susceptibility to SLE along with other factors.

1891W

Transcription factors are associated with disease risk loci: Epstein-Barr virus nuclear antigen 2 (EBNA2) is an environmental factor associated with multiple autoimmune diseases. *J.B. Harley[1,2,3,4], X. Chen[1], M. Pujato[1], D. Miller[1], A. Maddox[1], C. Forney[1], A.F. Magnusen[1], A. Lynch[1], K. Chetal[5], M. Yukawa[6,7], A. Barski[6,7], N. Salomonis[5], K.M. Kaufman[1,2,3,4,5], L. Kottyan[1,6], M.T. Weirauch[1,3,5].* 1) Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children;s Hospital Medical Center , Cincinnati, OH., USA; 2) Division of Immunobiology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 3) Division of Developmental Biology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 4) US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA; 5) Division of Biomedical Informatics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 6) Division of Allergy & Immunology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 7) Division of Human Genetics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

Mechanisms responsible for most complex disease genetic associations from GWAS projects appear to be regulatory. We reasoned that transcription factors (TFs) might operate across multiple loci in a given phenotype more frequently than expected by chance, which is now established by simulation for at least 92 disease and physiological human phenotypes involving 2,264 associations from 1,544 TF ChIP-seq data sets at a Bonferroni corrected $Pc<10E-8$. Among these, an environmental factor, Epstein-Barr virus (EBV) nuclear antigen 2 (EBNA2) ChIP-seq data identify multiple environment-risk gene locus interactions with autoimmune phenotypes, such as **Multiple Sclerosis** (44 of 109 loci evaluated, OR=4, Pc=6x10E-30), **Systemic Lupus Erythematosus** (26 of 53 loci, OR=6, Pc=1x10E-25), **Rheumatoid Arthritis** (25 of 84 loci, OR=4, Pc=2x10E-14), **Inflammatory Bowel Disease** (38 of 112, OR=3, Pc=1x10E-12), and **Type 1 Diabetes** (20 of 57, OR=4, Pc=6x10E-12), and others. These associations were also present when restricted to the open chromatin (DNase-seq) of EBV transformed B cells were considered. No association with EBNA2 was found with >150 other phenotypes. For each disease we observe a partially shared set of human TFs clustering with the EBNA2 associated loci that are also immunoprecipitated in EBV transformed B cells, but not generally in other cell types nor in B cell lines not infected with EBV. These relationships are genetic if and only if differential allelic mechanisms operate with respect to EBNA2. The data available to test this hypothesis are very sparse. However, the vitiligo risk locus at CD44 provides an example and a possible proof of principle. The EBNA2 ChIP-seq DNA binding at rs3794102 favors the risk allele, as evidenced by having >2-fold more risk than non-risk allele reads. In addition, CD44 gene expression favors the risk allele only when Ramos cells are EBV infected. This locus has properties of an "allelic switch" with more open chromatin (ATAC-seq) and activating TF ChIP-seq reads (EBF1, BATF, ATF2, & BCL3 in addition to EBNA2) on the risk allele, while suppressive TF reads (MTA3 & FTIC) dominate on the non-risk allele. Our approach nominates potentially important molecular relationships occurring in specific cell types that have the potential to reveal mechanisms of disease and, moreover, reveals a potentially important class of gene by environment interactions for a subset of autoimmune disorders that suggest molecular mechanisms involving EBNA2.

1892T

Identification of the primary functional variants in primary biliary cholangitis susceptibility gene loci *NFKB1/MANBA*. *Y. Hitomi[1], K. Nakatani[1], K. Kojima[2,3], N. Nishida[4], Y. Kawai[2,3], M. Kawashima[5], Y. Aiba[6], M. Nagasaki[2,3,7], M. Nakamura[6,8,9], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 3) Graduate School of Medicine, Tohoku University, Sendai, Japan; 4) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 5) Japan Science and Technology Agency (JST), Tokyo, Japan; 6) Clinical Research Center, National Hospital Organization, Nagasaki Medical Center, Omura, Japan; 7) Graduate School of Information Sciences, Tohoku University, Sendai, Japan; 8) Department of Hepatology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 9) Headquarters of PBC Research in NHOSLJ, Clinical Research Center, National Hospital Organization Nagasaki Medical Center.

Primary biliary cholangitis (PBC) is a chronic and cholestatic liver disease that is caused by the autoimmune destruction of intrahepatic small bile ducts, which eventually leads to liver cirrhosis and hepatic failure. We recently identified a susceptibility gene region for PBC in the Japanese population by our previous genome-wide association study (GWAS), including *nuclear factor kappa B subunit 1 (NFKB1)* and *mannosidase beta (MANBA)*. However, the primary functional variants in *NFKB1/MANBA* and their molecular mechanisms for disease susceptibility in PBC have not yet been clarified. In the present study, we attempted to identify the primary functional variants in *NFKB1/MANBA* by performing high-density association mapping based on an SNP imputation analysis by IMPUTE2 (ver. 2.3.1) using whole-genome sequence data from a reference panel of 1,070 Japanese individuals (1KJPN) and genotype data from our previous GWAS (PBC patients: n = 1,389; healthy controls: n = 1,508). From among the 83 SNPs in *NFKB1* and 52 SNPs in *MANBA* that showed a P value $< 5.0 \times 10^{-7}$ on high-density association mapping, candidate functional SNPs that potentially regulate gene expression efficiencies were selected by *in silico* analysis using the Regulome DB database and the UCSC Genome Browser. Two primary functional variants for disease susceptibility were identified by *in vitro* functional analysis (luciferase assay and electrophoretic mobility shift assay) using the human liver cell line HepG2 and the human T-cell line Jurkat. Additionally, e-QTL data from the GTEX portal database supported the potential role of these functional variants in regulating gene expression efficiencies ($P = 6.5 \times 10^{-5}$ in spleen; $P = 7.2 \times 10^{-8}$ in visceral adipose tissue). These results suggested that high-density association mapping based on SNP imputation analysis and on *in silico/in vitro* functional analysis can be used to identify primary functional variants in disease susceptibility genes. This study illustrated a systematic methodology for the identification of primary functional variants from a pool of suggestive SNP associations, and this methodology may eventually contribute to the development of novel diagnostic and therapeutic methods for PBC.

1893F

***TYK2* correlates with multiple sclerosis symptomatology at onset.** *J. Jiangyang[1], J. Yu[1], E. Utigard[2], D. Ontaneda[2], F.B.S. Briggs[1].* 1) Neuroimmunological Disorders Gene-Environment Epidemiology Lab, Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH; 2) The Mellen Center for Multiple Sclerosis, Cleveland Clinic, Cleveland, OH.

Background: Multiple sclerosis is a complex autoimmune diseases with inflammatory and neurodegenerative phases. The etiologic processes are not clear, but failures in endogenous repair mechanisms are thought to contribute to the manifestation of symptoms and disease progression. Tyrosine kinase 2 (*TYK2*) is an established MS risk locus, and functional analyses of the MS risk allele suggest a shift in T cell differentiation in favor of a Th2 phenotype. Interestingly, TYK2 is ubiquitously expressed, and likely to have functional roles in other systems, as evident by research demonstrating strong correlation between attenuation of *Tyk2* and enhanced CNS repair. **Hypothesis:** We hypothesize genetic variation in *TYK2* contributes to the presentation of symptoms at onset in persons with MS (PwMS), especially considering it is a known MS risk locus (therefore may also temporally affect early clinical manifestation). **Methods:** Genetic and survey data for 1,056 European Americans PwMS (onset age ≥18 yrs) from the Accelerated Cure Project for MS were available. DNA samples were genotyped on an Illumina HumanExome BeadChip with custom content (~270,000 autosomal SNPs) including 16 *TYK2* SNPs (MAF>0.01) residing within 2 haplotype blocks (solid spline of LD; D'>0.8). PwMS reported detailed symptoms experienced at onset (31 questions), which were classified into 11 impaired functional domain by two neurologists. We focused on the most prevalent impaired domains: sensory (50%), motor (46%), cerebellar (32%), brainstem/bulbar (31%), and optic nerve (28%); these domains minimally correlated ($r^2 < 0.25$), except for motor and cerebellar ($r^2 = 0.7$). Models were adjusted for gender, birth year, onset age, education, disease subtype, and genetic ancestry. A significance threshold of p<0.005 was imposed (2 blocks by 5 domains). **Results:** Based on the significance threshold, no SNP was associated with sensory or brainstem/bulbar impairment at onset. 2 missense SNPs were associated with increased optic nerve impairment (OR~1.4, p<0.003, MAF>0.25). 3 intronic SNPs were associated with increased cerebellar impairment (OR~2.5, $p=5 \times 10^{-4}$ to $p=6 \times 10^{-5}$, MAF>0.06), and marginally associated with motor impairment (OR~2.4, p<0.01). **Conclusions:** Uncovering genetic variants contributing to MS presentation and progression has proven challenging. Using a hypothesis driven approach we have identified *TYK2* as a strong candidate for future studies of cerebellar dysfunction (i.e. loss of coordination) in MS.

1894W

Genetic screening of Galectin-3 CRD variants in RA: A case-control association study. *T. Kaur[1], J. Singh[2], M. Kaur[1].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Department of Molecular Biology & Biochemistry, Guru Nanak Dev University, Amritsar, India.

  Statement of Purpose: Rheumatoid arthritis (RA) is a chronic, inflammatory, systemic disease condition characterized by destruction of peripheral joints, subsequently leading to deformity and disability. Various studies have demonstrated the pro-inflammatory role of galectin-3 in pathogenesis of RA. Human galectin-3 protein is encoded by a single gene *LGALS3* located on chromosome 14q22.3. Genetic variations in *LGALS3* can have profound effects on structure as well as function of galectin-3. So, the present study was designed for genetic screening of CRD variants of *LGALS3* in RA. **Methodology:** The present case-control study recruited 200 RA patients and a cohort of 200 age- and gender- matched controls. The study protocol was approved by Institutional Ethics Committee as per declaration of Helsinki. Genetic screening of rs2075602 and rs2075603 variants were carried out by Sanger sequencing. Serum galectin-3 were assessed in the studied participants using commercially available ELISA kits. Galectin-3 levels were stratified according to different genotypes and data thus obtained was subjected to suitable statistical analyses. **Results:** Comparison for both SNPs showed no significant ($P > 0.05$) difference in genotypic as well as allelic distribution between the studied groups. Significantly ($P < 0.001$) elevated serum galectin-3 levels were observed in cases (1.81 ± 0.96 ng ml^{-1}) as compared to controls (0.95 ± 0.51 ng ml^{-1}). Overall significantly ($P < 0.05$) elevated galectin-3 levels were observed in cases with each genotype of both variants than respective controls. However comparison of galectin-3 levels among various genotypes of both variants showed no significant ($P > 0.05$) difference in cases as well as in controls. **Conclusion:** The studied variants may not be associated with diseases susceptibility. Furthermore, studies involving other CRD variants are required to elucidate their association with disease susceptibility.

1895T

Genetic variation in the estrogen receptor alpha gene (*ESR1*) and susceptibility to rheumatoid arthritis. *S.E. Lofgren[1], I.A. Pereira[2], M.N. Drehmer[1].* 1) Biologia celular, Embriologia e Genetica, Universidade Federal de Santa Catarina, Florianopolis, SC, Brazil; 2) Hospital Universitário, divisão de reumatologia, da Universidade Federal de Santa Catarina, Florianopolis, SC, Brazil.

  Background: Estrogen is the main female hormone, which acts through its receptors ER-α and ER-β (ERs), codified by *ESR1* and *ESR2* genes, respectively. Today, several lines of evidence suggest a pivotal role of this pathway in several immune-related responses in the context of autoimmune diseases. Here wee analyzed the association of a genetic variant in the estrogen receptor α gene (*ESR1*) in a group of Brazilian patients with Rheumatoid arthritis (RA) for susceptibility and detailed clinical manifestations of RA. **Methods**: A total of 449 individuals (268 healthy controls and 181 RA patients) were genotyped for the polymorphism rs2234693 in the *ESR1* gene. Allele and genotype frequencies were calculated and analyzed using Unphased software. **Results:** The rs2234693 variant was associated with RA by allelic and genotypic analysis (OR = 1.31, $p = 0.048$ and OR = 1.57, $p = 0,036$, respectively). Stratifying by gender, in female RA patients and controls the association was more significant (OR = 1.43, $p = 0,016$ for allelic association and OR = 2.29, $p = 0,011$ by genotypic association). It was also associated with the presence of high blood pressure, osteoporosis and cardiopathy. A slight association was seen with anti-CCP antibodies when analyzing only females (OR= 1.19, $p = 0.041$). **Conclusion:** The rs2234693 variant was associated with the presence of RA, especially in females, and potentially associated with anti-CCP antibodies in female patients, as well as with other rather unspecific clinical manifestations of the disease.

1896F

Identification of rare variants in Italian multiplex families with multiple sclerosis using a next generation sequencing approach. *E. Mascia[1], A. Zauli[1], C. Guaschino[1], M. Sorosina[1], A.M. Osiceanu[1], S. Santoro[1], L. Ferrè[1,2], S. Peroni[1], D. Biancolini[3], S. Bonfiglio[3], D. Lazarevic[3], C. Basagni[4,5], N. Barizzone[4,5], G. Meola[6,7], V. Martinelli[2], G. Comi[2], G. Tonon[3], F. Esposito[1,2], S. D'Alfonso[4,5], F. Martinelli Boneschi[6,7].* 1) Laboratory of Human Genetics of Neurological Diseases, CNS Inflammatory Unit & INSPE – IRCCS San Raffaele Scientific Institute, Milan, Italy; 2) Department of Neurology, IRCCS San Raffaele Scientific Institute – Milan, Italy; 3) Center for Translational Genomics and Bioinformatics, IRCCS San Raffaele Scientific Institute – Milan, Italy; 4) Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy; 5) Department of Health Sciences, University of Eastern Piedmont, Novara, Italy; 6) Laboratory of Genetics of Complex Disorders and Department of Neurology, IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy; 7) Department of Biomedical Sciences for Health, University of Milan, Milan, Italy.

Multiple Sclerosis (MS) is a multifactorial inflammatory and neurodegenerative disease in which both genetic and environmental factors contribute to the pathogenesis. More than 200 common variants and several signals in the HLA region have been already identified to be associated to MS susceptibility, however it is still unclear the role of rare/private variants in raising disease risk. This study is aimed to explore the role of rare functional variants in MS by studying multiplex families. We recruited a total of 30 families with ≥ 3 affected relatives (2 of them with first-degree relationship). Agilent SureSelect Human All Exon QXT V5 kit was used to generate WES data, while one family (4MS + 4 healthy controls (HC)) was sequenced at whole genome level (Illumina TruSeq DNA Sample Preparation Kit). Sequencing was performed with paired-end (2x101bp) protocol on Illumina Hiseq2500 platform. Until now, sequencing and analyses have been performed for 13 families (41 MS and 32 HC). WES data generation for the remaining families is ongoing. An internal pipeline based on BWA aligner and GATK UnifiedGenotyper was followed for the alignment and calling steps, while variants annotation was performed using SnpEff suite. For each family we selected variants with: (i) high or moderate impact, (ii) allele frequency (AF) ≤ 5% according to public databases, (iii) AF ≤ 10% in Italian population according to our internal database of 62 HC. We then followed two models: (i) "sharing model" by selecting those variants shared by all cases within each family, and (ii) "segregation model" by selecting those variants that segregates according to the family pedigree. We generated 2 lists of genes specific for each family that were intersected in order to identify genes in common between families. Pathway analysis was also performed using WebGestalt tool. Following the "sharing model", we identified variants in 1926 genes, 148 of which showing variants in at least 3 families. Pathways analyses revealed that the identified genes were mainly involved in immunity and in autoimmune diseases. Following the "segregation model", only 12 genes were identified in common between at least 2 families. The same strategy will be used for the remaining families. We conclude that the application of WES in familial cases can be useful in the identification of novel genes and pathways involved in MS pathophysiology.

1897W

Genetic association between not related to HLA immune gene polymorphisms and development of specific autoantibody is limited to few genetic loci in patients with rheumatoid arthritis. *L. Padyukov[1], B. Brynedal[2], D. Ramsköld[1], L. Israelsson[1], M. Hansson[1], R. Holmdahl[3], L. Klareskog[1], V. Malmström[1], D. Gomez-Cabrero[4], K. Shchetinsky[1], RACI.* 1) Rheumatology Unit, Karolinska Inst, Stockholm, Sweden; 2) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Department of Medical Biochemistry and Biophysics Karolinska Institutet, Stockholm, Sweden; 4) Unit of Computational Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Development of anti-citrullinated peptide antibodies (ACPAs) is very specific event for rheumatoid arthritis (RA) and may inform about changes in biological pathways preceding the disease development. Since multiple genetic loci were found in relation to RA, we investigated association between genetic variations and development of several types of ACPAs in three big cohorts of RA patients from Sweden, USA and UK (EIRA, NARAC and WTCCC) with overall 5,668 individuals in relation to non-HLA genetic variations. Autoantibody detection for 16 ACPA specificities was performed by multiplex analytic microarray system and genotyping on Illumina Immunochip was used in our study. Logistic regression analysis in each cohort with following meta-analysis was employed as statistical method. We detected a limited number of associated loci that reached study-wide significant association with development of ACPAs. Development of multiple ACPAs associates with *PTPN22* genetic variants and development of anti-CEP1 antibody associates with locus at chromosome 13, corresponding to uncharacterized lncRNA gene *LOC105370177*. Most of RA associated loci, however, are not associated strongly with ACPA development in our study. Stratification by shared epitope alleles led to finding of residual association with *PTPN22* locus for vimentin peptide in SE positive cases, but no significant findings in SE negative cases, likely because of lack of statistical power. Our data suggest strong genetic regulation for development of multiple ACPA specificities in RA within *PTPN22* locus, but very specific regulation within locus at chromosome 13, which may characterise disease development in a subset of cases.

1898T

Variants near *HLA-DQA1* contribute to the development of antibodies to anti-TNF in Crohn's disease. *A. Sazonovs[1], N. Kennedy[2], L. Moutsianas[1], K. de Lange[1], C. Anderson[1], T. Ahmad[2], J. Barrett[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) Royal Devon and Exeter NHS Foundation Trust, Barrack Road, Exeter, Devon, United Kingdom.

Crohn's Disease (CD) is a chronic condition that causes inflammation of the gastrointestinal tract. While there is no cure for CD, the anti-TNF drugs infliximab and adalimumab are effective in treating patients with Crohn's disease refractory to conventional therapies. Clinical efficacy is associated with mucosal healing and improved quality of life. However anti-TNF drugs are expensive, 10-40% of patients fail to respond to 10-12 weeks of therapy (primary non-response - PNR), and a further 25-45% lose response (loss of response – LOR) within the 12 months of treatment. The PANTS (Personalised Anti-TNF Therapy in Crohn's Disease) study is a prospective observational UK-wide study designed to investigate the clinical, serological and genetic factors that determine primary non-response, loss of response and adverse drug reactions to anti-TNF drugs in patients with active CD. We have genotyped 1,404 patients to perform a genome-wide association study to uncover genetic variants that impact anti-TNF response. A comparison of patients who show no efficacy of anti-TNF (PNR) with those who achieve long-term remission did not show any significant associations. We next tested for association with the development of anti-drug antibodies, one of the key mechanisms of non-response. Using a Cox proportional hazards model, we uncovered two genome-wide significant associations with time from initial treatment to first detection of clearing anti-drug antibodies, as defined by a positive anti-drug antibody titre in the absence of detectable drug. One signal *(p = 5.8x10-9)* was near *HLA-DQA1*, a region for which other variants have been previously implicated in adverse events in response to other drugs. We next imputed HLA alleles at 4-digit resolution using HIBAG, but no single allele reached genome-wide significance, suggesting that this signal may not be driven by a classical HLA allele. These results represent the first robust associations to immunogenicity to one of the most widely prescribed anti-inflammatory medications. We are now carrying out whole genome sequencing in the same set of patients in order to expand these analyses from common to rare variants.

1899F

Longitudinal clinical, molecular, and immunohistochemistry changes during mycophenolate mofetil therapy in patients with systemic sclerosis. *D.M. Toledo[1], M. Hinchcliff[2], J. Taroni[3], T.A. Wood[1], J. Franks[1], S. Shah[4], R. Agrawal[4], L. Beussink-Nelson[4], M.A. Carns[4], S. Podlusky[4], B. Christensen[1], P.A. Pioli[1], M.L. Whitfield[1].* 1) Molecular & Systems Biology, Dartmouth College, Hanover, NH; 2) Feinberg School of Medicine, Northwestern University, Chicago, IL; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Northwestern University, Evanston, IL.

Background: Systemic sclerosis (SSc) is a rare autoimmune disease characterized by skin and internal organ fibrosis. The cause of SSc is largely unknown. We have shown that patients can be classified into four molecular gene expression subsets (inflammatory, fibroproliferative, limited, normal-like) based on gene expression from skin. In this study, we use gene expression data from a cohort collected as part of an investigator initiated clinical trial of mycophenolate mofetil/MMF to identify cell types and factors involved in disease progression and improvement during treatment. **Methods:** Skin biopsies and sera were obtained from controls and patients with SSc at base/ MMF naïve, 6, 12, 24, 36 months. DNA microarray, differential gene expression analysis, and gene set enrichment analysis was performed. Immunohistochemistry/IHC was performed for CD163 (macrophage MΦ marker), and CCL2 levels measured by qRT-PCR and ELISA. MΦ migration in response to SSc sera was also performed. Differentiation state and genome-wide DNA methylation was analyzed for SSc-like MΦs. **Results:** Six completers showed two distinct gene expression patterns between 24-36mo determined by MMF status. From base to 24mo, all six subjects on MMF showed a decrease in the inflammatory signature, in immune cell gene signatures, in MΦ IHC counts, and in CCL2 mRNA. MMF cessation at 24mo in three subjects resulted in a return of the inflammatory signature, increased MΦ IHC counts, increased CCL2 mRNA and worsening skin disease. In contrast, three subjects who continued MMF showed a persistent decrease in the inflammatory signature through 36mo with improved/stable skin disease. MMF continuers had lower CCL2 levels and decreased ability to recruit monocytes during treatment compared to base. **Conclusion:** A subset of patients with SSc lose their inflammatory skin gene expression signature and show decreases in skin MΦs and CCL2 mRNA while on MMF. Upon cessation, patients show rebounding inflammation, worsening skin disease, and increases in skin MΦs, while patients who remain on MMF do not. These results implicate a role for MΦs in skin improvement during MMF treatment. Collectively, these data support the hypothesis that MΦs may be the disease-potentiating cell type in SSc and MΦ modulation may be the key to disease improvement. We are further investigating the role of these cells in SSc by characterizing the SSc MΦ DNA methylation signature and using this to predict MΦ numbers in any SSc tissue sample.

1900W

Integrative analysis of genetic, gene expression and DNA methylation data on systemic lupus erythematosus (SLE). *W. Yang, T.Y. Wang, M. Guo, Y.F. Wang, H. Zhang, J. Yang, Y.L. Lau.* Paediatrics & Adolescent Med, Univ Hong Kong, Hong Kong, 10000, Hong Kong.

SLE is a prototype autoimmune disease with extremely complex etiology. There is a strong genetic component for SLE with increased disease risk for family members of affected. Genome-wide association studies (GWAS) have revealed 80 susceptibility loci in both European and Asian populations, pointing to the involvement of both innate and adaptive immunity in SLE pathogenesis. Gene expression and DNA methylation studies in SLE have detected important molecular aberrations involved in the disease process, such as a strong interferon signature in most SLE patients. However, the complex interplay of molecular changes and their functional impact at the cellular and intercellular level remains elusive, hindering an in-depth understanding of the disease mechanisms, which is essential for identifying new treatment targets and precise medical interventions. In this study, we try to integrate findings from genetics, gene expression, and DNA methylation on SLE and to identify hidden mechanistic changes for the disease. Genetics data are based on GWAS datasets on two Chinese populations and another one on European populations. Gene expression and DNA methylation data were extracted from a number of studies from public domains, followed by quality control measures and meta-analyses. Both protein-protein interaction (PPI) databases and information on transcription factor (TF) binding peaks from ENCODE are used to detect the intrinsic connections among different molecules. PPI network analysis was able to identify distinguished clusters (pathways) from these genes that probably play important roles in the disease. Prominent TF enrichment were detected for these SLE genes, such as IKZF1, STAT2, EBF1, and IRF4 being enriched in multiple gene sets. Interestingly, 'STAT' and 'IRF' TFs are found enriched for genes with hypomethylation but not those with hypermethylation. Findings on the estrogen receptor targets in T cells are particularly intriguing given the role of sex in SLE prevalence.

1901T

Identification of one novel IBD susceptibility locus through a genome-wide association study in Korean populations. *B.D. Ye[1], S. Jung[2], M. Hong[2], S.B. Lee[2], B.M. Kim[2], H.S. Lee[2], S.H. Park[1], S.K. Yang[1], K. Song[2].* 1) Department of Gastroenterology, Asan Medical Center, University of Ulsan Colleg, Seoul, South Korea; 2) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea.

Recent meta-analysis of the genome-wide association studies (GWAS) identified over 200 IBD associated regions. However, identified common variants account for only a fraction of IBD heritability. Moreover, despite of observed differences in clinical characteristics of IBD among different ethnicities, there have been limited studies in non-European populations. In a recent Korean GWAS on IBD in 1505 cases (922 CD, 583 UC) and 4041 controls for the discovery stage followed by replication in an additional 1,989 cases (993 CD, 996 UC) and 3,491 controls, we have reported 2 novel IBD susceptibility loci: rs3766920 in *PYGO2-SHC1* at 1q21 and rs16953946 in *CDYL2* at 22q13. In this study, we identified one additional locus at genome-wide significance through validation in additional 574 cases and 542 controls: rs3731257 at 9p21 (combined $P = 2.26 \times 10^{-8}$). rs3731257, located ~900 bp upstream of *CDKN2A-AS1,* is within a LD region of 173 kb that includes *CDKN2A-AS1 CDKN2A, CDKN2B-AS1,* and *CDKN2B*. This locus was reported to be associated with coronary artery disease, diabetes, and multiple cancers. The eQTL analysis showed marginal significance between rs3731257 and mRNA expression of *CDKN2A* ($P = 0.021$), in Japanese derived lymphoblastoid cell lines (http://www.genome.med.kyoto-u.ac.jp/SnpDB). Using knock-out mice, it was shown that CDKN2A plays a role as a modulator of macrophage activation and polarization via the JAK2-STAT1 pathway with possible roles in inflammatory diseases. Our results provide additional biological insight to IBD in Koreans.

1902F

The influence of human genetic variation on HIV related non-Hodgkin lymphoma. *C.W. Thorball[1,2], C. Hammer[1,2], P.J. McLaren[3,4], J. Fellay[1,2], the Swiss HIV Cohort Study.* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) JC Wilt Infectious Diseases Research Centre, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Canada; 4) Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg, Canada.

   Background. Despite effective antiretroviral treatment, individuals infected with human immunodeficiency virus (HIV) have a substantially increased risk of developing several cancers, especially non-Hodgkin lymphoma (NHL). This elevated cancer risk is partly due to virus-induced immunosuppression, co-infections with oncogenic pathogens and traditional risk factors such as smoking. However, all of these risk factors cannot entirely explain the excess cancer burden seen in the HIV-positive (HIV+) population. In this study, we sought to test whether common genetic variants contribute to the increased NHL risk in an HIV infected patient.   **Methods.** To search for human genetic determinants of HIV-associated NHL, we performed genome-wide genotyping of HIV+ patients in the Swiss HIV Cohort Study (SHCS) with a positive NHL diagnosis. To avoid bias introduced by differential treatment, we restricted the samples to individuals enrolled after 2000 (n = 161). We compared them to a control group consisting of age-matched HIV+ patients from the SHCS diagnosed with HIV before 2005 and with no registered cancer diagnosis as of September 2016 (n = 1184). Genotype data were imputed using the 1000 Genomes phase 3 reference panel. We used logistic regression with PLINK2 to identify associations with NHL while adjusting for sex and the first 5 principal components. Transcriptomic changes due to associated variants were imputed using MetaXcan with the GTEx whole blood and EBV transformed lymphocytes transcriptome datasets.   **Results.** Genome-wide associated variants (p < 5 x 10^{-8}) were observed at the *PAM16* locus with rs113974017 and rs201469434 having the lowest p-values (p = 3.90 x 10^{-8}). Analysis of publicly available transcriptomes with MetaXcan demonstrated that these variants significantly impact the expression of *PAM16* in whole blood (p = 3.82 x 10^{-6}) and *NMRAL1* in EBV transformed lymphocytes (p = 6.62 x 10^{-4}).   **Discussion and conclusion.** We have identified a novel NHL risk locus in an HIV+ population. The transcriptomic impact on *NMRAL1* is interesting, since this gene has previously been found to be a negative regulator of *TRAF3* by inhibiting its ubiquitination. The involvement of *TRAF3* in lymphoma is well established, as knockdown of *TRAF3* in mice leads to the spontaneous development of B cell lymphomas. On the other hand, overexpression of *TRAF3* leads to B cells becoming hyperactivity and thus promoting autoimmunity, inflammation and cancer.

1903W

***IL1RN* variants influence systemic juvenile idiopathic arthritis susceptibility and are a biomarker of non-response to treatment with anakinra.** *E.G. Shuldiner[1], V.L. Arthur[1], A. Hinks[2], P. Woo[3], W. Thomson[2], E.F. Remmers[4], M.J. Ombrello[1], International Childhood Arthritis Genetics (INCHARGE) Consortium.* 1) Translational Genetics and Genomics Unit, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD; 2) Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, University of Manchester, Manchester, United Kingdom; 3) Center of Paediatric and Adolescent Rheumatology, University College London, London, United Kingdom; 4) Inflammatory Disease Section, National Human Genome Research Institute, Bethesda, MD.

   Systemic juvenile idiopathic arthritis (sJIA) is a chronic childhood inflammatory disease that is poorly understood. Early amelioration of inflammation in sJIA is thought to improve long-term prognosis, but heterogeneous responses to anti-cytokine agents, such as anakinra (recombinant IL1RA [interleukin-1 receptor antagonist]), can delay achievement of this goal. Until recently genetic investigations of sJIA have studied candidate genes in small patient collections. These studies found only modest associations, yet they are regularly included in discussions of sJIA pathophysiology. Therefore we examined the 11 reported sJIA candidate susceptibility loci (*IL1A/B, GLI2, IL1RN, IL1R2, IL10/20, IL6, MVK, CCR5, MIF, SLC26A2, TAPBP*) in the largest sJIA study population ever assembled. Single nucleotide polymorphism (SNP) genotypes were evaluated in 770 sJIA cases and 6947 controls from the INCHARGE sJIA dataset. We examined 5479 SNPs in 11 candidate regions, among which 500 SNPs were independent (r^2<0.5), defining the study's significance threshold as p<1E-4. Logistic regression was performed in 9 case-control strata and association results were meta-analyzed. Association meta-analysis revealed a single association signal in the *IL1RN* promoter region, where 3 linked SNPs showed significant association with sJIA. The effect of sJIA associated SNPs on *IL1RN* expression was evaluated *in silico* in paired whole genome and RNA sequencing data from lymphoblastoid cell lines (LCL) of 1000 Genomes Project subjects. Analysis of LCL data showed that *IL1RN* expression is inversely correlated with sJIA risk. Moreover, sJIA associated SNPs are among the strongest published determinants of *IL1RN* and IL1RA levels, linking low *IL1RN* expression with increased sJIA risk. The relationship between sJIA associated SNPs and response to anakinra treatment was evaluated in 38 sJIA patients. The presence of homozygous *IL1RN* high expression alleles correlated strongly with non-response to anakinra (p=9.8E-4, OR 17.3). Although high expression *IL1RN* alleles were protective against sJIA, patients with 2 high expression alleles were significantly less likely to respond to anakinra treatment than those with 1 or 2 low expression alleles. This is the first report to link sJIA risk and response to anakinra treatment with genetically determined capacity to produce *IL1RN* or IL1RA. These SNPs are the first potential biomarker(s) capable of prospectively guiding therapeutic decision making in sJIA.

1904T

First report of the mutational and phenotypic spectrum of hereditary spherocytosis in Indian patients.

A. Aggarwal[1], M. Jamwal[1], P. Sharma[1], MUS. Sachdeva[1], P. Malhotra[2], D. Bansal[3], R. Das[1]. 1) Hematology, Post Graduate Institute of Medical Education and Research, Chandigarh, India; 2) Internal Medicine, Post Graduate Institute of Medical Education and Research, Chandigarh, India; 3) Pediatrics (Hemato-Oncology Unit) Post Graduate Institute of Medical Education and Research, Chandigarh, India.

Introduction Hereditary spherocytosis (HS) is a common inherited hemolytic anemia characterized by the presence of microspherocytes. The pathogenesis involves defects in any of the several genes coding for membrane proteins that link the membrane skeleton to the overlying lipid bilayer. Membrane proteins include ankyrin, band 3, β- and α-spectrin and protein 4.2. We studied the molecular spectrum and genotype-phenotype correlation of HS in Indian patients. **Methodology** Complete blood counts, incubated osmotic fragility test, antiglobulin test, eosin-5'dye-binding test were done in 76 cases from 50 families to diagnose HS. RBC membrane ghosts were prepared and were analysed on gradient gels (4-12%). Relative quantification of mRNA isolated from enriched reticulocytes was done by qRT-PCR. cDNA sequencing of *ANK1*, *SPTB*, *SLC4A1* and *EPB42* genes was done (ABI3130). TruSight One™ sequencing panel was used for preparing libraries in 11 cases that were sequenced on Miseq™ (Illumina). MiSeq Reporter™ and VariantStudio™ were used for analysis, characterization and annotation of variants. Possible pathogenic variants were validated by Sanger sequencing in cases and family members. G6PD-deficiency, α-thalassemia and *UGT1A1* polymorphism were studied as phenotype modifiers. **Result** SDS-PAGE, qRT-PCR and cDNA sequencing were not contributory in deciphering molecular pathologies due to instability of mutated RNA and compensary protein production by normal allele. NGS uncovered novel pathogenic mutations in *ANK1* and *SPTB* out of which 30% were splice site, 30% were small indels and 40% were nonsense mutations in 10 out of 11 patients. Inheritance was non-dominant in 50% and autosomal dominant in 30% cases. G6PD Mediterranean variant in four HS patients led to greater transfusion requirements. Genotyping for Gilbert syndrome showed homozygosity (TA7/TA7) for promoter variant of *UGT1A1* in 38% of the HS patients with significantly higher mean bilirubin level of 8.44 mg/dl and higher frequency of cholelithiasis (36%) (p<0.001). **Conclusion:** This first ever study on the molecular spectrum of HS from India revealed predominantly sporadic and dominantly-inherited defects in *ANK1* and *SPTB* in North Indian HS patients. Most of the cases (70%) presented at an early age with jaundice and anemia. Co-inherited G6PD deficiency and other genetic modifiers led to phenotypic variability. NGS provided a sensitive, cost-effective and rapid tool for understanding the molecular pathology of HS patients.

1905F

Autoimmunity genes are associated with acquired hypothyroidism in a large clinical cohort.

J. Freudenberg, C. Schurmann, C. O'Dushlaine, A. Li, N. Banerjee, RGC. GHS. DiscovEHR Collaboration, A.R. Shuldiner, J.D. Overton, J.G. Reid, A. Baras, T.M. Teslovich, O. Gottesman, F.E. Dewey, N.S. Abul-Husn. Regeneron Genetics Center, Tarrytown, NY.

Genome-wide association studies (GWAS) have identified a large number of genes as risk factors for a range of autoimmune diseases. Here, we evaluated the role of known autoimmunity genes in patients with Acquired Hypothyroidism from a large clinical care cohort. Autoimmune thyroiditis is a frequent and well established cause of hypothyroidism. Genes were evaluated based on their prior implication in Rheumatoid Arthritis, Type 1 Diabetes, Inflammatory Bowel Disease, Psoriasis, Multiple Sclerosis or Systemic Lupus Erythematosus in published GWAS. Exome sequencing and Illumina array genotyping data were analyzed for 12,040 hypothyroidism cases (defined by ICD-10 diagnosis code) and 41,318 controls of European ancestry from the Geisinger Health System DiscovEHR cohort. Genome-wide significant (P<5e-8) associations were observed for common variants in *PTPN22* (rs2476601, OR=1.29, P=3.3e-24), the *CTLA4* locus (rs231726, OR=1.13, P=2.29e-14), the *SH3B3-PTPN11* locus (rs3184504, OR=1.14, P=4.36e-18), and the *IZKF4* locus (rs10876864, OR=1.1, 7.68e-10). Within the HLA-region, significant associations were observed both in the class I region (rs3130977, OR=1.15, P=1.17e-18), as well as the class II region (rs1048372, OR=1.14, P=5.3e-18). These results indicate that Acquired Hypothyroidism, a diagnosis present in nearly 20% of individuals in our clinical care cohort, is driven to a notable extent by autoimmunity. Making use of the large sample size available for this trait, we further performed gene-based burden testing by aggregating rare (MAF<1%) protein-altering variants for each gene into a joint test. Interestingly, the strongest association was observed for the *IFIH1* gene, where 4.2% of chromosomes carry at least one such variant. In aggregate, these rare protein altering variants show a protective association with hypothyroidism (OR=0.85, P=4.61e-05) in our dataset. Rare *IFIH1* variants are already known to alter interferon signaling and to influence genetic risk for Type 1 Diabetes and Systemic Lupus Erythematosus. Our findings indicate that rare variants in *IFIH1* also contribute to hypothyroidism, consistent with autoimmunity as major underlying cause, and point to a role for interferon signaling in at least a subset of cases.

1906W

Trans-ethnic meta-analysis of fetal hemoglobin genome-wide association results identifies common variants at the *KLF1* locus. *Y. Ilboudo[1], C. Sidore[2], F. Cucca[2], G. Lettre[1]. 1) Department of Medicine, Université de Montréal, Montréal, QC H3T 1J4, Canada; Montreal Heart Institute, Montréal, QC H1T 1C8, Canada; 2) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.*

The clinical severity of sickle cell disease (SCD) and β-thalassemia, the major hemoglobinopathies that affect millions of patients worldwide, can be ameliorated by increased production of fetal hemoglobin (HbF). To date, genome-wide association studies have uncovered three main loci (*BCL11A*, *HBS1L-MYB*, and the *HBB* locus) that modulate levels of HbF in hemoglobinopathy patients as well as non-anemic individuals. Together, genetic variation at these three loci accounts for ~50% of HbF heritability.   To identify new HbF regulators, we performed a meta-analysis of genome-wide association results obtained from ~3,400 SCD patients of African ancestry and ~6000 healthy Sardinians. All association results were corrected for age, sex, and the main principal components, and were also conditioned on genotypes at *BCL11A*, *HBS1L-MYB*, and *HBB*. Meta-analysis results were well-calibrated (λ_{GC}~1.0) and were not enriched among genetic variants that map to erythroid regulatory elements defined using DNAse 1 hypersensitive site assays, histone tail modification profiling, or ATAC-seq experiments.   We identified one novel genome-wide significant locus on chromosome 19p13 (rs4804210, $P_{combined}$=6.1x10^{-9}, $P_{SardiNIA}$=1.6x10^{-6}, P_{SCD}=6.4x10^{-4}). HbF association results at this locus are independent from the previously reported HbF signal at the nearby *NFIX* gene (Danjou et *al*., *Nature Genet.*, 2015). Using the GTEx resource, we found that rs4804210 (and its linkage disequilibrium proxies) is an eQTL for *DNASE2* and *KLF1*. Further, this region physically interacts with the promoter of *CALR* as determined by Hi-C methodology. Although additional functional work is required, *KLF1* represents a strong candidate causal gene at this locus: (1) it encodes a key erythroid transcription factor, (2) it is mutated in patients with hereditary persistence of fetal hemoglobin, and (3) it regulates the expression of *BCL11A*.   Our results suggest that increasing sample size, even by combining individuals of different ancestry, is a promising strategy to find loci associated with HbF levels, a critical modifier of hemoglobin disease severity.

1907T

An intergenic variant between HLA-DRA and HLA-DRB contributes to the clinical course and long-term outcome of ulcerative colitis in Asians. *H.S. Lee[1], M. Hong[1], S. Jung[1], J. Hong[1], B.M. Kim[1], J.W. Moon[1], S.H. Park[2], K.J. Kim[2], B.D. Ye[2], S.H. Oh[3], K.M. Kim[3], Y.S. Yoon[4], C.S. Yu[4], J. Baek[1], C.H. Lee[5], B. Han[5], J. Liu[6], T. Haritunians[7], D.P.B. McGovern[7], S.K. Yang[2], K. Song[1]. 1) Department of Biochemistry and Molecular Biology, Department of Biochemistry and Molecular Biology, Seoul, South Korea; 2) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 3) Department of Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, Korea; 4) Colon and Rectal Surgery, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 5) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Seoul, Korea; 6) Human Genetics Group, Genome Institute of Singapore, Singapore; 7) The F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, United States of America.*

Most currently known susceptibility loci for ulcerative colitis (UC) are not associated with the prognosis of UC. To define the contribution of human leukocyte antigen (HLA) to UC prognosis, we performed HLA imputation on three independent discovery cohorts of 1,336 UC patients (972 with poor prognosis and 364 with good prognosis) using the SNP2HLA with the Korean HLA reference panel. Associations between UC prognosis and binary markers in HLA were analyzed by logistic regression. Validation was performed on an additional 625 patients (496 with poor prognosis and 129 with good prognosis), and those data were also analyzed by logistic regression. The three discovery cohorts and the replication cohort were then meta-analyzed using the inverse-variance weighted method, assuming a fixed-effect model. We found that an intergenic variant between HLA-DRA and HLA-DRB was associated with poor prognosis of UC at genome-wide significance (odds ratio = 1.72; 95% confidence interval = 1.43–2.07; $P_{combined-meta}$ = 1.04 × 10^{-8}), with effect size increasing incrementally according to worsening of prognosis, but showed no association with susceptibility. The presence of the risk variant had a sensitivity of 80.0% and specificity of 38.1% for colectomy. Our results provide new insights into prognosis-associated genetic variation in UC, which is distinct from the genetic contribution to disease susceptibility. These findings could be useful in identifying poor-prognosis patients who might benefit from early aggressive therapy.

1908F

Genome-wide association study of otitis media in children. *J. Li[1], G. van Ingen[2], Y.R. Li[1], A. Goedegebure[2], M.E. March[1], V.W.V. Jaddoe[3,4,5], F.D. Mentch[1], K. Thomas[1], Z. Wei[6], T. Chang[1], A.G. Uitterlinden[5,7], H.A. Moll[4], C.M. van Duijn[5], F. Rivadeneira[5,7], H. Raat[8], R.J. Baatenburg de Jong[2], P.M. Sleiman[1], M.P. van der Schroeff[2], H. Hakonarson[1,9,10].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Dept. of Otolaryngology, Head and Neck Surgery, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3000 CA, the Netherlands; 3) The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3000 CA, the Netherlands; 4) Dept. of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3000 CA, the Netherlands; 5) Dept. of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3000 CA, the Netherlands; 6) Dept. of Computer Science, New Jersey Institute of Technology, Newark, New Jersey; 7) Dept. of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3000 CA, the Netherlands; 8) Dept. of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam 3000 CA, the Netherlands; 9) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 10) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Otitis media (OM) is a common disease of early childhood. Interrelated phenotypes include acute OM (AOM), recurrent AOM (rAOM) and OM with effusion (OME). Its etiology is one of complex associations between host, pathogen, environmental and genetic risk factors. To unravel the genetic component to the etiologies of OM we conducted genome-wide association studies (GWAS) per otitis phenotype to identify genetic susceptibility loci to OM, next to underlining possible differences between phenotypes. Our cohort includes children of European ancestry at the Children's hospital of Philadelphia (n=10,634) and the Generation R Study, Erasmus University Medical Center in the Netherlands (n=2,669). We established 3 otitis media phenotypes, including AOM, rAOM, and OME, and 2 combined phenotypes that were previously used in literature, namely Tympanostomy, and an OM combined phenotype consisting of all cases. GWASs were performed per phenotype at both centers and then results were meta-analyzed. We identified a genome-wide significant locus for phenotype AOM at chr16p12.3 with top SNP rs76488276 (P-value=4.533e-08) which is located in an intergenic region. The closest gene is Glycoprotein 2 (*GP2*) which is involved in innate immune response. For the other phenotypes, we found more than 40 SNPs associated with each phenotype at significance level P<0.00001, though we did not identify individual genome-wide significant loci. Only a limited number of significant SNPs were discovered in more than one phenotype, suggesting different genetic profile for these affiliated phenotypes. The difference in individual SNP association, pathway analysis and protein-protein interaction network analysis results for each phenotype underline genetic complexity of this common disease and its interrelated subtypes.

1909W

Multiple HLA B*57 alleles, sharing the amino acid residue valine[97], are associated with drug-induced liver injury due to flucloxacillin in a European population. *P. Nicoletti[1], G.P. Aithal[2], S. Coulthard[3], R. Andrade[4], E. Bjornsson[5], J.F. Dillon[6], A.H. Maitland van der Zee[7], J.H. Martin[8], M. Molokhia[9], M. Pirmohamed[10], M. Wadelius[11], A. Floratos[12], Y. Shen[12], M.R. Nelson[13], A.K. Daly[3].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York; 2) NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and University of Nottingham, UK; 3) Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK; 4) UGC Digestivo, Instituto de Investigación Biomédica de Málaga (IBIMA), Hospital Universitario Virgen de la Victoria, Universidad de Málaga, Málaga, Spain; Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd); 5) Department of Internal Medicine, Landspitali University Hospital, Reykjavik, Iceland; 6) John F. Dillon, Medical Research Institute, University of Dundee,Ninewells Hospital, Dundee, UK; 7) Anke H. Maitland-van der Zee, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, Netherlands; 8) School of Medicine and Public Health, University of Newcastle, New South Wales, Australia; 9) Department of Primary Care and Public Health Sciences, King's College, London, UK; 10) Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, UK; 11) Department of Medical Sciences and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 12) Department of Systems Biology, Columbia University, New York; 13) Target Sciences, GSK, King of Prussia, Pennsylvania,.

Background: In Europe, flucloxacillin is a widely prescribed beta-lactam antibiotic but its use is associated with serious liver injury in approximately 1 in 12,000 patients. This liver injury shows strong genetic susceptibility driven by the *HLA-B*57:01* allele with carriers at 80 times higher risk. **Methods**: DILIGEN/iDILIC consortia collected 197 North European flucloxacillin-induced liver injury (Flu-DILI) cases between 2005 and 2013. The cases, matched with 6835 population controls, were genome-wide genotyped and SNPs, HLA alleles and amino acid residues were predicted to confirm and further investigate risk factors located within and outside the MHC region. **Results**: Our GWAS confirmed that *B*57:01* is the major risk factor (OR = 36.62; 95% CI [26.14-51.29]; P = 2.67×10^{-97}) with a carriage frequency of 84%. Other HLA alleles belonging to the *B*57:01* haplotype, like *C*06:02*, *DQB1*03:03*, *DRB1*07:01*, *DQA1*02:01* and *A*01:01*, also showed a genome-wide significant risk effect while *C*07:02*, *B*07:02* and *DQB1*03:01* were protective. Haplotype analysis showed that B*57:01-containing haplotypes confer risk while *B*57:01*-omitting haplotypes seems to be protective, suggesting that *B*57:01* and no other alleles within the haplotype is the main risk factor. *B*57:01* conditional analysis revealed that *B*57:03* was the most significant independent risk factor (OR = 79.21; 95% CI [13.57-462.4]; P = 1.2×10^{-6}). Interestingly, valine in position 97, a residue shared by several B*57 alleles including B*57:01, B*57:03 and B*57:02, had the highest effect size (OR = 38.1, 95% CI [27.07-53.62], P = 9.7×10^{-97}). Conversely, arginine[97] and serine[97] have significant protective effects (OR= 0.43 P= 5.13×10^{-14} and OR= 0.53, P=9.82×10^{-7}). Serine[97] is a characteristic residue for the protective B*07:02. No other signal was identified outside the MHC region. **Conclusion**: We found that *B*57:01* and *B*57:03* were significant independent risk determinants for Flu-DILI. The shared amino acid valine[97] showed the strongest risk association, while arginine[97] and serine[97] showed a genome wide significant protective effect. Abacavir hypersensitivity is associated only with *B*57:01,* but appears to have a different mechanism from Flu-DILI involving alteration of the peptide repertoire with aspartate [114] and serine[116]. The novel association between Flu-DILI and valine[97] could help explain the apparently different mechanisms involved in the adverse reactions to these two drugs.

1910T

Transcriptome analysis of systemic lupus erythematosus reveals distinct susceptibility, activity and severity signatures. *N. Panousis[1], G. Bertsias[2], I. Gergiannaki[2], M. Tektonidou[3], M. Trachana[4], A. Banos[5], A. Fanouriakis[6], C. Pamfil[7], D. Boumpas[5,6], E.T. Dermitzakis[1].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Department of Rheumatology, University of Crete, Faculty of Medicine, Crete, Greece; 3) Department of Rheumatology, University of Athens, Faculty of Medicine, Athens, Greece; 4) Department of Pediatrics, Aristotle University of Thessaloniki, Faculty of Medicine, Thessaloniki, Greece; 5) Biomedical Research Foundation of the Academy of Athens, Athens, Greece; 6) Department of Rheumatology, Attikon Hospital, Athens, Greece; 7) Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania.

Systemic Lupus Erythematosus (SLE) is a multi-organ, complex autoimmune disease with immense immunological and clinical heterogeneity. Understanding the molecular mechanisms of this variability is essential to facilitate early diagnosis and personalized therapy. We report an RNA-seq analysis of whole blood in 142 SLE patients and 58 matched healthy individuals. We defined the global transcriptional signature of SLE by identifying differentially expressed genes between SLE and healthy individuals. Correcting for the estimated proportions of different immune cell types revealed that the well-established mediator in the pathogenesis of SLE the Interferon (IFN) signaling pathway is robust and prevalent across different immune cells. We also defined a cell specific signature of SLE in immune cells and found that naive and memory B cells, CD4 memory resting T cells, CD8 T cells, and neutrophils have different cell composition effects on gene expression between SLE and healthy individuals. By differential expression analysis between active SLE, inactive SLE and healthy individuals, we defined a 'core/susceptibility' and a 'flare/activity' disease expression signature with oxidative phosphorylation, ribosome regulation and cell cycle being the biological processes enriched for the latter. We propose a measurement of disease activity/severity by combining the validated Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2k) with a new pseudo-phenotype, derived from PCA analysis that explains most of the gene expression variability correlated with this index. We also reveal unique transcriptional signatures across disease sub-phenotypes, showing that granulocytes/neutrophils and NETosis play a crucial role in lupus nephritis pathogenesis. The substantial differences (6730 differentially expressed genes 5% FDR) in gene expression between SLE and healthy individuals allowed as to build multiple classifiers that can discriminate disease versus healthy status with a median sensitivity of 83% and median specificity of 100%. Finally, we estimate the tissues harbour genetic causality of SLE by integration of GWAS signals for SLE and eQTLs from 44 tissues available from the GTEx project. Our study defines for the first time, a distinct susceptibility and activity/severity signature in SLE that may facilitate the identification of patients at risk for severe SLE and facilitate patient selection and monitoring for targeted therapies.

1911F

***NKG2D* variation and viral bronchiolitis.** *A. Pasanen[1], M.K. Karjalainen[1], L. Kummola[2], M. Ruotsalainen[3], T. Jartti[4], E. Goksör[5], G. Wennergren[5], I.S. Junttila[6], M. Hallman[1], M. Korppi[7], M. Rämet[1].* 1) PEDEGO Research Unit, Medical Research Center Oulu, University of Oulu, and Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland; 2) Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 3) Kuopio University Hospital, Pediatrics, University of Eastern Finland, Kuopio, Finland; 4) Department of Pediatrics, University of Turku and Turku University Hospital, Turku, Finland; 5) Department of Pediatrics, University of Gothenburg, Queen Silvia Children's Hospital, Gothenburg, Sweden; 6) Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland and Fimlab Laboratories, Tampere, Finland; 7) Center for Child Health Research, Tampere University and Tampere University Hospital, Tampere, Finland.

Bronchiolitis is a lower respiratory infection that is most often caused by respiratory syncytial virus. Annually, 2-3% of an age cohort develop bronchiolitis that requires hospital treatment. Genetic background of bronchiolitis is not well known. In the present study, our aim was to identify susceptibility loci by comparing a matched subset (n=187) of our previous bronchiolitis GWAS with known quantitative trait loci (QTLs). A Finnish-Swedish population with 700,000 genotyped variants was tested for association. To reconsider potentially omitted true associations, the variants below a permissive p-value threshold ($p<10^{-4}$) were analyzed for colocalization with known QTLs. Expression QTLs (eQTLs) were screened from Genotype-Tissue Expression (GTEx) data, and protein QTLs were examined in pGWAS server. Fluorescence-activated cell sorting (FACS) was used to confirm allele specific expression levels of the identified QTLs. The association of candidate loci was tested in two additional Finnish bronchiolitis case sets (n=124 and 101) that were tested against the GWAS controls. Of the 36 bronchiolitis GWAS SNPs with $p<10^{-4}$ (MAF>0.1), the variant rs10772271 resided within an eQTL associated to *NKG2D* expression ($p=2.2\times10^{-8}$, β=−0.36), and within a pQTL associated to NKG2D protein levels ($p=7.8\times10^{-9}$, β=−0.26). In the GWAS, the major allele (A) was the bronchiolitis susceptibility allele (OR=2.34, p=9.9×10⁻⁵). The association of *NKG2D* locus to bronchiolitis was nominally replicated in one population (OR=1.5, $p=0.049$), whereas the other showed directional consistency (OR=1.43). We confirmed the association of rs10772271 genotypes and NKG2D expression by FACS using NK cells extracted from healthy donors; the predisposing allele was associated to lower NKG2D expression ($p=6.0\times10^{-4}$). We identified *NKG2D* locus as a candidate for bronchiolitis susceptibility. The GWAS association was nominally replicated in one replication population. The susceptibility allele of rs10772271 of the *NKG2D* locus was associated to decreased gene and protein expression in QTL databases and in our flow cytometry experiments. NKG2D is a transmembrane receptor, of which ligands are expressed in response to stress, such as virus infection. We hypothesize that decreased NKG2D expression predisposes to bronchiolitis, but further studies are needed.

1912W

Human genetic variation impacts total IgA levels and pathogen-specific IgG levels. *P. Scepanovic[1,2], C. Alanio[3,4], C. Hammer[1,2], D. Duffy[3,4], J. Bergstedt[5,6], L. Abel[7,8], L. Quintana-Murci[9,10], M. Albert[3,4,11], J. Fellay[1,2], The Milieu Intérieur Consortium.* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Immunobiology of Dendritic Cell Unit, Institut Pasteur, Paris, France; 4) Center for Human Immunology, Institut Pasteur, Paris, France; 5) Department of Automatic Control, LTH, Lund University, Lund, Sweden; 6) International Group for Data Analysis, Pasteur Institute, Paris, France; 7) Laboratory of Human Genetics of Infectious Diseases, Necker branch, Inserm U1163, Paris Descartes University, Imagine Institute, Paris, France; 8) St Giles laboratory of Human Genetics of Infectious Diseases, Rockefeller branch, The Rockefeller University, New York, NY, USA; 9) Unit of Human Evolutionary Genetics, Department of Genomes and Genetics, Institut Pasteur, Paris, France; 10) Centre National de la Recherche Scientifique, URA 3012, Paris, France; 11) Department of Cancer Immunotherapy, Genentech, South San Francisco, CA, USA.

Introduction. Antibody isotype levels and humoral immune responses to latent infections and vaccines vary among individuals, and this is thought to be highly influenced by host genetics. To unravel the mechanisms of genetic control of antibody isotype levels and immunoglobulin G (IgG) responses to common pathogens and vaccines, we conducted genome-wide association studies (GWAS) in a well-characterized population of 1000 healthy individuals recruited by the *Milieu Intérieur* Consortium. Methods. The *Milieu Intérieur* (MI) project is a population-based study that aims at assessing the determinants of immunological variance within a healthy population. A total of 1000 individuals of French ancestry, with a 1:1 sex ratio, and stratified across five-decades of life (age 20 – 69), were recruited. Serum samples were used for measuring levels of four antibody isotypes (IgA, IgM, IgG and IgE) and for qualitative (serostatus: seropositive/seronegative) and quantitative assessment of IgG responses against Cytomegalovirus, Epstein-Barr virus (EBV), Herpes simplex virus 1 & 2, Varicella zoster virus, *Helicobacter pylori*, *Toxoplasma gondii*, Influenza A virus, Measles, Mumps, Rubella, and Hepatitis B virus. We used the Illumina OmniExpress and HumanExome arrays for genotyping. After imputation with IMPUTE2, logistic or linear regressions were performed to detect associations between 5 million human genetic variants and antibody levels. Association of rare variants was tested by using sequence kernel association test (SKAT). Results and discussion. Genome-wide significant associations ($P < 5 * 10^{-8}$) were observed between the levels of IgG mounted against EBV (EBNA antigen) and Rubella, and variants in the MHC locus on chromosome 6, confirming previous observations. By imputing classical HLA alleles and amino acids, we found that these associations rely on variations in amino acid composition of the HLA-DRβ1 and HLA-DPβ1 molecules. No genome-wide significant association was found for serostatus. Analysis of rare variants led to identification of seven new loci associated with levels of IgA. Together our results provide new possible insights into mechanisms determining IgA production and response to different viral antigens and encourage further genetic and functional work.

1913T

Characterising copy number variation at the Crohn disease-associated gene intelectin 1 (*ITLN1*). *F. Almalki[1,2], E.J. Hollox[1].* 1) University of Leicester, Leicester, United Kingdom; 2) Taibah University, Medina, Saudi Arabia.

Genome-wide association studies in Europeans have identified an allele at an intronic single nucleotide polymorphism (rs2274910) at *ITLN1* associated with Crohn disease (CD). The human intelectin protein 1 (hItln1), coded by the *ITLN1* gene, is expressed in the intestine and recognizes and binds with multiple glycans, in particular to the terminal acyclic 1,2 diol ligand that is found solely on bacteria. However the variant functionally responsible for the association with CD has not yet been identified. The Database of Genomic Variants (DGV) shows copy number variants overlapping *ITLN1* have been identified by either sequence read depth analysis of aligned short read sequences or by hybridisation intensity data from SNP chips. Our aim was to characterise these variants and explore their potential as variants that contribute to the risk of CD. We used a PCR-based method called the paralogue ratio test (PRT) or junction fragment PCR to characterise these copy number variants. A putative duplication and deletion was not validated, suggesting these were false positive calls from genomewide data. However a 4 kb deletion reported from the 1000 Genome Project was validated and confirmed that it is an in-frame deletion arisen as a result of a non-allelic homologous recombination event between two Alu elements. This deletion has only been found in Africans so far, and is rare, so does not explain the GWAS signal found in Europeans. Nevertheless it may represent an independent susceptibility allele to CD in non-European populations.

1914F

Genetic footprints and functional analysis of polymorphisms in the *PKLR* gene. *O.C.L. Bezerra[1], N. Mabunda[2], G. Salomé[2], L.E. Alvarado-Arnez[1], F.S.G. Kehdy[1], F.S. Neves-Manta[1], R.M. Silva[1], L.P. Ferreira[1], E. Tarazona-Santos[3], L. Gusmão[4], E.F. Carvalho[4], R.O. Pinheiro[1], A.G Pacheco[5], I.V. Jani[2], M.O. Moraes[1].* 1) Laboratory of Leprosy, IOC/FIOCRUZ, Rio de Janeiro (RJ), Brazil; 2) Laboratory of Molecular Virology, National Institute of Health, Maputo, Mozambique; 3) Departament of Geral Biology, Institute of Biological Science, Federal University of Minas Gerais (UFMG), Belo Horizonte (MG), Brazil; 4) DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro (RJ), Brazil; 5) Program of Scientific Computing (PROCC), FIOCRUZ, Rio de Janeiro (RJ), Brazil.

Pathogens exert a strong evolutionary force in humans. *PKLR* encodes for pyruvate kinase (PK), a key red blood cell enzyme that catalyzes the ATP output in glycolysis. PK defects, due to genetic mutations, cause nonspherocytic hemolytic anemia. This gene is under a strong selective pressure by Plasmodium in Africa, especially where malaria is endemic. Mutations responsible for PK deficiency have a dual role: association with resistance to malaria and susceptibility to *Salmonella typhimurium*. We hypothesized that malaria parasite selective pressure in Africa shaped *PKLR* resistant genotypes in humans and those could be an evolutionary trade off for intracellular pathogens. Thus, aiming to investigate the *PKLR* polymorphisms and its association to mycobacterial pathogens, we performed two independent case-control studies with two diseases caused by intracellular pathogens, leprosy and tuberculosis, in a Rio de Janeiro and Mozambique populations, respectively. Our results highlighted G allele of the four tested SNPs and the haplotype G/G/G/G as associated with leprosy susceptibility in individuals from Rio de Janeiro (haplotype frequency in cases – 0.39 – and controls – 0.27). Subsequently, association was confirmed for two of the four SNPs (G allele) in a Mozambican TB case-control study. The frequency of the haplotype G/G/G/G was higher in cases (0.48) than in controls (0.38), suggesting the same direction of association in Mozambique. Interestingly, the frequency of the susceptibility haplotype was augmented in African-ancestry cohorts from EPIGEN Consortium, Salvador (0.56), and 1000Genomes, YRI and LWK (0.87), compared to European-ancestry cohorts, Pelotas (0.32) and CEU (0.27). In addition, considering that iron overload is a secondary consequence of RBC lysis in PK deficient individuals, we evaluated the genotype-phenotype correlation by comparing iron parameters in association with *PKLR* SNPs. The GG-genotype (for all individual SNPs) was correlated to higher ferritin and haptoglobin loads in healthy subjects and leprosy patients. Our findings revealed a consistently genetic association between *PKLR* SNPs with leprosy and tuberculosis susceptibility and the functional analysis among G-carriers supported a relevance role of the *PKLR* gene with mycobacterial infections. Finally, we evidenced a locus of evolutionary trade off in which individuals carrying resistant selected genotypes to malaria infection were more susceptible to intracellular pathogens.

1915W

Exome sequencing identifies variants of the alkylglycerol monooxygenase gene (*AGMO*) as a cause of relapses in visceral leishmaniasis in Sudan. *S. Marquet[1], B. Bucheton[1,2], C. Reymond[1], L. Argiro[1], SH. El-Safi[3], JP. Desvignes[4], C. Béroud[4,5], A. Mergani[6], A. Dessein[1].* 1) INSERM UMR906, GIMP, Labex ParaFrap, Aix-Marseille University, MARSEILLE, France; 2) IRD-CIRAD 177, Montpellier, France; 3) Dept of Microbiology and Parsitology, Faculty of Medicine, Khartoum, Sudan; 4) INSERM UMR910, GMGF, Aix-Marseille University, Marseille, France; 5) AP-HM, Dept de génétique Médicale, Marseille, France; 6) College of Applied Mdical Sciences, Taif University, Turabah, Saudi Arabia.

Visceral leishmaniasis (or Kala-azar, KA) is the most severe form of leishmaniasis. It is characterized by fever, weight loss, hepatosplenomegaly and lymphadenopathy. It is caused by *Leishmania* parasite that multiplies in liver macrophages. During an outbreak of KA in Babar El Fugara, a Sudanese village, 5.7% of cured patients displayed relapses, with familial clustering in half the cases. We performed whole-exome sequencing on a total of 10 relapsing individuals and 11 controls from five families. Rare homozygous and compound-heterozygous nonsense (c.1213C/T, rs139309795, p.Arg405*) and missense (c.701A/G, rs143439626, p.Lys234Arg) mutations of the alkylglycerol monooxygenase gene (*AGMO*) were associated with KA relapse in three families. Sequencing in additional family members confirmed the segregation of these mutations with relapse and revealed an autosomal dominant mode of transmission. Moreover, these mutations were detected heterozygous in two subjects among 100 unrelated individuals with KA who never relapsed after cure, suggesting incomplete penetrance of *AGMO* deficiency. *AGMO* is expressed in hematopoietic cells in particular, and is strongly expressed in the liver. AGMO modulates PAF production by mouse macrophages, suggesting that it may act through the PAF/PAF receptor pathway previously shown to have anti-*Leishmania* activity. This is the first demonstration that relapses after a first episode of KA, are due to difference in human genetic susceptibility and not to modifications of parasite pathogenicity.

1916T

De novo mutations implicate novel genes with burden of rare variants in systemic lupus erythematosus. *A. Roberts[1], V. Pullabhatla[2], M. Lewis[3], D. Mauro[3], D. Morris[1], C. Odhams[1], P. Tombleson[1], U. Liljedahl[4], S. Vyse[1,7], M. Simpson[1], S. Sauer[5,8], E. de Rinaldis[2], A. Syvänen[4], T. Vyse[1,6].* 1) Medical & Molecular Genetics, King's College London, London, United Kingdom; 2) NIHR GSTFT/KCL Comprehensive Biomedical Research Centre, Guy's & St. Thomas' NHS Foundation Trust, London, UK; 3) Centre for Experimental Medicine and Rheumatology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 4) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 5) Max Planck Institute for Molecular Genetics, Berlin, Germany; 6) Division of Immunology, Infection and Inflammatory Disease, King's College London, London, UK; 7) Division of Cancer Biology, The Institute of Cancer Research, London, UK; 8) Max Delbrück Centre for Molecular Medicine (BIMSB/BIH), Berlin, Germany.

Background: The heritability of most complex diseases, including autoimmune disease Systemic Lupus Erythematosus (SLE), remains largely unexplained by common variation, and few examples of rare variant associations have been identified. Here, using complementary whole-exome sequencing (WES) and high-density imputation, we identify candidate genes through *de novo* mutation discovery and demonstrate collective rare variant associations at novel SLE-susceptibility genes. **Methods:** We selected SLE cases with a severe phenotype (young age of onset and clinical features associated with poorer outcome) and sequenced the exomes of 30 SLE parent-affected-offspring trios. Using conservative parameters across three bioinformatics tools, the WES data were scrutinized for *de novo* mutations. Genes with *de novo* mutations were used as candidate genes in subsequent analyses. In a follow-up cohort of 10,995 individuals of matched European ancestry, including 4,036 SLE cases, we imputed genotype data to the density of the combined UK10K-1000 genomes Phase III reference panel across the candidate genes and performed rare variant burden testing. In addition, we assayed the functional effects of the p.His198Gln *de novo* mutation within candidate gene *C1QTNF4* using HEK293-NF-κB reporter cell line and fibroblast L929 cell line. **Results:** We identified 14 genes with missense *de novo* mutations, none of which are within the >70 SLE susceptibility loci implicated through genome-wide association studies (GWAS). We identify a burden of rare exonic variants across *PRKCD* associated with SLE risk (*P*=0.0028), and across *DNMT3A* associated with two severe disease prognosis sub-phenotypes (*P*=0.0005 and *P*=0.0033). We further show no common variant association across these 14 loci. Additionally, we show the p.His198Gln *de novo* mutation within the candidate gene *C1QTNF4* inhibits NF-κB activation following TNF exposure. **Conclusion:** Exome sequencing studies typically lack power to detect rare variant associations for complex traits. In light of recent studies which have demonstrated the limitations of large-scale exome-wide case-control studies in detecting rare variant contributions, our results support extreme-phenotype sampling and *de novo* mutation discovery to aid a hypothesis-driven search for rare variation contributing to the heritability of complex diseases.

1917F

A long non-coding RNA in the rheumatoid arthritis risk locus at chromosome 18 is involved in T cell activities. *M. Houtman, K. Shchetynsky, L. Padyukov.* Rheumatology Unit, Department of Medicine Solna, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden.

Rheumatoid arthritis (RA) is a relatively common autoimmune disorder that is characterized by joint inflammation and destruction. Genome-wide association studies have enabled the discovery of common genetic variations in over 100 loci contributing to RA, but the precise targets of the majority of these associations are unknown. We aimed to identify the molecular mechanisms that connect variations in the protein tyrosine phosphatase non-receptor type 2 (*PTPN2*) locus with RA. We have found that previously identified RA-susceptibility variants in the *PTPN2* locus are associated with the expression of a long non-coding RNA *LINC01882* located 30kb downstream of *PTPN2* (whole blood; GTEx). To identify the function of this long non-coding RNA, we suppressed the expression of *LINC01882* using antisense oligonucleotides and RNA interference in Jurkat T cells. Twenty-four and 48 hours after knockdown, total RNA was extracted and ribosomal RNA was depleted. Sequencing libraries were generated using the Illumina TruSeq Stranded Total RNA kit and sequenced on an Illumina HiSeq 2500 platform, yielding about 25 million paired-end reads per sample. Differential expression analyses were conducted using DESeq2. We identified 12 deregulated genes after *LINC01882* knockdown (6 upregulated and 6 downregulated; FDR < 0.05). One of the downregulated genes is the zinc finger E-box-binding protein 1 (*ZEB1*), a repressor of interleukin-2 (*IL-2*) gene transcription, that points towards the involvement of *LINC01882* in T cell differentiation. Our data suggest that *LINC01882* might contribute to the molecular mechanism underlying RA. Overall, these findings point towards the potential role of *LINC01882* in the RA-susceptibility association attributed to the *PTPN2* locus.

1918W

WGS identifies rare variants influencing variation in blood cell traits in Mexican American families. *N.B. Blackburn[1], J.M. Peralta[1,2], L. Almasy[3,4], D.C. Glahn[5], R. Duggirala[1], J.E. Curran[1], J. Blangero[1].* 1) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA; 2) Menzies Institute for Medical Research, University of Tasmania, Hobart, Tasmania, Australia; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA.

The process of blood cell production, hematopoiesis, is a continuous and dynamic process influenced by a range of genetic and environmental factors. Variation in blood cell traits exists between healthy individuals and, importantly, is clinically relevant with extreme variation indicating a range of pathologies from anemia to infection or chronic hematological disease. It is established that blood cell traits are highly heritable and there have been a number of successes in localizing QTLs that influence variation in blood cell levels. While the key genes involved in hematopoiesis have been extensively characterized the underlying genetic basis of individual variation in blood cell traits remains relatively unknown. Here, we examine for the first time 16 blood cell traits in 1,462 Mexican American individuals from large pedigrees, each with WGS data available. These traits include count measures of RBCs, WBCs and platelets as well as measures of hemoglobin concentration and cell volumes. Using empirically calculated kinship probabilities, estimated using IBDLD from 100,000 LD pruned sequence variants, we identify that each of these traits is significantly heritable, for example 53% and 62% heritability for total WBC and RBC counts respectively (P=2.13×10^{-20} and P=1.51×10^{-28}). Using rare (MAF≤0.05 in public databases) nonsynonymous WGS variants predicted to be deleterious (CADD≥15) we conducted measured genotype associations for each blood cell trait in SOLAR. We identified multiple rare variants associated with variation in blood cell traits, including rs52833951 in *MMP24* (V344I) which was associated with a 2.38 SDU increase in RBC levels, with all carriers having RBC counts in the highest 1% of the sample. In addition, using this variant set, we assess previously identified genome-wide associations from population based cohorts. For example the variant rs34856868 in *BTBD8* (V60I) was previously associated with lower monocyte counts. In our study conversely, this variant was associated with a 0.48 SDU increase in monocyte levels (P=0.008). From the same population based cohort, rs3746072 in *S1PR4* (R375L) was previously associated with a decrease in total WBC count. Further supporting *S1PR4*'s role in WBC production, in our study we identify that rs61731111 (R243C) is associated with a 0.41 SDU decrease in total WBC count (P=0.02). Our results emphasize the utility of WGS in pedigrees for identifying rare variants influencing blood cell traits.

1919T

Trans-ethnic meta-analysis of the Korean, East Asian and European Immunochip data identifies three novel IBD susceptibility loci. *S. Jung[1], H.S. Lee[1], B. Kim[1], S.B. Lee[1], B.D. Ye[2], S.H. Park[2], S.K. Yang[2], K. Song[1].* 1) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea; 2) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.

To date, the 241 IBD susceptibility loci were reported in Europeans. Previously, we reported three new IBD as *NCF4-CSF2RB, CIITA* and *PSMA6-NFK-BIA* as an Asian-specific or Asian-dominant IBD susceptibility loci using a meta-analysis of Korean and East Asian IBD Immunichip. In this study, we performed a meta-analysis of European, Korean and East Asian Immunochip data sets including 42,311 IBD cases and 53,389 controls using Mantra/Metal. We identified three novel loci for IBD/CD at genome-wide significance (Bayes factor > 6 and $P < 5.0 \times 10^{-8}$): rs2624435 in *MYO10-BASP1* at 5p15 (Bayes factor=6.32; OR = 0.94, P_combined = 1.80×10^{-8}), rs2074023 in *TSPAN32* at 11p15 (Bayes factor=7.49; OR = 0.94, P_combined = 1.05×10^{-9}) for IBD and rs7170683 in *LRRK1* at 15q26 (Bayes factor=6.34; OR = 0.94, P_combined = 1.99×10^{-8}) for CD. Our results provide additional biological insight to IBD.

1920F

Associations between a polymorphism of the gene encoding the Toll like receptor and response to infliximab in Japanese patients with Crohn's disease. *S. Ma[1], Y. Fukumitsu[1], Y. Noma[1], T. Inamine[1], S. Kondo[1], S. Urabe[2], K. Matsushima[2], R. Uehara[2], T. Honda[2], H. Machida[2], N. Yamaguchi[2], K. Ohnita[2], F. Takeshima[2], H. Isomoto[3], K. Nakao[2], K. Tsukamoto[1].* 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki; 2) Dept Gastroenterology and Hepatology, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki; 3) Div Med and Clin Sci, Dept Multidisciplinary Int Med, Faculty of Med, Tottori Univ, Yonago.

   Purpose: Infliximab (IFX) is a chimeric monoclonal antibody which neutralizes tumor necrosis factor alpha and is widely used for the treatment of Crohn's disease (CD). However, some CD patients show loss of therapeutic response to IFX in about 20-30% of those at the 1-year treatment period although these patients had achieved remission. Therefore, we investigated an association of *TLR2*, *TLR4* and *TLR9*, which encode the Toll like receptors related to multiple inflammation signals, with response to IFX at the 1-year treatment period in Japanese CD patients. **Methods:** A total of 116 CD patients, which showed good response to IFX at the early period in Nagasaki University Hospital, were divided into two groups, responders and non-responders for IFX, according to the presence of IFX effect at the 1-year treatment period after IFX administration. Eight tag single nucleotide polymorphisms (SNPs) in these 3 genes were analyzed by PCR-restriction fragment length polymorphism, PCR-high resolution melting curves analysis and PCR-direct DNA sequencing. The frequencies of alleles and genotypes of each SNP between responders and non-responders at the period were compared in three different inheritance models using chi-squared test or Fisher's exact test. **Results:** Statistical analyses indicated the frequency of an A/A genotype of rs13105517 in *TLR2* in the minor allele recessive model was significantly decreased in responders at the 1-year treatment period in comparison to that in non-responders ($P = 0.004$, odds ratio = 0.195), implicating ~5.1-fold loss of response. Conversely, a G/G and G/A genotype of rs13105517 indicated ~5.1-fold response to IFX. **Conclusions:** Our results implicate the A/A genotype of rs13105517 in *TLR2* may affect the gain-of-function of TLR2, thereby leading to the activation of the downstream signals in the TLR signal pathway and eventually resulting in chronic intestinal inflammatory process in the intestines of CD patients. This is why these patients with this polymorphism of *TLR2* showed loss of response to IFX at the 1-year treatment period. Therefore, the polymorphism of *TLR2* may have the potential for being a new genetic biomarker to predict the long-term therapeutic effect and loss of response to IFX in Japanese CD patients, and for being a target molecule for new drugs overcoming loss of response to IFX.

1921W

GWA and MHC-fine mapping analyses of multiple sclerosis (MS) age at diagnosis (AAD) identify novel associations with an HLA-DQ heterodimer and *GZMA*. *P.G. Bronson[1], C. Tian[2], A. Dhroso[1], F.B. Briggs[3], K.D. Nguyen[1], M. Seldin[4], T. 23andMe Research Team[2], D. Hinds[2], A. Day-Williams[1], M. Penny[1], S. John[1], K. Estrada[1].* 1) Computational Biology & Genetics, Biogen, Inc., Cambridge, MA, USA; 2) 23andMe, Inc., Mountain View, CA, USA; 3) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 4) Department of Biochemistry, School of Medicine, University of California, Davis, CA, USA.

   Background: Genetic studies of MS have identified >100 non-MHC loci. The strongest association is with *HLA-DRB1*15:01* (OR=3.9, P<1e-600). In Europeans, this risk allele is in almost perfect LD with *HLA-DQB1*06:02*. The risk allele and *HLA-DQB1* are associated with younger age at onset (AAO) (β=-0.72 years, P=2e-8). An association between non-MHC loci and AAO has not been established. Objectives: To conduct case-only imputed HLA and GWAS studies of self-reported AAD in 23andMe MS cases (N=2,339, EUR). Methods: We developed a novel method to test the DNA sequence of HLA peptide binding grooves, based on cis or trans heterodimer configurations. AAD was analyzed using linear regression adjusted for gender and ancestry principal components (estimated from non-HLA SNPs). We tested all alleles, amino acid residues and peptide binding grooves for the classical HLA loci using a dominant model, whereas the standard additive model was used for the GWAS. The HLA significance threshold was P<1.7e-3 (false discovery rate, 0.05), and the GWAS utilized the standard genome-wide threshold. Results: We validated the association between AAD and the risk allele (*DRB1*15:01*) (β=-1.97, P=5.6e-5). However, an *HLA-DQ* peptide binding groove (*HLA-DQA1*05/DQB1*06:02*, in cis or trans) was even more strongly associated (β=-3.49, P=1.3e-6). When both the risk allele and *HLA-DQ* molecule were included in the same logistic regression analysis, association was much stronger for *HLA-DQ* (β=-2.72, P=5.3e-4) than for the risk allele (β=-1.26, P=0.02). *HLA-DQ* was still significant when tested only in the risk allele carriers (β=-2.68, P=4.4e-4). HLA was the only genome-wide significant peak in the GWAS. However, we observed an association with older AAD that was close to genome-wide significance in granzyme A (*GZMA*, 5q11). The peak variant was a common deletion (β=2.62, P=6.7e-8; rs33995145), and the region of association extended from *CCNO* to *DHX29*. Discussion: This is the first genetic HLA study to test the biological molecules (peptide binding grooves) to our knowledge, and we identified a novel association with an *HLA-DQ* molecule. This association was significant even after including *HLA-DRB1*15:01* as a covariate, and when examined only in *HLA-DRB1*15:01* allele carriers. We also identified a novel near-significant non-HLA association with *GZMA*, a protein produced by CD8+ T cells to induce apoptosis of tumor cells and virus-infected cells. Further research is ongoing to confirm our findings.

1922T

Pleiotropy analysis of penicillin and sulfa drug allergy in the Kaiser GERA cohort. *A. Majumdar[1], J. Hoffman[1], R. Melles[2], D. Ranatunga[3], J. Witte[1], E. Jorgenson[3].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, CA, USA; 2) Kaiser Permanente Northern California, Department of Ophthalmology, Redwood City, CA, USA; 3) Kaiser Permanente Northern California, Division of Research, Oakland, CA, USA.

Hypersensitivity reactions against drugs are governed by immunologic mechanisms. Penicillin and Sulfa are two common drug classes and allergies to these types of medications can cause skin rash, itching, breathing problems, headache, and anaphylaxis. Genome-wide association studies (GWAS) have identified genetic risk factors underlying drug allergies, but they have not been investigated as extensively as other complex traits, e.g., cancers, psychiatric disorders. GWAS have also detected loci associated with multiple different diseases (i.e. pleiotropy). Analyzing pleiotropy may provide a better understanding of pathways and biological mechanisms shared by different conditions. We undertake a GW pleiotropic association study of Penicillin and Sulfa drug allergy in the Kaiser Permanente GERA cohort. We consider 748 cases for Penicillin and 678 cases for Sulfa, with 95 common cases having both allergies and a shared pool of 15,894 controls. All the individuals studied here are of European ancestry. We computed the univariate summary statistics for each trait adjusting for relevant covariates, and considered 7,761,491 genotyped and imputed SNPs after various screening. Using the p-value cut-off $5×10^{-8}$, the univariate GWAS for Penicillin detected a novel signal at rs111969675 (P = $1.4×10^{-8}$) which is an intronic variant in the CSMD1 gene located at 8p23.2, but the GWAS for Sulfa did not identify any GW signal. We applied the subset-based meta-analysis ASSET for the two traits adjusting for the shared subjects. Including the univariate GW signal for Penicillin at 8p23.2, ASSET identified two more novel signals associated with both Penicillin and Sulfa that were missed by the univariate GWAS: rs143123254 at 7p11.2 (P = $4.3×10^{-8}$) and rs141475569 at 16q12.1 (P = $6.5×10^{-9}$). rs141475569 is a known eQTL for the ABCC11 gene in liver tissue. We then applied the conditional false discovery rate (cFDR) approach for the two traits. Considering Penicillin as the principal phenotype and Sulfa as the conditioning phenotype, the cFDR approach validated the three signals detected by ASSET using a 5% level of cFDR and Sulfa p-value cut-off as 0.01 and 0.001. The lead SNP of the largest associated LD block detected by the cFDR approach is rs57841279 at 1q23.3. ASSET also produced a suggestive evidence of association at rs57841279 (P = $2×10^{-5}$). rs57841279 is an intronic variant in the SDHC gene and a well-known eQTL. We attempt to replicate these signals in a separate dataset.

1923F

Genetic markers associated with leprosy susceptibility in a group of incident household contacts from Brazil. *L. Arnez[1], T. Gomes Toledo-Pinto[1], I. de Oliveira Moraes[1], T. Calvo Leal[1], P.T. Santos[2], P. Medeiros[3], R. Raju[1], N. Duppre[1], A.M. Sales[1], D. Garbin di Luca[1], E. Nunes Sarno[1], I. Goulart[4], A.G. Pacheco[5], A.C. Pereira[3], M. Ozório Moraes[1].* 1) Leprosy Laboratory, Fiocruz Institute, Rio de Janeiro, Rio de Janeiro, Brazil; 2) Molecular Carcinogenesis Program. National Cancer Institute, Rio de Janeiro; 3) Lauro de Souza Lima Institute, Bauru, Brazil; 4) National Reference Center for Sanitary Dermatology and Leprosy, Clinics Hospital, Federal University of Uberlandia, MG, Brazil; 5) Program of Scientific Computing, Fiocruz, Rio de Janeiro, Brazil.

We performed a Genome Wide Association Study (GWAS) in household contacts that developed leprosy (N=67) compared to contacts equally exposed to bacilli but that remained healthy (N=319) and did not progressed towards the disease. Using logistic regression we tested ~230.000 polymorphisms (IlluminaHumanCore-12) for association with leprosy per se as outcome. We prioritized significant SNPs with OR > 2 or < 0.5 to perform enrichment analysis in the genes were these top associated polymorphisms mapped to. We returned to genotyping data and evaluated the SNPs located in genes from the enriched processes and performed a second logistic regression tests by means of mixed models to assess leprosy. A genetic risk score was obtained and tested as a predictor in two different Brazilian populations with the same study design. Differential expression of gene candidates was assessed in THP-1, ST88-14 cell lines, blood and nerve biopsy samples from contacts and leprosy patients. Our results based on the top markers from the preliminary logistic regression and the enrichment analysis of associated SNPs pointed towards 31 genes that represented biological processes of neuron generation, regulation of neuromuscular junction, neurological system process and cell recognition ($P-value_{enrichment}$ < 0.001). The second logistic regression using mixed models indicated that 17 polymorphisms distributed in this specific set of genes were significantly associated either with susceptibility (N= 11 SNPs) or protection (N=6 SNPs) towards disease development even after correction for multiple comparisons ($P-value_{FDR}$<0.05). Combinatorial SNP analysis indicated that the group of incident household contacts had a higher genetic risk score (P<0.001), further by applying this score in different populations that had the same contact groups under comparison (develop disease versus remain healthy), we obtained probabilities that ranked higher in risk for the contacts that developed leprosy. Functional analysis evaluating the set of associated genes revealed differences in gene expression in nerve biopsies, blood samples and M.leprae-infected cell lines (THP-1, ST88-14). Interestingly, we determined novel SNPs in a group that exhibits the highest risk to progress towards the disease (contacts) indicating that they participate in pathogenesis and can be used as markers for disease outcome.

1924W

Genome-wide association study identifies *HLA-DR/DQ* region for child-hood nephrotic syndrome in Japanese. *X. Jia[1], T. Horinouchi[2], Y. Hitomi[1], A. Shono[2], S.S. Khor[1], Y. Omae[1], K. Nozu[2], K. Kojima[3,4], Y. Kawai[3,4], M. Nagasaki[3,4,5], K. Tokunaga[1], K. Iijima[2].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Pediatrics, Graduate School of Medicine, Kobe University, Kobe, Japan; 3) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 4) Graduate School of Medicine, Tohoku University, Sendai, Japan; 5) Graduate School of Information Sciences, Tohoku University, Sendai, Japan.

Nephrotic syndrome (NS) is the most common cause of kidney disease in children, but the underlying genetic etiology is not well understood. So far, no genome-wide association study (GWAS) has been reported for childhood NS in the world. To identify common genetic variations conferring the risk of childhood nephrotic syndrome, we performed a two-stage GWAS in the Japanese population. A total of 285 patients with childhood-onset NS and 419 adult healthy controls were genotyped using Affymetrix 'Japonica Array' in the discovery stage. *HLA*-genotyping for *HLA-DRB1/DQB1* was performed by sequence-specific oligonucleotide primed PCR (PCR-SSO) method using Luminex platform. In addition, we conducted high-accuracy imputation for both *HLA* class I and class II genes (*HLA-A, -B, -C, -DRB1, -DQB1* and *-DPB1*) based on Japanese population-specific reference. Whole-genome imputation was conducted using a phased reference panel of 2,049 healthy Japanese individuals (2KJPN panel). An independent cohort including 288 cases and 719 healthy controls were collected in the replication stage. Candidate SNPs detected in the discovery stage were genotyped by DigiTag2 or TaqMan assay in the validation and replication studies. In GWAS, the most significant signal was detected in *HLA-DR/DQ* region (rs3134996, P=2.39×10^{-12}, odds ratio (OR)=0.31). In the *HLA* allele analysis, *HLA-DRB1*08:02* (Pc=1.25×10^{-16}, OR=3.18, Pc denotes Bonferroni corrected probability value), *HLA-DRB1*13:02* (Pc=5.57×10^{-14}, OR=0.16) and *HLA-DQB1*06:04* (Pc=9.17×10^{-16}, OR=0.11) showed strong positive or negative associations with childhood NS in both GWAS and replication sample sets. *HLA-DRB1*08:02-DQB1*03:02* (Pc=6.25×10^{-18}, OR=4.50) and *HLA-DRB1*13:02-DQB1*06:04* (Pc=2.00×10^{-15}, OR=0.11) were identified as the strongest susceptible and protective haplotype, respectively. Moreover, the risk haplotype *HLA-DRB1*08:02-DQB1*03:02* also showed a higher frequency in patients with the history of steroid resistance than the ones without such history during the follow-up (Pc=2.01×10^{-2}). After whole-genome imputation, 9 loci in non-*HLA* region showed suggestive associations (P<1×10^{-5}). 18 SNPs from the 9 loci were selected as candidate SNPs for validation and replication, but none of the associations was replicated in the independent cohort.

1925T

A genome-wide association analysis identifies *NMNAT2* and *HCP5* as susceptibility loci for Kawasaki disease. *J. Kim, J. Lee.* Asan Institute for Life Sciences,Ulsan College of Medicine, Seoul, South Korea.

Kawasaki disease (KD), a systemic vasculitis of infants and children, manifests as fever and mucocutaneous inflammation. Although its etiology is largely unknown, epidemiological data suggest that genetic factors are important in KD susceptibility. To identify genetic variants influencing KD susceptibility, we performed a genome-wide association study (GWAS) in 249 children with KD and 1,000 controls in the Korean population, with replication in an independent 666 cases and 3,553 controls. Six single nucleotide polymorphisms (SNPs) in three loci were associated significantly with KD susceptibility (*P* < 1.0 × 10^{-5}), including the previously reported *BLK* locus (odds ratio (OR) = 1.52, *P* = 2.52 × 10^{-11}). The other two loci were newly identified: *NMNAT2* on chromosome 1q25.3 (OR = 1.33, *P* = 1.15 × 10^{-6}) and the human leukocyte antigen (HLA) region on chromosome 6p21.3 (*HLA-C, HLA-B, MICA*, and *HCP5*) (OR = 1.33 – 1.51, *P* = 8.93 × 10^{-6} – 5.24 × 10^{-8}). In a meta-analysis with Japanese data of 428 cases and 3,379 controls using imputed genotype data, the variant in *NMNAT2* remained significantly associated (OR = 1.27, *P* = 5.90 × 10^{-7}), whereas the HLA locus was not replicated in the Japanese population. These results provide new insights into the pathogenesis and pathophysiology of KD.

1926F

Immunochip v2.0 meta-analysis identifies shared genetic loci for inflammatory bowel disease in Korean population. *S.B. Lee[1], B.M. Kim[1], H.S. Lee[1], M. Hong[1], S. Jung[1], J. Hong[1], J.W. Moon[1], J. Baek[1], S.H. Park[2], B.D. Ye[2], S.H. Oh[3], B. Han[4], T. Haritunians[5], D.P.B. McGovern[5], S.K. Yang[2], K. Song[1]. 1) Biochemistry and Molecular Biology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea; 2) Department of Gastroenterology, Asan Medical Center, Seoul, South Korea; 3) Department of Pediatrics, Asan Medical Center Children's Hospital, Asan Medical Center, Seoul, South Korea; 4) Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul, South Korea; 5) The F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA.*

Current genome-wide association studies (GWAS) on inflammatory bowel diseases (IBD) have revealed more than 241 susceptibility loci specific to Caucasian population. These studies however, do not completely reflect the heterogeneity in genetic makeup and thus the clinical outcomes of IBD between different ethnicities. With limited non-European GWAS on IBD, our study aimed to identify novel IBD associated loci in Koreans. We newly genotyped 1,498 IBD cases (750 Crohn's disease, 748 ulcerative colitis) and 491 controls on an Immunochip v2.0 to be used as the discovery cohort. Replication of the candidate loci was performed using previously published Korean GWAS dataset of an independent cohort consisting of 1,337 IBD cases (905 CD, 432 UC) and 3,801 controls. Meta-analysis of the two datasets revealed no novel IBD susceptibility loci that exceed genome-wide significance. We compared our Korean data to the 241 previously reported IBD susceptibility loci in the European population and found that 182 SNPs were available in our data ($r^2 > 0.8$). A total of 20 loci exceeded a significant threshold of 2.07×10^{-4} (0.05/241) and an additional 52 loci showed association at the nominal significance ($P < 0.05$). Furthermore, rs76295456 in *CD28* at 2q33 (odds ratio = 1.25; 95% confidence interval = 1.17–1.32; $P_{combined-meta} = 1.08 \times 10^{-8}$) reached genome-wide significance for the first time in the Korean population. Our study does not provide any new insights on IBD susceptible loci but successful replication of previously reported loci highlights the shared genetic characteristics of IBD in trans-ethnic populations.

1927W

GWAS identified associations of *HLA-DRB1-DQB1* haplotypes and *BTNL2* gene with response to a hepatitis B vaccine. *N. Nishida[1,2], M. Sugiyama[1], H. Sawai[2], J. Ohashi[3], S-S. Khor[2], T. Tsuchiura[1], K. Tokunaga[2], M. Mizokami[1]. 1) Genome Medical Science Project, National Center for Global Health and Medicine, Ichikawa, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Japan; 3) Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Japan.*

Approximately 5-10% of individuals, who are vaccinated with a hepatitis B (HB) vaccine designed based on the HBV genotype C, fail to acquire protective levels of antibodies. Host genetic factors behind low immune response to the HB vaccine were investigated by GWAS and *HLA* association tests. GWAS and *HLA* association tests were carried out using a total of 1,193 Japanese individuals including 107 low responders (HBsAb ≤10 mIU/mL), 351 intermediate responders (10 mIU/mL< HBsAb <100), and 735 high responders (HBsAb ≥100 mIU/mL). Classical *HLA* class II alleles were statistically imputed using the genome-wide SNP typing data. GWAS identified independent associations of *HLA-DRB1-DQB1* haplotypes, *HLA-DPB1* alleles and *BTNL2* gene with immune response to the HB vaccine designed based on the HBV genotype C. Five *HLA-DRB1-DQB1* haplotypes and two *DPB1* alleles showed significant associations with response to the HB vaccine in a comparison of three groups of 1,193 HB vaccinated individuals. When frequencies of *DRB1-DQB1* haplotypes and *DPB1* alleles were compared between low immune responders and HBV patients, significant associations were identified for three *DRB1-DQB1* haplotypes, and no association was identified for any of *DPB1* alleles. In contrast, no association for *DRB1-DQB1* haplotypes and *DPB1* alleles was identified in a comparison between high immune responders and healthy individuals. The findings in this study clearly show the importance of *HLA-DR-DQ* (i.e. recognition of a HB vaccine related HBsAg by specific *DR-DQ* haplotypes) and *BTNL2* molecules (i.e. high immune response to the HB vaccine) for response to a HB vaccine designed based on the HBV genotype C.

1928T

Reduced severity of collagen-induced arthritis in peptidylarginine deiminase type 4 knockout mice. *A. Suzuki[1], Y. Kochi[1], K. Yamamoto[1,2].* 1) IMS, RIKEN, Yokohama City, Japan; 2) Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

Objective. Rheumatoid arthritis (RA) is one of the most common autoimmune diseases, which affects approximately 1% of the world population. RA is a chronic systemic inflammatory disease characterized by the inflammation of synovial joint tissues. ;Previously, we identified peptidylarginine deiminase type 4 (PADI4) as a susceptibility gene for RA by genome-wide association studies. PADI4 is highly expressed in immune cells, such as bone marrow, macrophages, neutrophils, and monocytes. Peptidyl-citrulline is an important molecule in RA because it is a target antigen of anti-citrullinated peptide antibodies (ACPAs), and only PADs (translated protein from PADI genes) can provide peptidyl citrulline via modification of protein substrates. The aim of this study was to evaluate the importance of the PADI genes in the progression of RA. *Methods.* We generated Padi4 knockout (Padi4−/−) DBA1J mice. Padi4−/− DBA1J and wild-type mice were immunized with bovine type II collagen (CII) to develop collagen-induced arthritis (CIA). We compared the incidence and severity score, and performed measurements of expression levels of various inflammatory cytokines and Padi genes in immune cells by real-time TaqMan assay. Also cytokine concentration and CII antibodies in sera were measured by enzyme-linked immunosorbent assay. We investigated PAD4 effect on the transcriptional pattern of macrophage from Padi4-/- mice by microarray. *Results.* We demonstrated that the clinical disease score was significantly decreased in Padi4−/− mice and Padi4 expression was induced by CII immunization. In Padi4−/− mice sera, serum anti-type II collagen (CII) IgM, IgG, and inflammatory cytokine levels were also significantly decreased compared with those in wild-type mice sera. Interestingly, Padi2 expression was compensationally induced in CD11b+ cells of Padi4-/- mice. We also identified Fus gene as a siginificant gene between WT and Padi4-/- mice by independent two microarray tests. *Conclusion.* On the basis of these studies, it appears that Padi4 enhances collagen-initiated inflammatory responses. Our results revealed that PAD4 affected on expression of various cytokines and also controlled Padi genes and Fus gene.

1929F

A locus on chromosome 5 shows African-ancestry-limited association with alloimmunization in sickle cell disease. *L.M. Williams[1], Z. Qi[2], K. Batai[3], S. Hooker[3], R. Machado[4], A. Chen[5], J. Moulds[6], S. Campbell-Lee[4], Y. Guan[2], R. Kittles[3], N. Hanchard[2].* 1) Biological Sciences, University of Botswana, Gaborone, Botswana; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) University of Arizona, College of Medicine, Tucson, AZ; 4) University of Illinois at Chicago, College of Medicine, Chicago, IL; 5) Gulf Coast Pathology, Houston, TX; 6) LifeShare Blood Centers, Shreveport, LA.

Red Blood Cell (RBC) transfusion remains a critical therapeutic intervention in Sickle Cell Disease (SCD); however, the apparent propensity of some patients to develop RBC alloantibodies presents a significant clinical challenge. Such alloimmunization 'responders' often form alloantibodies following each transfusion, limiting the availability of compatible blood. Donor and recipient factors have been implicated in this phenomenon, but there are no definitive predisposing genetic loci. We undertook a genome-wide association study of alloimmunization responder status in 267 SCD chronic transfusion recipients, using genetic estimates of ancestral admixture to bolster our findings. Analyses revealed SNPs on chromosomes 2 and 5 approaching genome-wide significance (GWS; min P-values= 2.0×10^{-8} and 8.4×10^{-8}, respectively), with local ancestry analysis demonstrating similar levels of admixture in responders and non-responders at implicated loci. Association at chromosome 5 was nominally replicated in an independent cohort of 150 SCD transfusion recipients, with META-analysis surpassing GWS (rs75853687, $P = 6.6 \times 10^{-9}$), and this association extended to individuals forming multiple (>3) alloantibodies (meta-analysis $P = 9.4 \times 10^{-5}$). Orthogonal genome-wide haplotype analyses implicated a rare haplotype of African origin at the same locus reaching GWS. The associated locus includes putative binding sites for immunoregulatory transcription factors predicted to affect *cis*-transcription of *ADRAB1* and the lncRNA *LINC01847* - both members of larger ontologies associated with immune regulation. Our findings provide insights to the pathophysiology underlying the development of alloantibodies, implicate non-RBC ancestry-limited loci in the susceptibility to alloimmunization, and may help to identify SCD transfusion recipients who are likely to become alloimmunization responders.

1930W

GWAS meta-analysis in Chinese and European populations identified a novel locus associated with systemic lupus erythematosus on Xp11. *H. Zhang[1], J. Yang[1], Y.-F. Wang[1], T.-Y. Wang[1], W. Yang[1], Y. Zhang[2,1].* 1) The University of Hong Kong, Hong Kong; 2) Guangzhou Institute of Pediatrics, Guangzhou Women and Children's Medical Center, Guangzhou Medical University.

Systemic lupus erythematous (SLE) is a complex autoimmune disease with extremely higher prevalence in females, especially for those in childbearing age. Taking advantage of high throughput genotyping technology, more than 80 loci have been identified as with SLE susceptibility, but only five of them are located on the X chromosome. Previously, we have identified three of the five X-linked SLE susceptibility genes in Chinese Han populations. In this study, making use of existing X chromosome genotype datasets on three trans-ethnic populations in Chinese and European, following by replication in three Asian cohorts, totally including 7,995 cases and 14,598 controls, we have successfully identified a novel risk-associated variant in Xp11 surpassing genome wide significance. We also further explored the established susceptibility loci in *L1CAM- MECP2* region, and showed evidence of a potential independent signal in Asians.

1931T

Targeted sequencing in 1000 SLE patients discovers regulatory alleles that downregulate DAP expression and promote autoimmunity. *P. Raj[1], R. Song[1], B.E. Wakeland[1], C. Liang[1], K. Viswanathan[1], C. Arana[1], B. Zhang[1], J. Zhou[1], B.R. Lauwerys[2], N.J. Olson[3], S.K. Nath[4], J.A. James[5], B.P. Tsao[6], P.M. Gaffney[6], D.R. Karp[7], Q. Li[1], E.K. Wakeland[1].* 1) Immunology, UT Southwestern Medical Center, Dallas, Texas, USA; 2) Pôle de pathologies rhumatismales, Institut de Recherche Expérimentale et Clinique,Brussels, Belgium; 3) Division of Rheumatology, Department of Medicine, Penn State Medical School, PA, USA; 4) Department of pathology, Oklahoma Medical Research Foundation, OK; 5) Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, OK; 6) Division of Rheumatology and Immunology, Department of Medicine, Medical University of South Carolina, Charleston, South Carolina, USA; 7) Rheumatic Diseases Division, University of Texas Southwestern Medical Center, Dallas Texas, USA.

Systemic Lupus Erythematosus (SLE) is an autoimmune disease with strong genetic predisposition. Analysis of common variants in GWASs have identified over 60 loci that account for significant amount of genetic heritability. However, large proportion of genetic heritability is still missing. Here we propose that at least part of the missing heritability of SLE can be explained by the genetic component of intermediate phenotypes that are not detectable through common variants at the level of the main complex trait. The idea is that certain loci might contribute intermediate phenotype to disease through low frequency functional alleles which are not captured on classical SNP arrays. Locus containing Death associated protein (DAP) gene was implicated in susceptibility to autoimmune diseases in past but was not investigated further. We hypothesize that DAP contribute intermediate phenotype to SLE through low frequency alleles and if so, targeted sequencing of the locus could reveal those causal variants. We performed deep targeted sequencing across entire DAP locus in 1000 SLE patients and identified a regulatory haplotype that pose strong [OR=1.4, p=4.82E-03] risk for SLE. RNA sequencing in healthy individuals carrying risk haplotype demonstrated significant downregulation of DAP expression before and after TLR stimulation. Multiple *cis* eQTLs were found embedded in the risk haplotype with most notable effect observed with C allele of rs2930047 SNP showing significant (p<0.01) reduction in DAP expression in multiple cell lineages. Furthermore, we show that SLE risk haplotype demonstrate high titers of antinuclear antibodies (ANA), especially, anti-sm IgGs. Transcriptome analysis in our eQTL panel and 1000 genome study sample showed upregulated expression of TNF in carriers of DAP risk genotype. Taken together, present study discovers a novel SLE risk haplotype that downregulate the expression of DAP gene, a well known negative regulator of autophagy. We propose that reduced DAP expression could potentially impair autophagy and apoptotic processes leading to dysregulated cellular homeostasis and development of autoimmunity.

1932F

Cross-genetic heritability of maternal and neonatal immune mediators during pregnancy. *M. Traglia[1], L.A. Croen[2], K.L. Jones[3,4], L.S. Heuer[3,4], C.K. Yoshida[4], R. Yolken[5], M. Kharrazi[6], G.N. DeLorenze[2], P. Ashwood[4,7], J. Van de Water[3,4], L.A. Weiss[1].* 1) Department of Psychiatry and Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 2) Divison of Research, Kaiser Permanente Northern California, Oakland, CA; 3) Department of Internal Medicine, Division of Rheumatology, Allergy, and Clinical Immunology, University of California, Davis,CA; 4) MIND Institute, University of California, Davis,CA; 5) Stanley Division of Developmental Neurovirology, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Division of Environmental and Occupational Disease Control, California Department of Public Health, 850 Marina Bay Pkwy, Bldg. P, Richmond, CA; 7) Department of Medical Microbiology and Immunology, University of California, Davis,CA.

Increasing evidence shows that during pregnancy, fetal genetics contributes to different aspects of maternal physiology, such as blood pressure, gestational diabetes, metabolism, and preeclampsia. We previously showed that maternal circulating toxicant levels in pregnancy are regulated in part by fetal genetic variation (Traglia et al., 2017). Here, we aim to clarify whether maternal and fetal genetics independently contribute to neonatal and maternal immune function.   We measured circulating levels of 22 mid-gestational maternal and 42 neonatal soluble immune mediators (SIMs) in EMA (Early Markers of Autism), a multi-ethnic population-based nested case-control study of autism including 790 genotyped women and 764 of their genotyped newborns (Croen, Autism Res 2008; Tsang, PLoS ONE 2013). We identified 5 neonatal SIMs with significant maternal heritability and 2 maternal SIMs with significant SNP-based fetal heritability. Because mother-offspring pairs share alleles via inheritance, we need to distinguish whether both maternal and fetal genomes contribute independently or only one individual's genome exerts influence (and the other is estimated to contribute to heritability due to sharing). To differentiate between the contributions of the maternal and fetal-genome, we identified the fetal alleles not inherited from the maternal lineage (fetal-specific alleles) and the maternal alleles not transmitted to the offspring (maternal-specific alleles). Using these derived datasets, we observed that most of the SIMs with significant cross-generational heritability estimates are independently influenced by the non-measured individual. For the same SIMs, we found low (albeit significant) correlation between the maternal-specific and fetal-specific SNP-based BLUP (best linear unbiased prediction) values ($\rho <$ 12%), suggesting that largely non-overlapping sets of maternal SNPs and fetal SNPs contribute to the heritability of the same SIM, consistent with our previous GWAS-identified loci (Traglia et al, ASHG 2016). Our results suggest that during pregnancy and immediately after birth, maternal and fetal genetics strongly mediate the variability of immune system status via distinct mechanisms, and that better understanding of the cross-generational regulation of maternal and neonatal phenotypes may shed light on healthy pregnancy and development of the immune system.

1933W

Association analysis of rheumatoid arthritis through whole-exome sequencing in a Singapore Chinese cohort. *V. Kumar[1], E.T. Koh[2], P. Cheung[3], M. Lahari[3], C.T. Ng[4], M.Y. Yong[2], C.M. Woo[2], K.P. Leong[2], S. Davila[1,5].* 1) Cancer and Stem Cell Biology, DUKE-NUS Medical School, SINGAPORE, Singapore; 2) Department of Rheumatology, Tan Tock Seng Hospital, Singapore; 3) Department of Rheumatology, National University Health System, Singapore; 4) Department of Rheumatology, Singapore General Hospital, Singapore; 5) SingHealth Duke-NUS Institute of Precision Medicine, Singapore.

Rheumatoid arthritis (RA) is a chronic inflammatory joint disease characterized by progressive bone and cartilage destruction. The contribution of genetic and environmental factors to RA pathogenesis is evident but their complex interaction remains elusive. About 100 genes have been found to be associated with RA. We aim to verify the known genes and to discover new associations in a Singaporean Chinese population. For this, we have done whole-exome sequencing in 600 Singapore Chinese patients with RA and 900 controls. Logistic regression and gene-based association analysis is currently underway. Besides investigating the genes that increase the risk of developing RA, we determine those that predict bad outcomes or poor response to medications.

1934T

Assessing the mechanisms of thymic involution in an animal model of multiple sclerosis. *S.G. Gregory[1,2], S.N. Giamberardino[1], S.K. Siecinski[1], E.A. Hocke[1], K. Abramson[1], S.F. Arvai[1].* 1) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA; 2) Discovery MS, David H Murdock Research Institute, Kannapolis, NC, USA.

   Purpose: The thymus is a primary lymphoid organ that plays a pivotal role in thymocyte differentiation and maturation into functional T-cells. The selection of the self-tolerant T-cell antigen receptor repertoire is important for establishing immune homeostasis and preventing autoimmunity. Interestingly, the thymus is a dynamic organ whose activities are influenced by pathophysiological challenge. These events can cause thymic involution that may affect its ability to appropriately mature T cells into self-tolerant mediators of immune system response. The purpose of this project was to assess the impact of an animal model of multiple sclerosis (MS), experimental autoimmune encephalomyelitis (EAE), on thymus cellularity and gene expression. **Methods:** EAE was induced in C57BL/6J wild type mice via injection of MOG_{35-55} and Complete Freund's Adjuvant emulsion and severity scored according to ascending paralysis clinical scores (CS) '0' asymptomatic to '5' moribund. All analyses were performed in thymi from naïve and peak EAE (CS 3.0-3.5) mice. Immune histochemical staining was performed using hematoxylin and eosin, anti-B220 antibody, or anti-cytokeratin antibody; flow sorting involved thymic disaggregation and CD45- depletion with CD45+ cells stained antibodies against CD3ε, CD19, CD4 and CD8b. CD4+, CD8+, or CD4CD8 double positive T cells were visualized from the CD3ε+ T cell gate. The same disaggregation approach was used on the thymi to generate Single Cell 3' libraries on the 10X Genomics Chromium Controller, prior to analysis using standard parameters in Seurat single cell analysis software. **Results:** We have established that the thymus severely atrophies during EAE due to significant loss of cellularity and disruption of its structure, with loss of CD4CD8 double positive T cells and, unexpectedly, increasing CD19+ B-cell numbers. In the naïve state, our single cell gene expression data include discrete epithelial, thymocyte, CD4+, and maturing CD8+ cells populations, while in the involuted thymus identified replicating thymic epithelial, thymic stromal cells, as well as T-cells and the B-cells observed in our IHC staining. **Conclusions:** We have generated the first single cell gene expression profile of an EAE induced involuted thymus in a mouse model of MS. The profile provides a glimpse into normal thymus function and during physiological stress, and identifies novel pathways that might be important in immune cell dysfunction in the development of MS.

1935F

Sex-specific transcriptional responses to lipopolysaccharide (LPS) in peripheral blood leukocytes (PBLs) in the Hutterite founder population. *M. Stein, K. Naughton, C. Billstrand, R. Nicolae, C. Ober.* Department of Human Genetics, University of Chicago, Chicago, IL.

   Many immune-mediated diseases show sex differences in prevalence, but the underlying genetic contributors to this phenomenon are still largely unknown. To address this question and characterize sex-specific differences in innate immune response, we studied the transcriptional response of PBLs to LPS in 47 male and 66 female members of the Hutterite community. Whole blood was incubated in a closed system (TruCulture®, Myriad RBM) for 30 hours with either media alone or media + 0.1µg/ml LPS, a component of the gram-negative bacterial cell wall. Genome-wide gene expression levels were measured by RNAseq. Genes were mapped to hg19 using STAR and normalized using voom, resulting in 10,425 autosomal genes detected as expressed. Using a linear model with age included as a covariate, and individuals as a random effect, we tested for differential gene expression between LPS-treated and untreated samples in males, females, and for differences between sexes. 6,933 (66.5%) genes were differentially expressed (DE) in response to LPS treatment in both males and females (FDR 5%). 1,479 DE genes significantly differed in magnitude and/or direction between males and females (FDR 5%). Notably, 89.4% (1,323) of these genes had a greater magnitude of response in males, including *IRF1* ($P_{ResponseDiff}=2.73 \times 10^{-7}$), a gene located 2.8kb downstream of rs2549003, which was previously associated with sex-specific asthma risk in European American males. In this gene set, pathway analysis of genes with a greater response in males identified enrichments for downstream activators IFNγ, IRF7, and STAT1. Additionally, 340 DE genes in males only included *MYD88*, a critical mediator in innate immune signaling ($P_{males}= 2.56 \times 10^{-6}$, $P_{females}=0.33$). In contrast, 194 DE genes in females only included an enrichment for genes downstream of IL-5, a type 2 T cell cytokine. These data suggest that there is a more robust response to LPS at 30 hours in males compared to females, particularly in IFN-mediated pathways. These results are consistent with previous reports of *IFNG* genotype-sex interactions in asthma risk, and highlight the potential link between sex-specific differences in immune response and risk for asthma and potentially other immune-mediated diseases with sex differences in prevalence. This work was funded by R01 HL085197.

1936W

Role of rare variants in progression form latent to active tuberculosis in Peruvian population. *S. Asgari[1], Y. Luo[1], R. Calderon[2], L. Lecca[2], S. Leon[2,3], J. Jimenez[2], M. Murray[3], S. Raychaudhuri[1].* 1) Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 2) Socios En Salud, LIma, Peru; 3) Department of Global Health and Social Medicine, Harvard Medical School, Boston, Massachusetts, USA.

   Background: Common variants (> 0.01 minor allele frequency (MAF)) have been associated with susceptibility to pulmonary tuberculosis (TB). Yet common variants explain only a portion of the total trait variance. The impact of rare coding variation on susceptibility to pulmonary TB has not been as thoroughly investigated. Using a costum-designed genotyping chip, here we utilize 1668 cases and 1326 matching controls to look for rare, functional variants that affect susceptibility to pulmonary TB in Peruvian population.   Methods: To discover rare variants we applied exome sequencing to 116 Peruvian TB cases. We conjectured that rare risk variants in Peruvian populations would be enriched among these cases. We then used the rare variant content from these samples, to design a genotyping array that included 712K markers including 320 markers derived from exome sequencing of 116 Peruvian TB cases. We used this custom array to then genotype 1668 Peruvians who developed active TB and 1326 matching controls from the same household who were positive for tuberculin skin test but did not develop active TB. We used Plink 1.9 for quality control. 8 cases and 4 controls were removed due to missingness > 0.05 or Hardy-Weinberg equilibrium p-value < 1e-5. Of the remaining individuals, 369 cases and 455 controls had identity by descent (IBD) > 0.2. Weighted sum method (Madesn and Browning, 2009) was used to collapse variants with MAF < 0.01 within each gene. Genetic relatedness matrix was calculated using GEMMA v0.94.1. We tested gene burden association, using GMMAT v0.7-1, by fitting a linear mixed model to the data with first 10 principal components and age as covariates.   Results: Analyses aimed at detecting the burden of rare variants within a gene showed no significant associations. Restricting this analysis to only variants that modify protein sequence did not reveal further signals. Possibly larger sample sizes are required to fully explore the role of rare coding variants in susceptibility to pulmonary TB.

1937T

Deciphering genetic susceptibility to tuberculous meningitis: Exome sequencing and a GWAS in a South African population. *M. Möller, N. Bowker, M. Salie, H. Schurz, P.D. van Helden, E.G. Hoal, C.J. Kinnear.* 1SA MRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical TB Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University.

   Tuberculous meningitis (TBM) is a type of extrapulmonary tuberculosis (TB) which leads to inflammation of the meninges through small lesions called Rich foci. TBM represents 1% of total TB disease, with the age of onset being around 2-5 years of age. Disease development remains poorly understood, including the mechanism of dissemination across the blood-brain barrier. This study concentrated on the involvement of the host genome in TBM susceptibility. We hypothesised that multiple common variants of moderate effect size are more likely to influence TBM susceptibility than rare variants of large effect. Ten TBM cases and 10 healthy controls were exome sequenced. A total of 123 TBM cases, 400 pulmonary TB (pTB) cases and 477 healthy controls were genotyped on the Illumina® Multi-Ethnic Genotyping Array (MEGA). Gene set association tests SKAT-O and SKAT Common Rare were used to assess the association of rare SNPs and the cumulative effect of both common and rare SNPs with susceptibility to TBM, respectively. The SKAT-O analysis included 8 322 gene sets comprising 16 728 SNPs, which did not yield any associations with TBM susceptibility. Ingenuity Pathway Analysis (IPA) of the top-hits of the SKAT-O analysis showed that two genes are both important in TBM pathogenesis and highlighted these as targets for future study. The SKAT Common Rare analysis included 13 270 gene sets comprised of 53 239 SNPs and one gene was associated with TBM susceptibility. In addition, a number of top-hit genes ascribed to the development of the central nervous system (CNS) and innate immune system regulation were highlighted. The role of common variants (>5%) in TBM susceptibility were assessed by conducting a genome-wide association study (GWAS). A GWAS comparing 114 TBM cases to 395 healthy controls showed no association with TBM susceptibility. A second analysis comparing 114 TBM cases to 382 pTB cases was conducted to investigate variants associated with different TB phenotypes. No significant associations were found with progression from pTB to TBM. This study reports the first exome sequencing and GWAS of a TBM cohort and has identified a single previously undescribed association with TBM susceptibility. These results further our understanding of TBM in terms of both SNPs and genes that influence susceptibility. In addition, a number of candidate genes involved in innate immunity have been identified using IPA for further genotypic and functional investigation.

1938F

Genome wide association in Peru demonstrates that progression to active tuberculosis is a polygenic and highly heritable trait. *S. Raychaudhuri[1,2], R. Calderon[3], L. Lecca[3], S. Leon[3], J. Jimenez[3], B.D. Moody[8], M. Murray[3,4,5,6,7], Y. Luo[1,2].* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Socios En Salud, Lima, Peru; 4) Partners In Health, Boston, MA, USA; 5) Division of Global Health Equity, Brigham and Women's Hospital, Boston, MA, USA; 6) Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, USA; 7) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 7) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 8) Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.

Since only ~10% of the 1.5 billion *Mycobacterium tuberculosis* (*Mtb*) infected individuals worldwide go on to develop active tuberculosis (TB), host genetic factors that control disease progression have emerged as a central question in TB pathogenesis. To dissect the genetic basis of TB progression, we conducted a genome wide association study (GWAS), comparing 1,676 primary pulmonary TB patients versus 1,330 individuals who did not develop TB despite *Mtb* exposure through a household contact. Contrasts with previous studies, here we explicitly control for *Mtb* exposure and infection status, thereby allowing us to investigate whether progression to active TB is indeed highly heritable. We first designed a customized array based on whole-exome sequencing data from 116 active TB cases to optimize the capture of genetic variations in Peruvians. This allowed detection of rare and low-frequency population-specific coding variants and those which predisposed to TB risk. We then genotyped 3,006 individuals at 712K genomic markers and imputed ~8 million variants using the 1000 Genomes Project Phase 3 reference panel. We used linear mixed models to test for single marker associations, which account for population stratification and relatedness in collected individuals. Our study identified a potentially new TB progression locus on chromosome 3q23, where 8 variants located in the downstream of the RASA2 gene had P-value<10^{-7}. We also saw association in a previously reported TB locus at HLA-DQA1 (P=1.6×10^{-7}, OR=1.27), but did not detect signals at other reported risk loci. This outcome suggests that previously reported loci may not be specific to TB progression, whereas HLA-association may affect both TB progression and *Mtb* infection. Next, we estimated genetic heritability of TB progression on the liability scale (h^2_g) using LD score regression (Bulik-Sullivan et al 2015). In total, we report 14.5% (s.e. 0.02) variation can be explained for TB progression. . One possible reason for the limited success of prior GWASes might be due to low heritability of traits pursued. Our study is the first to quantify h^2_g of TB progression, and indicates a strong genetic basis. This degree of heritability is comparable to other traits in which GWAS have been highly successful as sample size mounted. For example, estimated h^2_g for Crohn's disease in the WTCCC data is 22.0% (0.02). We are now undergoing a second phase of data collection, where this new risk locus will be further validated.

1939W

Genetic variation in *GLS2* is associated with development of complicated *Staphylococcus aureus* bacteremia. *W.K. Scott[1], F. Ruffin[2], B.S. Kuinkel[2], D.D. Cyr[3], S. Guo[1], D.M. Dykxhoorn[1], R.L. Skov[4], N.E. Bruun[5], A. Dahl[5], C.J. Lerche[6], V.G. Fowler Jr.[2], P.S. Andersen[4], DANSAB Study Group.* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Medicine, Duke University Medical Center, Durham, NC; 3) Duke Clinical Research Institute, Duke University Medical Center, Durham, NC; 4) Statens Serum Institut, Copenhagen, Denmark; 5) Gentofte University Hospital, Copenhagen, Denmark; 6) Rigshospitalet, Copenhagen, Denmark.

Staphylococcus aureus bacteremia (SAB) is a serious, common infection. Recent genome wide association studies (GWAS) have associated *S. aureus* infections with common variation in the HLA class II region. Rare variants are not well covered by GWAS and may be significant contributors to SAB risk or severity. We used whole exome sequencing (WES) to examine the cumulative effect of coding variants in each gene on risk of complicated SAB in a discovery set of patients, and then evaluated the nominally significant genes in a replication set of patients using custom-capture sequencing. The discovery set comprised 84 complicated SAB cases (endocarditis or bone/joint infection) frequency-matched by age (in deciles), sex, and bacterial clonal complex (CC5 or CC30 vs. CC8) to 84 uncomplicated SAB controls. All were white inpatients at Duke University. WES utilized Agilent SureSelect 72Mb capture kits, followed by sequencing on an Illumina HiSeq2000, alignment and base calling with a standard pipeline. The SKAT-O and EPACTS packages were used for gene-based association tests and logistic regression models with Firth bias correction, respectively. Both controlled for age, sex, and clonal complex as covariates. The replication set of 122 complicated SAB cases and 118 uncomplicated SAB controls was frequency matched by age, sex, and clonal complex. All were white Europeans collected by the Statens Serum Institut. An Agilent SureSelect 2Mb capture array captured genic sequence for 342 genes nominally associated with complicated SAB in discovery (SKAT-O p<0.035). Sequencing and data analysis proceeded as for WES. A Bonferroni-corrected gene-based test p-value of 1.5×10^{-4} determined significance in the replication set. One gene, *GLS2*, was significantly associated with complicated SAB in the replication set (p=1.2×10^{-4}). The strongest single-variant association in all 342 genes was rs2657878 in *GLS2* (p=5×10^{-4}). This variant is strongly correlated with a missense variant (rs2657879, p=4.4×10^{-3}) in which the minor allele (associated here with complicated SAB) has previously been shown to reduce circulating glutamine levels. Comprehensive examination of the coding sequence for association with complicated SAB in a two-stage discovery/replication design identified a novel candidate gene. *GLS2* is an interesting candidate for complicated SAB due to its role in regulating glutamine production, a key factor in activation of T-cell production.

1940T

High density imputation genome wide association study of spontaneous resolution of hepatitis C virus. *C.I. Vergara[1], C. Thio[1], R. Latanich[1], M.A. Taub[2], G.D. Kirk[2], S.H. Mehta[2], M. Busch[3], A.Y. Kim[4], G. Lauer[4], E. Johnson[5], A.H. Karl[5], M.G. Peters[6], M. Kuniholm[7], G.L. Wojcik[8], J.J. Goedert[9], A. Mangia[10], S.I. Khakoo[11], L. Alric[12], M.E. Cramp[13], S.M. Donfield[14], B.R. Edlin[15], G. Alexander[16], D.L. Thomas[1], P. Duggal[2].* 1) Department of Medicine, Johns Hopkins University, Baltimore, MD,USA; 2) Bloomberg School of Public Health,Johns Hopkins University, Baltimore, MD,USA; 3) University of California, San Francisco, CA,USA; 4) Massachusetts General Hospital and Harvard Medical School, Boston, MA,USA; 5) RTI International, Research Triangle Park, NC, Atlanta, GA, San Francisco, CA, USA; 6) Blood Systems Research Institute, San Francisco, CA,USA; 7) Department of Epidemiology and Biostatistics, University at Albany, State University of New York, Rensselaer, NY,USA; 8) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 9) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; 10) Liver Unit, IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Italy; 11) University of Southampton, Southampton General Hospital, Southampton, UK; 12) Toulouse III University, Toulouse, France; 13) South West Liver Unit, Plymouth, PL6 8DH, UK; 14) Rho, Inc., Chapel Hill, NC, USA; 15) SUNY Downstate College of Medicine, Brooklyn, NY, USA; 16) Department of Medicine, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.

The outcome of acute hepatitis C virus (HCV) infection is linked to variants in the MHC region and 2 variants in high LD in *IFNL4*. However, these two *IFNL4* variants do not fully explain the HCV outcomes and there is controversy about whether the *IFNL4* variants are associated with HCV outcome across ancestry groups. To address these issues, we performed a dense GWAS of spontaneous clearance of HCV compared to individuals who have persistent infection, in 4091 people in 3 different ancestry groups (1869 African Americans (AA, 340 clearance/1529 persistence), 1739 European Americans (EA, 701 clearance/1036 persistence) and 486 Hispanics (173 clearance/313 persistence). Samples were genotyped using the Illumina Human Omni-Quad and imputed to 1000genomes Phase 3 using Minimac3. Mach2dat was used to test the association of imputed SNP dosage with HCV clearance under a log-additive model. HIV infection status and 20 PCs were included as covariates, separately for each ancestry group and meta-analyzed. Significant associations were confirmed at the 2 known loci (*IFNL4,19q13.2* and *MHC,6p21.32*) for all 3 groups. On chromosome 6, the top SNPs were located near genes for class II MHC and included rs2647006 (A/C) (Meta-Pvalue: 5.15×10^{-20}). The most significant association on chromosome 19 was rs74597329 (Imputed; G/T; OR: 0.47, PValue: 9.8×10^{-16} for AA, OR: 0.44, PValue: 1.3×10^{-22} for EA, OR: 0.30, PValue: 2.84×10^{-12} for Hispanics, Meta-Pvalue: 8.17×10^{-46}). The SNP codes for change of Ala□Glu in position 22 of the protein but in conjunction with the -/T polymorphism (rs11322783) conforms the ΔG/TT polymorphism (rs368234815). The LD between previously identified *IFNL4* variants, rs12979860 and rs74597329 is high ($r_2 > 0.87$ for all 3 populations), and conditioning on either SNP removed significance of the other SNP. Interestingly, rs4803221 another top SNP not in high LD with the previous ones, had attenuated significance but was still present (PValue: 6.2×10^{-6} , 1.6×10^{-5}) in EA after conditioning on either of the first 2 SNPs. The associations on chromosomes 19 and 6 were independently associated with HCV clearance and additive. The *IFNL4* finding was confirmed in EA and observed for the first time with GWAS significance for AA and Hispanics. This study finds that in EA, rs4803221 is independently associated with HCV clearance at *IFNL4*; however in AA, both rs74597329 and rs12979860 best explain the association of HCV clearance at this locus.

1941F

Whole genome sequencing of pharmacogenetic drug response in racially and ethnically diverse children with asthma. *A.C.Y. Mak[1], M.J. White[1], Z.A. Szpiech[2], W.L. Eckalbar[1], S.S. Oh[1], M. Pino-Yanes[3,4], D. Hu[1], S. Huntsman[1], J. Galanter[1], D.G. Torgerson[1], B.E. Himes[5], S. Germer[6], K.L. Bunting[6], C. Eng[1], S. Salazar[1], K.L. Keys[1], T.A. Nguyen[1], P.-Y. Kwok[7,8], N. Ahituv[1], E. Ziv[1], M.A. Seibold[9], R.B. Darnell[6,10,11], N. Zaitlen[1], R.D. Hernandez[2,8,12], E.G. Burchard[1], The Trans-Omics for Precision Medicine Whole Genome Sequencing Program (TOPMed) Team.* 1) Department of Medicine, University of California San Francisco, San Francisco, California, USA; 2) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, USA; 3) Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; 4) CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; 5) Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 6) New York Genome Center, New York, New York, USA; 7) Cardiovascular Research Institute, University of California San Francisco, San Francisco, California, USA; 8) Institute for Human Genetics, University of California San Francisco, San Francisco, California, USA; 9) Center for Genes, Environment and Health, Department of Pediatrics, National Jewish Health, Denver, Colorado, USA; 10) Laboratory of Molecular Neuro-Oncology, The Rockefeller University, New York, New York, USA; 11) Howard Hughes Medical Institute, The Rockefeller University, New York, New York, USA; 12) Quantitative Biosciences Institute, University of California San Francisco, San Francisco, California, USA.

Introduction Albuterol, a bronchodilator medication, has been the standard of care for asthma treatment worldwide for the past three decades. Despite significant variation in bronchodilator response (BDR), albuterol is often the only medication prescribed for asthma, regardless of disease severity and the striking racial/ethnic disparity in BDR. Specifically, Puerto Rican and African American children with asthma have significantly lower BDR compared to Mexican American children. Previous studies have identified a genetic contribution to variation in BDR, yet much of the heritability of BDR remains to be explained. We hypothesize that rare, population-specific variants contribute to BDR in minority children with asthma. **Methods** We performed whole genome sequencing on 1,441 minority children with asthma from the tails of the BDR distribution. Our study included high and low drug responders from three racial/ethnic groups, African Americans (n=475), Puerto Ricans (n=483), and Mexican Americans (n=483). We performed logistic regression to identify associations with individual variants and BDR in each racial/ethnic group separately, and combined in a meta-analysis. We also performed association tests at pooled variants using SKAT-O in order to test for associations with both common and rare variants. All association models were adjusted for age, sex, BMI, and the top ten principal components. **Results and Conclusions** We identified population-specific and shared pharmacogenetic variants associated with BDR, including genome-wide significant and suggestive loci near genes previously associated with lung capacity (*DNAH5*), immunity (*NFKB1* and *PLCB1*), and β-adrenergic signaling pathways (*ADAMTS3* and *COX18*). Our study expands the understanding of pharmacogenetic analyses in racially and ethnically diverse populations and advances the foundation for precision medicine in at-risk and understudied minority populations.

1942W

ITPKC and SLC11A1 gene variations are associated with Kawasaki disease patients. *Y. Bae[1], K. Kim[2], W. Ji[3], D. Shin[1], H. Kim[2], D. Kim[2], K. Kim.* 1) Genome Research Center, Medizen Humancare, Seoul, South Korea; 2) Department of Pediatrics, Yonsei University College of Medicine, Severance Children's Hospital; 3) School of Medicine, Kyung Hee University, Seoul, Korea.

Kawasaki disease (KD) is an acute systemic vasculitis. Both the etiology of KD and erythema of Bacille Calmette-Guérin (BCG) injection site observed in the disease are poorly understood. The present study investigated the association between KD and single nucleotide polymorphisms (SNPs) in two candidate genes: inositol 1, 4, 5-triphosphate 3-kinase (*ITPKC*), one of the well-studied KD-associated genes, and solute carrier 11a1 (*SLC11A1*), which is associated with the hypersensitive reaction to the BCG strain in Koreans. Potential associations between BCG injection site erythema and SNPs in the *ITPKC* and *SLC11A1* genes were also evaluated. Gene-gene interactions between *ITPKC* and *SLC11A1* in KD and BCG injection site erythema were also analyzed. Three tagging SNPs in *ITPKC* and five tagging SNPs in *SLC11A1* were genotyped in 299 KD patients and 210 control children. SNP rs28493229 in *ITPKC* was associated with KD and coronary artery complications. SNP rs77624405 in *SLC11A1* was associated with KD. Comparisons of KD patients with and without BCG injection site erythema revealed that SNP rs17235409 in *SLC11A1* was associated with erythema, but no erythema-associated SNPs in *ITPKC* were identified. Interactions between *ITPKC* rs28493229_GG and *SLC11A1* rs17235409_GA and between *ITPKC* rs10420685_GG and *SLC11A*1 rs17235409_AA were strongly associated with BCG injection site erythema. In conclusion, this study identified several important variations in the *ITPKC* and *SLC11A1* genes in Koreans.

1943T

Location, location, location: Single cell gene expression of mucosal T cells vs peripheral blood T cells in Crohn's disease. *E.A.M. Festen[1,2], W.T.C. Uniken Venema[1], M.D. Voskuil[1], R. Alberts[1], A. Vich Vila[1], R.K. Weersma[1].* 1) University of Groningen and University Medical Center Groningen, Department of Gastroenterology and Hepatology, Groningen, The Netherlands; 2) University of Groningen and University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands.

Introduction Crohn's disease (CD) is a chronic inflammatory disease, that predominantly causes inflammation in the terminal ileum[1]. Genetic research has identified almost 200 genetic loci associated with CD, many of which are in genes involved in T cell pathways[2,3]. Furthermore, recent research has shown the importance of mucosa-specific T-cells[4]. Although studies have tried to characterize average expression of mucosal T cells using microarray expression data[5], subtype-specific differences have not been researched. **Aims & Methods** Our aim is to asses which T cell subpopulations are present in the intestinal mucosa of CD patients, and how their transcriptome differs from peripheral blood T cells. We performed single cell RNA sequencing (scRNAseq) of 5292 CD3+ T cells isolated from peripheral blood and from ileal mucosal biopsies of CD patients (paired). Isolation of mucosal T cells was performed with a protocol from Raine e.a.[5]. ScRNAseq was performed with a locally developed, plate based method, with 3' library generation and unique molecular identifiers, sequenced with Illumina HiSeq2500. **Results** Findings from previously published microarray expression data were replicated[5], identifying relative overexpression of *CXCR4*, *CD55*, *DUSP2*, *CD160*, *FASLG*, *ITGAE* and *TNFAIP3* in mucosal vs blood T cells. The relative overexpression of *TNFAIP3*, which promotes survival of CD4+ T-cells, is especially interesting, since variants near this gene are associated with CD and with response to anti-TNFa therapy[6,7]. Strikingly, a subgroup of mucosal CD8+ T-cells shows relative overexpression of three genes that are associated with CD: *TNFAIP3*, *FOS*, which promotes T-cell proliferation[2,8], and *PTGER4*, which encodes the prostaglandin receptor EP4 that promotes the immune response through Th1 and Th17 activation[6]. **Conclusion** We show that CD associated genes are significantly overexpressed in ileal mucosal T cells, mainly in CD8+ cells. The characterization of expression patterns in individual cells is promising for development of cell-specific therapies in CD patients. **References** 1. Baumgart *et al.*(*Lancet,* 2012). 2. Liu *et al.* (*Nat. Genet.,* 2015). 3. Farh *et al.* (*Nature,* 2015) 4. Cauley *et al.* (*Mucosal Immunol.,* 2013) 5. Raine *et al. (Gut,* 2014) 6. Jostins *et al.* (*Nature,* 2012) 7. Bank *et al.* (*Pharmacogenomics J.,* 2014) 8. Matsuzawa *et al.* (*Autophagy,* 2015).

1944F

Comprehensive bioinformatic characterization around RASGRP1 gene identifies multiple potential functional variants for lupus susceptibility.
J.E. Molineros[1], C. Sun[1], B. Singh[1], K.M. Kaufmann[2,3], J.B. Harley[2,3], S.K. Nath[1]. 1) Oklahoma Medical Research Foundation, Division of Arthritis and Clinical Immunology, Oklahoma City, OK; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229; 3) US Department of Veterans Affairs Medical Center, Cincinnati, OH, 45229.

Systemic Lupus Erythematosus (SLE) is a complex autoimmune disease with substantial genetic component. SLE disproportionately affects women and ethnic minorities. We recently identified a novel SLE susceptibility signal near Ras Guanyl releasing Protein1 (*RASGRP1*) in Asians [Sun et al, 2016, Nat Genet]. *RASGRP1* is a key signaling molecule, preferentially expressed in T-cells, and its deficiency in mice results in SLE due to improper splicing of the *RASGRP1* precursor transcript. However, the "true" functional variants and the genetic mechanisms by which the associated variants manifest SLE pathogenesis is largely unknown. The objective is to identify SLE functional variant(s) across multiple ethnically diverse populations that might influence *RASGRP1* expression. We conducted a comprehensive imputation-based (3347 SNPs) trans-ethnic meta-analysis from 6 cohorts (>9000 cases; 19000 controls) with Asian, European and African ancestries followed by additional genotyping, and selected 163 SNPs ($P<5x10^{-3}$) for further evaluation. We then assessed co-localization of peak SNPs with active histone marks (H3K27ac, H3K4me1, H3K3me3) and DNAse-I hypersensitivity sites across multiple tissues from ENCODE to further narrow the number of predicted functional SNPs associated with cis-regulatory modules. We further prioritized 91 SNPs ($P<5x10^{-4}$) using a reported Bayesian refinement method that combines functional information at each SNP and accounting for ethnic-specific linkage disequilibrium (LD). To assess possible regulatory roles of the candidate SNPs, we correlated targeted SNP genotypes with multiple tissue-specific eQTLs from Blood eQTL, GTEx, GEUVADIS. Several SNPs are identified as eQTLs for *RASGRP1*, and the strongest eQTL was rs7173565 ($P<10^{-19}$). Finally, we assessed enhancer-promoter interactions via capture Hi-C data from multiple cell types. In summary, using comprehensive bioinformatics, we identified 3 distinct signals ($P<5x10^{-8}$) from 3 LD-blocks containing several functional SNPs. While the first signal is intronic, the other 2 signals are intergenic. The rs7173565 was the most promising SNP identified as an eQTL and located in a super-enhancer region in CD4 and CD8 cells. Two signals were linked to enhancer-promoter interactions, as assessed by capture Hi-C. These results suggest that the susceptibility variants modulate *RASGRP1* expression through promoter-enhancer interaction, and its dysregulation and aberrant expression could increase the risk for SLE.

1945W

Male-specific association of the *FCGR2A* His167Arg polymorphism with Kawasaki disease.
Y. Kwon[1], J. Kim[1], S. Yun[2], J. Yu[3], K. Yoon[4], K. Lee[5], H. Kil[6], G. Kim[7], M. Han[8], M. Song[9], H. Lee[10], K. Ha[11], S. Sohn[12], R. Ebata[13], H. Hamada[14], H. Suzuki[15], K. Ito[16], Y. Onouchi[16,17], Y. Hong[12], G. Jang[11], J. Lee[1]. 1) Asan Institute for Life Sciences, Seuol, , South Korea; 2) Department of Pediatrics, Chung-Ang University Hospital, Seoul, Korea; 3) Department of Pediatrics, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea; 4) Department of Pediatrics, Kyung Hee University Hospital at Gangdong, Seoul, Korea; 5) Department of Pediatrics, The Catholic University of Korea, Daejeon St. Mary's Hospital, Daejeon, Korea; 6) Department of Pediatrics, Chungnam National University Hospital, Daejeon, Korea; 7) Department of Pediatrics, Seoul National University Children's Hospital, Seoul, Korea; 8) Department of Pediatrics, University of Ulsan, Gangneung Asan Hospital, Gangneung, Korea; 9) Department of Pediatrics, Inje University Paik Hospital, Busan, Korea; 10) Department of Pediatrics, Pusan National University Hospital, Busan, Korea; 11) Department of Pediatrics, Korea University Hospital, Seoul, Korea; 12) Department of Pediatrics, Ewha Womans University Hospital, Seoul, Korea; 13) Department of Pediatrics, Chiba-University Graduate School of Medicine; 14) Department of Pediatrics, Tokyo Women's Medical University Yachiyo Medical Center; 15) Department of Pediatrics, Wakayama Medical University; 16) Laboratory for Cardiovascular diseases, RIKEN Center for Integrative Medical Sciences; 17) Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan.

Kawasaki disease (KD) is an acute systemic vasculitis that can potentially cause coronary artery aneurysms in some children. KD occurs approximately 1.5 times more frequently in males than in females. To identify sex-specific genetic variants that are involved in KD pathogenesis in children, we performed a sex-stratified genome-wide association study (GWAS), using the Illumina HumanOmni1-Quad BeadChip data (249 cases and 1,000 controls) and a replication study for the 34 sex-specific candidate SNPs in an independent sample set (671 cases and 3,553 controls). Male-specific associations were detected in three common variants: rs1801274 in *FCGR2A* [odds ratio (OR) = 1.40, $P = 9.31 \times 10^{-5}$], rs12516652 in *SEMA6A* (OR = 1.87, $P = 3.12 \times 10^{-4}$), and rs5771303 near *IL17REL* (OR = 1.57, $P = 2.53 \times 10^{-5}$). The male-specific association of *FCGR2A*, but not *SEMA6A* and *IL17REL*, was also replicated in a Japanese population (OR = 1.74, $P = 1.04 \times 10^{-4}$ in males vs. OR = 1.22, $P = 0.191$ in females). In a meta-analysis with 1,461 cases and 5,302 controls, a very strong association of KD with the nonsynonymous SNP rs1801274 (p.His167Arg, previously assigned as p.His131Arg) in *FCGR2A* was confirmed in males (OR = 1.48, $P = 1.43 \times 10^{-7}$), but not in the females (OR = 1.17, $P = 0.055$). The present study demonstrates that p.His167Arg, a KD-associated *FCGR2A* variant, acts as a susceptibility gene in males only. Overall, the gender differences associated with *FCGR2A* in KD provide a new insight into KD susceptibility.

1946T

Local ancestry interaction models reveal a novel asthma association with asthma on chromosome 1q23.1 specific to people of African ancestry. *M.P. Boorgula[1], M. Daya[1], N.M. Rafaels[1], S. Chavan[1], T.M. Brunetti[1], A. Shetty[1], C.R. Gignoux[1], G. Wojcik[2], R. Johnston[3], E.G. Burchard[4], L. Caraballo[5], G.M. Dunston[6], M.U. Faruque[7], E.E. Kenny[8], J.M. Knight-Madden[9], D.A. Meyers[10], E.R. Bleecker[10], C. Ober[11], C. Rotimi[12], J.G. Wilson[13], L.K. Williams[14], H. Watson[15], Z.S. Qin[3], M.A. Taub[16], T.H. Beaty[17], I. Ruczinski[16], R.A. Mathias[17], K.C. Barnes[1], CAAPA Consortium.* 1) Colorado Center for Personalized Medicine, University of Colorado, Aurora, CO, USA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 3) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA; 4) Department of Bioengineering and Therapeutic Sciences , University of California, San Francisco, CA, USA; 5) Institute for Immunological Research, Universidad de Cartagena, Cartagena, Colombia; 6) Department of Microbiology, Howard University College of Medicine, Washington, DC, USA; 7) National Human Genome Center, Howard University College of Medicine, Washington, DC, USA; 8) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) Caribbean Institute for Health Research, The University of the West Indies, Kingston, Jamaica; 10) Department of Medicine, University of Arizona College of Medicine, Tucson, AZ, USA; 11) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 12) Center for Research on Genomics & Global Health, National Institutes of Health, Bethesda, MD, USA; 13) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 14) Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, MI, USA; 15) Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, Barbados; 16) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 17) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

RATIONALE: Asthma is a complex disease with striking disparities across racial/ethnic groups, which may be partly attributable to genetic factors. One of the main goals of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to discover genes conferring risk to asthma in populations of African descent. METHODS: We performed a genome-wide meta-analysis of asthma association across CAAPA studies, genotyped on the African Diaspora Power Chip (ADPC), and imputed using the WGS reference panel from CAAPA itself. RFMix was used to infer local ancestry of 6,686 African American and African Caribbean subjects. The 1000 Genomes Project CEU and YRI populations were specified as reference populations. A total of 273,738 phased SNPs passing quality control filters and overlapping in all the CAAPA and reference population data sets were used as input for local ancestry inference. SNPs with p-values <1e-5 were tested for interaction with the number of copies of segments of local African ancestry. The interaction p-value indicates whether a difference in allele frequency between cases and controls is observed in one class of local ancestry background more than another. RESULTS: An African-specific interaction that involved SNP rs11264909 on chromosome 1q23.1 was identified. This SNP is intronic to the KIRREL gene. A difference in allele frequencies was observed only in those subjects carrying 2 copies of African ancestry at this region, with the minor A allele being observed more frequently in cases (0.12) compared to controls (0.10), p=2.8e-5. SNP rs11264909 is ~0.9MB downstream from 2 SNPs intronic to the PYHIN1 gene previously identified by the NHLBI-supported EVE Consortium's asthma meta-analysis. Although the EVE SNPs were not polymorphic in Europeans, the A allele at SNP rs11264909 has a frequency of 0.20 in CAAPA and a much higher frequency (0.61) in Europeans. This A allele was observed more frequently in CAAPA cases (frequency of 0.22) compared to CAAPA controls (frequency of 0.19), and on average, cases have more European ancestry at this region compared to controls. These observations suggest the association observed with asthma is not due to an excess of African ancestry at this locus, but the association is specific to subjects of African ancestry at this region. CONCLUSIONS: SNP rs11264909 may have an African ancestry specific effect on risk to asthma, where the combination of African ancestry and the A allele increases risk.

1947F

A genome-wide association analysis of Hashimoto's thyroiditis. *V. Boraska Perica[1], L. Brčić[1], A. Barić[2], S. Gračan[2], D. Kaličanin[1], I. Gunjača[1], V. Torlak Lovrić[2], M. Brekalo[2], M. Radman[2], V. Škrabić[3], A. Miljković[4], I. Kolčić[4], S. Tokić[5], M. Štefanić[6], Lj. Glavaš-Obrovac[5], O. Polašek[4], T. Zemunik[1], M. Barbalić[1], A. Punda[2].* 1) Medical Biology, University of Split School of Medicine, Split, Croatia; 2) Department of Nuclear Medicine, University Hospital Split, Split, Croatia; 3) Department of Pediatrics, University Hospital Split, Split, Croatia; 4) Department of Public Health, University of Split School of Medicine, Split, Croatia; 5) Department of Medical Chemistry, Biochemistry and Clinical Chemistry, Faculty of Medicine, University of Osijek, Osijek, Croatia; 6) Department of Nuclear Medicine and Oncology, Faculty of Medicine, University of Osijek, Osijek, Croatia.

Background: Hashimoto's thyroiditis (HT) is the most common autoimmune thyroid disease characterized by chronic inflammation and reduced function of the thyroid gland. HT is the major cause of hypothyroidism in iodine-sufficient areas and it is highly prevalent in women. Although strong genetic susceptibility to HT has been proposed, underlying genetic factors are poorly known. Moreover, HT has never been analyzed on the genome-wide level. The aim of our study was to identify genetic variants associated with HT by performing the first genome-wide association analysis of this disease. **Methods:** A total of 869 individuals from Croatia, comprising of 430 HT cases and 439 controls, were included in this study. All HT cases met ETA recommendations and guidelines for Management of Subclinical Hypothyroidism. Control individuals were derived from the population-based "10,001 Dalmatians project" of which 349 individuals already had genome-wide data scanned by Illumina HumanOmniExpressExome array whereas additional 90 controls and HT cases were typed using Illumina Infinium HumanCoreExome array. Following quality control, two datasets were merged on an overlapping set of variants and imputed using 1000 Genomes reference panel. Association analysis was performed on 8,633,733 genetic variants under the univariate linear mixed model using GEMMA software. Binary disease status was treated as quantitative trait, and model was adjusted for age, sex, population stratification and relatedness. Reported investigations have been carried out following the principles of the Declaration of Helsinki. **Results:** We identified a missense variant inside *TET1* gene at the genome-wide significance level (β=0.365, se=0.066, p=4.21x10^{-8}) and a missense variant inside *UGT8* gene just below genome-wide significance (β=0,365, se=0,072, p=5.25x10^{-7}). A replication of these two signals along with 10 other suggestively associated variants (p<10^{-5}) is currently underway in an independent dataset of 303 HT cases and 302 controls. **Discussion:** TET1 is a methylcytosine dioxygenase that is involved in DNA demethylation and gene activation indicating a role of epigenetic events in HT susceptibility. UGT8 is an UDP-glycosyltransferase that has a role in the biosynthesis of galactocerebrosides of the nervous systems. **Conclusion:** In this first genome-wide analysis of HT to date, we have identified promising associations of biologically interesting genes pointing to a novel mechanisms underlying HT development.

1948W

HLA-DPB1 and Graves disease in Han Chinese. *Y. Lee[1,2,3,4,5], C. Huang[1,4,6], W. Ting[1,4,6], C. Lin[2], W. Lin[2], T. Chang[2], H. Yang[2].* 1) Department of Pediatric Endocrinology, MacKay Children's Hospital; 2) Department of Medical Research, MacKay Memorial Hospital Tamsui District; 3) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University; 4) Department of Medicine, Mackay Medical College; 5) Institute of Biomedical Sciences, Mackay Medical College; 6) MacKay Junior College of Medicine, Nursing, and Management, Taipei, Taiwan.

Graves disease (GD) is a multifactorial disease that appears to involve a complex interplay of multiple genetic and environmental influences. It affects up to 0.5% of general population. The risk to a female sib of a proband with GD (λs) is 10-15. The concordance rate is 20-35% in monozygotic twins but only 3-7% in dizygotic twins. These suggest that genetic factors play a role in the pathogenesis of this disorder. Because T-lymphocyte activating is dependent on recognition of MHC class II molecules. We reported association between DPB1 gene and GD in Han Chinese living in Taiwan. **Patients** The subjects were 256 GD patients (45 boys and 211 girls) aged 10.6 (SD 3.4), 2.7-19.0 years. GD was diagnosed on the basis of clinical manifestations and laboratory evidence. **Controls** The 381 control subjects consisted of 135 males and 246 females. They included hospital personnel and individuals who underwent routine health examinations or minor surgery. None had a history of autoimmune disease. All patients and controls were Han Chinese in Taiwan. Our institutional review board approved this study and all subjects or their guardians gave informed consent. **Genotyping of the DPB1 gene** We used sequencing-based typing to genotype the DPB1 gene. **Statistical analysis** Statistical difference in genotype, allele, and carrier distributions between patients and controls were assessed by the chi-square test. Odds ratios (OR) and 95% confidence intervals were also calculated. The Bonferroni correction, $Pc = 1 - (1-P)^n$, was used for multiple comparisons where Pc is the corrected P value, P the uncorrected value, and n the number of comparisons. In this study, n is 13 for each allele, or carrier. A Pc value of less than 0.05 was considered statistically significant. **Results** The DPB1*05:01 allele was the most frequent in patients (59.0%) and controls (46.7%). It was also significantly more frequent in patients than in controls, OR (95%CI)=1.64 (1.31-2.06), Pc = 2.27E-04. No other alleles were significantly different in frequency between patients and controls. **Conclusion** DPB1 was associated with T1D in Han Chinese in Taiwan. DPB1*05:01conferred susceptibility to GD. Extensive genotyping of other loci in the HLA complex is necessary to resolve the mechanism of the association between HLA and GD.

1949T

Genome-wide association study identifies candidate loci associated with intraoperative remifentanil requirements during laparoscopic-assisted colectomy. *D. Nishizawa[1], T. Mieda[2], M. Tsujita[3], H. Nakagawa[3], S. Yamaguchi[4], S. Kasai[1], J. Hasegawa[1], K. Fukuda[5], A. Kitamura[3], M. Hayashida[1,3,6], K. Ikeda[1].* 1) Psychiatry and Behavioral Science (Addictive Substance Project), Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Anesthesiology, Saitama Medical University Hospital, Saitama, Japan; 3) Department of Anesthesiology, Saitama Medical University International Medical Center, Saitama, Japan; 4) Department of Gastroenterological Surgery, Saitama Medical University International Medical Center, Saitama, Japan; 5) Department of Oral Health and Clinical Science, Tokyo Dental College, Tokyo, Japan; 6) Department of Anesthesiology & Pain Medicine, Juntendo University School of Medicine, Tokyo, Japan.

Purpose: Opioid analgesics are commonly required during and after laparoscopic surgery. Further, opioid requirements during and after laparoscopic surgery are highly variable among patients. We conducted a genome-wide association study (GWAS) in patients undergoing laparoscopic-assisted colectomy (LAC) to identify potential candidate single nucleotide polymorphisms (SNPs) that may significantly contribute to such individual differences. Methods: We conducted a three-stage GWAS by using whole-genome genotyping arrays with more than 900,000 markers in 351 patients, who underwent LAC under general anesthesia using sevoflurane and remifentanil, and received intravenous fentanyl patient-controlled analgesia for postoperative pain control. During surgery, the end-tidal concentration of sevoflurane was fixed at 1.2 %, and the infusion rate of remifentanil was titrated so as to maintain systolic blood pressure between 90 and 120 mmHg, heart rate below 75 beats per minute, and electroencephalographic bispectral index between 40 and 60. We further investigated whether some of the best candidate SNPs identified with the GWAS significantly affects opioid sensitivity in 355 patients undergoing mandibular sagittal split ramus osteotomy (SSRO) to corroborate potential associations found in patients undergoing LAC. Results: As a result of GWAS in surgical patients, a nonsynonymous SNP in the *TMEM8A* gene region, rs199670311, and intronic SNPs including rs4839603 SNP in the *SLC9A9* gene region, had highly significant associations with the intraoperative average remifentanil infusion rate; the subjects with the A and T alleles required higher average remifentanil infusion rates for intraoperative anesthesia, compared with those without these alleles in the rs199670311 and rs4839603 SNPs, respectively. In patients undergoing SSRO, the postoperative and total perioperative fentanyl requirements were greater in the A and T allele carriers compared with non-carriers in the rs199670311 and rs4839603 SNPs, respectively. Conclusions: Our findings provide valuable information for personalized pain control by opioids during LAC and during/after SSRO, in which the A and T alleles of the rs199670311 and rs4839603 SNPs, respectively, are associated with lower opioid sensitivity and require more opioids during LAC and during/after SSRO.

1950F

NELFCD and CTSZ loci are associated with jaundice-stage progression in primary biliary cholangitis in the Japanese population. *M. Kawashima[1,2], N. Nishida[2,3], Y. Hitomi[2], K. Kojima[4,5], Y. Kawai[4,5], Y. Aiba[6], H. Nakamura[6], M. Nagasaki[4,5,7], K. Tokunaga[2], M. Nakamura[6,8,9,10].* 1) National Bioscience Database Center, Japan Science and Technology Agency, Chiyoda-ku, Japan; 2) Department of Human Genetics, Graduate School of Medicine, the University of Tokyo; 3) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine; 4) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University; 5) Graduate School of Medicine, Tohoku University; 6) Clinical Research Center, National Hospital Organization, Nagasaki Medical Center; 7) Graduate School of Information Sciences, Tohoku University; 8) Department of Hepatology, Nagasaki University Graduate School of Biomedical Sciences; 9) Headquarters of PBC Research in NHOSLJ, Clinical Research Center, National Hospital Organization Nagasaki Medical Center; 10) Headquarters of PBC Consortium in Japan (PBCJPN), National Hospital Organization (NHO) Study Group for Liver Disease in Japan (NHOSLJ), and gp210 Working Group in Intractable Liver Disease Research Project Team of the Ministry of Health and Welfare in Japan.

Aim and Methods: Approximately 10-20% of patients with ;primary biliary cholangitis (PBC) progress from early stage to jaundice stage regardless of treatment with ursodeoxycholic acid. In this study, we performed a GWAS and validation study to identify genetic factors associated with progression to jaundice stage in PBC using a total of 1,125 PBC patients (975 early-stage and 150 jaundice-stage). Then, a high-density association mapping for the surrounding region of the top SNP detected from the GWAS and validation study was performed using 1,375 PBC patients (1,202 early-stage and 173 jaundice-stage), including GWAS set samples. Functional variants were predicted by in silico analysis using the Regulome DB database, UCSC genome browser, the SNP Function Prediction (FuncPred), and the PolymiRTS database. To assess the biological effects on mRNA expression, we compared the endogenous expression levels in whole blood, transformed fibroblasts, liver, spleen, and EBV-transformed lymphocytes of CTSZ and NELFCD using the GTEx portal database.   Results and discussion: A SNP located in the 3'UTR of cathepsin Z (CTSZ) showed the strongest association (odds ratio [OR] = 2.15, 95% confidence interval [CI] = 1.58-2.94, P = 7.62 × 10^{-7}) with progression to jaundice stage in GWAS, and this association reached borderline significance in the validation study (OR = 2.15, 95% CI = 1.61-2.87, P = 1.36 × 10^{-7}). High-density association mapping around CTSZ and negative elongation factor complex member C/D (NELFCD), which are located within a strong linkage disequilibrium block, revealed that another intronic SNP-X of CTSZ showed statistical significant association with jaundice-stage progression (OR = 2.16, 95% CI = 1.62-2.87, P = 8.57 × 10^{-8}). Among SNPs that are strong linkage disequilibrium (r^2 > 0.8) with the SNP-X, four SNPs and two SNPs were predicted to be located in transcriptional regulatory elements and microRNA binding sites, respectively. Additionally, an eQTL analysis indicated that SNP-X was correlated with both NELFCD and CTSZ mRNA levels in whole blood and transformed fibroblasts. These results indicate that SNPs at *NELFCD/CTSZ* loci might be involved in progression to jaundice stage in PBC via alteration of these mRNA levels.

1951W

Exome-wide association study of kidney function in 55,041 participants of the DiscovEHR cohort. *C. Schurmann[1], M.K. Puurunen[1], S. Bruse[1], A. Li[1], N. Banerjee[1], J. Penn[1], A. Lopez[1], K. Devalaraja[1], M.D. Ritchie[2], D.J. Carey[2], J.G. Reid[1], J.D. Overton[1], A. Chang[2], K. Ho[2], L. Morton[1], A. Baras[1], F.E. Dewey[1], O. Gottesman[1], T.M. Teslovich[1], N.S. Abul-Husn[1], Geisinger-Regeneron DiscovEHR Collaboration.* 1) Regeneron Pharmaceuticals, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

Background: Chronic kidney disease (CKD), a progressive loss of renal function, is a major health burden affecting 5-10% of the world's population. Risk factors include diabetes mellitus, hypertension, cardiovascular disease, and family history. Serum creatinine levels and estimated glomerular filtration rate (eGFR) are established measures of kidney function. Genome-wide association studies (GWAS) mainly focusing on common variants have identified more than 60 loci associated with renal function, but less than 4% of the estimated heritability of eGFR has been explained so far. **Aim**: The aim of our study was to evaluate the impact of rare sequence variation on kidney function in a large clinical care cohort. **Methods**: We utilized exome sequencing and phenotypes derived from electronic health records for up to 55,041 European samples [60% female, median age 61 years, 26% type 2 diabetic, 64% hypertensive] from the DiscovEHR cohort. We tested single variants and genes for association with median serum creatinine and eGFR levels using linear regression models with adjustment for age, age[2], sex, and genetic ancestry. **Results**: Preliminary association results for serum creatinine and eGFR include the replication of previously reported GWAS loci (*GCKR, GATM, UMOD*/*PDILT*, *CCDC158*, *ALMS1,* and *SLC47A1*). Furthermore, we identified ten gene-based associations at suggestive significance level (P<1x10^{-5}), all of which were shared between serum creatinine and eGFR. These associations include aggregates of rare variants in known CKD-associated genes *SLC22A2* (Serum creatinine: B=0.14 standard deviations [SD], P=2.3x10^{-5}; eGFR: B=-0.12 SD, P=5.6x10^{-4}) and *SLC47A1* (Serum creatinine: B=0.15 SD, P= 4.8x10^{-13}; eGFR: B=-0.17 SD, P=6.5x10^{-16}). The *SLC47A1* gene-based signals for serum creatinine and eGFR are largely driven by a single rare variant (rs111653425, MAF=0.9%) with B=0.19 SD and P=2.9x10^{-9} for serum creatinine and B=-0.20 SD and P=2.6x10^{-10} for eGFR, respectively. The *SLC47A1* gene, previously associated with eGFR, creatinine levels, and CKD, is highly expressed in the kidney and liver and plays a role in treatment response to metformin, a drug used to improve blood glucose control in individuals with type 2 diabetes. **Conclusion**: These findings extend our understanding of the genetic architecture of renal function.

1952T

Alternative splicing of *ICAM3* in Crohn's disease. *I. Arijs[1,2], S. Verstockt[3], M. Vancamelbeke[2], B. Verstockt[2], F. Schuit[4], M. Ferrante[2], S. Vermeire[2], I. Cleynen[3].* 1) Department of Physiology - Biochemistry - Immunology , UHasselt - Jessa Hospital, Diepenbeek, Belgium; 2) Department of Clinical and Experimental Medicine, Translational Research Center for Gastrointestinal Disorders, KU Leuven, Leuven, Belgium; 3) Department of Human Genetics, KU Leuven, Belgium; 4) Department of Cellular and Molecular Medicine, Gene Expression Unit, Leuven, Belgium.

Crohn's disease (CD) is reflected by differential mucosal expression of genes involved in immune response, antimicrobial response, tissue remodeling and cell adhesion. The majority of protein-encoding genes, including those underlying susceptibility to CD, undergo alternative splicing, hereby influencing human physiology, development and disease. As differential splicing patterns on a genome-wide scale have not yet been studied in CD, we wanted to search for CD-associated alternative splicing events. Genome-wide alternative splicing was studied with Affymetrix Human Exon 1.0 ST Arrays on total RNA from mucosal biopsies from inflamed colon of 19 CD patients, and from normal colon of 6 controls.. Data were analyzed with R/Bioconductor (aroma.affymetrix and limma packages), and with Exon Array Analyzer (EAA). Significant (false discovery rate <5% and >2-fold change) alternative splicing events from both programs were compared. Bio-functional analysis was performed with Ingenuity Pathway analysis (IPA). We identified 350 differential splicing events, representing 243 different genes, in both R and EAA. IPA analysis highlighted cellular movement, cell death and survival, and immune cell trafficking among the most enriched biological functions. When comparing only the top 50 most significant splicing events from both analysis programs, 20 were overlapping. Of these, five are known alternative splicing events, located in *ICAM3*, *AKAP9*, *AKAP13*, *SDR16C5* and *IGSF9*. Cluster analysis based on the expression of the corresponding exon probe sets distinguished CD patients from controls. We found alternative splicing of many transcripts in colonic biopsies of CD patients, many of which are involved in biological functions related to the pathogenesis of CD. *ICAM3* encodes an adhesion molecule regulating leukocyte trafficking. It represents a very promising alternative splice event in CD, as it is located in a known susceptibility region for CD. Currently, we examine the role of the identified *ICAM3* genetic variants on alternative splicing by quantitative RT-PCR.

1953F

Chronic obstructive pulmonary disease subtyping through multiple -omics data integration. *B.D. Hobbs[1,2], J.D. Morrow[1], B.R. Celli[2], R. Bueno[3], G.J. Criner[4], D.L. DeMeo[1,2], C.P. Hersh[1,2], E.K. Silverman[1,2], M.H. Cho[1,2].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA; 3) Division of Thoracic Surgery, Brigham and Women's Hospital, Boston, MA; 4) Division of Pulmonary and Critical Care Medicine, Temple University School of Medicine, Philadelphia, PA.

Rationale: Chronic obstructive pulmonary disease (COPD), the third leading cause of death in the United States, is a highly heterogeneous disease. Some therapies for COPD appear to be most beneficial in specific subsets of patients. However, these subsets have often been based on a limited set of clinical factors and expert opinion. We hypothesized that biologically similar COPD subtypes can be identified by leveraging machine learning on integrated gene expression (transcriptomic), DNA methylation (epigenomic), and genetic variation data. Further, we hypothesized that subtypes derived from integrated -omics will associate with COPD clinical and radiographic features. **Methods**: We performed gene expression and methylation profiling and genome-wide genotyping on 105 lung tissue samples from former smokers of European ancestry with severe COPD. After quality control, in 84 samples with overlapping -omic data, we constructed subject distance (similarity) networks for each -omic data type (limiting genetic data to only the top 64 COPD-associated regions from a recent COPD GWAS) and merged them using the similarity network fusion (SNF) method. We used spectral clustering to identify subtypes of COPD patients from the fused patient similarity network. We performed differential gene expression and genetic association studies of the COPD subtypes to evaluate the gene expression and genetic differences between clusters. **Results**: We identified 4 clusters (subtypes) of subjects sharing molecular information. Although all groups had severe emphysema, clusters were significantly different in quantitative lung density (percent emphysema) on CT scan. The most severe subtype (cluster 2) had 38% emphysema and the least severe group had 28% emphysema. Lung function was not significantly different across clusters. A variant (rs17015027) at the *FAM13A* locus was significantly associated (P = 2.03×10^{-4}) with cluster 2 membership (compared to all other clusters). Cluster 2 showed differential expression (FDR < 10%) of *KPNA7*, *MMP10*, *LINC00479*, *GOLM1*, and *KRT17* compared to all other clusters where MMP10 levels have been associated with quantitative imaging emphysema subtypes. **Conclusions**: An integration of multi-omic data from lung tissue of severe COPD cases identified four subtypes with significant clinical and genetic differences. Multi-omic integration has the potential to discover novel subtypes of COPD patients with similar underlying disease pathogenesis.

1954W

Identifying genetic determinants of age at menarche and age at menopause in the Japanese population. *M. Horikoshi[1], F.R. Day[2], Y. Kamatani[1,3], K. Matsuda[4], M. Hirata[4], K.K. Ong[2], M. Kubo[1], J.R.B. Perry[2].* 1) RIKEN Centre for Integrative Medical Sciences, Yokohama Japan; 2) University of Cambridge, Cambridge, UK; 3) Kyoto University Graduate School of Medicine, Kyoto, Japan; 4) University of Tokyo Institute of Medical Science, Tokyo, Japan.

Population genetic studies over the past decade have been highly successful in elucidating the genetic architecture of reproductive ageing. However, those studies have been largely limited to individuals of European ancestry, restricting the generalizability of findings and benefits of trans-ethnic approaches. To address this, we performed the largest genome-wide association study (GWAS) to date outside of European ancestry groups, comprising up to 67,030 women of Japanese ancestry collected by the Biobank Japan, for the ages at menarche (a marker of puberty timing) and menopause, imputed to 1000 Genomes reference variants. Genotyping array estimated heritabilities for age at menarche (13%) and age at menopause (10%) are 3-times smaller than those reported in Europeans. For menopause, 16 independent signals reached genome-wide significance ($P<5 \times 10^{-8}$), 8 of which are novel and not previously reported in Europeans. For menarche, 10 independent signals reached genome-wide significance, 2 of which are novel regions, and a third represents a novel Japanese-specific signal in a known European region ($r^2 \sim 0$) near *PTPRD*. Of the 12 novel signals for the two traits, 5 showed significantly larger effects in Japanese than in Europeans ($P_{heterogeneity} < (0.05/12)$). A further 3 signals were likely not identified in previous GWAS due to markedly lower allele frequencies in Europeans: *EIF4E* (rs199646819, minor allele frequency in Japanese vs. Europeans: 39% vs. 2%), *NKX2-1* (rs2076751: 25% vs. 7%) and *THOC1* (rs77001758: 40% vs. 0.1%). Highlighted mechanistic insights include a SIFT-predicted deleterious missense variant in *GNRH1* (encoding gonadotrophin releasing hormone, and a known signal for menarche) newly associated with age at menopause ($P=3 \times 10^{-13}$) and enrichment of menarche signals near genes that encode for receptor tyrosine phosphatases. Furthermore, genes involved in basic molecular processes: histone modification (*H1FX*), mRNA translation (*EI4FE* and *EIF4EBP1*) and DNA damage response (*RAD21*) were implicated in menopause timing. In summary, we report the first genomic study outside of Europeans to identify genome-wide significant loci for reproductive ageing. Widespread differences in allele frequencies and effect estimates present challenges for such large-scale trans-ethnic approaches, but allow new mechanistic insights.

1955T

Markers of the adaptive immune response are associated with progressively worse chronic kidney disease status. *D.C. Crawford[1], W.S. Bush[1], J.N. Cooke Bailey[1], K. Miskimen[1], P. Miron[1], J. O'Toole[2], J.R. Sedor[2,3].* 1) Institute for Computational Biology, Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Department of Medicine, Case Western Reserve University, Cleveland, OH; 3) Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, OH.

Germline and somatic genomic variation represent the bulk of 'omics data available for precision medicine research. These data, however, may fail to capture the dynamic biological processes that underlie disease development, particularly for chronic diseases of aging such as chronic kidney disease (CKD). To demonstrate the value of additional dynamic precision medicine data, we sequenced somatic T-cell receptor rearrangements, markers of the adaptive immune response, from genomic DNA collected during a clinical encounter from 15 participants with CKD and associated co-morbidities. Participants were consented as part of a larger precision medicine research project at the MetroHealth System, a large urban public hospital in Cleveland, Ohio. Genomic DNA was extracted from whole blood, and T-cell receptors were sequenced with six replicates per sample using Adaptive Biotechnologies' immunoSEQ assays coupled with the Illumina NextSeq PE. All sequences were assembled using Adaptive Biotechnologies' ANALYZER bioinformatics pipeline. Demographic and clinical data closest to the time of blood draw were extracted from the electronic health record. Overall, the average age of patients was 61.73 years, more than half (60%) were female, and the majority were African American (80%). T-cell receptor diversity was estimated using productive clonality, a measure ranging from 0 (polyclonal samples) to 1 (monoclonal or oligoclonal samples). Productive clonality in this sample ranged from 0.0151 to 0.2565 with a mean of 0.1030 (standard deviation or SD=0.0669). Average productive clonality did not statistically differ by sex: females = 0.0811 (SD=0.0533) and males = 0.1358 (SD=0.0764). We then tested for correlations between T-cell receptor diversity and biomarkers of CKD, including disease status calculated using the CKD-EPI equation. Reduced T-cell diversity was associated with increased creatinine ($R^2=0.0995$), BUN ($R^2=0.0258$), and eGFR ($R^2=0.066$), but not with white blood cell count ($R^2=0.0004$). Reduced T-cell diversity was also associated with worsening CKD status ($R^2=0.2362$), with a higher on average productive clonality (0.0488) among stage 4 patients (n=5) compared with stage 3 (0.0330; n=8) and stage 2 patients (0.0149; n=2). These data suggest an association between advanced CKD and premature aging of the adaptive immune system and highlight the potential of dynamic 'omic data to generate novel hypotheses about disease mechanisms.

1956F

Effect of CAG repeat length in the *androgen receptor* gene on hirsutism among healthy Israeli women of different ethnicities. *S. Ben-Shachar[1,3], M. Slakman[3], G. Israeli[2], O. Eyal[2,3], N. Weintrob[2,3].* 1) Genetic Institute, Tel Aviv Medical Center, Tel Aviv, Israel; 2) Pediatric Endocrinology and Diabetes unit, Dana Hospital, Tel Aviv Medical Center, Tel Aviv, Israel; 3) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

Introduction: The differences in the levels of hirsutism between women of different ethnicities may stem from multiple etiologies. Polymorphism of the CAG repeats of the *androgen receptor* (*AR*) gene is common. Shorter length of the repeats are associated with increased activity of the receptor. We hypothesized that there are ethnic differences in the degree of hirsutism among Israeli women, and that the CAG repeat length may contribute to these differences. **Methods:** Healthy Israeli Jewish women age 18-45 years of Ashkenazi and non-Ashkenazi (North African and Near-Eastern) origin were recruited. Hirsutism was assessed using the simplified Ferriman–Gallwey (sFG) score. Demographic data (weight, body mass index [BMI], and age) were collected, and serum total testosterone levels were determined. The CAG repeat length was measured by PCR. Methylation-sensitive methods were used to detect the fractional activity of each allele in blood, and the weighted arithmetic mean was calculated for the CAG repeat length. **Results**: One-hundred and eight women were recruited (49 Ashkenazi and 59 non-Ashkenazi). There was a significant negative correlation between the number of CAG repeats in the *AR* gene and the sFG score. The number of repeats was not related to testosterone levels. Ashkenazi women had a lower degree of hirsutism ($p<0.001$), a lower BMI (21.4 vs 24 for non-Ashkenazi women, $p=0.003$), and lower testosterone levels (0.015 vs 0.022 ng/ml, respectively, $p=0.017$). The length of the shorter CAG repeat allele and the weighted CAG repeat length were greater among the Ashkenazi women (21.1±2.3 vs 19.6±2.6, $p=0.03$ and 22.2±2.4 vs 21.1±2.3, $p=0.015$, respectively). **Conclusions**: There is a significant difference in the degree of hirsutism between Ashkenazi and non-Ashkenazi women in Israel. The length of CAG repeats partially explains some of these differences. These results suggest that the ethnic difference in the degree of hirsutism is the result of a complex genetic trait.

1957W

Rare variation associated with immunosuppressant drug concentrations: Moving beyond common SNPs in predicting drug metabolism. *A.A. Seyerle[1], D. Schladt[2], W. Guan[3], R. Remmel[4], W.S. Oetting[5], C. Dorr[6], B. Wu[3], A. Israni[6], P.A. Jacobson[5].* 1) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 2) Department of Nephrology and Chronic Disease Research Group, Minneapolis Medical Research Foundation, Hennepin County Medical Center, Minneapolis, MN; 3) Division of Biostatistics, University of Minnesota, Minneapolis, MN; 4) Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN; 5) Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN; 6) Department of Nephrology, Minneapolis Medical Research Foundation, Department of Medicine, University of Minnesota, Minneapolis, MN.

Tacrolimus (TAC), an immunosuppressive agent commonly used in solid organ transplant recipients, has a narrow therapeutic index requiring therapeutic drug monitoring in clinical care. Common genetic variants in cytochromeP450 3A4/5 are well known to affect the TAC metabolism and blood concentrations. Common variants and clinical factors explain ~50% of interpatient variability in blood trough concentrations. We examined whether rare genetic variants were associated with TAC troughs to explain additional variability. Among European American (EA; N=1,354) and African American (AA; N=299) kidney transplant (tx) recipients in the Deterioration of Kidney Allograft Function (DeKAF) Genomics study, longitudinal TAC troughs were measured in the first six months post-tx in each subject. A simple spline method was used to model a day 9 TAC trough concentration. Genotyping on recipient DNA was performed on a custom exome-plus Affymetrix TxArray SNP chip. For SNPs with a minor allele frequency < 5%, race-stratified statistical analyses were done using both gene-based burden testing and sequence kernel association testing (SKAT). Among EAs, the burden test identified ten genes on chromosome seven associated with TAC residuals (*PDAP1, ARPC1B, CYP3A5, ZNF655, ZSCAN25, ATP5J2, ZNF394, CYP3A4, ARPC1A, PTCD1*) and one gene on chromosome 21 (*CNN2P7*). SKAT identified *CYP3A43* and *CNN2P7* associated with TAC residuals in EAs. Burden testing in AAs identified the genes *ZNF138* and *ZNF107* associated with TAC residuals while SKAT testing did not identify any genes significantly associated with TAC residuals. Adjustment for common, well-known variants associated with TAC, *CYP3A5*3* and *CYP3A4*22* in EAs and *CYP3A5*3, *6, and *7* in AAs, eliminated significant findings on chromosome seven, although *CNN2P7* remained. These results suggest that low frequency variants in the genome may modify TAC blood concentrations. Our results also indicate that better characterization of the low frequency variants surrounding the *CYP* genes is needed to fully account for the role of genetics in TAC metabolism. Better understanding the role low frequency variants play in TAC metabolism could ultimately aid in more personalized and effective TAC dosing and may be relevant to hundreds of other CYP3A drug substrates.

1958T

NUDT15 variants contribute to thiopurine-induced myelosuppression in European populations. *M.D. Voskuil[1], G.J. Walker[2,3], J.W. Harrison[4], G.A. Heap[2,3], J. Koskela[5,6], E.A.M. Festen[1], M.J. Daly[5,6], R.K. Weersma[1], R. Ward[4], M.N. Weedon[4], J.R. Goodhand[2,3], N.A. Kennedy[2,3], T. Ahmad[2,3], IBD Pharmacogenetics Study Group.* 1) Department of Gastroenterology and Hepatology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; 2) Department of Gastroenterology, Royal Devon and Exeter Hospital NHS Foundation Trust, Exeter, UK; 3) IBD Pharmacogenetics Group, University of Exeter, Exeter, UK; 4) University of Exeter Medical School, Exeter, UK; 5) Broad Institute of Harvard University and MIT, Cambridge, MA, USA; 6) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA.

Thiopurines are commonly used in the maintenance treatment of inflammatory bowel disease (IBD) but this is limited by myelosuppression in 7% of patients. Thiopurine S-methyltransferase (*TPMT)* variants only explain 26% of thiopurine-induced myelosuppression (TIM) in Europeans suggesting the presence of other genetic determinants. We aimed to identify the clinical features of TIM and to identify novel variants associated with TIM through a genome wide association study (GWAS) and whole exome sequencing. We recruited 491 IBD patients with TIM from 82 hospitals. Inclusion criteria included drug exposure in the 7 days prior to TIM, white blood cell count $<2.5 \times 10^9$/L, or neutrophil count $<1.0 \times 10^9$/L and that TIM necessitated dose reduction/drug withdrawal. After expert adjudication, 329 cases assigned a high likelihood of causality were genotyped and exome sequenced with 635 thiopurine-tolerant IBD controls for use in the final analyses. We first confirmed an association of TIM with *TPMT* in an initial GWAS. We then, using exome sequencing, discovered a novel 6bp in-frame deletion (rs746071566; AGGAGTC/A, p.Gly17_Val18del) at position 48611918 of chromosome 13 in exon 1 of *NUDT15* (5.8% of cases, 0.2% of controls OR = 38.2; $P = 1.3 \times 10^{-8}$). We replicated this in an independent cohort (2/73 [2.7%] cases vs. 2/840 [0.2%] thiopurine-tolerant IBD controls; OR=11.8, P=0.034). All *NUDT15* coding variants were validated by Sanger sequencing. Multivariable logistic regression demonstrated that weight adjusted dose (OR=2.2, P=9.4×10^{-11}), *NUDT15* genotype (OR 21.7, P=2.8×10^{-8}) and *TPMT* genotype (OR 2.2, P=2.6×10^{-4} for MUT/WT and OR 51.2, P=1.8×10^{-4} for MUT/MUT) were independently associated with TIM. These are the largest ever clinical and genetic analyses of thiopurine-induced myelosuppression. We identified *NUDT15* coding variants, including a novel 6bp deletion; carriage of any coding variant confers a 22-fold increase in the odds of TIM, independent of *TPMT* genotype and thiopurine dose. Although *NUDT15* variants are less common than *TPMT* variants, their effect size for heterozygotes is greater. The number of patients needed to genotype to prevent TIM due to *NUDT15* coding variant heterozygosity is 100. The estimated absolute risk of TIM in *NUDT15* heterozygotes is 59%. A future clinical decision tool incorporating *NUDT15* and *TPMT* genotypes will offer personalized thiopurine therapy and prevent the considerable morbidity and costs associated with severe bone marrow toxicity.

1959F

Exome sequencing highlights novel DNA variants with a potential role in polycystic ovary syndrome. *S.G. Wilson[1,2,3], R.S. Smith[1,4], P.J. Campbell[1], K.A. Burns[1,2], B.H. Mullin[1,2], R.J. Mead[4], E.M. Lim[1,5,6], B.G.A. Stuckey[1,2].* 1) Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA, Australia; 2) School of Medicine and Pharmacology, University of Western Australia, Crawley, WA, Australia; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 4) School of Biological Sciences, Murdoch University, Murdoch, WA, Australia; 5) PathWest Laboratory Medicine, Nedlands, WA, Australia; 6) School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, WA, Australia.

Polycystic ovary syndrome (PCOS) is the leading cause of anovulatory infertility in women and affects 5%–10% of women in the developed world. Symptoms include hyperandrogenism, metabolic disturbances and chronic anovulation. The pathophysiology of PCOS is complex, but there is strong evidence of a genetic component. There is substantial heterogeneity in the clinical presentation of PCOS suggesting that the disorder is multifactorial, with multiple gene and environmental factors contributing to disease susceptibility. Numerous genome-wide association studies (GWAS) of PCOS have highlighted variants (MAF > 0.05) at 20 potentially relevant loci, however these common SNPs are unlikely to be the causal variants and together explain only a fraction of the heritability of PCOS. The aim of this project was to focus on variants at the lower end of the allelic frequency spectrum, which had predicted functional consequences on the protein product of the gene. We recruited first degree relative pairs and trios with physician diagnosed PCOS from a clinical practice using the NIH diagnostic criteria. Genomic DNA was extracted from the blood of 33 study subjects and Human exome capture and sequencing was undertaken using Agilent SureSelect XT Human All Exon +UTR v5 on Illumina HiSeq 2500. Targeted mapped coverage was 50x and SNV and Indel detection were performed. Data was then processed using ANNOVAR to annotate functional consequences of the genetic variants. We studied individual PCOS families, retaining only those variants that were present in all available affected individuals in the pedigree. Loci were filtered based on the following criteria: SIFT <= 0.05, PolyPhen-2 > 0.85, MAF < 0.01 and mutation of the murine ortholog reported to cause a reproductive phenotype (where data was available). Among the missense variants that were observed, SNV in the genes *DNAH11, INSR, FZD1* may be of particular interest. In summary, exome sequencing appears to be an efficient technique for the study of rare variants segregating in families with PCOS. The study identified a number of variants of potential relevance to disease susceptibility and sets the stage for future functional studies aimed at increasing understanding of the pathophysiology of the disorder. If these data can be replicated, or functional relevance established, the findings may be of potential direct clinical value in molecular screening protocols and in the creation of genetic risk scores for patients.

1960W

Search for genetic factor associated with right-sided colonic diverticula in Korean population: Genome-wide association study. *E. Choe[1], E. Shin[2], J. Kim[2], S. Chung[3], S. Yang[3], Y. Kim[3], J. Bae[3]. 1) Surgery, Seoul National University Hospital, Healthcare System Gangnam Center, Seoul, South Korea; 2) DNA Link, Inc., Seoul, South Korea; 3) Internal Medicine, Seoul National University Hospital, Healthcare System Gangnam Center, Seoul, South Korea.*

Background : Diverticular disease is one of the most frequent problem in gastrointestinal tract. But since most of the colonic diverticula are asymptomatic, the exact prevalence is not well known and as such, single nucleotide polymorphism (SNP) based genetic study is scarce. In Asian population, diverticula is more frequently found in right colon with younger age, compared to western population. This early onset feature implies the high genetic predisposition for the diverticula in right colon. We performed a genome-wide association study to identify genetic factors associated with right-sided colonic diverticula in Korean population. Method : Genome-wide association study was conducted in 8,072 Korean population, who had colonoscopy during routine health check-up, using the Affymetrix Axiom® Customized Biobank Genotyping Arrays for right-sided colonic diverticula. After excluding those in age under 50 years and did not have diverticula, data from total 5444 samples were analyzed. Control group (test set, n=3396; replication set, n=1003) was defined as those in age≥50 years and did not have diverticula; case group (test set, n=678; replication set, n=367) was defined as those in any age and had right-sided colonic diverticula. Multiple logistic regression was used for analysis with age and gender adjustment. Result : In the test set (age = 57±6.68 years; Male : Female = 2588 : 1486), there were no SNPs that had significant association with right-sided colonic diverticula after the bonferroni correction (threshold for P value were < 10^8). So we investigated the regional plotting for SNPs that had P value less than 10^3. There were 3 blocks that had SNPs aggregated regionally with distances less than 100 base pair apart. 31 SNPs were identified in 3 blocks. In the replication set, 17 SNPs among 31 SNPs were replicated and the genes associated with each block were CDC42BPA; RPS6KA2; and TEX10, MSANTD3 gene. Most significant SNPs in each block were rs2718207 (CDC42BPA, P=9.2x10^4), rs79706510 (RPS6KA2, P=6.5x10^4) and rs2417084 (TEX10, MSANDTD3; P=7.4x10^4). Conclusion : We demonstrated that the CDC42BPA, RPS6KA2, TEX10 and MSANTD3 are associated with the presence of right-sided colonic diverticula in Korean population. These findings are a novel finding that provide the genetic role in the pathogenesis of colonic diverticula. This should be validated in a larger size of diverticula cohort with variable ethnicity.

1961T

Dysregulated gene and miRNA expression in different stages of Crohn's disease. *I. Cleynen[1], S. Verstockt[1], G. De Hertogh[2], J. Van der Goten[3], M. Vancamelbeke[3], B. Verstockt[3], L. Van Lommel[4], F. Schuit[4], P. Rutgeerts[3,5], M. Ferrante[3,5], S. Vermeire[3,5], I. Arijs[3,6]. 1) Department of Human Genetics, KU Leuven, Leuven, Belgium; 2) Department of Morphology and Molecular Pathology, University Hospital Gasthuisberg, Leuven, Belgium; 3) Translational Research Center for Gastrointestinal Disorders (TARGID), Department of Clinical and Experimental Medicine, KU Leuven, Leuven, Belgium; 4) Gene Expression Unit, Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium; 5) Department of Gastroenterology and Hepatology, University Hospital Gasthuisberg, Leuven, Belgium; 6) Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium; Jessa Hospital, Hasselt, Belgium.*

Crohn's disease (CD) is characterized by chronic inflammation of the gut, progressing to stricturing and/or penetrating complications in most patients. Effective intervention before bowel damage, and thus in the early phase of the disease, will be required to optimize patient outcomes. We therefore aimed to define the molecular landscape of early CD, by performing transcriptomics on ileal mucosal biopsies of CD patients with early (<18 months after ileo-colonic anastomosis) post-operative recurrence (POR; Rutgeerts' score i2b, i3 or i4; n=24). We also included biopsies of newly diagnosed (<18 months) CD patients (n=19); late (>3 years after diagnosis or ileo-colonic anastomosis) CD patients with active disease (n=14); and 12 normal controls. Gene and microRNA (miRNA) expression was studied using Affymetrix Human Gene 1.0 ST and miRNA 2.0 arrays, and validation using quantitative reverse transcriptase-PCR. mRNA dysregulation was most pronounced in new and late CD with 608 and 614 significantly (false discovery rate (FDR) <5% and >2-fold change) differentially expressed gene probe sets, while 353 in POR CD versus controls, and an overlap of 261. The top dysregulated gene in POR CD was *FOLH1* (p=8.07x10^-23; FC=42.15). In contrast, miRNA expression was most pronounced in POR CD, with 13 significantly (FDR<5% and >1.5-fold change) differential signals, while five and one in new and late CD. Integrating dysregulated genes and miRNAs found 64 miRNA-mRNA pairs with negative correlation in expression in POR CD, five of which experimentally supported in literature: hsa-let-7g-5p/*PRDM1* and hsa-let-7g-5p/*PTGS2*, hsa-miR-30d-5p/*SLC7A11* and hsa-miR-30d-5p/*WNT5A*, and hsa-miR-196a-5p/*ANXA1*. To be sure that POR expression is not influenced by previous disease (before surgery), we looked at gene expression in three CD patients with uninflamed post-operative ileum (i0), and six with POR i1 (≤5 aphthous lesions after surgery). There was only one significantly dysregulated gene (*WNT5A*) in i0, and 123 (including *WNT5A*) in i1 CD. *WNT5A* moreover had an increased expression with an increasing Rutgeerts' score (p<0.0001), and was dysregulated throughout different disease stages. WNT5A is a secreted glycoprotein that signals through the non-canonical pathway, and promotes crypt regeneration after tissue injury. Immunohistochemical staining for WNT5A is currently ongoing to narrow down which ileal regions show increased expression to help elucidate its role in early CD.

1962F

Comparative bacteria communities between gallbladder bile and gallstone in gallstone disease patients in Taiwan. *H. Yang[1], P. Yang[2,3,4], C. Tsai[2], C. Tsai[2], T. Liu[2], K. Ho[2], J. Lee[2], M. Li[2], D. Chu[2], S. Shyr[2], S. Lee[2], C. Chung[2], Y. Wu[2], Y. Lien[1], S. Shih[3,5,6], Y. Lee[1,7,8,9].* 1) Medical Research, Mackay Memorial Hospital, New Taipei City, Taiwan; 2) Department of Surgery, MacKay Memorial Hospital, Taipei, Taiwan; 3) Department of Medicine, MacKay Medical College, New Taipei City, Taiwan; 4) Department of Pharmacology and Graduate Institute of Medical Sciences, Taipei Medical University, Taipei, Taiwan; 5) Division of Gastroenterology, Department of Internal Medicine, MacKay Memorial Hospital, Taipei, Taiwan; 6) MacKay Junior College of Medicine, Nursing and Management, Taipei, Taiwan; 7) Department of Pediatric Endocrinology, MacKay Children's Hospital, Taipei, Taiwan; 8) Institute of Biomedical Sciences, MacKay Medical College, New Taipei City, Taiwan; 9) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan.

   Background and Goal: Gallstone disease (GSD) is one of the most common gastrointestinal disorders in Taiwan. This complex disorder most resulting from interactions between genetics and environmental factors. The incidence of GSD is about 5-14.3% in adult and correlated with age positively. It costs a lot in clinical and surgical treatment. Previous research indicated that lifestyle and other environmental factors are one of the risk factor in GSD. Recently, genetic factors are also affecting in pathogenesis of gallstone. Cholesterol-, mixed- and pigmented-stone are major types in GSD. Mixed- and pigmented-stone may correlate with bacterial infection. But, the relationship between bacteria communities and bile in GSD are not well discussed in Taiwan. In this study, we try to find the possible risk factors in bacteria affect in GSD. **Methods:** We enrolled 60 patients (40 females) in this study. Gallbladder bile and gallstone total DNA were extracted by QIAamp DNA stool mini kit. 16S rDNA sequencing will be performed by MiSeq. Preliminary analysis using MiSeq reporter. Statistical analysis using SPSS. **Results:** Although the majority of identified bacteria were diminished in bile samples, Enterobacteriaceae genera and Pyramidobacter were abundant in bile. **Conclusion:** Our study provides database for the potential source of gallbladder bile and gallstone bacteria, and illustrates the influence of GSD formation on bacterial communities.

1963W

101 novel loci and novel associations with gene expression detected in transethnic genome-wide study of estimated glomerular filtration rate in over 270,000 participants: The Million Veteran Program. *A.M. Hung[1], A. Giri[2], J. Hellwege[2], O. Wilson[1], E. Torstenson[2], C. Kovesdy[3], K. Birdwell[1], M. Li[9], E. Siew[1], M. Matheny[5], C. Chung[4], Y.V. Sun[6], P.W.F. Wilson[7], P.S. Tsao[8], C. Roumie[4], E. Velez Edwards[2], C. O'Donnell[10], T.L. Edwarrds[2] on behalf of the VA Million Veteran Program.* 1) Nephrology & Hypertension, VA TVHS & Vanderbilt University , Nashville, TN; 2) Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN; 3) Nephrology Section, Memphis VA Medical Center, Memphis TN; 4) VA TVHS Nashville, Department of Medicine, Vanderbilt University, Nashville, TN; 5) Department of Bioinformatics, Vanderbilt University, Nashville, TN; 6) Department of Epidemiology, Emory University Rollins School of Public Health; Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA; 7) Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA, 30233; 8) VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine; 9) Division of Nephrology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT 84132; 10) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School.

   Estimated glomerular filtration rate (eGFR) is a measure of kidney function and defines renal dysfunction and stages of chronic kidney disease (CKD). CKD affects over 14% of the US population, with higher prevalence in African Americans than non-Hispanic Whites. We have conducted the largest trans-ethnic genome-wide association study (GWAS) of eGFR to date in a well-defined cohort of African American (N = 56,253) and European American (N = 216,585) veterans participating in the Million Veteran Program (MVP). Genotypes were imputed based on the 1000 Genomes reference. eGFR was regressed onto additively encoded genetic variants, age, body mass index, sex, and 5 principal components. All analyses were performed in strata of hypertension status, diabetes status, and race, followed by fixed effects meta-analysis. We identified 13,261 SNPs in 157 loci at the significance threshold of $p = 5 \times 10^{-8}$ explaining 4.3% of the phenotypic variance, with the strongest signal replicating the previously detected association at *UMOD/PDILT* (p-value = 4.3×10^{-145}). Of these, 56 represented previously reported loci from GWAS of kidney function or CKD, and 101 are novel and explain 2.1% of eGFR variance. The strongest evidence for novel signals were near *TPPP* (rs454190; p-value = 2.26×10^{-19}) and near *NRIP1* (rs2823139; p-value = 1.61×10^{-18}). Mutations in *NRIP1* have been implicated in congenital anomalies of the kidney and urinary tract. We also performed genome-wide association analysis of genetically predicted gene expression (GPGE) and eGFR using the MetaXcan method in multiple tissues derived from the Genotype-Tissue Expression (GTEx) project. Predicted expressions of 209 genes were significantly (p-value < 5×10^{-8}) associated with eGFR in one or more tissues. Predicted expression of previously suspected genes *GATM* and *SPATA5L1* were strongly associated with eGFR in many tissues, notably in thyroid (p-value = 1.27×10^{-82}). Corresponding to novel signals from GWAS, associations between GPGE of nearby genes such as *TMUB1* in stomach, *SLC28A2* in nerve, and *SLC30A4* in subcutaneous adipose tissue were observed. As the largest transethnic GWAS of eGFR to date, we replicate over 50 previously identified loci, identify 101 novel loci and associate predicted expression of genes across several tissues with kidney function.

1964T

An analysis of Crohn's disease genes in the French-Canadian population. *B.E. Avila[1,2], M. Rivas[1,3], C. Stevens[1], J. Rioux[4,5], P. Goyette[4], G. Boucher[4], A. Bitton[6], M. Daly[1,2], NIDDK IBD Genetics Consortium.* 1) Medical and Population Genetics, Broad Institute, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Department of Biomedical Data Science, Stanford, CA, USA; 4) Montreal Heart Institute, Montreal, Quebec; 5) Université de Montréal, Montreal, Quebec; 6) McGill University Health Centre, Montreal, Quebec.

Exome sequencing has been powerful in determining variants with increased prevalence and effect size in contributing to risk of Inflammatory Bowel Disease (IBD) among population isolates. The French-Canadian population of Quebec has seen an intense founder effect more recently than many population isolates commonly studied today. In this study, we analyse 2,701 samples (with 667 cases, 1,020 controls, and 1,014 samples from complete trios) of French-Canadian origin for exome variants associated with Crohn's Disease in this population. Further, enrichment analysis based on a categorical Bayesian model allows us to determine which risk alleles are present in increased or decreased frequency due to the recent founder effect. The 338 complete trios give us the first opportunity to look sensitively at ultra-rare de novo variation in IBD, dissected by tolerance of genes to mutation and whether the gene has been potentially implicated based on proximity to one of the 250 confirmed IBD-associated loci discovered by GWAS. Similarly, we identify thousands of alleles which are specifically enriched from the founding population bottleneck, and explore their contribution to IBD in the same fashion.

1965F

Evaluation of candidate genes for Hirschsprung disease using target sequencing. *W.Y. Lam[1], C.S. Tang[1], P.C. Sham[2,3], P.K.H. Tam[1], S.S. Cherny[2], M.M. Garcia-Barcelo[1].* 1) Department of Surgery,The University of Hong Kong, Hong Kong; 2) Department of Psychiatry,The University of Hong Kong, Hong Kong; 3) Center for Genomic Sciences,The University of Hong Kong, Hong Kong.

Hirschsprung Disease (HSCR), or congenital aganglionic megacolon, is a relatively rare and complex genetic disorder characterised by the congenital absence of enteric neurons along a variable length of the distal intestine. Patients present with bowel obstruction and the disease may be fatal unless treated surgically. There is significant population variation in the incidence of the disease, and its incidence in Chinese is one of the highest in the world (1.4/5,000 live-births). Through the International Hirschsprung Disease Consortium exome sequencing project, candidate genes recurrently mutated with either *de novo* and/or inherited damaging variants were identified. In this follow-up study, we investigate if these genes are definitive genetic risk factors for HSCR in a larger number of patients. We screened 12 newly identified genes for mutations in 1200 Chinese and Caucasian HSCR patients, performed target sequencing using xGen®Lockdown® Custom Probes for enrichment of the sequences of the selected genes and high-throughput sequencing with depth 700X. Latest findings of this study will be presented to provide insights on whether the mutated genes identified through our exome study really involved in the pathogenesis of HSCR. This project is supported by the HMRF grant 02131866 to MMGB. .

1966W

Identification of blood UMOD and HER2 as causal mediators of chronic kidney disease using Mendelian randomization in the ORIGIN trial. *J. Sjaarda[1,2], H. Gerstein[1,2], S. Yusuf[1,2], J. Mann[3], D. Treleaven[1], M. Walsh[1,2], S. Hess[4], G. Pare[1,2].* 1) McMaster University, Hamilton, Ontario, Canada; 2) Population Health Research Institute, Hamilton, Ontario, Canada; 3) University of Erlangen-Nuremberg, Bavaria, Germany; 4) Sanofi Aventis, Frankfurt, Germany.

   BACKGROUND: Identification of biomarkers for chronic kidney disease (CKD) may lead to important advances in both prevention and treatment, particularly if they are causally linked. Many biomarkers have been epidemiologically linked with CKD, however, the possibility that such associations may be due to reverse causation or confounding limit their utility. This limitation can be overcome using a Mendelian randomization (MR) approach. We therefore used this technique to identify novel, causal mediators of CKD. **METHODS:** MR was performed by first identifying genetic determinants of 227 protein biomarkers assayed in 4,147 ORIGIN trial (Outcome Reduction with Initial Glargine Intervention) participants, and assessing their effects on CKD in the CKDGEN consortium (N=117,165, 12,385 cases) using the inverse-variance weighted method. The relationship between the serum concentration of each biomarker identified using this approach and incident CKD in ORIGIN participants was then estimated. **FINDINGS:** Uromodulin (UMOD) and human epidermal growth factor receptor 2 (HER2) were identified as novel, causal mediators of CKD using MR (UMOD: OR=1.30 per SD; 95% CI 1.25 to 1.35; p<5x10^{-20}; HER2: OR=1.30 per SD; 95% CI 1.14 to 1.48; p=8.0x10^{-5}). Consistent with the MR findings, blood HER2 was also associated with CKD events in ORIGIN (OR=1.07 per SD; 95% CI, 1.01 to 1.13; p=0.01). Additional exploratory MR analyses identified angiotensin converting enzyme (ACE) as a regulator of HER2 levels (β=0.13 per SD, 95% CI 0.08 to 0.16, p=2.5x10^{-7}). This latter finding was corroborated by an inverse relationship between ACE-inhibitor use and HER2 levels (β=-0.25 SD, 95% CI -0.30 to -0.20, p<5x10^{-16}). **INTERPRETATION:** UMOD and HER2 are independent causal mediators of CKD in humans, and HER2 levels are regulated in part by ACE. Both these biomarkers are potential therapeutic targets for CKD prevention.

1967T

Mendelian randomization shows non-linear causality between vitamin D levels and kidney function. *A. Teumer[1,2], G. Gambaro[3], C. Pattaro[4], A. Köttgen[5], P.M. Ferraro[3] on behalf of the CKDGen Consortium.* 1) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 2) DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany; 3) Division of Nephrology, Fondazione Policlinico Universitario A. Gemelli, Catholic University of the Sacred Heart, Rome, Italy; 4) EURAC Research Institute for Biomedicine, Bolzano, Italy; 5) Institute of Genetic Epidemiology, Medical Center - University of Freiburg, Germany.

   Vitamin D is important for regulating the concentration of calcium in the bloodstream and for healthy growth of bones. The prohormone vitamin 25(OH)D (calcidiol) is produced in the liver and converted in the kidney to the active form 1,25(OH)$_2$D (calcitriol). Vitamin 25(OH)D is measured in serum as a marker for vitamin D. Observational studies suggest an association between lower levels of vitamin D and impaired kidney function, although the observed effect direction varied. Confounding might at least partly explain such findings. We performed Mendelian randomization (MR) analysis to assess whether circulating vitamin D levels are causally associated with renal function, quantified by the estimated glomerular filtration rate (eGFR) and urinary albumin-to-creatinine ratio (UACR). We applied a two-sample MR triangulation analysis with standard errors estimated by the delta method (Thomas *et al.*, AEP 2007) using rs10741657 (*CYP2R1*), rs12785878 (*DHCR7*) and rs2282679 (*GC*) as instruments. Effect estimates of these SNPs were obtained from published GWAS meta-analysis results of vitamin D (Wang *et al.* 2010, n=33,868), eGFR (Pattaro *et al.* 2016, n=133,723) and UACR (Teumer *et al.* 2015, n=54,450), respectively. All traits were log-transformed. Overall, a significant inverse causal effect of vitamin D levels on kidney function could be shown for eGFR (beta=–0.013, p=0.003). This result was validated in a one-sample 2-stage least-squares MR analysis in 15,481 individuals not included in the published vitamin D meta-analysis. Stratification of observational data by chronic kidney disease (CKD), defined as eGFR <60 ml/min/1.73m², revealed an inverse vitamin D - eGFR association in 16,902 non-CKD individuals (beta=-0.022, p=6.6E-60) and a positive association in the 1,275 CKD individuals (beta=0.015, p=6.5E-4). Vitamin D was not causally associated with UACR (beta=0.059, p=0.179). The overall finding of an inverse causal relationship between vitamin D and eGFR is opposite to findings from observational studies, but the causal effect is very small: 10% vitamin D increase leads to 0.1% eGFR decrease. The direction of observational associations of vitamin D and eGFR changed in those with low kidney function, validating the contradicting results of the MR analyses and former observational studies. Our study gives insight into the relation of vitamin D and kidney function. Further investigation is warranted to fully understand the biology leading to the causal association.

1968F

Differential metagenomic analysis associated between alcoholic and non-alcoholic fatty liver disease using 16S rRNA gene sequencing.
Y. Yun[1], H. Kim[1], Y. Chang[2], S. Ryu[2], H. Shin[2], H. Kim[1]. 1) Department of Biochemistry, Ewha Medical Research Institute, School of Medicine, Ewha Womans University; 2) 2Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, School of Medicine, Sungkyunkwan University, Seoul, South Korea.

Alcoholic liver disease (ALD) and non-alcoholic fatty liver disease (NAFLD) share similar histological findings and many common pathogenic mechanisms, however which occur as a result of different causes, alcohol and obesity, respectively. Therefore, it's important to identify the differential disease progress which ends up to the same consequence. Emerging data demonstrate dysbiosis of the gut microbiota to be linked with ALD and NAFLD. But there is no report of comparing gut microbiota of these two condition in human. We used the data from a cross-sectional study of the Kangbuk Samsung Hospital Health Screening Center, between 23 June 2014 and 5 September 2014, which comprise three groups of 760 normal control, 297 NALFD, and 162 ALD. The fecal microbiota community of total 1219 samples was analyzed by pyrosequencing of the V3-V4 domain of the 16S rRNA genes. Alpha-diversity of ALD and NAFLD by phylogenetic diversity (PD) whole tree analysis was similar, while both were significantly lower compared to control ($p<0.001$). Three group showed significant far distant phylotypes between each other by ANOSIM test ($p<0.001$) of weighted UniFrac analysis. The decrease of Cyanobacteria and the increase of *Fusobacterium* were commonly associated in ALD and NAFLD. In contrast, Paraprevotellaceae and *Desulfovibrio* in ALD and *Methanobrevibacter* in NAFLD showed negative association uniquely. This might suggest a key distinguishable pathogenic mechanism between ALD and NALFD, which will contribute therapeutic development. [This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03035501, NRF-2016R1A6A3A1193 2719), Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare (HI14C0072), and Intramural research support program of Ewha Womans University School of Medicine.].

1969W

Discovery of health disparities among African American patients at Vanderbilt University Medical Center.
K. Actkins[1,3], P. Straub[2,3], N. Cox[2,3], L. Davis[2,3] on behalf of the Precision Medicine and Health Disparities Collaborative. 1) Department of Physiology, Meharry Medical College, Nashville, TN; 2) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN.

Knowledge of population-specific disease risk is critical to the mission of personalized medicine. Health conditions known to disproportionly affect African Americans including stroke, heart disease, and type 2 diabetes are not only more common, but are also more likely to result in complications and morbidity. Unrecognized health disparities pose a significant challenge to the promise of personalized medicine. We sought to discover novel health disparities in an African American patient population using the VUMC biobank (BioVU), a repository of DNA extracted from discarded blood collected during routine testing and linked to de-identified electronic medical records (EMR). We conducted a study of 29,370 individuals genotyped on the Illumina Human Exome Bead Chip and EIGENSTRAT was used to calculate principle components based on 4,388 ancestry informative markers. The first principal component (PC1) was found to represent the proportion of African ancestry. We performed a phenome-wide association study (PheWAS), with PC1 as the dependent variable (smaller values indicting greater African ancestry) and sex as a covariate, to determine the association between African ancestry and 924 clinical phenotypes (each with at least 100 cases) extracted from the EMR. Individuals were considered "cases" if their EMR included least 3 occurrences of a qualifying ICD code. A total of 232 phenotypes exceeded the Bonferroni corrected p-value (5.41E-5) for statistical significance. Hereditary hemolytic anemias (including sickle cell anemia) (OR = 0.92, p = 1.76E-100) were the most significantly associated phenotypes. Additional known associations with African ancestry included asthma (OR = 0.97, p = 1.57E-57) and, among adults aged 18 or older, end stage renal disease (OR = 0.96, p = 9.03E-70), and chronic renal failure (OR = 0.97, p = 1.74E-68) were also confirmed. We also identified a protective effect of African ancestry for "malignant melanomas" (OR = 1.07, p = 8.00E-39). Finally, we discovered several novel associations with African ancestry including "fever of unknown origin" (OR = 0.98, p = 8.97E-78), "acute upper respiratory infections" (OR = 0.99, p = 2.02E-72), and "inflammatory diseases of female pelvic organs" (OR = 0.96, p = 9.91E-42). An important limitation is the inability to distinguish genetic from environmental mediators of our reported associations. Future directions include incorporation of socioeconomic variables which may be mediating our results.

1970T

Fetal but not maternal APOL1 genotype is associated with increased risk for preeclampsia among African-Americans. *C.L. Simpson[1], K.J. Reidy[2], R. Horton[2], A.Z. Rosenberg[12], S.D. Rosenblum[4], C. Kovesdy[6], F. Tylavsky[7], J. Myrie[5], B. Ruiz[5], K. Mozhui[1,7], S. Haque[5], M. Suzuki[8], S.E. Reznik[9], F. Kaskel[2], J.B. Kopp[11], C.A. Winkler[10], R.L. Davis[3,11].* 1) Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN; 2) Department of Pediatrics, Division of Pediatric Nephrology, Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY; 3) Center for Biomedical Informatics, Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN; 4) Department of Pediatrics, Division of Neonatology, Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY; 5) Albert Einstein College of Medicine, Bronx, NY; 6) Department of Medicine, University of Tennessee Health Science Center, Memphis, TN; 7) Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN; 8) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 9) Departments of Pathology and Obstetrics & Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, NY; 10) Basic Research Laboratory, Center for Cancer Research, National Institutes of Health, Frederick, MD; 11) Kidney Disease Section, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Bethesda, MD; 12) Department of Pathology, Johns Hopkins University School of Medicine,Baltimore MD.

Background: African Americans are at increased risk for preeclampsia. Genetic variants in the apolipoprotein L1 (*APOL1*) gene account for a substantial fraction of increased risk of kidney disease among African Americans. APOL1 is expressed in human placenta, and transgenic mice expressing APOL1 developed preeclampsia. The role of APOL1 variants in human preeclampsia has not been studied. Methods: Two studies were performed evaluating maternal and fetal *APOL1* genotypes in African American women with preeclampsia. At Albert Einstein College of Medicine (AECOM) Affiliated Hospitals, we studied 122 pregnancies in African American women with preeclampsia. At University of Tennessee Health Science Center (UTHSC), we studied 93 pregnancies in African American women with preeclampsia compared to 793 control birth mothers and infants. Results: In both studies, fetal *APOL1* high risk (HR) genotype comprising 2 risk alleles was associated with preeclampsia in their mothers, relative risk at AECOM 1.65 (95% CI 1.11, 2.44) and odds ratio at UTHSC 1.92 (1.05,3.49). In both studies, maternal *APOL1* HR genotypes were not associated with preeclampsia. Gestational age, birth weight and Cesarean section did not vary by *APOL1* genotype. Fetal *APOL1* HR genotype births with preeclampsia were not more likely to have severe preeclampsia, but those mothers were more likely to have cerebral or visual disturbances (63% versus 37%, p = 0.04) and infants had lower APGAR scores at 5 minutes (8.0 versus 9.0, p = 0.01). Conclusion: Fetal *APOL1* high-risk genotype confers an increased risk for preeclampsia, likely by adversely affecting placental function. *APOL1* genetic testing may have a clinical role to predict and perhaps improve pregnancy outcomes.

1971F

Patterns of *APOL1* G1 association with kidney function in young adults in five African diaspora populations. *B. Tayo[1], H. Kramer[1], A. Luke[1], P. Bovet[2], T. Forrester[3], E. Lambert[4], J. Plange-Rhule[5], R. Cooper[1], R. Durazo-Arvizu[1].* 1) Public Health Sciences, Loyola Univ Chicago, Maywood, IL; 2) Institute of Social & Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland & Ministry of Health, Republic of Seychelles; 3) Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston, Jamaica; 4) Research Unit for Exercise Science and Sports Medicine, University of Cape Town, Cape Town, South Africa; 5) Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

Introduction: The excess risk of end-stage renal disease in African Americans is now known to be attributable in part to *Apolipoprotein L1* gene (*APOL1*) risk variants G1 and G2 which are associated with higher risk of chronic kidney disease (CKD) initiation and progression in African Americans. Although these variants are specific to African ancestry populations, their frequencies vary among Afro-origin populations. It can be assumed that native African and other African diaspora populations share the increased predisposition to CKD as result of a common genetic background. **Goal:** The aim of the present study was to evaluate association of *APOL1* G1 with urine albumin and estimated glomerular filtration rate (eGFR) as markers for CKD in young adults from five African Diaspora populations. **Methods:** We measured urine albumin, serum creatinine and serum cystatin C in a cohort of 1561 non-diabetic unrelated adults from Ghana (n=211), Jamaica (n=383), Seychelles (263), South Africa (302) and the USA (402). Urine albumin-to-creatinine ratio (UACR), creatinine-based eGFR (eGFR-Scr) and cystatin C-based eGFR (eGFR-Scys) were calculated. All subjects were genotyped for *APOL1* rs73885319 and rs60910145 that were used to determine G1 haplotype. We tested for association of *APOL1* G1 with natural logarithm transformed UACR and eGFR under recessive genetic model using linear regression and controlled for age, sex and study site. Population stratification was controlled for by inclusion of principal components estimated from ancestry informative markers. CKD could not be analyzed as a dichotomous variable because there were few cases. **Results & conclusion:** Mean age and range in the cohort were 34.4±6.2 and 23–45 years. Mean UACR, eGFR-Scr and eGFR-Scys were 12.9±71.3 mg/g, 143.7±18.5 and 124.3±15.4 mL/min/1.73m^2, respectively. Frequency of *APOL1* G1 was 21.2% overall; and 46.1%, 28.4%, 4.6%, 10.0% and 21.2%, in Ghana, Jamaica, Seychelles, South Africa, and USA, respectively. Associations with *APOL1* G1 in the cohort were significant for eGFR-Scr ($p=1.2 \times 10^{-4}$) and UACR ($p=0.0281$), and marginally for eGFR-Scys ($p=0.0697$). There were significant associations ($p<0.05$) with eGFR-Scr and UACR in the Ghana samples. These data represent the first report of patterns of *APOL1* G1 associations among young non-diabetic adults in African Diaspora populations. Additional studies that include both *APOL1* G1 and G2 risk variants are required to confirm these findings.

1972W

Integrated linkage and rare variant association tests reveal rare variants associated with elevated androgen levels in polycystic ovary syndrome. *M. Dapas[1], R. Sisk[1], R.S. Legro[2], M. Urbanek[1], A. Dunaif[1,3], M.G. Hayes[1].* 1) Northwestern University, Chicago, IL; 2) Penn State College of Medicine, Hershey, PA; 3) Icahn School of Medicine at Mount Sinai, New York, NY.

Polycystic ovary syndrome (PCOS) is a common endocrine disorder affecting up to 15% of premenopausal women worldwide. Common genetic susceptibility variants for PCOS account for little of the observed heritability. To test whether rare genetic variants were responsible for this heritability deficit, we implemented a novel bioinformatics pipeline that integrated linkage and rare variant association tests using whole genome sequencing data from 76 two-generation families with ≥1daughters with PCOS. Variants were filtered according to allele frequency (MAF≤2%), call quality, observed inheritance, predicted deleteriousness, and evidence for linkage. Associations between sets of rare variants and PCOS and its quantitative hormonal traits were assessed using burden tests, grouping variants at the gene-level (including 3' UTR and 7.5kb upstream of 5' TSS) and in sliding windows across the genome, accounting for relatedness, and adjusting for age and BMI. After correcting for multiple testing (P_c), associations with dehydroepiandrosterone levels were found for rare variants in *FANCL* ($P=2.61×10^{-6}$, $P_c=0.030$), *PTCH1* ($P=6.57×10^{-8}$, $P_c=0.0081$), and an intergenic region upstream of *SFRP5* ($P=2.38×10^{-8}$, $P_c=0.0029$). Associations were also found with luteinizing hormone levels for rare variants in *RHOD* ($P=3.54×10^{-6}$, $P_c=0.041$). The strongest association of rare variants with testosterone (T) levels was found in the PCOS candidate gene *SRD5A2* ($P=1.03×10^{-5}$, $P_c=0.12$). No significant associations were found with PCOS diagnosis, fasting insulin or sex hormone binding globulin levels. *FANCL* encodes a protein that is part of the Fanconi anemia (FA) core complex. The FA core complex also includes FANCC, whose gene is located in a PCOS GWAS susceptibility locus on chr. 9. The GWAS risk alleles in the *FANCC* locus are also in linkage disequilibrium with the rare variants we identified in nearby *PTCH1*, suggesting these variants contribute to that GWAS signal. *SFRP5*, which encodes a Wnt inhibitor, is differentially expressed in women with PCOS. *RHOD* is a regulator of actin cytoskeleton dynamics, a pathway linked to PCOS insulin resistance. *SRD5A2*, which metabolizes T, has previously been implicated in PCOS pathogenesis. Our study is the first to find evidence that rare variants contribute to elevated androgen levels in PCOS. Moreover, the variants identified in this study all lie in regions that correspond with relevant disease pathways, GWAS loci and/or established PCOS candidate genes.

1973T

Genome-wide CNV analysis identifies *TGFBR3* as a candidate causal gene for endometriosis and infertility. *E. Hatchwell[1], K.T. Zondervan[2,3], P.S. Eis[1].* 1) Population Bio, Inc., Melville, NY; 2) Endometriosis CaRe Centre, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, John Radcliffe Hospital, Oxford, UK; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

Endometriosis is a gynecological complex disease characterized by chronic inflammation, pelvic pain, and reduced fertility. It is estimated to occur in 5-10% women during their reproductive years and is frequently undiagnosed (e.g., the degree of pain does not necessarily correlate with the severity of the disease). Endometriosis heritability has been estimated at ~50%. Linkage studies on >1,100 affected sister-pair families with a history of endometriosis revealed significant linkage peaks mapped to chromosomes 7 (in 2007) and 10 (in 2005) but the variants underlying these signals have not yet been found. The most recent GWAS on >17,000 cases vs. 191,000 controls revealed 14 genetic loci, but the associated non-coding SNPs have small effect sizes, that collectively account for only 5% of disease risk, and leave as yet unanswered questions regarding the associated causal mechanisms. We used a copy number variant (CNV) method to identify additional genes/variants underlying the pathogenesis of endometriosis. Our CNV-based disease gene discovery method has two advantages over SNV-based methods: 1) variant interpretation is simpler because there are >1,000-fold fewer CNVs in the population as compared to SNVs, and 2) the large size of CNVs often unambiguously results in loss of function of the implicated gene. We performed genome-wide CNV analysis of blood-derived genomic DNA from endometriosis patients (all diagnosed via laparoscopic surgery) using array-based comparative genomic hybridization (aCGH). Interpretation of benign vs. potentially pathogenic CNVs was performed using our in-house CNV database and algorithms. The top gene that carried a potentially pathogenic CNV identified in our study was transforming growth factor, beta receptor III (*TGFBR3*, aka betaglycan). An intronic 8Kb deletion, which impacts predicted transcription factor binding sites, was found in 5 of 911 cases vs. 0 of 1,000 controls (OR = 12.1 (95% CI: 0.67-219.9), $P_{exact} = 0.024$). In the subset of moderate/severe cases, the deletion was found in 4 of 449 cases (OR=20.2 (95% CI: 1.08-376.2; $P_{exact}=0.009$). Segregation analysis of the *TGFBR3* deletion correlated with disease in at least 3 of 4 families that could be tested. Supporting biology for *TGFBR3* includes high gene expression in reproductive tissues (uterus, ovary, and placenta), involvement in the activin/inhibin pathway, and links to infertility, reproductive tissue cancers (breast, endometrial, and ovarian), and pain.

1974F

Identification of lung cell populations from single-cell RNA-seq profiling of murine emphysema model. *J.H. Yun[1,2], C.H. Lee[3], R. Kirchner[4], S. Boswell[5], P. Castaldi[2], L. Pinello[6], X. Zhou[2].* 1) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Harvard Chan Bioinformatics Core, Boston, MA; 5) Single Cell Core, Harvard Medical School, Boston, MA; 6) Department of Pathology, Massachusetts General Hospital, Boston, MA.

Background: Emphysema is a chronic lung disease characterized by destruction of alveolar space and capillaries that constitutes chronic obstructive pulmonary disease (COPD) along with chronic bronchitis. Hhip (Hedgehog Interacting Protein) is a well-replicated GWAS-identified gene associated with COPD, and we have shown that *Hhip* haploinsufficient mice (*Hhip[+/-]*) developed age related emphysema and increased lymphocyte aggregates, similar to emphysema in human patients. The lung is composed of over 40 cell types, and multiple cell types are implicated in the development of emphysema. As traditional bulk RNA profiling approaches are limited in understanding the specific role of each cell types in emphysema pathogenesis, we performed single cell RNA sequencing (scRNA-seq) in wild type and *Hhip* heterozygous mice before and after the onset of emphesema.   Methods: Lungs isolated from 2 and 8 months old *Hhip[+/-]* and its littermate C57BL6 wild type mice (n=2 each) were mechanically minced, enzymatically dissociated and gradient centrifuged to generate epithelial-cell enriched single cell suspension. scRNA-seq was then performed on 3,000 unsorted cells per murine lung using inDrops, where the cell capture, lysis and reverse transcription of RNA from single cells occurs in droplets. cDNA libraries were constructed and paired end sequencing was performed with Illumina NextSeq. Reads were filtered and aligned using established inDrops processing pipeline. Data analysis was performed in R using the Seurat package.   Results: Unsupervised clustering revealed clusters with expression patterns consistent with known lung tissue cell types such as alveolar cell types (Type I, II), endothelial cells, fibroblasts, and immune cells (macrophages, myeloid cells) as well as distinct clusters of apparently novel cells types with distinct molecular cell profiles. With the relatively large number of cells profiled, potential rare and novel cell populations may be identified and validated by histology. Older *Hhip[+/-]* mice exhibited a higher proportion of immune cell types than their wildtype controls, consistent with pathologic observations of increased lymphoid aggregates in the older *Hhip[+/-]* mice. In addition, cell types enriched with Hhip expression could be identified.   Conclusion: Unbiased single cell RNAseq in murine lung identifies known and potentially novel cell types and confirms an increased proportion of inflammatory cells in the aged lung of Hhip heterozygous mice.

1975W

The multi-phenotype derived Nephrotic Syndrome Severity (NS2) score empowers genomic discovery. *C.E. Gillies[1], K. Yasutake[1], X. Wen[2], M.G. Sampson[1].* 1) Department of Pediatrics-Nephrology, University of Michigan , Ann Arbor, MI; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Background: Among patients with nephrotic syndrome (NS), we are interested in discovering genomic factors associated with disease severity over time. This requires creating accurate models to represent NS biology and a statistical strategy that takes advantage of deep phenotypic data but accounts for modest sample sizes and multiple testing. Thus, we developed a Nephrotic Syndrome Severity ("NS2") score for patients in the Nephrotic Syndrome Study Network (NEPTUNE).   **Methods:** NEPTUNE is a prospective, longitudinal study of NS enrolling affected adults and children receiving a clinically indicated biopsy. Rich demographic and clinical data are collected at baseline and over time. Genomic and histologic data are collected at baseline. We used the following parameters to create the multi-phenotype NS2 score: interstitial fibrosis, eGFR, protein-to-creatinine ratio, eGFR slope, time to complete remission, and time to a composite endpoint. We modeled the relationships between these variables and meaningful covariates using a Bayesian network (BN). The NS2 score represents a latent factor explaining the correlations in the observed data. The BN's parameters were inferred from 616 patients NEPTUNE participants using Markov chain Monte Carlo.   **Results:** Compared to existing multi-phenotype methods, NS2 score increased power for discovery without inflating Type I error. With regards to known biomarkers of NS severity, a worse NS2 score was significantly associated with the *APOL1* high-risk genotype in black patients (p< 2.2 e-5) and lower tubulointerstitial expression of *EGF* (p< 5.7 e-10). After FDR control, 1,040 glomerular transcripts were significantly associated with NS2 score. Using geneset enrichment analysis, kidney development genes were among the most enriched NS2-associated glomerular transcripts in adults (p<5.3 e-9), including 15 known Mendelian SRNS genes. In children, "TNF alpha induced protein 3" was the gene whose expression was most associated with NS2 score.   **Conclusion:** The NS2 score is a robust metric created by capitalizing on extensive clinical data and NS-specific knowledge. As a robust multi-phenotype method, it improved statistical power for discovery without inflating false positives and replicates known genomic associations. Thus, using NS2 score as an outcome measure in analyses ranging from gene expression correlation to GWAS may empower genomic discoveries.

1976T

Transcriptomic analysis of the ratio of serum aspartate transaminase to serum alanine transaminase (Ast/Alt ratio) using a genotype-by-diet interaction model identifies a number of potentially important genes for liver disease in the San Antonio Family Heart Study. *V.P. Diego, J.M. Peralta, M. Almeida, H.H.H. Göring, M. Johnson, J.E. Curran, S. Williams-Blangero, J. Blangero.* South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX.

The ratio of serum aspartate transaminase to serum alanine transaminase (Ast/Alt ratio) is a well-known biomarker of liver function in general and nonalcoholic fatty liver disease (NAFLD) in particular. The latter, NAFLD, has risen in prevalence largely in conjunction with the rise of metabolic syndrome, type 2 diabetes (T2D), and obesity, and this is especially true in Mexican Americans. We therefore sought to better understand the genetic determinants of NAFLD in the Mexican American study participants of the San Antonio Family Heart Study. Several studies have shown that diet significantly modulates liver function and Ast/Alt ratio. Thus, in our model we accounted for any potential diet effect by way of a genotype-by-diet interaction (GDI) model. To model a dietary effect we estimated the Mahalanobis distance (MD) kernel over all vectors of dietary data in our study, and used this MD kernel to estimate a corresponding dietary variance component (VC). The dietary variables consist of major macronutrients (i.e. mono- and polyunsaturated fats; saturated fats; carbohydrate; proteins; alcohol; and simple sugars) measured by percent of caloric intake and by estimated grams of the macronutrient. Under the GDI model, we modeled interaction by taking the Hadamard product of our genetic relationship kernel with the MD kernel and using this new kernel to estimate a GDI VC. With this model as our foundation, we then tested for the significance of a transcript by adding a single transcript to the model as an independent predictor variable. This analysis was performed over 22,353 gene transcripts found to be significantly heritable in our previous work. We found 10 gene transcripts to be significantly associated with Ast/Alt ratio, where their p-values fell under the corrected significance threshold. Under a less stringent correction, we found 29 more gene transcripts to be suggestively associated with Ast/Alt ratio. These results show that there are important genetic determinants underlying the variation in Ast/Alt ratio and, presumably, are involved in the progression to NAFLD.

1977F

Chromatin interactions reveal novel gene targets for drug repositioning in rheumatic diseases. *P. Martin[1], K. Duffus[1], A. McGovern[1], A. Yarwood[1], A. Barton[1,2], J. Worthington[1,2], S. Eyre[1], G. Orozco[1].* 1) Arthritis Research UK Centre for Genetics and Genomics, University of Manchester, Manchester, United Kingdom; 2) NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom.

The treatment of rheumatic diseases can be both expensive and ineffective with up to 1/3 of patient's failing to respond to current treatments and costing the NHS around $4.5 billion annually. There is therefore a need to identify new treatments and to target these to individual patients. Although genetic studies have been successful in identifying common variation associated with disease susceptibility, a large proportion of these lie outside protein-coding regions. Many show enhancer activity but it is often unclear which gene(s) they regulate and how they contribute to disease. Chromatin folding brings linearly distant areas of the genome, such as promoters and enhancers, into close proximity, driving gene expression. Capture Hi-C (CHi-C) interrogates these interactions in a high-throughput, high-resolution manner, linking implicated enhancers to causal genes. We explored our existing CHi-C data on 3 rheumatic diseases, rheumatoid arthritis, juvenile idiopathic arthritis and psoriatic arthritis, targeting all known genetic associations, for the potential to identify candidate causal genes that are targets for existing drugs, which could be repositioned for use in these diseases. Chromatin interaction data for T- and B-cells in the 3 diseases was analysed using CHiCAGO v2, interactions between disease enhancer regions and active promoters were identified using BEDTOOLS v2.21.0 and intersected with drug targets from DrugBank v5.0.1. Overall 454 genes were identified as interacting with a disease associated region showing enhancer activity. Of these, 41 are existing drug targets (92 drugs), defined by the presence of the relevant name in the 'indication' field, and 9 are existing therapies used in the treatment of RA. Following pathway analysis of the implicated genes we identified a potential 275 drugs (202 genes) which could be repositioned. Importantly, over 90% of the drugs currently used to treat RA were identified showing the effectiveness of this approach. Our study also identified potential new targets within enriched pathways. We have identified genes which are implicated in disease, are the target of existing drugs and offer the potential for drug repositioning as well as targets involved in the same pathway. This data shows a novel insight into how functional annotation of genetic associations in rheumatic diseases can provide gene targets for re-positioned therapies.

1978W

Genetic burden contributing to extremely low or high bone mineral density in a senior male population from MrOS study. *S. Chen[1], M. Jain[2], S. Jhangiani[5], Z. Coban Akdemir[1], P. Campeua[3], B. Klein[4], C. Nielson[4], D. Muzny[6], E. Boerwinkle[7], R. Gibbs[1,5], E. Orwell[4], J. Lupski[1,5,6], J. Posey[1], B. Lee[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Osteogenesis Imperfecta Clinic at Kennedy Krieger Institute, Baltimore, MD; 3) University of Montreal Medical Genetics Service, Montréal QC Canada; 4) Oregon Health & Science University Portland, Oregon; 5) Human Genome Sequencing Center Baylor College of Medicine, Houston, TX; 6) Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, TX; 7) Human Genetics Center and Dept. of Epidemiology , UTHealth School of Public Health, Houston, TX.

   Purpose of the study Worldwide, 1 in 5 men aged over 50 will experience osteoporosis and a clinical bone fracture and suffer from a greater fracture related mortality rate than women. However, the mechanism(s) underlying osteoporosis in men remains unknown. We aimed to identify genes associated with extremely low or high bone mineral density (BMD) as a potential means to elucidate contributing gene variants involved and possibly gain an understanding of biological processes underlying bone density loss that may point to potential therapeutic targets for increasing bone mass. **Methods** Subjects (<75) from the MrOS cohort were screened for BMD abnormalities consistent with phenotypic extremes. Low BMD is defined as having both hip DXA and vertebral QCT values in the 1st or 2nd decile while high BMD is defined as having both values in the 9th or 10th decile. To reduce inadvertent inclusion of low BMD resulting from a Mendelian disease, those with BMD beyond 3 sd were excluded. Controls were from CEU population of 1000 Genomes. In total, 90 low BMD, 99 high BMD and 82 controls were identified for WES at the Baylor HGSC as part of the BHCMG program. WES data were filtered and parsed for rare variants with allele frequency < 1% in Caucasian population. Fisher's Exact test was performed to compare low BMD or high BMD subjects to controls for single-gene associations. Additionally, sets of genes causative of either Osteogenesis Imperfecta (OI) or Ehlers-Danlos syndrome (EDS) were grouped together for multi-gene analyses. **Results** No single gene associations with rare variants were found for either low BMD (n=30) or high BMD (n=18). Of note, rare variants in *FKBP10, SERPINF1, SERPINH1, SP7, TMEM38B* and *WNT1* were only detected in low or high BMD individuals but not in controls. By grouping 'OI genes' (*CRTAP, FKBP10, LEPRE1, PPIB, SERPINF1, SERPINH1, SP7, TMEM38B, WNT1*) together, we were able to detect a significant 3-fold increased accumulation of rare variants in low BMD subjects compared to controls (p=0.009). Additionally, genes associated with EDS (*SLC39A13, B4GALT7, COL3A1, COL5A1, COL5A2*) demonstrated a significant 2-fold increased frequency in low BMD subjects compared to controls (p=0.03). Overall, our study demonstrated an association between the accumulation of rare variants with low BMD, by using a multi-gene signature implicated in OI or EDS. These findings can direct a gene-panel approach to screen for additional multi-variant associations in a larger cohort.

1979T

An integrative analysis of gene expression profiling and genome-wide DNA methylation datasets shows a different underlying molecular mechanism between Kashin-Beck disease and osteoarthritis. *Y. Wen, P. Li, W. Wang, J. Hao, J. Han, A. He, Y. Du, L. Liu, X. Liang, X. Guo, F. Zhang.* School of Public Health, Health Science Center, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

   Background: Kashin-Beck disease (KBD) is an endemic and chronic osteochondropathy with unknown etiology, which mainly occurs to children aged from 3 to 13. Osteoarthritis (OA) is a kind of degenerative joint disease distributing worldwide, sharing some characteristics on the manifestation and pathological changes of the articular with KBD. This study is to investigate the differences in gene expression profiling and DNA methylation level of articular cartilage from KBD versus those from OA. Methods: In this study, we conducted a genome-wide DNA methylation study in the cartilage chondrocytes of KBD compared to those of OA with Illumina Infinium HumanMethylation450 BeadChip. Furthermore, by utilizing a gene expression profiling dataset in the articular cartilage of KBD compared to OA published previously, an integrative analysis was conducted using the two datasets in IncroMap. Result: By performing integrative enrichment analysis, 19 KEGG pathways were identified to be alerted in the cartilage chondrocytes of KBD relative to OA both in the gene expression and DNA methylation levels, including mTOR signaling(P = 0.0301), glycosaminoglycan biosynthesis-chondroitin sulfate/dermatan sulfate(P = 0.0391), glycosaminoglycan biosynthesis-keratan sulfate(P = 0.0278) and PI3K-Akt signaling(P=0.0243). Several genes, such as *TGB3* (expression ration = 0.6643, P_{methy} = 0.0164), *TGFBR2* (expression ration = 1.5432, P_{methy} = 0.0370), *JAK1* (expression ration = 1.5282, P_{methy} = 3.97E-03) were also found to be commonly differently methylated and expressed in KBD relative to OA. Conclusion: Our study results illustrated a different molecular mechanism and biological networks underlying KBD compared to OA.

1980F

Identification of genetic variants associated with knee OA in patients with knee replacement surgery. *Y. Zhang[1], S.A. Lietman[2], S.A. Pendergrass[3], G. Bang[3], J.E. Miller[3], D. Kim[3], K.A. Kost[4], M.T.M. Lee[1].* 1) Geisinger Genomic Medicine Institute, Geisinger Health System, Danville, PA; 2) Orthopedic Department, Geisinger Health System, Danville, PA; 3) Biomedical & Translational Informatics Institute, Geisinger Health System, Danville, PA; 4) Phenomics Core, Geisinger Health System, Danville, PA.

Background: GWAS have identified approximately 25 susceptibility loci associated with osteoarthritis (OA). However, only a few of them have been replicated. For knee OA, most studies used Kellgren and Lawrence (K-L) grade as a way to define cases (>=2) and controls (0 and 1). However, the use of K-L grade is very subjective and neglects the variability of OA phenotypes. In addition, existing studies have not controlled for many risk factors during study design, such as injury or trauma history, hip dysplasia, and age. Thus, heterogeneity of the disease and inadequate study designs may cause the lack of replication. To overcome these limitations, we conducted a knee OA GWAS in patients with severe (end stage) OA who underwent knee replacement surgery from the Geisinger Health System MyCode® Community Health Initiative. Method: Illumina OMNI Express array genotyping data (MAF>1%) from the DiscovEHR collaboration of the MyCode Community Health Initiative at Geisinger and Regeneron Genetics Center was used for this study. EHR data was used to retrospectively identify cases and controls. Patients with rheumatoid arthritis and knee trauma/injury were excluded from analysis. Cases were patients who had knee OA diagnosis and underwent total knee joint replacement. BMI and gender matched MyCode participants who were > 65 years old and without any history of OA in any joints were selected as controls. 1938 cases and 3293 controls from 59,499 genotyped participants in MyCode were identified. Models were adjusted for age, sex, BMI, smoking, diabetes status, and the first two principle components. Results: Seven SNPs with suggestive association ($p<1\times10^{-5}$) were identified. The most significant SNP (rs2174299, $p=4.97\times10^{-7}$) was in an intergenic region overlapping a long noncoding RNA. One SNP (rs743742, $p=4.28\times10^{-6}$) was found near the gene *TIMP3*. *TIMP3* belongs to the TIMP gene family, which are inhibitors of the matrix metalloproteinases, a group of peptidases involved in degradation of the extracellular matrix. This may involve in the degradation of cartilage in OA. Additionally, some SNPs reside in gene regulatory regions as defined by RegulomeDB, thus providing insight as to how the genetic variants may have functional consequences. Conclusion: Our results highlight the importance of performing GWAS for OA using different phenotypic information that better reflects OA. Future work will include larger sample size and further exploration of OA phenotypic variability.

1981W

Evaluation of shared genetic aetiology between osteoarthritis and bone mineral density identifies *SMAD3* as a novel osteoarthritis risk locus. *S. Hackinger[1], K. Trajanoska[2], U. Styrkarsdottir[3], E. Zengini[4,5], J. Steinberg[1], G.R.S. Ritchie[1], K. Hatzikotoulas[1], A. Gilly[1], E. Evangelou[6,7], J.P. Kemp[8,9], D. Evans[8,9], T. Ingvarsson[10,11,12], H. Jonsson[11,13], U. Thorsteinsdottir[3,11], K. Stefansson[3,11], A.W. McCaskie[14], R.A. Brooks[14], J.M. Wilkinson[4], F. Rivadeneira[2], E. Zeggini[1], arcOGEN Consortium, GEFOS Consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, 3000 CA, Netherlands; 3) deCODE Genetics, 101 Reykjavik, Iceland; 4) Department of Oncology and Metabolism, University of Sheffield, Beech Hill Road, Sheffield S10 2RX, United Kingdom; 5) 5th Department, Dromokaiteio Psychiatric Hospital, Athens 124 61, Greece; 6) Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110, Greece; 7) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom; 8) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 9) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia; 10) Department of Orthopedic Surgery, Akureyri Hospital, 600 Akureyri, Iceland; 11) Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland; 12) Institution of Health Science, University of Akureyri, 600 Akureyri, Iceland; 13) Department of Medicine, Landspitali, The National University Hospital of Iceland, 101 Reykjavik, Iceland; 14) Division of Trauma & Orthopaedic Surgery, University of Cambridge, Box 180, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, United Kingdom.

Osteoarthritis (OA) is a common complex disease with a high public health burden and no curative therapy. The risk of developing OA is correlated with high bone mineral density (BMD), suggesting a shared underlying biology. However, despite a large body of epidemiological research into this association, the genetic overlap between OA and BMD has not yet been assessed on a genome-wide scale. Here, we investigated the shared genetics of OA and BMD using data from the arcOGEN and GEFOS consortia – the two largest genome-wide association studies (GWAS) for these respective traits published to date. We performed pairwise comparisons of summary statistics from the GEFOS consortium for lumbar spine (n=31,800) and femoral neck (n=32,961) BMD, and from the arcOGEN consortium for three OA phenotypes (hip, n_{cases}=3,498; knee, n_{cases}=3,266; hip and/or knee, n_{cases}=7,410; $n_{controls}$=11,009). Using LD score regression to estimate genetic correlation on a genome-wide scale, we found a significant correlation between the combined OA phenotype (hip and/or knee) and lumbar spine BMD (r_g=0.18, $p=2.23\times10^{-2}$). To identify distinct variants exerting cross-phenotype effects, we searched for shared association signals across different p-value cut-offs and performed a Bayesian co-localisation analysis to search for genomic regions harbouring pleiotropic signals. From these analyses we took forward 143 variants with evidence of cross-phenotype association for replication in independent large-scale OA datasets, and subsequent meta-analysis with arcOGEN for a total sample size of up to 23,425 cases and 236,814 controls. We found robustly replicating evidence for association with OA at rs12901071 (OR 1.08 95% CI 1.05-1.11, p_{meta}=3.12$\times10^{-10}$), an intronic variant in the *SMAD3* gene, which is known to play a role in bone remodeling and cartilage maintenance. Our findings provide the first systematic evaluation of pleiotropy between OA and BMD, highlight genes with biological relevance to both traits, point towards shared biological mechanisms involving both cartilage and bone homeostasis, and establish a robust new OA genetic risk locus at *SMAD3*. This work exemplifies the added power that can be afforded by considering shared genetic determinants for complex traits of medical relevance.

1982T

Identification of *IRF4*, *NOTCH4* and *RPS12* genes for non-syndromic sagittal craniosynostosis in a genome-wide association study using logistic regression. *H. Sung[1], CM. Justice[1], L. Brody[2], PA. Romitti[3], SA. Boyadjiev[4], AF. Wilson[1].* 1) Genometrics Section, NIH/NHGRI/CSGB, Baltimore, MD; 2) Medical Genomics and Metabolic Genetics Branch, NHGRI, NIH, Bethesda, MD, USA; 3) Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, IA, USA; 4) Section of Genetics, Department of Pediatrics, University of California Davis, Sacramento, CA, USA.

Craniosynostosis, the premature closure of the cranial vault sutures, is a common congenital anomaly with sagittal non-syndromic craniosynostosis (sNSC) being the most common form of craniosynostosis. Justice et al. [2012] identified candidate loci for sNSC near *BMP2* and in *BBS9* by performing a Transmission Disequilibrium Test (TDT) on case-parent trios, which were recruited through the International Craniosynostosis Consortium [genetics.ucd-mc.ucdavis.edu/icc.cfm]. Because there is low correlation between association results from logistic regression analysis and the TDT [Sung et al. 2008 and Eu-ahsunthornwattana et al. 2014], logistic regression analysis was performed on case-control data with the affected probands from Justice et al. [2012] as cases and healthy individuals from GWAS Trinity study [Desch et al. 2013] as controls. Both the case and control data were previously genotyped on the Illumina 1M human Omni1-Quad array. Samples with a genotyping rate less than 80% and Single Nucleotide Polymorphisms (SNPs) with a call rate less than 98% and a Minor Allele Frequency (MAF) less than 1% were excluded from the combined data for further analysis. Logistic regression in PLINK was performed on 748,100 SNPs in 127 sNSC children and 2,359 Trinity controls. In addition to the SNPs in or near *BMP2* and *BBS9* found by TDT analysis, logistic regression analysis identified additional SNPs associated with sNSC at the genome-wide significance level after Bonferroni correction, in or near genes including *IRF4, NOTCH4 and RPS12.* Each of these genes is located on chromosome 6 and the most significant SNP identified for each gene was rs12203592 in the intronic region of *IRF4* (p < 1e-18), rs371156 in the intergenic region between *NOTCH4* and *c6orf10* (p < 1e-10) and rs9493468 in the intergenic region of the *RPS12* and *LINC00326* (p < 1e-13). The *IRF4* gene is known to be one of the top 50 genes which were differentially expressed by prematurely fused human sutures [Coussens et al. 2008]. The *NOTCH4* gene is involved in embryonic development where NOTCH plays a critical role in skeletal development and homeostasis [Zanotti and Canalis, 2010 and 2016]. The *RPS12* gene is found to be one of the enriched genes in expression analyzing transcription profiles of human pharyngeal arch 1 [Cai et al., 2005]. Further investigation of these genes will be necessary to better understand the etiology of sNSC.

1983F

Copy number variant causes the mis-calculation of single nucleotide polymorphism in association analysis. *Z. Yan[1,2,3], J. Liu[1,2,4], Y. Zhou[2,3,5], S. Liu[1,2,3], N. Wu[1,2,3,7], X. Song[7], X. Wang[2,6], Z. Akdemir[7], W. Chen[1,2,3], R. Du[7], Y. Zuo[1,2,3], Z. Liu[2,8], Z. Wu[2,3,9].* 1) Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China; 2) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, China; 3) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, China; 4) Department of Breast Surgical Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 5) Department of Internal Medicine, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China; 6) State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China; 7) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA; 8) Department of neurosurgery, Xuanwu hospital, Capital Medical University, Beijing, China; 9) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China.

Background As genomic study continues to develop recently, genetic variates related to the diseases of human-beings are widely reported, including the copy number variants (CNVs) and single-nucleotide polymorphisms (SNPs). Nonetheless, the association analysis does not ensure successful identification of the causal variants. Incurred by the coexistence of CNVs and SNPs in the same genomic region, the probable bias of pathogenicity assignment and bioinformatic analysis has not been carefully examined. Characterized with the compound inheritance model which includes a common hypomorphic allele and a deletion of the TBX6 gene, congenital scoliosis (CS)'s genetic etiology has been reported recently. Methods To simulate the scenario, a data set which has been experimentally validated was adopted as a realistic model from a CS cohort, which including 161 unrelated sporadic CS patients enrolled from the Peking Union Medical College Hospital (PUMCH) in China between October 2010 and June 2014 and 166 unrelated healthy controls. Results Per the association analysis, the apparent prevalence of the TBX6 hypomorphic haplotype can be enhanced by the deletion in the population of patients, which resulted in overestimation of the SNPs' pathogenic contribution in the same region. Moreover, we found that around 10% of significant SNPs as identified by genome-wide association studies were located in the same region with common CNVs (frequency > 1%) and 99.69% overlapping the pathogenic CNVs. It shows that miscalculation widely exists. In addition, a generalized model was introduced to show the mechanism through which CNVs result in the association power's miscalculation. Conclusion A conceptual pipeline is put forward about how to prevent miscalculation in the association analysis of SNPs. Also, to reveal the potential inheritance model, SNPs should be assessed in relation to other forms of genetic variates.

1984W

Osteoporosis-associated risk variant in distal enhancer at 1p36.12 regulates expression of long noncoding RNA through long-range loop formation. *Y. Guo, X.F. Chen, D.L. Zhu, M. Yang, B.J. Lu, Y.Y. Duan, Y. Rong, S.S. Dong, T.L. Yang.* School of Life Science and Technology, Xi'an Jiaotong University, Key Laboratory of Biomedical Information Engineering of Ministry of Education, Xi'an, Shaanxi, China.

   Genome-wide association studies (GWASs) have reproducibly associated variants within noncoding regions of chromosome 1p36.12 with osteoporosis, but the molecular mechanisms of how specific risk variants functionally contribute to the underlying pathogenesis are unknown. Traditional identification of pathogenic genes usually focus on genes closest to the GWAS hits, which may not be the true target. The aim of this study was to identify the target gene(s) and the functional variant(s) underlying the association between SNPs at 1p36.12 and osteoporosis, through integrative analysis combining data from expression quantitative trait locus (eQTL), genomic chromatin interaction (Hi-C), epigenomic profiles and functional assays. We demonstrate that three functional intergenic SNPs (rs6426749, rs6684375, rs34963268) at 1p36.12 act as distal allele-specific enhancers regulating expression of a lncRNA (*LINC00339*) via long-range chromatin loop formation. This loop is regulated by CTCF occupied in the loop boundaries. Specifically, rs6426749-G allele could bind to transcription factor TFAP2A, which efficiently elevates the enhancer activity and increases the *LINC00339* expression. Downregulation of *LINC00339* significantly increase the expression of *CDC42* in osteoblast cells, which is a pivotal regulator involved in bone metabolism. Together, our findings elucidate a potential mechanistic basis that a functional noncoding SNP rs6426749 obliquely regulates *CDC42* via *LINC00339*. Our study provides a novel mechanistic insight into how a noncoding SNP affects osteoporosis by long-range interaction, which would be promising therapeutic targets for osteoporosis.

1985T

Broad mutation spectrum of *FBN1* gene implicated in a cohort of idiopathic scoliosis and its related genotype-phenotype correlation study. *M. Lin[1,2], Y. Chen[1,2], W. Chen[1,2], Y. Zuo[2], J. Liu[3,4,5], G. Liu[2], S. Liu[2], G. Qiu[2,4], N. Wu[2,4,5], Z. Wu[1,4,6].* 1) Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, Beijing, ChinaBeijing, China; 2) Department of Orthopaedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 3) Breast Surgical Oncology, Cancer Hospital of Chinese Academy of Medical Sciences, Beijing 100021, China; 4) Beijing Key Laboratory for Genetic Research of Skeletal deformity; 5) Medical Research Center of Orthopaedics, Chinese Academy of Medical Sciences, Beijing China; 6) Department of Central Laboratory, Peking Union Medical College Hospital, Beijing 100730, China.

   Background Idiopathic scoliosis (IS) causes spinal deformity in 3% of children. Notwithstanding a strong genetic basis, few genes have been reported associated with IS and the pathogenesis remains to be elucidated. Etiology of IS is still unclear. We previously demonstrated a compound heterozygous model in which a null allele mutation in combination with a common haplotype of *TBX6* causes congenital scoliosis, suggesting that genetic factors play a significant role in vertebral malformation(VM). Hence, we established an in silicon designed target panel comprising genes previously known to cause VM. *FBN1* is included in the panel. Fibrillin-1 encodes a large glycoprotein, fibrillin constitutes a cluster of large extracellular proteins that form the core of highly ordered extended and universally distributed aggregates, termed microfibrils. In this study, we report that the loss of function mutations of *FBN1* gene may contribute to IS. Methods We enrolled 102 patients diagnosed as IS in Peking Union Medical College Hospital (PUMCH) in China between October 2010 and November 2015. The IS diagnosis were confirmed by radiological imaging. Whole genome sequencing (WGS) was performed in 102 sporadic IS cases with additional 325 Han Chinese patients radiologically confirmed with vertebral malformation Target sequenced by an in silicon designed target panel. Results We identified two novel *FBN1* frameshift mutations (c.7785delC, p.Y2596Tfs*86; c.8275_8291delGAGAAGACAGCCATCTT, p.E2759Cfs*9) in two unrelated patients by panel sequencing, respectively. These two variants did not exist in ExAC, 1000 Genome, ESP6500 database, universal mutation database (UMD) or any literature. The variants were possibly contributing to the phenotypes, but of uncertain significance. Moreover, the two patients were diagnosed Marfan syndrome based on the revised Ghent criteria. Furthermore, we also identified three *FBN1* missense mutations in highly conserved domain and one novel rare *FBN1* splicing defect by WGS. Conclusions Up to our knowledge, our findings indicated that loss of function mutations in *FBN1* might contribute to spinal deformity formed of IS. Our data showed that IS patients were inclined to harbor *FBN1* mutations, recapitulating the vital role of *FBN1* in the pathogenesis of IS.

1986F

PheWAS meta-analyses on bone microarchitecture phenotypes assessed by HR-pQCT and CRISPR/Cas9 gene-editing in zebrafish identify novel genetic risks of osteoporosis and fractures: The Bone Microarchitecture International Consortium (BoMIC). *Y.H. Hsu[1,2,3], K. Kinyua[3], C. Liu[4], M. Nethander[5], E. Biver[6], EJ. Atkinson[7], E. SornayRendu[8], C. Watson[9], AKO. Wong[10], S. Amin[7], E. Lim[4], R. Joehanes[1,3], E. Lespessailles[11], S. Demissie[4], D. Karasik[3], B. Merle[8], D. Mellström[5], B. Richards[12], D. Goltzman[13], DA. Hanley[14], SK. Boyd[14], ML. Bouxsein[15], RY. Kwon[9], M. Lorentzon[16], C. Ohlsson[17], P. Szulc[8], S. Ferrari[18], RD. Chapurlat[8], S. Khosla[7], DP. Kiel[1,2,3].* 1) Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Hebrew SeniorLife, Institute for Aging Research, Boston, MA; 4) Boston University School of Public Health, Boston, MA; 5) University of Gothenberg, Gothenberg, Sweden; 6) University of Geneva, Geneva, Switzerland; 7) Mayo Clinic College of Medicine, Rochester, MN; 8) INSERM and University of Lyon, Lyon, France; 9) Orthopaedics and Sports Medicine, Univ. Washington, Seattle, WA; 10) Toronto General Research Institute, Univ. Health Network and McMaster University, Toronto, Canada; 11) University of Orleans, Orléans, France; 12) Lady Davis Institute and Dept. of Human Genetics, McGill University, Montreal, Canada; 13) Departments of Medicine and Physiology, McGill University, Montreal, Canada; 14) Cumming School of Medicine, Univ. of Calgary, Calgary, Canada; 15) Beth Isreal Deaconess Medical Center, Boston, MA; 16) Sahlgrenska University Hospital, Gothenburg, Sweden; 17) Center for Bone and Arthritis Research at the Sahlgrenski Academy, Univ Gothenberg, Gothenberg, Sweden; 18) Geneva University Hospital, Geneva, Switzerland.

Bone microarchitecture assessed by high resolution peripheral quantitative computed tomography (HR-pQCT) has been reported to be heritable even after adjusting for DXA-based areal BMD (aBMD), and has been associated with fracture risks. Genetic studies of these phenotypes may reveal unique, novel genes that contribute to skeletal integrity not identifiable in previous genome-wide association studies (GWAS) of aBMD. Thus, our objective was to conduct such a whole genome association study. We performed an expanded whole genome association meta-analysis on both rarer functional variants and common variants in 5,692 adult Caucasians obtained from 7 cohort studies (FHS, Mayo Clinic, GeRiCo, Lyon OFELY, Lyon STRAMBO, MrOS Sweden and GOOD Study) and replicated on top associated SNPs in 2,100 independent samples (CaMOS and Lyon Qualyor studies). HR-pQCT (XtremeCT, Scanco Medical AG) bone structure traits were assessed (Table 1). All studies were genotyped by SNP chips and whole genome imputation to ~40 million SNPs was performed based on the Haplotype Reference Consortium. We used either linear regression or mixed-effects model (family-based) for single SNP-phenotype associations. Additive genetic effect models were applied and adjusted for age, sex, weight and genetic ancestry. For less common variants (MAF < 1%), we performed gene-based Optimal SKAT test in each study. Meta-analysis was performed to combine association results. Of the 17 genome-significant loci (386 SNPs with $p < 5 \times 10^{-8}$), 12 were not reported by previous aBMD GWAS meta-analyses: SIX2, PNPT1, chr4q35.2, KCNIP1, chr5q11.2, chr7q31.31, CSMD1, NRG1, PTHLH-CCFC91, FAM155A, C3 and FMN2 loci. The majority of these loci were specifically associated with μFEA-FL, TtAr, CtTh, TbSp, TbN or CtBMD. The most significantly associated SNP, rs10254825, was positively associated with FL ($p=2.6 \times 10^{-22}$) and previously associated with fracture in GWAS meta-analyses. The remaining 5 loci (ZBTB40, SPTBN1, MEF2C-AS1, WNT16-FAM3C and AKAP11-TNFSF11) were previously reported by aBMD GWAS and are associated with multiple HR-pQCT phenotypes at both radius and tibia (e.g., TbBMD, TtBMD, TbSp). Our findings suggest that the genetic control of bone microarchitecture may provide additional insights into the pathogenesis of skeletal integrity not available from aBMD. Genome editing in zebrafish using the CRISPR/Cas9 on selected genes from the novel loci is underway to dissect the potential functional involvement in bone biology.

1987W

Novel genetic risk factors identified from a genome-wide association study for lumbar disc degeneration in Southern Chinese. *Y. Li[1,4], P. Gao[2], X. Zhou[2,4], K.M.C. Cheung[3], K.D.K. Luk[3], J. Karppinen[5,6], Y.Q. Song[2], K.S.E. Cheah[2], D. Chan[2], P.C. Sham[1,4].* 1) Center for Genomic Sciences, The University of Hong Kong, Hong Kong, SAR, China; 2) School of Biomedical Science, The University of Hong Kong, Hong Kong, SAR, China; 3) Department of Orthopedics and Traumatology, The University of Hong Kong, Hong Kong, SAR, China; 4) Department of Psychiatry, The University of Hong Kong, Hong Kong, SAR, China; 5) Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland; 6) Finnish Institute of Occupational Health, Health and Work Ability, and Disability Prevention Center, Oulu, Finland.

Background Lumbar disc degeneration (LDD) is a major etiological risk factor for low back pain (LBP), which is a common and disabling condition, affecting 80% of the general population at some time in life. LDD has been reported to have a substantial genetic component with up to 74% heritability. Most reported LDD associated genes are from candidate gene studies, only one meta-analysis was published based on European individuals. There is much missing heritability need to explained and all candidate genes needs to be tested in Southern Chinese population using an unbiased genome-wide approach. **Method** This study aimed to identify genetic variants associated with LDD in Southern Chinese population. Both single and combined MRI phenotypes were used for genome-wide association study (GWAS) in 2,200 subjects with Southern Chinese origin. Moreover, to improve the power of GWAS, cross disorder analyses were employed to leverage on pleiotropy with bone mineral density. **Results** One genome-wide significant signal for modic change was identified on Chromosome 15 (rs10214211, p = 9.53e-9) at gene *ADRA1B*. Another genome-wide significant signal in Chromosome 15 (rs1902430, P = 1.57e-8) was identified while combining two most related phenotypes, disc bulging and signal intensity loss. Two genes, *SCG3* and *LYSMD2*, were significant in this region. An additional low frequency missense variant, which only exists in LDD cases, was further identified in *SCG3* from whole exome sequencing analysis (rs146162352, P = 0.005). Moreover, two loci in gene *KCND3* and *SPC24* were identified to be pleiotropic between BMD and LDD (SFDR q<0.05). **Discussion** Phenotype definitions are essential for genome-wide association studies. This study has provided several successful cases that detailed definitions of LDD could boost the power of GWAS to identify disease associated markers. Using the definition of modic change, we identified one genome-wide significant signal. While combining highly correlated phenotypes, signal intensity loss and disc bulging, the power of GWAS also boosted. Leveraging on the pleiotropic effects from bone mineral density, small p-values enrichments were identified on two genetic loci. Together, this study presents a catalog of candidate genetic variants associated with lumbar disc degeneration in Southern Chinese population.

1988T

The genetic architecture of osteoarthritis: Insights from UK Biobank.
E. Zeggini[1], E. Zengini[2,3], K. Hatzikotoulas[1], I. Tachmazidou[1], J. Steinberg[1], S. Hackinger[1], U. Styrkarsdottir[4], D. Suveges[1], B. Killian[1], A. Gilly[1], T. Ingvarsson[5,6], H. Jonsson[6,7], G. Babis[8], U. Thorsteinsdottir[4,6], K. Stefansson[4,6], J. Wilkinson[2]. 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) University of Sheffield, Sheffield, UK; 3) Dromokaiteio Psychiatric Hospital, Athens, Greece; 4) de-CODE genetics, Reykjavik, Iceland; 5) Akureyri Hospital, Akureyri, Iceland; 6) University of Iceland, Reykjavik, Iceland; 7) The National University Hospital of Iceland, Reykjavik, Iceland; 8) National and Kapodistrian University of Athens, Athens, Greece.

Osteoarthritis (OA) is the most prevalent musculoskeletal disease and a leading cause of disability worldwide. It affects 40% of individuals over the age of 70 and is associated with an increased risk of comorbidity and death. The health economic burden of OA is rising, commensurate with longevity and obesity rates, and there is currently no treatment. Half of the variation in risk of OA is due to genetic factors. OA has been relatively refractive to GWAS approaches, with only 18 established loci to date. These traverse hip, knee and hand OA, with limited overlap. Here we conduct the largest OA GWAS to date, using genotype data from the UK Biobank and linkage to Hospital Episode Statistics data to define cases and controls. We compare and contrast this hospital diagnosed (HD, n=10,083 cases) GWAS to self-reported (SR, n=12,658 cases) OA GWAS drawn from the same UK Biobank dataset (non-OA controls selected to be ~4x the number of cases). We find power advantages with the SR dataset, indicating that the increase in sample size overcomes phenotype uncertainty. We identify nine novel OA loci, which robustly replicate in an independent dataset from Iceland (in up to 18,069 cases and 246,293 controls). For example, rs3771501 (OR 0.94 (95% CI 0.92-0.96), p=1.66x10^{-8}) resides in an intron of *TGFA*, which has been associated with cartilage thickness, is differentially expressed in OA cartilage lesions and contributes to the development of chondrocyte clusters. rs11780978 (OR 1.13 (95% CI 1.08-1.17), p=1.98x10^{-9}) resides in *PLEC*, which plays a role in skeletal muscle disorders. rs2521349 (OR 1.13 (95% CI 1.09-1.17), p= 9.95x10^{-10}) lies within *MAP2K6*, which is differentially expressed in OA synovial tissue. Several of the novel signals reside in the same regions as variants associated with metabolic and anthropometric traits. Through LD score regression, we find significant positive genetic correlation between OA and body mass index (rg=0.42, p=9.8x10^{-9}), waist circumference (rg=0.34, p=2.0x10^{-6}), hip circumference (rg=0.24, p=1.3x10^{-3}) and a negative correlation with educational attainment (rg=-0.49, p=7.9x10^{-6}). All constitute established epidemiologic risk factors for OA. All novel signals are at common frequency variants and confer small effects, in line with a highly polygenic model underpinning OA risk.

1989F

Longitudinal genome-wide association analyses and heritability estimates of pediatric bone mineral density. *D. Cousminer[1,2], A. Chesi[1], J. Mitchell[1,2], H. Kalkwarf[3], J. Lappe[4], V. Gilsanz[5], S. Oberfield[6], J. Shepherd[7], A. Kelly[1,2], S. McCormack[1,2], B. Voight[1,2], B. Zemel[1,2], S. Grant[1,2].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA; 3) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Creighton University, Omaha, NB; 5) Children's Hospital Los Angeles, Los Angeles, CA; 6) Columbia University Medical Center, New York, NY; 7) University of California San Francisco, San Francisco, CA.

Background: SNP heritability (h^2_{SNP}) assesses the trait variance attributed to genotyped genetic factors. h^2_{SNP} for bone mineral density (BMD) is known to vary across skeletal sites and with age, and assessments of h^2_{SNP} and genome-wide association studies (GWAS) have typically been performed using cross-sectional data in adults. In contrast, systematic longitudinal assessment of h^2_{SNP} in the pediatric context has not yet been conducted. Thus, we sought to determine h^2_{SNP} estimates for BMD and bone mineral content (BMC) across specific clinically important skeletal sites using modeled longitudinal pediatric growth curves. We also performed complementary GWAS in search of novel bone-related loci. **Methods**: In the multicenter, multiethnic Bone Mineral Density in Childhood Study (BMDCS) of healthy children aged 5 to 20 years old with up to 7 annual measurements each, BMD and BMC of the spine, total hip, femoral neck, distal radius and total body less head were measured in up to 1,876 boys and girls. We used GCTA-Restricted Maximum Likelihood (GREML) to estimate h^2_{SNP} for Super Imposition by Translation and Rotation (SITAR)-modeled longitudinal growth parameters (*size*, *tempo*, and *velocity*) at each skeletal site, adjusting for population ancestry. Subsequently, sex and ancestry-specific standardized SITAR-modeled growth parameters were subjected to GWAS analyses using GEMMA, which also accounts for population ancestry. **Results**: In general, h^2_{SNP} was higher for the *size* parameter than for the *tempo* and *velocity* parameters, and was comparable to previously published cross-sectional estimates. h^2_{SNP} for *size* ranged from 0.52 (SE=0.093, *P*=2.0x10^{-8}) for femoral neck BMD to 0.20 (SE=0.09, *P*=0.011) for radius BMD. GWAS recapitulated the known signal at *CPED1* for female radius BMD as associated with the radius *size* parameter, but also yielded a number of novel genetic signals associated with *size*, *tempo*, and *velocity* that varied across skeletal sites. **Conclusions**: We generated skeletal site-specific longitudinal estimates of h^2_{SNP} for pediatric BMC and BMD that are comparable to previous estimates but expand to more detailed skeletal resolution and capture growth longitudinally. Both h^2_{SNP} and the most prominent GWAS-implicated loci vary across skeletal sites. These results provide new insights into the process of skeletal bone acquisition, which is vital to optimizing the reduction in later-life osteoporosis risk.

1990W

Exome sequencing in idiopathic scoliosis families implicates cilia genes in disease etiology. *E.E. Baschal[1], E.A. Terhune[1], C.I. Wethey[1], R.M. Baschal[1,2], K.D. Robinson[1], M.F. Tjandra[1], S. Pradhan[1], B.S. Sutphin[1], M.R.G Taylor[3], K. Gowan[4], C.G. Pearson[5], L.A. Niswander[4], K.L. Jones[4], N.H. Miller[1,2].* 1) Department of Orthopedics, University of Colorado Anschutz Medical Campus, Aurora, CO; 2) Musculoskeletal Research Center, Children's Hospital Colorado, Aurora, CO; 3) Department of Cardiology, University of Colorado Anschutz Medical Campus, Aurora, CO; 4) Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO; 5) Department of Cell and Developmental Biology, University of Colorado Anschutz Medical Campus, Aurora, CO.

Idiopathic scoliosis (IS) is a structural lateral spinal curvature ≥10° that affects 3% of healthy children, with females at the greatest risk for severe progression. Severe progressive curvature can lead to life-long problems of cosmetic deformity, operative spinal fusion, back pain, and degenerative disease. IS frequently runs in families, with most cases strongly suspected to be genetic. Previous efforts to understand the genetic etiology of IS have led to the identification of multiple genes that may contribute to the IS phenotype, including *LBX1*, *GPR126*, *FBN1*, *HSPG2*, *POC5*, and musculoskeletal collagen genes. Despite these findings, the majority of genetic factors contributing to IS remain undiscovered. To identify rare variants associated with IS, we performed exome sequencing on 5 multigenerational families with IS (3-4 individuals per family). Illumina HiSeq reads were mapped to the human reference genome (hg38) using GSNAP, and variants were called using FreeBayes. Variants were annotated using SnpEff and dbNSFP. We filtered the list of variants, requiring that each variant was present in all sequenced members of the family, was predicted to be damaging, and had an ExAC frequency of less than 5%. This filter resulted in a total of 352 variants across the 5 families, with 23 to 120 variants identified in each family. Seventeen genes were shared by at least 2 families. We combined the variant lists from all 5 families, and entered the corresponding genes into PANTHER. We used PANTHER Gene Ontology (GO) term enrichment analysis, with the annotation dataset "Cellular Component Complete." This analysis revealed an enrichment of genes associated with the GO term "Cell Projection Part" (p = 0.0243 with Bonferroni correction, p = 1.84E-5 without Bonferroni correction, n = 34 genes). The "Cell Projection Part" GO term includes genes involved in cilia, actin-based processes/projections (lamellopodia and filopodia), and neuronal projections. Previous studies of IS have implicated variants in cilia genes, including *PTK7* and *POC5*. We looked at cilia in more detail, and found that "Ciliary Part" was significantly enriched in our gene list without Bonferroni correction (p = 5.41E-4, n = 16 genes). Further investigation and functional studies will be required to fully elucidate the role of cilia in IS pathogenesis. Our studies highlight a new set of cilia genes for further consideration in the etiology of IS.

1991T

Association of TGF β1 gene polymorphisms with primary knee osteoarthritis in Asian Indians. *Q. Hasan, S. Poornima, K. Subramanyam, G. Iyer, M. Darooei, K. Konda, S. Daram, A. Priya, S. Kumari.* Kamineni Hospitals, Hyderabad, India.

Primary Osteoarthritis (OA) is slowly progressive and irreversible pathology, which is considered as a part of the ageing process. It affects all the joints of the body like hip, spine but predominantly of the knee. It is estimated that ~10 % of the world's aged population ≥60 years have symptomatic OA and present therapy for it includes the use of pain relieving medicines, physiotherapy and joint replacement surgery. Early onset OA is currently being observed frequently in Asian Indians, with individuals below 50 years presenting with the pathology. Current evidence indicates an important role of gene polymorphisms in the etiology of complex disease arthritis. Transforming Growth Factor β1 (TGFβ1) plays a role in development, homeostasis of various tissues by regulating cell proliferation, differentiation, apoptosis, migration and bone remodeling. Two SNPs of TGFβ1 gene have been selected for the evaluation which were identified in GWAS. The aim of the study was to analyse TGFβ1 –C509T and T869C polymorphisms in Asian Indian primary knee OA patients below the age of 50 years. This study was approved by Institutional Ethics Committee. The study involved 200 OA cases, which were clinically diagnosed and radiologically confirmed along with age and gender matched controls. Genomic DNA was isolated from blood, PCR amplification by specific primers after restriction digestion was carried out. Both SNPs appeared to be in HWE in Asian Indian. Results showed that TGFβ1 T allele of rs1800469 (C509T) was significantly associated with the disease (OR =4.08,95% CI=3.03–5.50, p <0.0001). Dominant (OR=4.31,95% CI=2.84–6.56, p<0.0001) and recessive modes of inheritance (OR=4.15,95%CI=2.68–6.45,p=0.0001) also showed statistical significance. In T869C polymorphism the C was significantly associated with the disease (OR=3.76, 95%CI = 2.79–5.06,p <0.0001).Dominant (OR=4.54,95% CI = 2.98–6.93,p<0.0001), recessive (OR=3.51,95%CI=2.24–5.49,p=0.0001) and co dominant modes of inheritance (OR=1.81 95%CI=1.10-2.99,p=0.01) were also associated with disease. These SNPs, rs1800469 and rs1982073, are associated with young onset primary knee OA and could be used as potential biomarkers to identify individuals/family members, who are at risk of developing knee OA to plan preventative strategies in our population, after confirming in a larger study group. This will help reduce morbidity and disability associated with this pathology.

1992F

The genetic architecture and phenotypic spectrum in the skeletal ciliopathies. *W. Zhang[1], S.P. Taylor[3], H. Ennis[1], K.N. Forlenza[2], I. Duran[2,5], B. Li[1], J.A. Ortiz Sanchez[1], L. Nevarez[1], D.A. Nickerson[6,7], M. Bamshad[6,7,8], R.S. Lachman[9], D. Krakow[2,3,4,9], D.H. Cohn[1,2,9], University of Washington Center for Mendelian Genomics.* 1) Department of Molecular, Cell, and Developmental Biology, University of California at Los Angeles, Los Angeles, California, 90095, USA; 2) Department of Orthopaedic Surgery and Orthopaedic Hospital Research Center, David Geffen School of Medicine at UCLA, Los Angeles, California, 90095, USA; 3) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, California, 90095, USA; 4) Department of Obstetrics and Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, California, 90095, USA; 5) Networking Research Center on Bioengineering, Biomaterials and Nanomedicine, (CIBER-BBN). University of Malaga, Spain; 6) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 7) University of Washington Center for Mendelian Genomics, University of Washington, Seattle, Washington, 98195, USA; 8) Department of Pediatrics, University of Washington, Seattle, WA 98195 and Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA 98105, USA; 9) International Skeletal Dysplasia Registry, University of California, Los Angeles, Los Angeles, California, 90095, USA.

Defects in the biosynthesis and/or function of primary cilia cause a spectrum of disorders collectively referred to as ciliopathies. A subset of these disorders is distinguished by profound abnormalities of the skeleton that include a long narrow chest with markedly short ribs, extremely short limbs, and polydactyly as a common finding. These include the perinatal lethal short-rib polydactyly syndrome (SRPS) phenotypes and the phenotypically less severe asphyxiating thoracic dystrophy (ATD), Ellis van Creveld (EVC) syndrome and cranio-ectodermal dysplasia (CED). To identify new genes and define the spectrum of mutations in the skeletal ciliopathies, we analyzed 152 unrelated families with SRP, ATD and EVC. Causal variants were discovered in three newly associated genes, *TRAF3IP1*, *FUZ*, and *LBR*, and in eleven known skeletal ciliopathy genes, in 120 of the families. The results established the molecular basis of SRPS type IV, in which we identified mutations were in four different ciliary genes. The data have provided systematic insight regarding the genotypes associated with a large cohort of these highly genetically heterogeneous phenotypes and identified the ciliary components required for normal skeletal development.

1993W

Targeted sequencing of sagittal nonsyndromic craniosynostosis in regions on chromosomes 3, 7, and 20. *C.M. Justice[1], A. Cuellar[2], W. Lattanzi[3], P.A. Romitti[4], S.A. Boyadjiev[2], A.F. Wilson[1].* 1) Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Baltimore, MD; 2) Section of Genetics, Department of Pediatrics, University of California Davis, Sacramento, CA; 3) Università Cattolica del Sacro Cuore, Rome, Italy; 4) Department of Epidemiology, University of Iowa, Iowa City, IA.

Craniosynostosis (CS) is a common congenital malformation where one or more of the infant skull sutures (metopic, coronal, sagittal, or lambdoid) fuse prematurely. The prevalence of the most common type of CS, sagittal non-syndromic craniosynostosis (sNCS), is estimated to be about 2 per 10,000 live births. We previously performed the first genome-wide association study (GWAS) for sNCS, and discovered robust associations to a 120 kb region located 345kb downstream from *BMP2* on 20p12.3 and in a 167kb region intronic to *BBS9* on 7p14.3. Analyses of 52 imputed variants in *DLG1* on 3q29 were found to be genome-wide significant. In order to identify variants in the coding and noncoding regions responsible for the sNSC phenotype, targeted Next Generation Sequencing (NGS) was used to sequence the associated regions on 3q29, 7p13.2, and 20p12.3 and two additional genes in close proximity to these associated regions (*BMP2* and *BMPER)* in 100 case-parent trios. Association analysis resulted in significant markers on chromosomes 3, 7 and 20, as expected, given our previous findings. The most significant association on chromosome 3 was for rs3773855 ($p=1e^{-5}$), on chromosome 7 for rs10228486 ($p=6.11e^{-06}$) and on chromosome 20 for rs1124471 ($p=6.4e^{-16}$). Thirty-two exonic variants were identified, 16 of which were non-synonymous (7 in *DLG1*, 4 in *BBS9*, 4 in *BMPER* and 1 in *BMP2*). None appeared causative, because of either the lack of prediction of being a damaging allele (by Sift or Polyphen2), or because the variant was found in unaffected parents and not transmitted to the affected offspring. No *de novo* variants were located in exonic regions. Of the 32,604 interaction tests performed between variants on *BBS9* and *BMP2*, none reached significance after adjustment for multiple tests . The top three interactions (p < 0.004) were between rs142092 on chromosome 20 and 3 SNVs on chromosome 7 (rs3823763, rs6944723 and rs4604335). Variation in regulatory region(s) may be involved in the phenotype, and allele frequency differences between cases and the reference genome may indicate which alleles may be of interest for functional studies using zebrafish.

1994T

Shared and subtype-specific genetic variation define the genetic susceptibility of juvenile idiopathic arthritis. *Y. Li[1], J. Li[1], J. Bradfield[1], R. Chiaroni-Clarke[3], M. Mohebnasab[1], D. Abrams[1], C. Kim[1], F. Mentch[1], E.T. Luning Prak[2], P. Sleiman[1,4], J. Ellis[3], E. Behrens[1], H. Hakonarson[1,4,5].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Perelman School of Medicine, Department of Pathology, Philadelphia, Philadelphia, PA; 3) Murdoch Childrens Research Institute, The Royal Children's Hospital, Flemington Rd Parkville, Victoria 3052, Australia; 4) Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA; 5) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.

Juvenile Idiopathic Arthritis (JIA) encompasses a heterogeneous group of rare, immune-mediated joint disorders that afflicts ~1 in 10,000 children in North America and Europe and causes severe pain and physical debilitation. There are 7 JIA subtypes, with the most prevalent being oligoarthritis (OLG) and Rheumatoid Factor-negative polyarthritis (PRFN), however this classification is controversial as it does necessarily reflect disease biology, in part because known genetic risk factors explain only a fraction of total disease heritability (~20-30%). To identify genetic risk factors for JIA, we performed a heterogeneity-sensitive GWAS with over 1250 pediatric cases from all 7 clinical subtypes and over 9000 population-based pooled controls all genotyped on the Illumina 550/610 platforms and imputed, using IMPUTE2.We adapted a previously published method for pooled inverse-chi square analysis (Li and Hakonarson, 2014) to identify both shared and subtype-specific JIA susceptibility loci, adjusting for the use of disparate subtype sample sizes and common, population-based controls.. After association testing and quality filtering, we identified 32 genome-wide significant (GWS; P_{meta}<5E-8) as well as several dozen loci reaching marginal significance (GWM; P_{meta}<1E10-5). Overall, we replicated 87.5% of GWS and 72.7% of GWM loci reported by prior publications (P_{GWS}<0.05). Twenty (21) of the GWS loci have been reported in one or more autoimmune or inflammatory diseases/traits including *ANKRD55* (P_{meta}<4.6E-10; previously reported in JIA) and *IL22RA1* (P_{meta}<6.9E-11; psoriasis and chronic rhinosinusitis). Many of the other novel candidates have relevant biological annotation, for example, *LPHN2*, which encodes a G-protein coupled receptors (GPCR) that mediates neutrophil chemotaxis and is associated with all 7 JIA subtypes. The most robust association was observed for OLG/PRFN and enthesitis subtypes, P_{unadj}<4.4E-22, 3.2E-15 and 5.6E-08, respectively. Another novel candidate is *RCN1* (P_{meta}<4.2E-08), encodes a calcium-binding protein in the lumen of the endoplastic reticulum and expressed on bone endothelial cells in response to by TNF-alpha. Importantly, JIA-associated GWS candidate genes show enrichment for protein-protein enrichment as compared to a genome-wide background (p<0.012). Independent replication analysis is underway as well as gene network, functional and multi-omic analysis for expression enrichment in relevant disease tissues.

1995F

Characterizing the molecular biology of systemic sclerosis with RNA-Seq of both skin and PBMCs. *E. Roberson[1,2], L. Cao[1], D.J. Morales-Heil[1], M. Carns[3], B. Korman[3], J. Varga[3].* 1) Department of Medicine, Division of Rheumatology, Washington University, St. Louis, MO; 2) Department of Genetics, Washington University, St. Louis, MO; 3) Feinberg School of Medicine Scleroderma Program, Northwestern University, Chicago, IL.

Background: Systemic sclerosis (SSc) is a complex and heterogeneous orphan disease with multiple organ involvement. Treatment is complicated by a highly variable clinical course, lack of biomarkers of disease activity, and absence of FDA approved treatments. In most patients, skin and subcutaneous tissue are affected, and circulating immune cells are activated. We performed one of the first RNA-Seq analyses on simultaneously obtained skin and peripheral blood mononuclear cells (PBMCs) from patients with clinically well-characterized SSc. **Methods:** We obtained 3mm skin biopsies and PBMCs from healthy controls (n=14) and individuals with systemic sclerosis (n=21 with n=10 6 month follow-ups), including both limited cutaneous systemic sclerosis (lcSSc), diffuse cutaneous systemic sclerosis (dcSSc), scleroderma sine scleroderma (SSS), and Very Early Diagnosis of Systemic Sclerosis (VEDOSS). We performed stranded, total RNA-Seq for transcriptome profiling. Modified Rodnan skin score (MRSS) was evaluated at each visit, and most cases also had longitudinal pulmonary function data and high-resolution CT scan data to detect interstitial lung disease. **Results:** In skin the most significant increases were in the genes *COMP*, *COL4A4*, and *IGFBP4*. Genes significantly increased in SSc skin are enriched for AP-2 alpha transcription factor sites in their promoter, and functional categories were enriched for both extracellular matrix and immune system processes. In contrast, we detected a marked decreased expression of *SPAG17*, part of the ciliary axoneme central pair complex, in SSc skin. Decreased genes were significantly enriched for POU4F1 binding sites, and modestly enriched for selenocysteine synthesis. **Conclusion & future directions:** SSc samples are clinically characterized by fibrosis, but little inflammatory cell infiltrate. The skin transcriptomes demonstrate increased expression of genes involved in extracellular matrix, but also immune system processes. Few changes are detectable in PBMCs, but the presence of some significant changes supports the possibility of a blood biomarker of activity. The increased inflammatory signaling is in contrast to the general lack of immune cell infiltrate. A future direction will be to use single-cell RNA-Seq of dissociated skin to identify the cell lineage driving inflammatory gene expression. Overall this study will provide a rich resource for continued study, matching in-depth clinical phenotyping with transcriptome data.

1996W

Developement of the scleroderma genotype-phenotype map for assessing phenotypic relationships and drug repurposing. *K. Chesmore[1], C. Guoshuai[1], M. Whitfield[1], S. Williams[2]. 1) Department of Genetics, Dartmouth College, Hanover, NH; 2) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH.*

Scleroderma, an inflammatory disease of the skin, is a rare and often debilitating disorder thought to affect between 1/8,000 to 1/33,000 people, although this number can vary drastically based on race and gender. Little is known about the underlying genetic/molecular components that contribute to disease risk of scleroderma, and as such, efforts to develop or repurpose drugs for the treatment of scleroderma has yielded little. In this study, we explore the GWAS catalog to assess the relationships of phenotypes to scleroderma based on shared gene and pathway associations. Through the identification of closely related phenotypes, we gain insight into the disease etiology of scleroderma and from this we are able to predict the repurposing of drugs with potential therapeutic value in the treatment of this disease. Such methods can provide an invaluable tool to improve the treatments of many human diseases and can have a particularly profound effect on rare and understudied diseases for which little information is known about the underlying biology and potential drug targets.

1997T

Polymorphisms of genes involved in extracellular matrix homeostasis may play a role in the risk to develop anterior cruciate ligament and medial meniscus tears. *L. Casilla[1,2], M.F. Leal[1,2], A.M. Oliveira[1,2], S.E.B. Santos[3], A.K.C Ribeiro-dos-Santos[3], G. Arliani[1], D. Astur[1], C. Franciozi[1], P. Debieux[1], M.C. Smith[1,2], A. Pochini[1], C. Andreoli[1], B. Ejnisman[1], M. Cohen[1]. 1) Ortopedia e Traumatologia, Universidade Federal de São; Paulo, São Paulo, Brazil; 2) Morfologia e Genética, Universidade Federal de São Paulo, São Paulo, Brazil; 3) Laboratório de Genética Humana e Médica, Universidade Federal do Pará, Brazil.*

Anterior cruciate ligament (ACL) and medial meniscus are the main injured structures in knee, mainly during high-impact sporting activities. The disequilibrium between extracellular matrix synthesis and degradation may be involved in the tear of ligament and meniscus. Recent studies suggest polymorphisms in genes of extracellular matrix in the ACL tear may be associated with the multifactorial etiology of tear. However, few studies evaluated polymorphisms in meniscus tear. We originally evaluated whether 32 polymorphisms in *COL1A1*, *COL1A2*, *COL2A1*, *COL3A1*, *COL5A1*, *COL12A1*, *FN1*, *TNC*, *TNXB*, *COMP*, *PLOD1*, *PLOD2*, *LOX*, *TGFβ1*, *TGFβR1*, *MMP2*, *MMP3*, *MMP9*, *MMP13* and *TIMP2* were associated with the risk of ACL or meniscus tears in a Brazil population. We genotyped 177 patients with ACL tear, 309 ACL paired controls, 132 patients with meniscus tear and 275 meniscus paired controls using TaqMan assays in a real-time PCR system. 62 molecular markers were used to determine the genetic ancestry in the studied admixed population. Univariate and multivariate logistic regression were used to investigate if the polymorphisms were associated with the risk to develop ACL and meniscus tears. The models were adjusted for gender, age and molecular ancestry. The distribution of genotypes was in Hardy-Weinberg equilibrium. We observed that the rs185819 (*TNXB*) CC carriers had a lower risk to develop ACL tear (p=0.034). A *TNXB* TT haplotype (p=0.002) and a *MMP9* AA haplotype (p=0.045) may be a protective factor to ACL tear. Moreover, rs3106796 (*COL3A1*) GG carriers had an increased risk to develop meniscus tear (p=0.021) and rs522616 C allele (*MMP3*) carriers may be a protective factor for meniscus tear (p=0.0023). The *TGFβ1* GG haplotype (p=0.0055) may be a protective factor for meniscus tear. The *MMP3* CT haplotype (p=0.009), *COL5A1* CT (p=0.0098) and AC (0.0073) haplotypes may have a risk to develop meniscus tear. Therefore, while *TNXB* and *MMP9* genetic variants may play a role in the etiology of ACL tears, *COL3A1*, *MMP3, TGFβ1* and *COL5A1* genetic variants may be associated with meniscus tear etiology.

1998F

Filaggrin variations associated to atopic dermatitis in Ecuadorian pediatric population. *V. Romero[1], I. Inoue[1], H. Nakaoka[1], S. Palacios[2], J. Cabezas[2].* 1) Human Genetics, National Institute of Genetics-Japan, Mishima, Shizuoka, Japan; 2) Centro de la Piel, Quito, Ecuador.

   Atopic dermatitis (AD) is the most common form of eczema in childhood, manifestations as a chronic, relapsing itchy rash that usually starts in early life and, in many children, wanes in severity later in childhood. Early-onset AD is caused by a skin barrier abnormality frequently caused by filaggrin gene (*FLG*) mutations in Northern Europeans and Asians. AD was positively correlated with greater distance from the equator and negatively correlated with mean annual outdoor temperature. However, globally Ecuador has the second highest country prevalence (22.5%) for children between 6 to 7 years, and the eight highest one for children between 13 to 14 years (16.5%). Furthermore, within Latin America, Quito has the highest risk of current symptoms of eczema in children between 6-7 years and the fifth highest risk for children between 13-14 years; but there are no studies analyzing a genetic relationship. We obtained samples from Centro de la Piel of Ecuadorian pediatric population with and without history of atopic dermatitis and focused on the mutations between groups. We identified several variations found in cases that differed from the most common ones described in Europe and Asia. Furthermore, we analyzed the difference of distribution of disease-cause variations among these populations.

1999W

Differential alternative splicing of *MAPT* in brains supports its role in the pathogenesis of Parkinson disease. *L. Wang[1,2], S. Lang[1], G. Garbiso[1], S. Sivasankaran[1], L. Wang[1,3], J. Vance[1,2], J. Young[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Biostatistics, University of Miami, Miami, Fl.

   Background: Microtubule-associated protein tau (MAPT) is an abundant brain protein that binds to microtubule and plays an important role in axonal stability and dynamics. Aggregation of the protein Tau in the brain has been implicated in many neurodegenerative diseases including Alzheimer's disease and frontotemporal dementia (FTD). Genome-wide association studies (GWAS) have repeatedly identified *MAPT* as a susceptibility gene for Parkinson disease (PD). The molecular mechanisms underlying the genetic association, however, remain unclear. *MAPT* has multiple alternative transcripts: the majority of them are generated via alternative usage of exon 2 (E2), E3, and E10 in the adult human brain. Herein, we investigated the role of *MAPT* alternative splicing in PD etiology.   Methods: The dorsal motor nucleus of the vagus nerve (DMV) region is where the earliest sign of PD pathology is observed in the brain. Postmortem DMV tissues from four PD cases (age: 72 +/- 6 years) and four sex-matched controls (age: 72 +/- 8 years) were obtained. Total RNA was extracted and measured by RNA-seq. Differential alternative splicing was analyzed using the rMATs software. Alternative splicing events with a false discovery rate < 0.05 were regarded as significant. Quantitative reverse transcript-PCR (qRT-PCR) were conducted in additional samples (10 cases and 4 controls) to validate significant RNA-seq findings.   Results: In the comparison of cases and controls, we confirmed three significant alternative splicing events in *MAPT*: 1). Decreased inclusion of E2 (22.9% vs 56.2%); 2). Increased inclusion of E10 (52.2% vs 20.4%); 3). Decreased retention of a 136bp fragment of intron 2 (5.4% vs 12.7%). E2 encodes a repeat domain that reduces the filament length of Tau-aggregation. Therefore, less inclusion of E2 could lead to longer filament. E10 encodes a microtubule binding domain which further stabilizes microtubules. Increased inclusion of E10 has been reported in FTD patients and has been associated with the H1 haplotype, which is associated with increased genetic risk for PD. The intron 2 retention results in a premature stop codon (76 amino acids vs 352~441 amino acids in wild type Tau). The truncated Tau may play a protective role by reducing Tau aggregation.   Conclusion: Our findings support a role of *MAPT* in the pathogenesis of PD. Future studies are needed to evaluate if alternative splicing is a mechanism to confer the genetic risk of *MAPT* in PD.

2000T

Mapping functional regulatory variants at Alzheimer's disease risk loci.
M. Allen[1], X. Wang[2], M.M Carrasquillo[1], D.J Serie[2], S. Lincoln[1], T. Nguyen[1], K. Malphrus[1], D.W Dickson[1], Y. Asmann[2], N. Ertekin-Taner[1,3]. 1) Dept Neuroscience, Mayo Clinic, Jacksonville, Jacksonville, FL; 2) Dept Health Science Research, Mayo Clinic, Jacksonville, Jacksonville, FL; 3) Dept Neurology, Mayo Clinic, Jacksonville, Jacksonville, FL.

Genome-wide association studies have identified common variants associated with risk for Alzheimer's disease (AD). We and others have shown that some of these variants also associate with brain expression levels of near-by genes. Importantly these findings implicate the biological mechanism of action and the likely influenced gene(s), and nominate transcriptional regulation as an important factor in AD pathogenesis. Despite these advances, the precise regulatory mechanism and functional genetic variants are largely not yet known. Furthermore, much of the heritability of AD remains unexplained, indicating additional genetic risk loci remain to be identified. We hypothesize that a systems biology approach may nominate additional AD candidate genes, and when combined with in-depth sequence analysis, identify functional regulatory risk variants at both known and novel loci. We previously generated transcriptome-wide brain gene expression measures from two brain regions (temporal cortex and cerebellum), and genome-wide genotypes, for ~ 200 AD subjects and ~200 subjects with other pathologies (nonAD). Gene expression levels were measured using Illumina's WG-DASL array and genotypes were generated using the Illumina Hap300 chip. We performed eQTL analysis using PLINK, differential expression analysis using R statistical software, and co-expression network analysis using WGCNA, in both brain regions. We are performing targeted sequencing, in these same samples, across selected loci using the Agilent Haloplex target enrichment system and will annotate and test identified variants to validate the functional impact on gene expression. Our prior work indicated that AD index risk variants were associated with expression of *CLU*, *ABCA7* and *PILRB*. Transcription profiling analyses identified 10 genes that were consistently differentially expressed (DEG) between AD and nonAD subjects in both the temporal cortex, which is affected by AD pathology, and the cerebellum, which is largely spared. Co-expression network analysis identified 8 hub genes with high connectivity in "immune" networks enriched for AD candidate genes. Sequencing is being performed across both coding and non-coding regions for the 3 eQTL candidate loci and 12 genes nominated by the systems biology approach. The identification of functional, AD risk, regulatory variants is expected to provide novel insights into the pathophysiology of this disease and nominate therapeutic targets.

2001F

Whole-exome sequencing analyses for late-onset Alzheimer's disease in Japanese. *Y. Asanomi[1], D. Shigemizu[1,2,3], Y. Nagata[1], R. Mitsumori[1], T. Mori[1], S. Niida[1], K. Ozaki[1,4].* 1) Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Alzheimer's disease (AD), the most common form of dementia, has been thought to have familial and common form (late-onset AD; LOAD). With an aging population, the prevalence of LOAD is expected to dramatically increase. LOAD is thought to result from complex interactions among multiple genetic and environmental factors. Although recent advances in genetic studies have discovered several genetic risk factors for LOAD, more than the half of the heritability of the disease remains unclear. To explain the missing heritability, several efforts that identify rare variants to have strong genetic effect are currently underway. Although a risk rare-variant in *TREM2* was found with relatively strong effect in a Caucasian study, the carrier of this variant is hardly detectable in the Japanese population. To identify the rare variants that would explain a part of genetic architecture for LOAD in Japanese, we have performed whole-exome sequencing analyses with 202 LOAD individuals. We narrowed down the candidate variants and the corresponding genes that may relate to LOAD by the definitions as below; 1) the genetic variants to have minor allele frequency less than 1%, 2) the scores for combined annotation dependent depletion (CADD), a tool for scoring the deleteriousness of variants, are convincing, 3) the genes are ordinary expressed in the brain, 4) the genes carry a number of the candidate variants. All variants called from the exome data of the patients were filtered through these definitions. We filtered the data of allele frequencies in public database, such as 1000 Genomes Project, Exome Sequencing Project, Exome Aggregation Consortium, and Japanese population reference panel (2KJPN). We excluded the genes hardly expressed in the brain on Genotype Tissue Expression (GTEx) database. In addition, we excluded the variants identified from our in-house exome data of 176 normal-cognitive elderlies. We are now focusing on the candidate rare variants and investigating the genetic pathway implicated in LOAD.

2002W

Identification and validation of novel key drivers of Alzheimer's disease from multiscale causal networks that integrate large-scale DNA, RNA, and proteomic data. *N.D. Beckmann[1,2], W.J. Lin[3], M. Wang[1,2], A. Cohain[1,2], G.M. Belbin[1], W. Ma[1], P. Wang[1], M. Ehrlich[4,5], B. Zhang[1,2], S. Salton[3], E.E. Schadt[1,2].* 1) Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Departments of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY.

Alzheimer's Disease (AD) is the prevailing cause of dementia, affecting more than 28 million worldwide. Further, its incidence is projected to double in the next 20 years. Its phenotypic characteristics include progressive loss of cognitive function, memory, ability to think and reason. AD patients present with senile plaques in the brain that are composed of accumulated Amyloid Beta (ABETA) and neurofibrillary tangles (hyperphospho-TAU protein). To date, several genetic components of familial and late onset AD and many biological pathways have been implicated in the disease. Nonetheless, causal genes, pathways/networks and mechanisms of the disease continue to elude the scientific community, with no effective drugs developed over the last decade to prevent, halt or reverse the onset or progression of disease. Here, we study gene and protein expression from the anterior prefrontal cortex in a large cohort of post-mortem samples (n>300), across the complete range of neuropathologies (from controls to definite AD; AMP-AD consortium). Using whole exome sequencing (WES) with the multiscale expression data, we identified 4224 eQTLs and 158 pQTLs, with 74 gene and protein products having both eQTLs and pQTLs. Differential expression analyses revealed 788 differentially expressed (DE) genes and 1016 DE proteins between AD cases and controls, highlighting mitochondrial and immune processes. As expected, ABETA was the protein with the highest log fold change in AD patients versus controls. Coexpression network analyses highlighted 4 gene modules enriched for DE genes, which were enriched for known disease pathways. We then constructed 3 Bayesian Causal Networks (BN): a gene BN seeded with an expanded gene list from the DE genes and 4 modules of interest; a protein BN seeded with all 2694 proteins; and a multiscale BN, comprised of all genes and proteins present in the other networks. We ran key driver (KD) analyses for DE genes and DE proteins in all networks, and identified important regulators of the disease. Strikingly, we identified a novel KD not previously associated with AD that was the only KD conserved across all 3 networks. When overexpressed in AD mouse model, we find that this KD lowered overall Abeta plaque and phospho-Tau level, validating our network findings and the importance of this gene as a driver of AD. Additionally, the genes and clinical features linked to this KD provide novel insights into the mechanisms underlying AD risk and pathogenesis.

2003T

Whole-genome sequencing in non-Hispanic white familial late-onset Alzheimer's disease identifies rare variation in AD candidate genes. *G.W. Beecham[1], B. Vardarajan[2], E. Blue[3], S. Barral[2], J.L. Haines[4], W. Bush[4], C.M. van Duijn[5], E.R. Martin[1], G.D. Schellenberg[6], R. Mayeux[2], E. Wijsman[3], M.A. Pericak-Vance[1], Alzheimer's Disease Sequencing Project.* 1) Dr. John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst.for Human Genomics, University of Miami, Miami, FL, USA; 2) Columbia University, New York, NY, USA; 3) University of Washington, Seattle, WA, USA; 4) Case Western Reserve University, Cleveland, OH, USA; 5) Erasmus MC, Rotterdam, Netherlands; 6) University of Pennsylvania, Philadelphia, PA, USA.

The Alzheimer's Disease Sequencing Project (ADSP) is an initiative to identify rare genetic variation influencing Late Onset Alzheimer's Disease (LOAD) risk. As part of the ADSP, we performed whole-genome sequencing (WGS) in 44 non-Hispanic white (NHW) extended families multiply affected by LOAD, followed by extensive quality control, variant filtering, and gene-based association tests. WGS data were generated for 197 persons from 44 NHW families, including both AD and cognitively intact relatives. Alignment was performed using the Burrows-Wheeler algorithm, followed by genotype calling using a consensus calling pipeline that used both GATK genotype calls and ATLAS genotype calls. Variants were annotated for allele frequency and predicted functional impact, categorized into damaging (e.g., loss of function, high CADD scores, etc) and likely damaging (e.g., non-synonymous, moderate CADD, etc). We performed gene-based association testing of 32 known AD candidate genes, accounting for family structure using the FSKAT software. Association was performed using rare variants (MAF<0.01) and two models (damaging set only; damaging and likely damaging). Examination of the 32 known LOAD and EOAD genes (largely identified by GWAS-based meta-analyses) confirmed the role of rare functional variation in a number of genes, including FERMT2 (p =0.001) and SLC24A4 (p=0.009 in NHW). The association in FERMT2 survives a Bonferroni correction for 32 genes tested (p-value = 0.05/32 = 0.00156). Both genes still showed association after adjusting for age, sex, and principal components (FERMT2 p-value=0.002; SLC24A4 p-value=0.023). The PICALM gene also showed nominal association after adjusting for age, sex, and principal components (p-value = 0.032; unadjusted p-value = 0.111). SLC24A4 also showed nominal association in the damaging variant only analysis (p-value=0.026). These three genes still showed evidence of association after including index SNP genotypes as covariates (FERMT2 p-value=0.002, SLC24A4=0.015, PICALM p-value=0.021). This indicates that the associated variation in the genes is likely independent of the common variants that initially implicated the genes. These results suggest rare, functional variation may influence LOAD risk in multiplex families, even among genes identified through common variation. Variants are currently being validated using other technologies, and follow-up and replication analyses are ongoing.

2004F

Novel candidate AD-risk loci identified through whole exome sequencing in African Americans. *M.M. Carrasquillo[1], C.C.G. Ho[1], A. N'Songo[1], X. Wang[2], J.D. Burgess[1], M. Allen[1], T. Nguyen[1], G.D. Bisceglio[1], S.J. Lincoln[1], K.G. Malphrus[1], Y.W. Asmann[2], S.G. Younkin[1], R. Duara[3], M.T. Greig Custo[3], N.R. Graff-Radford[4], N. Ertekin-Taner[1,4].* 1) Department of Neuroscience, Mayo Clinic, Jacksonville, FL; 2) Department of Health Science Research, Mayo Clinic, Jacksonville, FL; 3) Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL; 4) Department of Neurology, Mayo Clinic, Jacksonville, FL.

Alzheimer's disease (AD) is twice as prevalent in African-American individuals, yet most AD studies have been conducted in Caucasians. Whole exome sequence data from 131 African-American AD patients and 107 non-demented African-American elderly controls were tested for association with disease status, both at the single variant level using a linear regression model adjusted for appropriate covariates, and at the gene level using SKAT-O. The SKAT-O analysis was performed using default weights and various minor allele frequency (MAF) upper and lower bounds (none, MAF ≥ 5%; 5% > MAF ≥ 0.5%, MAF < 0.5%), and the test was only performed when 2 or more variants could be analyzed in each gene. We followed up in an independent set of African-American AD cases and controls (67 AD vs. 234 controls) associations that met any of the following criteria: 1) 25 variants that have the smallest AD association P-values; 2) AD association with P<0.05, prioritizing 25 variants with the largest CADD Phred score; 3) variants not observed in our controls, prioritizing 25 variants with the largest minor allele counts in ADs and that have a MAF<0.001 in public datasets; 4) variants within the gene that had the best P-value for gene-based association. We identified several variants with nominally significant associations with AD status. A non-synonymous variant in the gene encoding the junction-mediating and -regulatory protein (*JMY*) had the best P-value in the single variant analysis in the replication cohort (p=0.035, odds ratio=2.36), and remains significant after Bonferroni correction for all 97 variant tested when the discovery and follow-up cohorts were analyzed jointly (p=0.00004, odds ratio=3.16). *JMY* has been determined to be a modulator of neuritogenesis and more recently has been shown to play an important role in autophagy. Interestingly, *JMY* and 4 other genes with a variant whose AD-risk association improves in the joint analysis show significant differential gene expression (FDR q-value <0.05) in the cerebellum of AD vs. controls in the Mayo RNAseq AMP-AD cohort (syn7332090). Employing whole exome sequencing in a cohort of African-Americans, we have identified associations with potentially functional coding variants at genes not previously associated with AD-risk. Our results warrant replication in independent cohorts, given the paucity of studies in African-Americans, and the need to identify risk factors that may be more relevant to AD in this minority population.

2005W

A patient-derived iPSC model of a rare *TTC3* mutation segregating with Alzheimer's disease. *H.N. Cukier[1,2], F.S. Johnson[1], C. Garcia-Serje[1], R.M. Carney[1,3], J.M. Vance[1,2,4], M.L. Cuccaro[1,4], M.A. Pericak-Vance[1,2,4], D.M. Dykxhoorn[1,4].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, University of Miami, Miami, FL; 3) Miami Veterans' Affairs Medical Center, Miami, FL; 4) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL.

While 20+ loci have been associated with late onset Alzheimer's disease (LOAD), a large number of genetic factors still have yet to be revealed. We recently performed whole exome sequencing on a non-Hispanic white LOAD family with 11 AD individuals with a range of onset between 70–85 years old (average 75.7 years) that was negative for alterations in known AD candidate genes (Kohli, et al, 2016). A single rare, nonsynonymous variant was identified in all affected individuals in the *tetratricopeptide repeat domain 3* (*TTC3*) gene. This missense alteration, rs377155188 (p.S1038C), is predicted to be deleterious by five *in silico* algorithms and extremely rare in the ExAC database (1.96x10^{-5}). Studies have reported that *TTC3* expression is reduced in LOAD patients and negatively correlated with AD neuropathology, making this gene an interesting LOAD candidate to further investigate. To understand the mechanism by which the *TTC3* alteration may be acting to contribute to LOAD risk, induced pluripotent stem cells (iPSC) lines were developed to examine cellular consequences in neuronal tissue, which can otherwise only be collected postmortem. Peripheral blood mononuclear cells (PBMCs) were extracted from the whole blood of two LOAD individuals bearing the *TTC3* change, as well as two aged, and ethnically matched control individuals. iPSC reprogramming was performed using Sendai virus. Each iPSC line generated was validated for pluripotency through immunocytochemical staining and shown to be negative for any large-scale chromosomal abnormalities via karyotyping. These iPSC lines will be differentiated into forebrain neurons to generate a disease relevant tissue to assess. There is evidence that modulation of *TTC3* affects neurite growth (Berto, et al, 2007); therefore, morphological measures of axon and synapse formation will be assessed in differentiating neuronal cultures using live cell imaging. Alzheimer-specific phenotypes such as the levels of secreted beta amyloid and intercellular whole and phosphorylated tau will also be measured. Utilizing patient-specific iPSC lines will permit assessing the cellular and molecular consequences of genetic alterations that may contribute to LOAD risk. We will present the morphological and cellular phenotypes of these lines carrying a potential risk factor for AD.

2006T

Genome-wide association study of brain amyloid deposition as measured by PiB-PET imaging and assessment of the genetic variance of amyloid deposition. *F.Y. Demirci[1], Q. Yan[1], K. Nho[2], J. del-Aguila[3], X. Wang[1], S. Risacher[2], B. Snitz[1], H. Aizenstein[1], C. Mathis[1], O. Lopez[1], E. Feingold[1], W. Klunk[1], A.J. Saykin[2], C. Cruchaga[3], M.I. Kamboh[1].* 1) University of Pittsburgh, Pittsburgh, PA; 2) Indiana University, Indianapolis, IN; 3) Washington University, St. Louis, MO.

Genomic efforts mainly through large-scale genome-wide association studies (GWAS) have identified over 20 risk loci for late-onset Alzheimer's disease (AD). However, known common AD variants in these loci account for only ~30% of the AD genetic variance. Genetic studies focusing on AD-related quantitative endophenotypes may help to identify additional AD-related genes. One such AD-related endophenotype is deposition of amyloid-beta (Aβ) in the brain, which is one of the two main pathologic hallmarks of AD; the other being the formation of tau tangles in the brain. The *in vivo* detection of Aβ deposition in the brain, as measured by PET scanning with ^{11}C-labeled Pittsburgh Compound-B (PiB) and the increased retention of PiB in AD than cognitively normal brains has been confirmed in many studies. Here we have used PiB-PET as an endophenotype to identify novel genetic loci for AD pathology using the GWAS approach, the first to our knowledge, using the largest sample (n=1,019) with the PiB-PET imaging from University of Pittsburgh (PITT), Washington University (WU), Alzheimer's Disease Neuroimaging Initiative (ADNI) and Indiana University (IU). In the GWAS meta-analysis, the PiB uptake values from 4 cortical regions were averaged in each subject to calculate a mean global score as the quantitative endophenotype. With the exception of *APOE*, none of the other known AD risk loci showed significant association with PiB score. The *APOE* region showed the most significant association where several SNPs surpassed the genome-wide significant threshold *(P*<5E-08) with *APOE*4* most significant (*P*-meta = 9.09E-30; β= 0.18). Interestingly, after conditioning on *APOE*4,* 14 SNPs remained significant at *P*<0.05 in the *APOE* region that were not in linkage disequilibrium with *APOE*4,* including *APOE*2*. Outside the *APOE* region, the meta-analysis revealed 15 non-*APOE* loci with *P*<1E-05 on 9 chromosomes, with two most significant SNPs on chromosomes 8 (*P*=4.87E-07) and 3 (*P*=9.69E-07). The top 15 non-*APOE* SNPs along with *APOE*4* and *APOE*2* explained 35%, 25% and 31% of the amyloid variance in the PITT, WU and ADNI/IU datasets, respectively; of which 14-17% was explained by *APOE*4* alone. In conclusion, we have identified novel signals in the *APOE* and non-*APOE* regions, in addition to *APOE*4* that affect the Aβ deposition in the brain. Our data also highlights the presence of yet to be discovered variants that may be responsible for the unexplained genetic variance of Aβ deposition.

2007F

Complement receptor 1 (*CR1*) intragenic duplication and Alzheimer's disease. *E.J. Hollox[1], E. Kucukkilic[1], K. Brookes[2], K. Morgan[2].* 1) University of Leicester, Leicester, United Kingdom; 2) University of Nottingham, Nottingham, United Kingdom.

Genomewide association studies have highlighted SNP alleles within and surrounding genes involved in the immune response with late-onset Alzheimer's disease (LOAD), suggesting an important role of the immune response in the etiology of Alzeimer's disease. One of the strongest associations surrounds the complement receptor 1 gene (*CR1*), which encodes a gene important in the innate immune complement pathway. The associated SNP alleles form a risk haplotype, and several variants on this haplotype have been suggested to be the causative variant underlying the increased risk of LOAD. One of these variants is a 18 kb duplication that affects several exons of the *CR1* gene, causing an in-frame insertion of coding sequence and a consequent extra protein domain binding complement C3/C4. Here, we develop a robust approach to type *CR1* intragenic copy number, including the common duplication (*CR1*B) and deletion (*CR1*C) alleles, using a PCR-based approach called the paralogue ratio test (PRT). We also develop a locked nucleic acid junction-fragment PCR approach to validate duplication *CR1*B alleles in European samples. Applying these methods, we fail to show association of the *CR1*B allele with early-onset Alzheimer's disease partients (EOAD, p=0.755 OR=1.05, 95% CI 0.77 to 1.43) in a UK cohort of 449 cases and 184 controls, but confirm the association in LOAD patients (p=0.015 OR=1.21 95% CI 1.04-1.42) in a UK cohort of 1180 cases and 1005 controls. We suggest that the *CR1*B allele is the functional basis of the association, perhaps by altering the number of C3/C4-binding domains encoded by the protein and thereby modulating the immune response to neurodegeneration. We suggest further functional comparison of proteins encoded by *CR1*B risk alleles with those encoded by *CR1*A non-risk alleles to understand the mechanism of this association.

2008W

Genome-wide association study for Alzheimer's disease in a Puerto Rican dataset. *J. Jaworski[1], B. Feliciano[2], K. Celis[1], F. Rajabli[1], K.L. Hamilton-Nelson[1], LD. Adams[1], AR. Betancourt[2], JM. Vance[1], ML. Cuccaro[1], GW. Beecham[1], MA. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) PRADI and Universidad Central del Caribe, School of Medicine, Caguas, PR, USA.

Background: There are over 5 million individuals of Puerto Rican (PR) origin, making up over 1.5% of the US population and representing the 2nd largest Hispanic/Latino population in the continental US. The population is highly admixed, with average ancestry values of 64% European, 21% African, and 15% Native American, but individual ancestries are highly variable. The number of Alzheimer's disease (AD) cases on the island of PR is estimated at 65,000. Genetic studies have been limited to a few rare families with early onset AD (PSEN mutations). Our goal in this study is to identify the genetic risk factors associated with AD. **Methods:** We have ascertained 395 individuals into the PR AD Initiative (PRADI) study including 97 families and 124 unrelated individuals for a total of 271 affected and 124 cognitively intact individuals. We performed genome-wide genotyping on the IlluminaMEGA and GSA arrays. Association was tested using the GWAF geepack package in R which fit logistic regression via Generalized Estimation Equation (GEE) for the additive model, treating each family as a cluster. **Results:** The top SNP observed, rs35293463, had a p-value of 2.4×10^{-7} which passed a Bonferroni corrected significance level. This SNP is located on Chromosome 2 (2p12) in the intron region of the predicted gene LOC105374814. APOE was nominally significant with a p-value of 0.013 and OR of 1.5. This is not as strong an effect as seen in non-Hispanic Whites (NHW)(See Rajabli et al abstract). A look at GWAS candidate AD loci observed in NHW datasets showed two genes were nominally significant; SLC24A4 (rs7150295, p-value 0.004) and CASS4 (rs7274581, p-value 0.007). In addition, APP (rs3991, p-value 0.007) was also nominally significant. **Conclusions**: The Puerto Rican population is a growing yet understudied population in the US. Understanding the etiology of Alzheimer's disease in this unique population will unlock keys to the disease in the population as a whole. This preliminary analysis of GWAS association suggest that associations observed in AD, while present, are not as strong in PR AD. We observed a novel top SNP on Chromosome 2 in an intronic region of a predicted gene. Genotyping more families and replicating the analysis is warranted.

2009T

A whole exome study of Alzheimer's disease which is augmented by population data found the noble AD risk genes. *J. Kim[1,3], M. Fernandez[1,2], J. Budde[1,2], H. Oscar[1,2], C. Cruchaga[1,2].* 1) Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid Ave. B8134, St. Louis, MO 63110, USA; 2) Hope Center for Neurological Disorders. Washington University School of Medicine, 660 S. Euclid Ave. B8111, St. Louis, MO 63110, USA; 3) Department of Neurology, Ilsan hospital, Koyang-shi.

Alzheimer's disease (AD) has high heritability. The currently known risk genes includes ABCA7, APOE, BIN1, CD33, CD2AP, CLU, CR1, EPHA1, MS4A6A/MS4A4E, PICALM, PLD3, SORL1, and TREM2 between others, and explain only part of AD heritability. Recently, a large scale whole exome study called as Alzheimer's Disease Sequencing Project (ADSP), aimed to find additional AD risk genes harboring low frequency and rare coding variants. However, although ADSP includes as much as 10,909 participants, it may be still underpowered for very rare alleles. The ExAc database which includes genetic information on 60,706 individuals, and should be a good representation of the allele frequencies in the population. In this study, we augmented statistical power by using the ExAc database as comparison group for the ADSP cases and tried the noble AD risk and protective genes. In order to avoid the population bias, we included only Non-Finnish European population in ExAc. In ADSP, we used European ancestry and excluded outliers by the plot of the first two principal components. We selected variants with minor allele frequencies (MAF) between that of the ADSP cases and ADSP controls. The MAF differences were tested by Pearson's chi-square tests. The MAF of gene-based analysis were obtained by collapsing MAFs of variants within gene boundaries. In the variant level, the well-known AD risk genes–TOMM40, TREM2, and MS4A6A–are replicated. The noble genes with genome wide significant level (P = 5×10^{-8}) were 38 genes including C10orf2, CHRD, and CROCC. In the gene-based level using burden test, we selected genes with P < 1×10^{-6}, considering the number of human genes. The well-known AD risk genes–ABCA7, SORL1, and TREM2–are replicated. The significant noble genes were 69 genes includeing AGAP1, ALYREF, and TYRO3. We not only replicate known AD risk genes–ABCA7, MS4A6A, SORL1, and TREM2, but also found the noble candidate AD risk genes. We need to replicate these noble risk genes in other genetic studies. Our augmentation methods can be applied to the whole exome sequencing studies on other diseases.

2010F

Alzheimer's disease exome sequencing study in the Finnish population isolate. *M.I. Kurki[1,2,3], S. Helisalmi[4], K.M Mattila[5], A.S. Havulinna[3,6], J. Krüger[8,9], T. Lehtimäki[5], M. Perola[6,10], V. Salomaa[6], A. Remes[7,8,9], M. Hiltunen[11], H. Soininen[12], M.J. Daly[1,2,3,13], A. Palotie[1,2,3].* 1) 1)Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 2) Stanley Center, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 4) Institute of Clinical Medicine, University of Eastern Finland; 5) Faculty of Medicine and Life Sciences, University of Tampere, Finland; 6) THL-National Institute for Health and Welfare, Helsinki, Finland; 7) Department of Neurology, Institute of Clinical Medicine; 8) Medical Research Center, Oulu University Hospital, Oulu, Finland; 9) Unit of Clinical Neuroscience, Neurology, University of Oulu, Oulu, Finland; 10) Estonian Genome Center, University of Tartu, Tartu, Tartumaa 51010, Estonia; 11) Institute of Biomedicine, University of Eastern Finland; 12) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 13) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

Both common and rare variants contributing to Alzheimer's disease (AD) have been identified. High impact variants explaining the majority of the heritability for familial early onset Alzheimer's disease (EOAD) have been identified in genes related to amyloid-β processing. For the most common late onset Alzheimer's disease (LOAD), genome wide association studies have identified mostly low effect common variants in > 20 genomic loci (APOE*ε4 being the only high risk exception) suggesting many different pathways like metabolism, inflammation, synaptic activity and intracellular trafficking. To enrich for deleterious low frequency variants in our study sample, we investigated AD cases in a population isolate. Finland is the largest genetic population isolate in Europe where multiple genetic bottlenecks and genetic drift have increased the frequency of some deleterious alleles with negative selection. Previously we have identified Finnish enriched variants associated with schizophrenia, intellectual disability and cardiovascular diseases. So far, whole exome and targeted sequencing studies have identified predisposing rare coding variants (e.g. *APP*, *TREM2*, *PLD3*) with risk increases on par with that of APOE*ε4. Still the majority of the estimated heritability of LOAD (60-80%) remains unexplained and further studies are needed to pinpoint specific pathomechanisms to identify novel targets for treatment development. Here we describe an ongoing exome sequencing case control study of AD in the Finnish population. To date we have sequenced 1592 Finnish AD cases and 18186 controls. 1501 cases and 14983 controls passed the quality control process. We identified rare (MAF < 1%) damaging missense variants in known AD genes (e.g. *TREM2* AF: 3.36E-05, OR : 20.9 , p: 0.02; *SORL1*, AF (controls): 0, p 6.4e-4; *CR1* AF (controls) :0, p: 6.7E-3; *APP* , AF (controls): 3.3e-3, OR: 0.17, p: 3.96E-3) nominally associated with AD as well as new candidates not previously implicated by GWAS. To identify novel genes we are increasing the number of samples exome sequenced or genotype imputed using cases and controls identified from Finnish electronic health care registries. Finnish registry data allows exploration of these rare AD variants across hundreds of disease endpoints and clinically relevant phenotypes, better articulating their biological impact.

2011W

Performance of a genetically-based biomarker risk algorithm for an Alzheimer's disease prevention screening study. *M.W. Lutz[1,2], C.L. Dunn[3], J.S. Cranston[3], D.J. Hatch[2,4], D.K. Burns[5], A.M. Saunders[5], B.L. Plassman[1,2,4], K.A. Welsh-Bohmer[1,2,4].* 1) Department of Neurology, Duke University School of Medicine, Durham, NC 27710; 2) Duke Bryan Alzheimer's Disease Research Center, Duke University School of Medicine, Durham, NC 27705; 3) Duke University, Durham, NC 27708; 4) Department of Psychiatry, Duke University School of Medicine, Durham, NC 27710; 5) Zinfandel Pharmaceuticals, Chapel Hill, NC 27517.

The efficiency of clinical trials for prevention or delay-of-onset of Alzheimer's Disease (AD) is greatly increased if an enrichment strategy successfully identifies individuals likely to convert from normal cognition to AD or mild cognitive impairment (MCI) due to AD in the time frame of the trial. A genetic biomarker risk algorithm (GBRA) has been developed to enrich a double-blind, delay-of-onset clinical trial of the efficacy of low dose pioglitazone, a thiazolidinedione used for the treatment of Type 2 diabetes, to delay the onset of MCI due to AD (TOMMORROW, ClinicalTrials.gov identifier, NCT00106899). The GBRA is based on age and two genetic variants that are easily genotyped in blood: APOE and TOMM40 '523. The GBRA has been designed to allow subject enrichment across the ages of 65-83 years and was calibrated with data from the Duke Bryan ADRC Memory, Health and Aging Study. Other studies have compared the performance of the GBRA with imaging and biofluid biomarkers. The present study evaluates the GBRA for a longitudinal community based AD prevention screening study with measures of cognitive function, e.g. Montreal Cognitive Assessment (MoCA). The GBRA was tested with data from the Duke Bryan ADRC Prevention Screening Study and Database/Repository (PREPARE); comprised of a non-demented socioculturally diverse volunteer cohort from Durham, NC. The GBRA was used to stratify subjects at baseline into high or low risk groups. Individuals who had baseline and follow-up neurocognitive evaluations were included, N=423, mean age 69 years, sd 9.5. 303 individuals were classified as low risk at baseline, 120 as high risk. The difference in change between baseline and follow-up MoCA scores between the risk groups was 0.5 with the high risk group showing the largest decline (mean 1.1, sd 0.2) compared to the low risk groups (mean 0.6, sd 0.1). This difference was statistically significant (p=0.02) after adjusting for different follow-up intervals. Individuals showing both cognitive decline and cognitive improvement were observed in both risk groups. Cognitive, genetic and demographic data are presented to characterize the risk groups. The results of this study characterize the performance of the GBRA for measures of cognitive decline. When used in the context of a clinical trial, the GBRA has potential to significantly reduce the numbers of study subjects required and, in the case of a delay-of-onset study, reduce trial duration.

2012T

Novel Alzheimer disease loci identified in subsets of whole exome sequencing data stratified by APOE genotype. *Y. Ma[1], X. Zhang[1], J. Chung[1], A.C. Naj[2], K. Hamilton[3], E.R. Martin[3], B.W. Kunkle[3], R. Mayeux[4], J.L. Haines[5], M.A. Pericak-Vance[3], G.D. Schellenberg[2], K.L. Lunetta[1], L.A. Farrer[1], Consortium for Alzheimer Sequencing Analysis, the Alzheimer Disease Sequencing Project.* 1) Biomedical Genetics, Boston University, Boston, MA; 2) University of Pennsylvania, Philadelphia, PA; 3) University of Miami, Miami, Florida; 4) Columbia University, New York, New York; 5) Case Western Reserve University, Cleveland, Ohio.

Background: Previous genome-wide association studies identified novel Alzheimer disease (AD) loci that were detected only when stratified by presence or absence of the *APOE* ε4 allele. We evaluated whole exome sequence (WES) data from the Alzheimer Disease Sequencing Project (ADSP) to identify additional APOE-stratified associations with infrequent and rare single nucleotide variants (SNVs). **Methods:** Association with SNVs in the ADSP WES sample of unrelated non-Hispanic white AD cases and controls was evaluated in subgroups of *APOE ε4* carriers and non-carriers: *ε4+* (2377 cases, 706 controls) and *ε4-* (3145 cases, 4213 controls). We used EPACTS score test adjusting for principal components of ancestry in Model 1 and with additional adjustment of age and sex in Model 2. Analyses were restricted to SNVs with a minor allele count (MAC) $\geq$ 10, resulting in 87,405 and 123,178 tests in the ε4+ and ε4- subgroups, respectively. Genotype data that were imputed using the Haplotype Reference Consortium panel for ADGC participants not included in the ADSP were used for replication. **Results:** Among ε4+ subjects, novel associations reached genome-wide significance (GWS) ($P<6 \times 10^{-7}$) with *C4ORF17* (rs41275705, $P=5.6 \times 10^{-7}$ in Model 1) and *SNPH* (rs367658252, $P=5.0 \times 10^{-7}$ in Model 2), but neither of them were replicated. Two GWS ($P<4 \times 10^{-7}$) novel associations were found in *AC099552* (rs536940594, $P=2.2 \times 10^{-7}$ in Model 1) and *GPAA1* (rs138412600, $P=2.7 \times 10^{-7}$ in Model 2) among ε4- subjects. The effect of *GPAA1* rs138412600 was stronger in *ε3/ε3* (OR=2.28, $P=6.8 \times 10^{-7}$) than *ε2/ε2 and ε2/ε3* subjects (OR=1.49, $P=0.25$). Association with this SNV was confirmed with the same model and effect direction in the *ε4-* replication sample ($P=0.03$ and meta-$P=7.7 \times 10^{-8}$). Replication results were unavailable for the *AC099552* SNV because it was too rare (MAF=0.0007) to be imputed accurately. **Conclusion:** We identified a novel robust association with *GPAA1* in subjects lacking the *APOE* ε4 allele. *GPAA1* encodes glycosylphosphatidylinositol (GPI) anchor attachment protein, which has a male-specific function in memory formation. Further studies are necessary to determine the functional consequence of this non-synonymous mutation and its mechanism of action in relation to *APOE* genotype.

2013F

Ambidexterity and Alzheimer's disease risk. *E.E. Mlynarski[1,2,3], M. Tang[1,2,3], A. Amlie-Wolf[1,2,3], L. Qu[1,2,3], A. Kuzma[1,2,3], L. Cantwell[1,3], J.M. Ringman[4,5], G.D. Schellenberg[1,3], L.-S. Wang[1,2,3].* 1) Penn Neurodegeneration Genomics Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Philadelphia, PA; 2) Institute for Biomedical Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 4) Memory and Aging Center, Keck School of Medicine of University of Southern California, Los Angeles, CA; 5) Department of Neurology, University of Southern California, Los Angeles, CA.

Previous studies have investigated the association between left handedness and Alzheimer's disease, however these studies either excluded ambidextrous individuals, and/or did not include genome-wide genetic analyses. Therefore, we performed association analyses to examine the prevalence of ambidexterity and left handedness in late-onset Alzheimer's disease. More than 7000 individuals from the NACC database were included in the analysis, all of which were of European descent and had available handedness, Alzheimer's disease case/control status, birth year, sex and education level phenotype data. Interestingly, our analysis revealed an association between Alzheimer's disease and ambidexterity, but not left handedness. In fact, the results suggest that ambidexterity is protective for late-onset Alzheimer's disease (OR = 0.591, 95% CI = 0.396–0.875, P = 9.26 x 10^{-3}). A multi stage GWAS was then performed for each of the four different handedness splits: ambidextrous vs left or right (A vs L/R), left vs right or ambidextrous (L vs R/A), right vs left or ambidextrous (R vs L/A), and left vs right (L vs R). The discovery analysis consisted of 4151 individuals with handedness phenotype data from six previously published Alzheimer's GWAS datasets, the replication was a meta-analysis of more than 4000 additional subjects from several independent Alzheimer's disease cohorts, and a meta-analysis of the discovery and replication cohorts was also performed. In the discovery, only one locus in the A vs L/R analysis reached genome-wide significance (p < 10^{-8}). The one genome-wide significant signal (p = 4.07 x 10^{-9}), located at 10p12.33, replicated in the A vs L/R meta-analysis (p = 3.35 x 10^{-9}). The top SNP from the meta-analysis mapped to an intergenic region ~16 kb distal to the 3' end of *ST8SIA6*. Notably, there were no significant signals for any of the other three handedness splits (L vs R/A, R vs L/A, and L vs R), and the few suggestive loci (p < 10^{-5}) that were identified in the discovery analysis did not replicate. Together, these results indicate that ambidexterity is associated with a lower risk of AD and the locus on 10p12.33 may play a role. Additional functional analyses are currently underway to identify the specific causal variant for the GWAS signal and elucidate the underlying genetic mechanisms.

2014W

Variant prioritization by pedigree-based haplotyping in an Alzheimer's disease pedigree. *R.A. Nafikov[1], A.Q. Nato[1], H. Sohi[1], A.R. Horimoto[1], S. Ahmad[2], N. Amin[2], S. van der Lee[2], E. Blue[1], B. Wang[3], M. Saad[1,4], P. Navas[1], M. Dorschner[5], G. Beecham[6], B.N. Varadarajan[7], S.M. Barral[7], J.L. Haines[8], W.S. Bush[6], E.R. Martin[6], G.D. Schellenberg[9], R.P. Mayeux[7], M.A. Pericak-Vance[6], A.L. DeStefano[10], S. Seshadri[11], E. Boerwinkle[12], C.M. van Duijn[2], E.M. Wijsman[1,4] for the Alzheimer's Disease Sequencing Project (ADSP).* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 2) Department of Epidemiology, Erasmus MC, Rotterdam, Netherlands; 3) Department of Statistics, University of Washington, Seattle, WA, USA; 4) Department of Biostatistics, University of Washington, Seattle, WA, USA; 5) Department of Pathology, University of Washington, Seattle, WA, USA; 6) Department of Human Genetics, University of Miami Health System, Miami, FL, USA; 7) Department of Neurology, Columbia University, New York, NY, USA; 8) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 9) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA; 10) Department of Biostatistics, Boston University, Boston, MA, USA; 11) Department of Neurology, Boston University, Boston, MA, USA; 12) Department of Human Genetics, UT Health Science Center at Huston, Huston, TX, USA.

  Disease-associated linkage regions contain large numbers of variants. This complicates determining which are associated with a disease. In the absence of recurrent mutation, a new risk allele will occur on a single haplotype. Identification of variants unique to a haplotype driving a linkage signal could therefore provide a solution to the problem. Existing population-based haplotype phasing methods, however, require large reference samples and dense SNP marker panels for optimal performance. These are often unavailable in family-based studies. To address a variant prioritization problem in a linkage region, we developed a pedigree-based haplotyping method and applied it to Alzheimer's disease (AD) data from the ADSP. Our study sample consisted of subjects from recent generations of a population isolate of European decent with deep shared ancestry. The subjects were grouped into two pedigrees (P1 and P2), with 30 of 189 and 5 of 47 members, representing AD cases, respectively, and with WGS data available on 30 AD subjects from P1 and P2, combined. Using the MORGAN package, inheritance vectors representing chromosomal decent within pedigrees were generated using sparse SNP array data on 25 and 5 subjects in P1 and P2, respectively, with subsequent computation of LOD scores. A LOD score of 3.3 on chr5 was identified in P1. A modified version of the GIGI program was used to perform pedigree-based phasing and imputation of 49,740 WGS variants from the linkage region. This yielded a most-likely haplotype across the region, shared by 10 AD cases in P1. Identification of the risk haplotype responsible for the linkage signal reduced the number of variants of interest to 24,521. Further reduction to 276 variants was achieved by identifying variants that were unique to the risk-haplotype as opposed to being on haplotypes from subjects who did not carry the risk-haplotype in P1 and P2. The number of variants potentially associated with AD can be reduced further to 38 or 21 by applying minor allele frequency thresholds of 5% or 1%, respectively. These results show that pedigree-based haplotyping with subsequent identification of a haplotype driving a linkage signal may be a useful tool for WGS variant prioritization and for follow up in other samples.

2015T

Contribution to Alzheimer's disease risk of rare variants in *TREM2, SORL1* and *ABCA7* in 1,779 cases and 1,273 controls. *G. Nicolas[1,2], C. Bellenguez[3,4,5], C. Charbonnier[1], B. Grenier-Boley[3,4,5], O. Quenez[1], K. Le Guennec[1], G. Chauhan[6], D. Wallon[1], S. Rousseau[1], A.C. Richard[1], A. Boland[7], G. Bourque[8], H.M. Munter[8], R. Olaso[7], V. Meyer[7], A. Rollin-Sillaire[9], F. Pasquier[9], L. Letenneur[6], R. Redon[10], J.F. Dartigues[6], C. Tzourio[6], T. Frebourg[1], M. Lathrop[8], J.F. Deleuze[7], D. Hannequin[1], E. Génin[11], P. Amouyel[3,4,5,12], S. Debette[6], J.C. Lambert[3,4,5], D. Campion[1,13], CNR MAJ collaborators.* 1) Normandie Univ, UNIROUEN, Inserm U1245 and Rouen University Hospital, Department of Genetics and CNR-MAJ, F 76000, Normandy Centre for Genomic and Personalized Medicine, Rouen, France; 2) Radboud University Medical Center, Genome Research, Department of Human Genetics, Nijmegen, The Netherlands; 3) Inserm, U1167, RID-AGE - Risk factors and molecular determinants of aging-related diseases, F-59000 Lille, France;; 4) Institut Pasteur de Lille , F-59000 Lille, France; 5) University Lille, U1167 - Excellence Laboratory LabEx DISTALZ, F-59000 Lille, France; 6) Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, UMR1219, F-33076 Bordeaux, France; 7) Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Evry, France; 8) McGill University and Génome Québec Innovation Centre, Montréal, Canada; 9) CNR-MAJ; and Department of Neurology, Université de Lille, CHU, Inserm UMR-S 1171, Lille, France; 10) Inserm UMR-1087 / CNRS UMR 6291, l'institut du thorax, Univ. Nantes, Nantes, France; 11) Inserm UMR-1078, CHRU Brest, Univ. Brest, Brest, France; 12) Centre Hospitalier Universitaire de Lille, Epidemiology and Public Health Department, F-59000 Lille, France;; 13) Department of Research, Centre hospitalier du Rouvray, Sotteville-lès-Rouen, France.

  Alzheimer disease (AD) is a complex disorder with high genetic component. Exome-wide significant association of *TREM2, SORL1* and *ABCA7* rare variants with AD risk has recently been reported, however with limited panel of variants or sample sizes, restricting the possibility to fully assess their effect. We took advantage of the ADES-FR dataset made up of 1,779 AD cases and 1,273 controls from France. We focused on *TREM2, SORL1* and *ABCA7* but also included *PLD3*, *AKAP9* and *UNC5C* genes, for which association of rare variants with AD risk has recently been reported. We generated whole genome (n=955) or whole exome sequencing data (n=2,106) and performed multiple-sample calling followed by stringent individual and variant quality controls. We then performed gene-based burden association analysis from exome data of 927 late-onset AD (LOAD, onset >65years) cases, 852 early-onset AD (EOAD, onset ≤65) cases and 1,273 controls. When aggregating rare (MAF<1%) protein truncating (PT) and missense variants predicted damaging by three bioinformatic tools (strictly damaging, SD), we found exome-wide significant association (p<2.5e-6) between EOAD risk and rare variants in *TREM2* (OR=6.3, 95%CI=[2.9-13.9], p=2.3e-7), *SORL1* (OR=3.4 [2.1-5.6], p=1.6e-7) and *ABCA7* (OR=2.5 [1.7-3.8], p=9.3e-7). The association with *TREM2* rare variants was robust to the exclusion of the p.R47H variant, previously described as associated with AD risk. No exome-wide significant signal was obtained in the LOAD sample while significance of the order of 10^{-6} was observed in the whole AD group for *TREM2* (OR=4.8 [2.3-10.1], p=6.2e-6). No association was found with *PLD3*, *UNC5C* and *AKAP9* rare variants. With a cumulative MAF of 0.63% among controls and an OR of 6.3, *TREM2* rare PT+SD variants explained 1.17% of EOAD heritability. *SORL1* rare PT+SD variants explained 1.42% of it (OR of 3.4, cumulative MAF of 1.89%), and *ABCA7* rare PT+SD variants, 1.33% (OR of 2.5 and cumulative MAF of 3.22%). In comparison, 9.12% of EOAD heritability was attributable to *APOEε4*. Our study confirms previous gene-level results for *TREM2, SORL1* and *ABCA7* and provides a clearer insight into the classes of rare variants involved, including enlargement of the association to *TREM2* PT+SD variants. Despite different effect sizes and cumulative MAFs, *TREM2, SORL1* and *ABCA7* PT+SD rare variants contribute similarly to EOAD heritability with 1.1 to 1.5% of EOAD heritability attributed to each.

2016F

Single variant and polygenic score analysis of whole exome data from the Alzheimer's Disease Sequencing Project (ADSP). *J.S. Reddy[1,2], J.M. Biernacka[3], G.D. Jenkins[3], S.K. McDonnell[3], V. Sarangi[3], J.P. Sinnwell[3], Y.W. Asmann[2], M.M. Carrasquillo[1], C.P. Pottier[1], Y. Ren[1,2], C.S. Younkin[1], R. Rademakers[1], O.A. Ross[1], N. Ertekin-Taner[1], S.G. Younkin[1].* 1) Neuroscience, Mayo Clinic, Jacksonville, FL; 2) Health Sciences Research, Mayo Clinic, Jacksonville, FL; 3) Health Sciences Research, Mayo Clinic, Rochester, MN.

Alzheimer's Disease (AD), the most common form of dementia, is known to have a strong genetic component. To evaluate the association of exonic variants with AD, we analyzed data from the Alzheimer's Disease Sequencing Project, which consists of whole exome sequence (WES) from 10,929 subjects generated by three large scale genomic sequencing and analysis centers (LSACs). WES fastq files obtained from dbGaP (phs000572.v4.p2) were processed using Mayo Clinic's GenomeGPS pipeline, which aligned reads with Novoalign, and variants were identified using GATK's Best Practices workflow. Quality scores were assigned to variants using the GATK's variant quality score recalibration (VQSR) tools. Pooling information from public databases, variants were annotated using BioR and ANNOVAR. Rigorous quality control (QC) measures were implemented to retain a set of high quality, unrelated samples with relatively homogenous population substructure. Samples were QCed for coverage, call rate, missingness, TiTv ratio, sample contamination, sex check, sample duplication, relatedness, principal component analysis and APOE genotypes. Autosomal variants that passed VQSR, had a 95% call rate, met HWE criteria in controls, and had a minor allele count of at least 20 were analyzed. After sample and variant QC, 9904 samples and 102,828 variants remained. Single variant analysis, conducted by logistic regression using an additive genetic model adjusting for sex, APOE4 dose, APOE2 dose, LSAC, and three principal components addressing population substructure, yielded 5 variants with study-wide significance. Remarkably, additional QC to assess heterogeneity in minor allele frequency (MAF) across the three LSACs revealed that four of the five variants were likely false positives with markedly different MAFs across the LSACs (P.lsac<1E-34). Only rs75932628 encoding TREM2 p.R47H, which is known to associate with AD, achieved study-wide significance after this additional QC measure, which left 102,150 variants (P.lsac>4.8E-7). Logistic regression was also performed on variants with r²<0.2 in random sample(s) of 4952 subjects. Using the resultant coefficients, polygenic risk scores analyzed using the remaining half of subjects showed highly significant association with AD. Polygenic scores from relatively rare exonic variants (MAF < 1%) and from common variants (MAF>10%) were both significant. Polygenic scores from protein-altering variants and from synonymous SNVs were also both significant.

2017W

The roles of *CD33* and *TREM2* in neurodegeneration associated with Alzheimer's disease (AD) and frontotemporal dementia (FTD). *A. Rendina[1], S. Napoletano[1], F. Tripodi[2], D. Drongitis[3], A. Postiglione[4], G. Milan[2], E. Vitale[1].* 1) Institute of Protein Biochemistry (IBP), National Research Council (CNR), Naples, Naples, Italy; 2) Geriatric Clinic "Frullone" ASL Napoli 1, Naples, Italy; 3) Department of Biology, University of Naples Federico II, Naples, Italy; 4) Department of Clinical Medicine & Surgery, University of Naples "Federico II", Naples, Italy.

Frontotemporal dementia (FTD) and Alzheimer's disease (AD) are two multifactorial, heterogeneous and genetically complex neurodegenerative disorders. Despite massive research and drug development efforts, there are still no therapies that slow down or stop their progression. Several genes have been associated with both pathologies, suggesting that an as yet unknown common molecular pathway could exist. Mutations in genes related to neuroinflammation may be predisposing factors in these diseases. Coding and non-coding polymorphisms in genes expressed in microglia, including *CD33* and *TREM2*, are risk factors and could be entry points for therapeutic intervention. Two SNPs in *CD33* and one in *TREM2* were described to be risk factors associated with Late Onset Alzheimer disease (LOAD). Specifically, *CD33* SNPs rs3856444 and rs12459419 in minor alleles were found to confer strong protection while conferring elevated risk of LOAD in major alleles. These SNPs directly modulate the *CD33* exon 2 splicing efficiency. In addition, the rare heterozygous missense variant rs75932628-T in *TREM2* exon 2 was strongly associated with the ability of *TREM2* to activate microglial cells. In order to assess the presence of these polymorphisms in an FTD cohort, we analyzed 216 Caucasians, mean age 60–85 years, diagnosed with LOAD, and 50 age-matched healthy controls. We used High Resolution Melting analysis (HRM) on genomic DNA from the whole blood of patients as a screen for mismatches. We sequenced the DNA from individuals with different melting curves using the Sanger method and used these results as a reference in our analysis. Our patients exhibited the coinheritance of SNPs rs12459419 and rs3865444, being heterozygous in 40%, homozygous for the major allele in 56.48% and homozygous for the minor allele in 3.7% of individuals. In addition, we identified a third SNP in *CD33* exon 2, rs2455069, which belongs to a previously identified LD SNP block associated with an increased rate of cognitive decline. We found that all patients analyzed for SNP rs75932628 in the *TREM2* gene are homozygous for the wild-type allele. However, some individuals are heterozygous for the nearby SNP rs143332484, which could be potentially associated with AD in our population. Further investigations are in progress to understand the mechanism of action of these two genes.

2018T

Targeted sequencing of deep-phenotyped individuals for Alzheimer's disease susceptibility prediction. *J. Seo[1], M. Byun[2], D. Lee[3], M. Choi[1], The KBASE consortium.* 1) Biomedical Sciences, Seoul National University, Seoul, South Korea; 2) Institute of Human Behavioral Medicine, Medical Research Center Seoul National University, Seoul, Republic of Korea; 3) Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Republic of Korea.

Alzheimer's disease (AD), the most common form of dementia, is often classified on the basis of the onset age; late-onset AD (LOAD) is defined as onset of 65-year-old and more. Although having multifactorial origins, LOAD is clearly a genetic disorder, with heritability known to be 60-80%. Therefore, understanding the genetic etiology of LOAD will provide valuable insights into the pathogenesis of AD. Attempts to identify common genetic variants that are associated with LOAD susceptibilities have generated a number of candidate genes and variants with questionable functional implications. Here, we set up a targeted sequencing analysis of the coding and UTR regions of 132 LOAD-associated genes to determine if they carry previously uncharacterized rare variants with minor allele frequency less than 5%. Prior to the data production, we optimized the variant calling pipeline through a series of pilot experiments to ensure 93.5% sensitivity and 94.0% specificity in variant calling. The evaluation of imaging, psychological, psychiatric and experimental assessments in the KBASE (Korean Brain Aging Study for Early Diagnosis and Prediction of Alzheimer's disease) cohort provided more than 46 traits that can be utilized for genetic association test for 336 control, 137 mild cognitive impairment (MCI) and 84 LOAD cases. The cohort was analyzed by the Ion Torrent panel sequencing at a mean coverage depth of more than 400x. The initial association tests yielded 38 novel signals that exceeded the multiple-test corrected cutoffs from 20 traits. To validate the results, a larger replication set containing 3,113 control, 606 MCI and 845 LOAD cases was recruited. This study is an efficient and powerful genome analysis on common complex diseases to understand disease-causing mechanism and provide early-on disease prognosis prediction.

2019F

Genetic markers in LUZP2 and FXBO40 genes are associated with the normal variability in cognitive performance in the elderly. *V.A. Stepanov[1], K.V. Vagaitseva[1], A.V. Bocharova[1], A.V. Marusin[1], V. Markova[2], L.I. Minaicheva[1,2], O.A. Makeeva[1,2].* 1) Institute of Medical Genetics, Tomsk National Medical research Center, Tomsk, Russian Federation; 2) Nebbiolo LLC, Tomsk, Russian Federation.

Cognitive performance is an important endophenotype for various neurodegenerative and neuropsychiatric diseases. In the present study two genetic variants in leucine-zipper protein (LUZP2) and F-box 40 protein (FBXO40) genes, previously reported to be genome-wide significant for Alzheimer's diseases and schizophrenia, were examined for association with cognitive abilities in normal elderly from the Russian population. Rs1021261 in LUZP2 and rs3772130 in FBXO40 were genotyped by multiplex PCR and MALDI-TOF mass spectrometry in a sample of 708 normal elderly subjects tested for cognitive performance using the Montreal Cognitive Assessment (MoCA). Association of genetic variability with the MoCA scores was estimated by parametric and non-parametric analysis of variance and by frequency comparison between upper and lower quartiles of MoCA distribution. Significantly higher frequency of 'TT' genotype of rs1021261 in LUZP2 gene, as well as "A" allele and 'AA' genotype of rs3772130 in FBXO40 gene were found in a subsample of individuals with MoCA score less than 20 comparing to fourth quartile's subsample (MoCA > 25). The data of the present study suggests that genetic variability in LUZP2 and FXBO40 loci associated with neurodegenerative and neuropsychiatric diseases is also contributed to normal variability in cognitive performance in the elderly. This work was funded by the Russian Science Foundation (project No. 16-15-00020).

2020W

Whole genome sequence analysis of Caribbean Hispanic families with late onset Alzheimer's disease. *B. Vardarajan[1], S. Barral[1], G.W. Beecham[2], E. Blue[3], D. Reyes-Dumeyer[1], J. Haines[4], W. Bush[4], C. Van Duijn[5], E. Martin[2], G. Schellenberg[6], E. Wijsman[3], M. Pericak-Vance[2], R. Mayeux[1], Alzheimer s Disease Sequencing Project.* 1) Neurology, Columbia University, New York, NY; 2) Neurology, University of Miami, Miami, FL; 3) Medical Genetics, University of Washington, Seattle, WA; 4) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 5) Department of Epidemiology & Biostatistics, Erasmus University Medical School Rotterdam, The Netherlands; 6) Department of Pathology and Laboratory Medicine University of Pennsylvania School of Medicine, Philadelphia, PA.

   Background. Late-onset Alzheimer's disease (LOAD), the most common form of dementia, lacks effective treatment and is an enormous emotional and financial burden to society. Genetic research in early onset forms of the disease were the basis for current treatment strategies, but genome-wide association and exome sequencing studies of late onset disease have identified several novel susceptibility loci clustering in unique pathways that could lead to additional approaches. In this study we aim to determine whether family-based sequencing can provide a comprehensive and detailed knowledge of rare genetic variation leading to LOAD. **Methods:** We analyzed whole genome sequencing in members of 115 Caribbean Hispanic families with multiple relatives affected by LOAD. Recruitment for the study was based in medical clinics in the Dominican Republic, Puerto Rico and New York. Family ascertainment and follow-up has been in progress since 1998. All patients met recommended research criteria for LOAD, and controls were tested to be dementia free. We investigated rare genetic variants segregating within families or associated with LOAD. Single variant and gene-based association tests were performed, accounting for family structure. **Results.** A variant in *AKAP9,* p.R434W, segregated in two large families and was nominally associated with disease (OR=5.75, 95% CI 1.07–31.06, p=0.040). In addition, missense mutations in *MYRF* and *ASRGL1* under previously reported linkage peaks at 7q14.3 and 11q12.3 segregated completely in one family and in follow-up genotyping were nominally significant (p<0.05) in association with LOAD. We also identified rare variants in a number of genes implicated by GWAS, including *CR1* (p=0.049), *BIN1* (p=0.0098) and *SLC24A4* (p-value=0.040). **Conclusions and Relevance.** These results suggest rare variants in multiple genes influence LOAD risk in multiplex families. The results also support the view that rare variants may underlie loci established by GWAS.

2021T

A new CAG repeat disease responsible for X linked cerebellar ataxia? *L. Parodi[1], ML. Monin[1,2], M. Raynaud[3], G. Stevanin[1,2,4], A. Brice[1,2], A. Durr[1,2].* 1) ICM Institut du Cerveau et de la Moelle, INSERM U1127, CNRS UMR7225, Sorbonne Universités – UPMC Université Paris VI UMR_S1127, Paris, France; 2) APHP Department of Genetics, Groupe Hospitalier Pitié-Salpêtrière, Paris France; 3) CHRU de Tours, Service de Génétique, 37000 Tours ; INSERM U930, 37000 Tours; 4) Ecole Pratique des Hautes Etudes (EPHE), Paris Sciences et Lettes (PSL) Research University, Neurogenetics team, 75013 Paris, France.

   Spinocerebellar ataxias (SCA) are inherited disorders predominantly affecting afferent and efferent cerebellar connections, resulting in a loss of gait and coordination as a disease hallmark. Among dominantly inherited SCAS the most prevalent group is due to pathological coding CAG repeat expanions in *ATXN1, 2, 3, CACAN1A* and *TBP* genes. We present a French family in which known CAG repeat expansions have been excluded, presenting with a spastic ataxia in two brothers at age 25 and 30. There was also apallesthesia, progressive dystonic head tremor and progressive cognitive impairment. Both manifested severe cerebellar atrophy. Both their mother and maternal grand-mother, even in absence of any functional complain, manifested moderate cerebellar dysarthria. The less severe presentation in the mother was in favor of mitochondrial or X-linked transmission. Since mitochondrial work-up was negative (normal muscular biopsy and respiratory chain levels, mtDNA quantification and sequencing) we performed Whole Exome Sequencing of the affected brothers and their father. Bioinformatic analysis was carried out in order to detect maternally inherited mutations. We identifed a duplication (p.Gln2109_Gln2115dup) harbored in *MED12* C-terminal region, predicted to add 7 Gln residues to the Glutamine-rich Opa domain resulting in a stretch of 14 consecutive CAG codons. Situated on the X chromosome, *MED12* encodes for the subunit 12 of a multiprotein complex necessary for mRNA transcription, crucial during neuronal development and differentiation and playing a fundamental role in both, Wnt and Shh pathways. Sanger sequencing validation showed that the duplication was carried in an heterozygous state by the mother and in hemizygous state by the two affected brothers, being absent in the father and in an unaffected daughter. Moreover, it is not present in public databases (ExAc, gnomAD) and absent in 324 French additional controls. Polyglutamine expansions in known SCAs genes are larger (21 CAG repeats in *CACAN1A* as smallest) but polyalanine expansions responsible for *OPMD* exhibit only 10-15 repeats. We are suspecting another mechanism than polyglutamine toxicity in this family and functional studies are currently underway. *MED12* involvement in critical aspects of neural development and the absence of the duplication in large control populations, are strongly supporting the hypothesis that the detected *MED12* rearrangement is a new cause of X-linked SCA.

2022F

Neurobehavioral traits in family members inform gene discovery in ASD.

S. Luzi[1], M.L. Cuccaro[1,2], E.R. Martin[1,2], L. Gomez[1], A.J. Griswold[1], H.N. Cukier[1], P.L. Whitehead[1], J.L. Haines[3], J.P. Hussman[4], M.A. Pericak-Vance[1,2]. 1) Human Genomics, University of Miami - Miller School of Medicine (HIHG), Miami, FL; 2) Dr. John T. Macdonald Department of Human genetics, university of Miami, Miami, FL; 3) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio; 4) Hussman Institute for Autism, Baltimore, MD.

Autism (AUT) is multi-dimensional phenotype with a complex genetic architecture. Prior studies have shown that both neuropsychiatric (NPD) and neurodevelopmental (NDD) disorders (e.g., schizophrenia, ADHD, and bipolar disorder) have overlapping features and may have a shared molecular basis. We hypothesize that the presence of NPD, NDD, and subclinical AUT features within the same family concentrates autistic liability across generations with common genetic drivers. Using items from the ADI-R, we developed the Quantitative Autism Score (QAS) and applied it to create a phenotypically homogenous AUT cohort. We divided this cohort (N=520 individuals with AUT from the Simons Simplex Collection and Hussman Institute for Human Genomics) into two groups, based on family history of NPD, NDD and subclinical autistic features in relatives. The High Burden group (HB; N=185) consisted of individuals with AUT from families with a strong presence of NPD, NDD, and subclinical AUT features. The Low Burden group (LB; N=335) consisted of individuals with AUT from families with no family history of NPD, NDD or subclinical features. Using the QAS as our outcome measure, we conducted a SKAT-O gene-based test on WES data. In the HB group, *LMAN1L*, *GREB1L* and *EIF4A2* were significant (33513 genes analyzed, p-value=9.7E-06; 6.4E-05 and 6.8E-05). *LMAN1L* was also the most significant when only damaging variants were analyzed, surviving Bonferroni correction (18917 genes; p-value=2.4E-05). *LMAN1L* exerts its function in glycoprotein transportation and organelle targeting. *EIF42A* has been reported to be overexpressed in frontotemporal dementia and down regulated in schizophrenia. In the LB group *PMFBP1*, *MIS18A* and *RPL14* were significant (p-value=3.2E-06; 7.2E-05 and 9.9E-05). When only damaging variants were analyzed, *SLC35F1* was the most significant (p-value=5.7E-05). *PMFBP1* is involved in the general organization of the cellular cytoskeleton and localized in an AUT CNV enriched region. *RPL14* is reported to interact with other genes in mice models of AUT, whereas *SLC35F1* is highly expressed in the fetal brain. Our results show that when we held the autism phenotype constant through the QAS, we were able to detect differences in autistic liability derived from family history. Our results point towards the presence of different mechanisms/different genes leading to ASD and suggest that family liability may bolster AUT gene discovery.

2023W

Genome-wide analysis in pediatric-onset multiple sclerosis (MS) confirms a role for adult MS risk variants and reveals new candidates.

L.F. Barcellos[1], X. Shao[1], J. Graves[2], A. Waldman[3], T. Lotze[4], T. Schreiner[5], B. Greenberg[6], B. Weinstock-Guttman[7], G. Aaen[8], J.M. Tillema[9], J. Ness[10], M. Candee[11], L. Krupp[12], M. Gorman[13], L. Benson[13], T. Chitnis[14], S. Mar[15], I. Kahn[16], C. Casper[11], J. Rose[11], M. Moodley[19], M. Rensel[19], M. Rodriguez[9], Y. Harris[10], A. Belman[18], S. Roalstad[11], M. Goyal[15], C. Schaefer[17], E. Waubant[2]. 1) University of California, Berkeley, Berkeley, CA; 2) University of California, San Francisco, San Francisco, CA; 3) Children's Hospital of Philadelphia, Philadelphia, PA; 4) Texas Children's Hospital, Houston, TX; 5) University of Colorado School of Medicine, Aurora, CO; 6) University of Texas Southwestern, Dallas, TX; 7) SUNY Buffalo, Buffalo, NY; 8) Loma Linda University, Loma Linda, CA; 9) Mayo Clinic, Rochester, MN; 10) University of Alabama at Birmingham, Birmingham, AL; 11) University of Utah, Salt Lake City, UT; 12) NYU Langone Medical Center, New York, NY; 13) Boston Children's Hospital, Boston, MA; 14) Brigham and Women's Hospital, Boston, MA; 15) Washington University St. Louis, St. Louis, MO; 16) Children's National Medical Center, Washington, DC; 17) Kaiser Permanente Division of Research, Oakland, CA; 18) SUNY Stony Brook, Stony Brook, NY; 19) Cleveland Clinic, Cleveland, OH.

Multiple sclerosis (MS [MIM 126200]), a complex autoimmune disorder, is the leading cause of non-traumatic neurological disability in young adults. Disease onset typically occurs between the ages of 20 and 40; however, ~5% of all MS patients have symptoms before 18 years. Exact mechanisms involved in pediatric MS pathogenesis remain unclear; however, both genetic and environmental factors have been implicated in disease susceptibility. The largest GWAS performed in adult MS recently identified 200 non-MHC MS variants (IMSGC, unpublished). Here, we describe the first GWAS study in white pediatric-onset MS cases and controls using the Illumina and Affymetrix platforms. Whole genome profiles for 500 pediatric MS cases and 11,577 controls were studied. Genetic imputation was performed using IMPUTE2, using the 1000 Genomes Phase 3 integrated variant set (March, 2012). Similar to adult onset MS, *HLA-DRB1*15:01* showed a strong association in pediatric MS (OR=2.71, 95% CI 2.34-3.14, p=6.9 x 10^{-40}). Within the MHC region, after conditioning on *HLA-DRB1*15:01*, strong associations within the class I region genes (*HLA-B, HCG27*, and *PSORS1C1*; OR=2.12, p= 1.78 x 10^{-5}) and class II region genes (*HLA-DPA1* and *MIR689*; OR=2.31, p=1.0 x 10^{-5}) remained. Top established non-MHC adult MS risk variants associated with pediatric MS included rs6498163 (OR=1.29, 95% CI 1.14-1.46, p=8.2 x 10^{-5}) and rs2286974 (OR=1.28, 95% CI 1.12-1.48, p=2.3 x 10^{-4}) within *CLEC16A*; rs2150879 within *VMP1* (OR=1.29, 95% CI 1.14-1.47, p=8.9 x 10^{-5}) and rs9878602 within *FOXP1* (OR=1.24, 95% CI 1.09-1.42, p=0.0009). More than 20% of tested adult MS risk variants demonstrated p<0.05 and had a mean OR=1.2 in pediatric-onset MS compared to OR=1.1 in adult MS. Furthermore, we searched the entire genome for new risk variants in pediatric-onset MS. Several gene regions reached genome-wide significance: SNPs clustering in genes *CHRD* and *THPO* on chromosome 3q27.1 (OR=2.92, 95% CI= 2.13, 4.18, p=1.7 x 10^{-10}) and SNPs near or within gene *NOD1* on chromosome 7p14.3 (OR= 3.14, 95% CI=2.12, 4.65, p= 9.9 x 10^{-9}) were observed. Results overall suggest variants within genes involved in adult-onset MS are also risk factors in children and suggest that similar biological processes are present in both groups.

2024T

Multiple sclerosis in Orkney: The contribution of common variants to excess prevalence. *C.L.K. Barnes[1], P.K. Joshi[1], T.S. Boutin[2], C. Hayward[2], H. Campbell[3], J.F. Wilson[1,2,3].* 1) Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, University of Edinburgh, Institute of Genetics and Molecular Medicine, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK; 3) Centre for Global Health Research, The Usher Institute, University of Edinburgh, 9 Little France Road, Edinburgh, EH16 4UX, UK.

The Orkney Islands, located north of mainland Scotland, have the highest prevalence of Multiple Sclerosis in the world (4.02 / 1000 individuals)[1]. The prevalence of MS in Orkney is significantly higher than that found in mainland Scotland (a maximum of 2.29 / 1000 in Aberdeen)[1], and higher than Shetland, a group of islands 50 miles north of Orkney (3.02 / 1000)[1]. Although various genetic and environmental risk factors for MS have been established, the reason for this excess prevalence of MS in Orkney is not clear. Several causes have previously been studied in Orkney, including vitamin D deficiency[2] and homozygosity[3], but neither explain this excess prevalence. Compston et al (1981)[4] suggested that the reason behind this excess prevalence was due to a higher frequency of common risk variants in Orkney, occurring by chance due to the founders of the islanders' gene pool having higher frequencies of these variants. Here, we aim to explore this suggestion by comparing the common risk variants in Orkney to those in mainland Scotland to ask if there is a difference in their frequency, and if so, how that translates into disease prevalence. We calculated polygenic risk scores (PGRS, from an MS GWAS not including Orcadians) in controls from Orkney and Shetland (using the ORCADES and VIKING cohorts) and compared these to controls from Lowland Scotland (using individuals from Generation Scotland). When comparing PGRS between Orkney, Shetland and mainland Scotland, there was a small, but significant difference between the means of each population, with Orkney having the highest risk. However, this did not translate into a meaningful difference in prevalence that would be expected if common risk variants were the cause of the excess prevalence of MS in Orkney. This suggests that another cause, possibly rare variants, are more likely to be behind the excess prevalence of MS in Orkney. <u>References</u> 1. Visser, E. M. *et al*. A new prevalence study of multiple sclerosis in Orkney, Shetland and Aberdeen city. *J. Neurol. Neurosurg. Psychiatry* **83**, 719–24 (2012). 2. Weiss, E. *et al*. Farming, Foreign Holidays, and Vitamin D in Orkney. *PLoS One* **11**, e0155633 (2016). 3. McWhirter, R. E. *et al*. Genome-wide homozygosity and multiple sclerosis in Orkney and Shetland Islanders. *Eur. J. Hum. Genet*. **20**, 198–202 (2012). 4. Compston, A. Multiple sclerosis in the Orkneys. *Lancet* **318**, 98 (1981).

2025F

Elucidating the role that genetic ancestry plays on the impact of variation in the major histocompatibility complex on risk of multiple sclerosis. *A. Beecham[1,2], L. Amezcua[3], A. Chinea[4], G. Beecham[1,2], S. Delgado[5], N. Patsopoulos[6], J. Oksenberg[7], J. McCauley[1,2].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 3) Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 4) San Juan MS Center, Guaynabo, Puerto Rico, USA; 5) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA; 6) Department of Neurology and Ann Romney Center for Neurological Diseases, Brigham and Women's Hospital, Boston, MA, USA; 7) Department of Neurology, University of California at San Francisco, San Francisco, CA, USA.

The Major Histocompatibility Complex (MHC), located on chromosome 6p21.3, has long been known to demonstrate the strongest genetic association with multiple sclerosis (MS) risk, on its own explaining ~10% of the sibling recurrence risk. The most extensively studied and replicated association has been seen with HLA-DRB1*15:01, demonstrating the strongest genetic effect in Europeans and African Americans. However, the complex genomic structure of the region has made it difficult to comprehensively identify risk alleles across populations. The region has been most studied in individuals of Northern European ancestry, with additive and dominant effects for at least 13 different alleles. However, MHC variation across populations is high, with various population-specific alleles and allele frequency differences noted, and so those alleles may not be the most relevant in all populations. To assess MHC variation in Hispanic populations, we performed association analyses using genotype data from a custom Illumina array which were available on 1222 Hispanic MS cases and 1339 controls from sites across Puerto Rico and both the east and west coasts of the United States. Single Nucleotide Polymorphisms (SNPs) which tag ($R^2>0.2$) 9 of the 13 published alleles were available. Association with MS was confirmed for 5 of the 9 in this Hispanic sample (p ≤ 0.05) including HLA-DRB1*15:01, HLA-DQA1*01:01, HLA-DQB1*03:02, HLA-A*02:01, and rs9277565 (correlated with HLA-DPB1). However, association was not confirmed (p > 0.05) for HLA-DRB1*08:01, HLA-B*55:01, HLA-B*44:02, and rs2229092 (missense in *LTA*). Utilizing conditional linear modeling (p ≤ 1.0E-05) within the extended MHC on chromosome 6 from 27 to 34 Mb, the three most strongly associated independent signals observed were HLA-DRB1*1501 (OR = 2.45, p = 1.06E-22), rs9277565 (OR = 1.55, p = 5.80E-08), HLA-A*02:01 (OR = 0.75, p = 9.42E-07), and an intronic SNP in NFKB inhibitor like 1 (*NFKB1*) which serves as an eQTL for both DExD-box helicase 39B (*DDX39B*) and natural cytotoxicity triggering receptor 3 (*NCR3*) in lymphoblastoid cell lines (OR = 0.66, p = 5.80E-06). Further work is being done to characterize the local ancestry across each of these loci. This detailed investigation of the MHC provides a first step towards uncovering the MHC risk for MS in admixed Hispanic populations and chronicling the ancestral lineage of MHC risk haplotypes.

2026W

Genetics of vaccination related narcolepsy. *H.M. Ollila[1], A. Wennerstrom[2], M. Partinen[2], E. Mignot[1], J. Saarela[3], P. Saavalainen[4], O. Vaarala[5], P. Tienari[6], M. Perola[2].* 1) Psychiatry, Stanford University School of Medicine, Palo Alto, CA; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 4) Research Programs Unit, Immunobiology, University of Helsinki, Helsinki, Finland and Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland; 5) Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland; 6) Research Programs Unit, Molecular Neurology, Biomedicum, University of Helsinki, Helsinki, Finland; Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland.

Type 1 Narcolepsy is a severe hypersomnia caused by a specific loss of neurons producing hypocretin/orexin in the hypothalamus and affecting 1/3000 individuals. In 2009/2010 immunization towards pandemic H1N1 Influenza-A campaign was launched and the vaccine used in Northern European countries associated with increased risk for narcolepsy. We built a multilocus genetic risk score with established narcolepsy risk variants using whole genome sequencing and exome sequencing data from 7,000 individuals to examine the genetic load for narcolepsy in 81 individuals with vaccination related narcolepsy, a third of all cases from Finland. Previously discovered risk variants had strong predictive power ($P<2.2*10^{-16}$) in the vaccine related narcolepsy cases with only 4.9% of cases being assigned into the low risk category. Furthermore, at genome-wide level the strongest association was seen in *IL12RB1* (rs17885060, $P=2.93*10^{-7}$) and *ZXDC* a gene required for HLA class II transcription (rs11715293 $P=7.06*10^{-7}$). Our findings suggest that genetic risk factors have a major role in regulating predisposition for narcolepsy and genetic risk score gives high predictive power to detect individuals with high risk for narcolepsy.

2027T

Assessment of genomic variations in multiple sclerosis patients identifies mutations in *ADAMTS14*, *IL22RA2*, *HNRNPA1* and *TNPO1* genes indicating the existence of molecular mimicry and cytokine/interferon receptor pathway disruption inducing autoimmunity. *A.M. Veerappa.* Department of Studies in Genetics and Genomics, University of Mysore, Mysuru, India.

Multiple Sclerosis (MS) is a rare inflammatory autoimmune demyelinating and neurodegenerative disorder in which axonal injury, neuronal loss, and atrophy of the central nervous system leads to permanent neurological and clinical disability. Immune-mediated destruction of myelin is considered to be the major cause of MS. Etiology of MS involves the interplay between genetic and environmental factors. To understand the role of genetic factors; we performed Whole Exome Sequencing (WES) on five Relapsing Remitting Multiple Sclerosis (RRMS) cases using Illumina HiSeq 2500. Focusing on deleterious/damaging mutations, identified mutations in *HNRNPA1*, *TNPO1, ADAMTS14* and *IL22RA2* playing a role in pathogenesis of RRMS. Mutations in *ADAMTS14* and *IL22RA2* were found present in all samples. Mutations in *ADAMTS14* contributes to disruption of balance between disintegrating proteinases and the corresponding inhibitors in favor of proteolysis leading to pathologic brain ECM destruction and MRI lesions. *ADAMTS14* is also known to be involved in the regulation of inflammation and acquired immunity. *IL22RA2* on the other hand, is involved in cytokine and interferon receptor activities inducing neuro-inflammation. However, in one of the cases, along with *ADAMTS14* and *IL22RA2* mutations, damaging variants were also identified in *HNRNPA1* and *TNPO1* disrupting the binding of TNPO1 with nucleoporin, affecting the transport of HNRNPA1 leading to its accumulation in cytoplasm resulting in increased cellular stress causing cellular death. Further analysis revealed the presence of 62% homology between Glycoprotein E (gE) of VZV virus and M9 domain of HNRNPA1. This molecular mimicry prompts the production of auto-antibodies resulting in autoimmunity. Thus we hypothesize that the molecular mimicry between gE of VZV virus and M9 domain of HNRNPA1 leads to accumulation in cell's cytoplasm triggering the auto-reactive immune response in neuronal cells causing neuro-degeneration.

2028F

Genomic and functional evaluation of the role of the TNFSF14-TNFRSF14 pathway in susceptibility to multiple sclerosis. *M. Zuccala[1], N. Barizzone[1], E. Boggio[1], L. Gigliotti[1], M. Sorosina[2], R. Bordoni[3], F. Clarelli[2], S. Anand[3], E. Mangano[3], D. Vecchio[4], F. Esposito[2], C. Basagni[1], E. Corsetti[1], G. Predebon[1], R. Cantello[4], V. Martinelli[2], G. Comi[2], M. Leone[5], G. de Bellis[3], U. Dianzani[1], F. Martinelli-Boneschi[2,6], S. D'Alfonso[1].* 1) Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Avogadro University, Novara, Italy; 2) Laboratory of Human Genetics of Neurological Diseases, San Raffaele Scientific Institute, Milano, Italy; 3) National Research Council of Italy, Institute for Biomedical Technologies, Segrate, Milano, Italy; 4) MS Centre, SCDU Neurology, AOU Maggiore della Carità, Novara, Italy; 5) SC Neurologia, Dipartimento di Scienze Mediche, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy; 6) Department of Biomedical Sciences for Health, University of Milan, Milan, Italy.

Multiple Sclerosis (MS) is a complex disease. Over 100 MS susceptibility genes were identified. Among these, the strongest non-HLA signal in the Italian population maps in the Tumor Necrosis Factor (ligand) superfamily member 14 (*TNFSF14*) gene. *TNFSF14* encodes for LIGHT, a transmembrane glycoprotein expressed on various immune cells and involved in immune modulation and dendritic cells (DC) maturation. We demonstrated that an intronic variant (rs1077667) is the primarily associated one through a fine-mapping approach. Cis-eQTL analysis data from different public databases showed that carriers of MS risk allele have a lower *TNFSF14* RNA expression in EBV-transformed lymphoblastoid cell lines (Geuvadis, Bioportal, Gtex) and in PBMCs (Gtex), but not in brain cells (Braineac). These data are consistent with the observation that in heterozygous individuals the allelic expression is unbalanced in favor of the allele with higher eQTL expression (p<0.0001, RNAseq on 97 lymphoblastoid cells, Geuvadis). We confirmed these data by RT-PCR expression analysis on PBMC of 84 Italian MS and 80 healthy controls (HC): individuals with the MS risk genotype produced lower levels of *TNFSF14* transcript (p=1.1e-4). In addition, we observed that MS patients show significantly lower expression levels compared to HC (p=0.031). Preliminary analysis performed on peripheral blood of HC by flow cytometry showed that in myeloid DC (CD11c+) the homozygous individuals for the risk allele (N= 10) had a higher percentage of LIGHT positive cells compared to the others (N=10, p-value = 0.02). We compared the expression of LIGHT and its receptor HVEM (*TNFRSF14*) in 5 populations of monocyte-derived DC (MDC). LIGHT is expressed in all types of immature MDCs and was substantially upregulated in the mature (LPS-activated) MDCs, particularly in those derived using GM-CSF+IL-15 (p=0.04). Conversely, in this same population, HVEM was downregulated (p=0.04). In conclusion, we propose that an altered *TNFSF14* expression in immune cells driven by the intronic variant can contribute to the pathogenesis of MS. In particular, this variant seems associated with a low *TNFSF14* RNA expression in a mixed population of PBMCs and with a higher percentage of *LIGHT* positive cells in myeloid dendritic cells (representing only 1 – 2% of PBMCs) suggesting a cell specific influence of this variant on *LIGHT* expression at the protein level, which might have distinct effects on the surface and secreted forms of the protein.

2029W

Predicted expression of *TMEM163* is associated with traumatic brain injury risk in a biobank population. *J. Dennis, X. Zhong, N. Cox.* Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN.

Over 8 million Americans sustain a traumatic brain injury (TBI) annually, with sequelae ranging from persistent headaches to dementia and death. Identification of genes predisposing to TBI and its late effects could help identify those at highest risk of adverse outcomes. We report the first GWAS of TBI risk, conducted in the Vanderbilt University Biobank, BioVU, which includes DNA samples linked to electronic medical records in >270,000 patients. We conducted gene-based analyses in 9155 European-ancestry patients using PrediXcan, a computational tool that predicts gene expression levels across tissues from genome-wide genotype data in the target sample (BioVU) and a reference transcriptome dataset (GTEx). We tested 11005 genes in whole blood and 9164 genes in brain cortex for association with TBI defined by CDC criteria as one or more ICD codes for 'Cerebral laceration and contusion', 'Concussion', 'Intracranial hemorrhage following injury', 'Skull fracture', and 'Other intracranial injury'. Analyses were adjusted for sex, age (at first TBI or at last ICD code in the medical record), and principal components 1-3. TBI was diagnosed in 503 of 9155 patients. The median age at first TBI was 51.7 years, 6.8% of patients were <18 years of age at first TBI, and 57.8% of TBI patients were male. No genes expressed in whole blood were associated with TBI or any of its sub-categories after applying Bonferroni correction. Conversely, expression of *TMEM163* in cortex was associated with 1.59-fold increased risk of 'Other intracranial injury' (p=2.13x10-7), and was marginally associated with TBI overall (p=0.001). *TMEM163* is a transmembrane zinc transporter implicated by meta-GWAS in Parkinson disease (PD) risk. PD was diagnosed in 1.8% of TBI cases vs. 0.8% of controls (p=0.04), while 'Other neurodegenerative diseases' were diagnosed in 72.5% TBI cases vs. 35.1% of controls (p<2.2x10-16). Nonetheless, the association between *TMEM163* expression and TBI persisted when we stratified by neurodegenerative disease, and *TMEM163* expression was only marginally associated with other neurodegenerative disease (odds ratio=1.12, p=0.08) in 3755 cases and 5575 controls. Our analyses suggest that *TMEM163* expression in cortex is associated with the risk of TBI, and perhaps with the risk of neurodegeneration in general. Further research is warranted to determine whether *TMEM163* is independently associated with PD risk, or whether the association is mediated by TBI.

2030T

A genome-wide screen to identify suppressors of neurodegeneration in Gaucher disease. *S.U. McKinstry, K. Adams, I. Tsai, N. Katsanis.* Center for Human Disease Modeling and Department of Cell Biology, Duke University School of Medicine, Durham, NC.

   Gaucher disease (GD) is the most common lysosomal storage disorder with a broad range of symptoms including hepatosplenomegaly, blood and skeletal disorders, and neurological defects, including neurodegeneration. Gaucher is caused by mutations in *GBA,* leading to a loss of its glucosylceramidase activity and an accumulation of glucocerebroside in the lysosome. GBA enzyme replacement therapy is effective for many GD symptoms, but the enzyme cannot cross the blood-brain barrier and thus the neurological symptoms are left as an unmet need. While the precise mechanisms thorough which lysosome dysfunction leads to neurological defects is unknown, the link of *GBA* to a more common neurodegenerative disorder, Parkinson disease (PD), offers an important clue. Data from idiopathic PD patients and pathogenic single-gene mutations in PD has highlighted defects in mitochondria quality control through mitophagy, the process by which cells degrade damaged mitochondria, as potential pathomechanism in neurodegeneration. Importantly, defects in mitochondria quality control have been identified in GD patient cells, suggesting this mechanism could underlie neurodegeneration in GD as well. Disease severity in GD can vary widely suggesting that modifier genes play an important role in disease penetrance. In some cases, additional mutations ameliorate the disease severity, suggesting that genetic suppressor could be viable therapeutic targets. In this study, we conducted a genome wide RNAi screen to identify suppressors of mitochondria dysfunction as potential therapeutic targets for GD neurological dysfunction. We employed a dual track approach of using human cell lines to identify genetic suppressors of mitochondrial dysfunction and developing *in vivo* zebrafish disease models for target validation.

2031F

Genome editing by CRISPR-Cas9 followed by RNA sequencing to identify the transcriptional regulatory role of MEIS1 in restless legs syndrome. *F. Sarayloo[1,2], H. Catoire[2], D. Rochefort[2], P. Dion[2,3], G. Rouleau[2,3].* 1) Human Genetics, McGill University , Montreal , Canada; 2) Montreal Neurological Institute, McGill University, Montreal (QC), Canada; 3) Department of Neurology and Neurosurgery, McGill University, Montreal (QC), Canada.

   Restless legs syndrome (RLS), a frequent movement disorder in the elderly in western populations, is characterized by an irresistible urge to move the legs and abnormal sensations in the lower limbs. A successful GWAS on RLS in 2007 identified common variants associated with this disorder located in intron 8 of *MEIS1* gene and the intergenic region between *MAP2K5* and *SKOR1*. These results were also replicated by other independent studies. Our group subsequently showed that *MEIS1* risk haplotype is associated with decreased mRNA and protein expression of this gene in RLS patients' LCL and thalamus samples. To further explain the role of other RLS associated genes, we examined the *SKOR1* mRNA expression in patient cells with the *MEIS1* risk haplotype and observed a significant decrease in its expression, suggesting a regulatory role of MEIS1 on *SKOR1* expression. To further confirm this regulatory role, a luciferase reporter assay followed by electro mobility shift assay showed a direct interaction between MEIS1 and *SKOR1* promoter in two different regions. Given the fact that MEIS1 has a transcriptional regulatory function, we hypothesis that this gene has an indirect regulatory role in RLS pathways. To further investigate that, we are overexpressing MEIS1 and also using CRISPR-Cas9 to knockout this gene, separately in different human cell lines (HEK293 and SK-N-SH cells, a neuroblastoma cell line). Whole transcriptome of these sets of cells as well as wild type cells with normal endogenous expression are extracted followed by RNA-Seq experiment. We believe that differential expression analysis of RNA-Seq data will lead to identify genes that are in fact regulated by these transcription factors, some of which might be novel genes with direct roles in RLS pathological pathways.

2032W

Investigation of rare variations in four SLI candidate genes in Pakistani SLI population. *E.M. Andres[1], M.A.R. Basra[2], H. Hafeez[2], F. Kausar[2], S. Riazuddin[3], M.L. Rice[1], M.H. Raza[1].* 1) University of Kansas, 1000 Sunnyside Avenue, Lawrence, KS 66045; 2) Institute of Chemistry, University of the Punjab, Lahore 54590, Pakistan; 3) Alama Iqbal Medical Research Center, Lahore 54550, Pakistan.

Specific language impairment (SLI) refers to a language deficit, which cannot be explained by any other neurological disorder. Both receptive and expressive language are often impaired in people with SLI. The estimated prevalence of SLI is 7% in kindergarten children. The causes of SLI are not well understood but it is clear that genetic factors are involved. Twin studies have shown that SLI is highly heritable and aggregates in small to medium size families. Many candidate genes have been associated with language impairment but there has been very little replication of rare coding variations between populations. Four candidate genes previously reported for language impairment (*NFXL1, TM4SF20, ATP2C2,* and *CNTNAP2)* were selected to sequence for the coding exons in probands of 29 SLI families ascertained from Pakistan. The probands were classified for SLI based on their family history and performance on the Peabody Picture Vocabulary Test-4th Edition (PPVT-4). Coding exons of all four candidate genes were sequenced in 29 probands through Sanger DNA sequencing. The sequencing data were analyzed to report the rare coding variants (less than 0.5% in the 1000 Genome). In total, one predicted splice site and eight rare coding variations were found within Pakistani SLI probands across four genes: *TM4SF20* (rs137891000), *NFLX1* (rs370816326, rs749383964, rs775580487), *CNTNAP2* (rs150918383, rs138517537, an unknown SNP), and *ATP2C2* (rs531589939, rs551531561, rs758765955, rs757310826). Our data show that previously reported SLI genes *NFXL1, TM4SF20, ATP2C2,* and *CNTNAP2* might be implicated in SLI in the Pakistani population when the sample size is increased.

2033T

Defining the critical region for brain malformations in 6q27 microdeletions. *M.D. Dias Hanna[1], P.N. Moretti[1], R.S Bonadio[1], M.T.A.S Rosa[2], M.S. Cordoba[3], S.F. Oliveira[1], A. Pic-Taylor[1], J.F. Mazzeu[4].* 1) Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brasil; 2) Secretaria de Estado de Saúde do Distrito Federal, Brasília, Brasil; 3) Hospital Universitário de Brasília, Universidade de Brasília, Brasília, Brasil; 4) Faculdade de Medicina, Universidade de Brasília, Brasília, Brasil.

Terminal microdeletions of the long arm of chromosome 6 have been rarely described and the phenotype includes multiple brain malformations, intellectual disability and epilepsy. A smaller region of overlap of 1,7 Mb has been proposed to contain a gene responsible for the brain anomalies. We studied a 10-year-old girl with moderate intellectual disability and multiple brain alterations including periventricular nodular heterotopia, hydrocephalus, incomplete hippocampal inversion, hypoplastic vermis and reduced mesencephalic volume. Chromosome microarray analysis was perfomed using Cytoscan 750K (Affymetrix). Array revealed a 325kbp deletion on chromosome 6q27, smaller than the previous microarray refined microdeletions described on the literature, comprising six genes: *DLL1, FAM120B, MIR4644, PSMB1, TBP* and *PDCD2.* The deletion was inherited from her affected mother. *THBS2, PHF10, DLL1,* and *C6orf70* genes mapped at 6q27 have been previously proposed as candidate genes for the brain malformations in 6qter deletion carriers. The patient here described has a smaller deletion including only one of the previous candidate genes: *DLL1.* DLL1 is involved in cell adhesion regulation, cell-cell communication and cell-fate determination and *DLL1* haploinsufficiency has been linked to holoprosencephaly in humans. *TBP* is a possible candidate gene for its expression in the brain and association to spinocerebelar ataxia and Parkinson's disease. *TBP* knock-out mice are not viable but tbp heterozygous showed no abnormalities. However, its effect is possibly greater for intellectual disability, but not for brain development and its contribution could be due to position effect, which can span as far as 5 Mb. *PDCD2* is associated with programmed cell death in thymocytes and *PSMB1* is a subunit of a proteasome complex. Both are tightly linked to *TBP. FAM120B* is not considered a candidate due to its expression in all adult and fetal tissues. MIR4644 is a microRNA and the only study suggested its potential use as a biomarker in pancreatobilliary tract cancer. Study using bioinformatics tools to prioritize candidate genes for brain development pointed DLL1 as the best candidate, but TBP could also be considered as a potential candidate. Grant support: FAPDF.

2034F

Genomic variants related to verapamil response in the treatment of migraine. *F.M. Cutrer[1], C.J. Klein[1], E.J. Atkinson[2].* 1) Neurology, Mayo Clinic, Rochester, MN; 2) Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN.

At present, there is no biologically based rationale for drug selection among at least five pharmacologically distinct classes of prophylactic treatment in migraine, a disorder that afflicts over 40 million people in the United States. Whole Exome sequencing (WES) was performed in 22 patients who were highly responsive to Verapamil (mean 77% decrease) and in 15 patients who were poorly responsive (mean 3% decrease). After filtering out synonymous variants, we identified 588 correlating SNP's with p< 0.01. We then genotyped 188 different patients from whom we had Verapamil monotherapy data using the 524 most significant SNP's identified by WES and tested for a correlation with reduction in headache days. We then used all SNP's that correlated with Verapamil treatment response (p<0.05) in a pathway analysis to identify potential functional molecular cascades carrying a disproportionate number of Verapamil-migraine implicated SNP's. We assessed the change in the number of headache days using the percentage change model (Pre-treatment – Post treatment /Pre-treatment values. The table shows the SNP's with a p-value<0.01.) **Pre- Post Treatment change** (Percentage reduction) We carried out a pathway analysis using the SNP's which were most highly correlated with change in headache days after Verapamil monotherapy treatment (p < 0.05). Two pathways were implicated. When SNP's with p<0.05 correlation is used, the myo-inositol pathway is implicated. When the SNP's are further restricted to those with a p<0.01 then the phospholipase C signaling cascade is implicated. We propose that response to prophylactic treatment is an element of phenotype that is informative of the molecular pathophysiology of migraine susceptibility in individuals whose migraine is suppressed by a specific drug. We have demonstrated that using WES in highly responsive vs non-responsive subjects we can identify variants that implicate functional molecular cascades that are relevant to the anti-migraine action of the drug investigated. To our knowledge, this is the first work of its kind in migraine.

2035W

Targeted sequencing of migraine-epilepsy susceptibility locus on chromosome 12q. *M.E. Hiekkala[1], M.A. Kaunisto[2], A. Siren[3], V. Artto[4], S. Vepsäläinen[4], P. Häppölä[2], P. Palta[2], E. Hämäläinen[2], S. Lindh[1], M. Färkkilä[4], A. Palotie[2,5,6,7,8,9], M. Kallela[4], A.E. Lehesjoki[1,10], M. Wessman[1,2].* 1) Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland; 2) Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Helsinki, Finland; 3) Outpatient Clinic for Persons with Intellectual Disabilities, Tampere University Hospital, Finland; 4) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland; 5) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, and Harvard Medical School, Boston, USA; 6) Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA; 7) Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, USA; 8) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, USA; 9) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, USA; 10) Neuroscience Center and Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland.

Migraine and epilepsy are both common disabling episodic disorders with a large number of similar triggers, medication and overlapping genetic contributions. Furthermore, the incidence of migraine in patients with epilepsy is almost twice of that in individuals without epilepsy (Ottman and Lipton 1994). We hypothesize that identifying predisposing variants to migraine and epilepsy will provide important insights into shared pathophysiology underlying the diseases. We have identified a migraine-epilepsy locus on chromosome 12q in a large Finnish migraine-epilepsy family (Polvi et al. 2012). Using SNP genotyping (48 individuals) we were able to restrict the 13.5 Mb large locus to a region of only 2 Mb. The region contains several potential candidate genes, such as transcriptional corepressor gene *NCOR2* (nuclear receptor corepressor 2) that has an important role in neuronal function. In order to identify susceptibility variants on 12q we have sequenced the 2 Mb region (GRCh37/hg19: chr12:123.750.000-126.250.000) in 48 individuals from the migraine-epilepsy family. So far, we have focused our analyses on *NCOR2* gene. Our preliminary analysis did not reveal any rare or low frequency and functionally interesting *NCOR2* variant shared by the affected family members. Next we will expand our variant analyses on other genes in the region and pay special attention to small deletions and insertions. In addition, we will utilize the genome-wide SNP data to perform copy-number variation analysis identifying possible larger deletions or duplications within the area.

2036T

RNA-seq analysis after moderate blast exposure in peripheral blood samples. *H. Kim[1], K. Edwards[1], V. Motamedi[1], S. Yun[2], J. Gill[1].* 1) SML, NINR/NIH, Bethesda, MD; 2) Yotta Biomed, Bethesda, MD.

Blast induced traumatic brain injuries are a signature injury of recent war campaigns; yet little is known about the biological mechanisms underlying blast exposures. In a group of military personnel (N = 69), prior to, and on each day of a blast training program (10 days total), blood was collected; throughout training, blast exposure measures were detected by helmet sensors to determine the mean peak pressure in pounds per square inch (psi). On day 7, some participants (n = 29) sustained a moderate blast (mean peak pressure = 7.9 psi) and were matched to participants with no/low blast exposure during the training (n = 40). Headaches and concentration problems occurred on the day of the moderate blast exposure, and the headaches continue on the following day. Total RNA was isolated from blood samples and processed for Illumina's HiSeq 2500 sequencing system to measure expression of genes across the entire genome. Based on the dysregulation data from RNA-sequencing and top gene networks identified using Ingenuity Pathway Analysis, a subset of genes were chosen to validate with Nanostring's nCounter Digital Analyzer system. In this sample, day 10 samples from moderate blast cases (n = 29) were compared to each participant's own baseline, leading to the identification of 1,803 genes that were differentially expressed after moderate blast exposure (746 up-regulated genes and 1,058 down-regulated genes). IPA suggests dysregulation of networks for both cell death and survival as well as cellular development and function with the most altered network being related to amyloid precursor protein (APP).

2037F

Very rare homozygous variants: A flashlight to possible involvement in ALS? *A. Orr-Urtreger[1,2], O. Goldstein[1], M. Gana-Weisz[1], M. Kedmi[1], S. Twito[1], O. Nayshool[1], B. Vainer[3], B. Nefussy[3], V.E. Drory[2,3].* 1) Gen Inst, Tel-Aviv Sourasky Med Ctr, Tel Aviv, Israel; 2) Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel; 3) Neuromuscular Service, Department of Neurology, Tel-Aviv Sourasky Medical Center, 6 Weizmann Street, Tel Aviv 64239, Israel.

North Africa Jews (NAJ) are considered a population with a genetic architecture that is relatively homogenous due to a history of religious and cultural restrictions. We previously showed that Amyotrophic Lateral Sclerosis (ALS) is observed in NAJ in a similar worldwide incidence but at an earlier age-of-onset, and that some ALS patients from this origin carry the homozygous *OPTN* 691_692insAG founder mutation . We aimed to study the possibility that additional rare homozygous genetic variations may contribute to familial and sporadic ALS in NAJ patients. Methods: Whole Exome Sequencing (WES) of 43 NAJ individuals affected with ALS was carried out. Filtering was done to identify very rare recessive *in-silico* damaging variants, with allele frequencies of less than 0.002 in the gnomAD browser, annotated to ALS-associated cellular pathways. Two variants were subsequently genotyped in additional 70 unrelated NAJ ALS patients and 400 ethnically matched controls. Results: We identified 34 rare homozygous variants in genes associated with pathways that are known to be involved in ALS, including autophagy, mitochondria, RNA-binding and cytoskeleton. These included genes previously reported as upregulated (*SFXN4*) or downregulated (*ARMC4* and *MTHFSD*) in ALS patients, and genes previously associated with other neurodegenerative or neuromuscular diseases, such as *HTT* (Huntington's disease), *ATM* (Ataxia-Telangiectasia), *ZFYVE26* (Spastic Paraplegia 15) and *MFN2* (Charcot-Marie-Tooth disease type 2A). The homozygous variations that were detected in *MFN2 and NEK1* (p.Arg663Cys and p.Arg261His) were further evaluated. Their allele frequencies in ALS patients were 11.2 and 1.9 times higher than in the ethnically matched controls, (p=0.031 and not significant, respectively), with no homozygotes observed in controls for both. Overall, 10 ALS patients (23%) were homozygotes for more than one rare variant. In conclusion, we demonstrate that the unique homogenous population of North Africa Jews has the power to identify rare homozygous variants in ALS, suggesting that additional genes are involved in this disease. We report for the first time the association of *MFN2* with ALS, and suggest that the *NEK1* p.Arg261His is a risk factor for this disease. In addition, *MTHFSD*, which is involved in stress granules, might be a candidate for ALS. Our data also support oligogenic inheritance in ALS and highlight the importance of autophagy, mitochondria and RNA homeostasis in ALS pathogenesis.

2038W

Whole-genome sequencing points to *SV2A, DENND4B, MIB2, SPTBN2,* and *APP* as new genes of interest in two individuals sporadically affected with childhood apraxia of speech. *B. Peter[12], C. Vose[1], E. Groff[1], L. Eng[1], M. Naymik[3], I.B. Schrauwen[345].* 1) Speech & Hearing Sci, Arizona State University, Tempe, AZ; 2) Communication Sciences and Disorders, Saint Louis University, Saint Louis, MO; 3) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 4) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 5) Division of Natural Sciences, Arizona State University, Tempe, AZ.

Childhood apraxia of speech (CAS) is a severe speech sound disorder that interferes with a child's ability to control the articulatory machinery towards intelligible speech. Since the discovery of the *FOXP2* gene in a family with syndromic CAS, additional evidence that CAS has a genetic etiology has been emerging. Aside from a few recurring genomic changes such as 6p11.2 deletions, the genetic etiology appears to be heterogeneous and includes inherited as well as sporadic changes, point mutations as well as copy-number variations (CNVs), and single as well as multiple hits. Many genes of interest are highly expressed in cerebellum, consistent with the phenotype. Here, we present two trios with affected offspring and unaffected parents. Patient 1, age 19 years, still exhibits residual signs of CAS in the form of monotone pitch and robotic-sounding speech rhythm. His identical twin has the same clinical history. Patient 2, age 6, shows active CAS signs including vowel errors and inconsistent word productions. In both cases, family history was negative for severe speech disorders except for Patient 1's twin brother. Whole-genome sequences obtained from blood samples were analyzed for *de novo* variants. Patients 1 and 2 carry 5 and 12 *de novo* CNVs, respectively, with no overlaps. None of the CNVs overlap with previously published regions of interest and all with but one gene-containing CNVs are reported in the Database of Genomic variants. CNV causality is unlikely. The two patients share 22 variants, all rare, in 20 genes. One of the most damaging of these is a *SV2A* (synaptic vesicle glycoprotein 2A) E138G missense mutation. *SV2A* is highly expressed in brain, especially cortex and cerebellum, and influences motor regulation. Shared variants in *MIB2* (Mindbomb E3 Ubiquitin Protein Ligase 2) and *DENND4B* (DENN Domain Containing 4B) are also deleterious and expressed in cerebellum. The two patients carry nonidentical deleterious variants in several genes that are highly expressed in cerebellum, influence cell-cell adhesion, and/or are associated with motor discoordination. These include *PPFIA4, BSN, TNK2, PCDHA8, PCKCNH2, MAMDC4, CACNA1B, CCDC88B, SPTBN2, NCOR2,* and *KCNC3*. Both patients have inversions disrupting *APP,* which influences neurite growth, neuron adhesion, and axonogenesis. Results are consistent with the view that CAS is heterogeneous and may result from multiple hits. Future work includes investigating rare inherited variants in these trios.

2039T

Mutation identification for epilepsy in the U.S. Latino population using whole exome sequencing. *C. Xu[1], A. Garza[1], Y. Ramirez[1], J. Ordonez[2], C.X. Mao[1], X.M Gong[2], R. Brower[2].* 1) University of Texas Rio Grande Valley, One West University Blvd, Brownsville, TX, 78520; 2) TTUHSC-El Paso.

Introduction: Epilepsy, characterized by recurrent unprovoked seizures, is a common neurological disorder related to a wide variety of genetic, developmental and acquired brain conditions. Genetically determined epilepsies are associated with a multiplicity of potential genetic variants identified in predominantly Caucasian populations. However, there has been limited research in the US Hispanics/Latino populations. Whole Exome Sequencing (WES) is a technique that is used to discover causative mutations for a variety of diseases and has enabled to identify pathogenic mutations in patients with well-characterized neurological disorders, including epilepsy reported from the non-Hispanic population. Moreover, *De Novo-* and transmitted mutations are emerging as important causes of neurological diseases, including epilepsy, and WES in case-parent trios design is a powerful method to detect them. Here, we report the findings of WES in two US Latino families with epilepsy. Methods: We performed WES in a total of 10 subjects who were recruited from the Department of Neurology at the Texas Tech University Health Sciences Center in El Paso, Texas. The genomic DNA was extracted from blood using standard methods, as in our previous study [20]. A total of 50 ng DNA from each patient was used for WES analysis using Illumina Nextera Rapid Capture Exome Enrichment kits (Illumina, FC-140-1001) on NextSeq 500 (Illumina) sequencer. Results: After the quality control and filtering, we identified three *de novo* mutations in these six affected subjects with epilepsy from two Hispanic families and they were non-synonymous rare mutations in CACNA2D1, O4C3 and WDR16 genes, including NM_000722.2:c.1956-5_1956-4dupTT in the CACNA2D1 gene with potential splice-site alteration; NM_001004702.1:c.169_177delTTGCTGATC (9 base pair deletion in coding region, including p.Leu57_Ile59) in the OR4C3 gene and NM_145054.4:c.1159delA (p.Ser388AlafsTer21) of the WDR16 gene. In addition, we identified shared mutations between two families and homozygous mutations in affected subjects. These mutations are predicted to be pathogenic impact and may result in the in-frame deletion and frameshift respectively. Discussion: This WES led to the discovery of additional epilepsy genes in the U.S. Latino population; however, future validation and segregation analysis using a Sanger sequencing and more cases are needed to confirm the current findings.

2040F

Identification of somatic mutations in malformations of cortical development. *N.G. Griffin[1], S. Garcia-Tarodo[2], G. Von Allmen[2], G.A. Grant[3], B. Porter[3], M.A. Mikati[4,5], N.M. Walley[6], C.R. Muh[7], C.M. LaCoursiere[8], K. Keever[3], C. Shain[8,9], E. Yang[10,11], P.B. Crino[12], A. Poduri[8], E.L. Heinzen[1,13].* 1) Institute for Genomic Medicine, Columbia University, New York, NY; 2) Department of Pediatrics, McGovern Medical School, University of Texas in Houston, Houston, TX; 3) Department of Neurosurgery, Stanford University School of Medicine, Stanford, CA; 4) Division of Pediatric Neurology, Duke University Medical Center, Durham, NC; 5) Department of Neurobiology, Duke University, Durham, NC; 6) Division of Medical Genetics, Department of Pediatrics, Duke University School of Medicine, Durham, NC; 7) Department of Neurosurgery, Duke University Medical Center, Durham, NC; 8) Epilepsy Genetics Program, Department of Neurology, Division of Epilepsy and Clinical Neurophysiology, Boston Children's Hospital, Boston, MA; 9) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 10) Harvard Medical School, Boston, MA; 11) Department of Radiology, Boston Children's Hospital, Boston, MA; 12) Department of Neurology, University of Maryland, School of Medicine, Baltimore, MD; 13) Department of Pathology and Cell Biology, Columbia University, New York, NY.

   Malformations of cortical development (MCD) comprise several phenotypes affecting the structure and organization of the cerebral cortex. MCD result from defects in neuronal migration and differentiation during development, leading to intractable epilepsy and intellectual disability. Because the majority of cases of MCD cannot be explained by an underlying genetic germline mutation, it has been hypothesized that *de novo* somatic mutations arising in neural progenitor cells lead to MCD. We and others have demonstrated that approximately 30% of cases of MCD can be explained by somatic mutations in the PI3K-AKT-mTOR pathway. To identify somatic mutations in MCD, we performed paired exome sequencing and/or targeted sequencing of surgically resected brain tissue and leukocytes from 30 individuals with MCD including hemimegalencephaly, focal cortical dysplasia and polymicrogyria. Sequencing libraries were prepared with either the Illumina TruSeq or Kapa library preparation kit; targeted capture of the exome was performed with either Illumina TruSeq or Roche SeqCap EZ Exome v3.0 exome enrichment kits, while targeted capture of the 192 genes in the PI3K-AKT-mTOR pathway was carried out with a custom Roche SeqCap EZ kit. Paired-end sequencing reads were generated on either an Illumina HiSeq 2000 or HiSeq 2500. The sequencing coverage was greater than 200-fold and 800-fold, on average, for the samples sequenced with exome sequencing and targeted sequencing, respectively. Somatic single nucleotide variants and indels were called with Mutect2 and Strelka; somatic loss of heterozygosity was evaluated with Varscan-2. Likely pathogenic variants, i.e., those that were predicted to have a deleterious effect on the final protein and that were absent from in-house controls and public databases, were confirmed with Sanger sequencing or digital droplet PCR. In our cohort, we have identified and confirmed five somatic point mutations in known genes. Two were confirmed in *MTOR* (NM_004958.3:c.133C>A;p.Gln45Lys and NM_004958.3:c.4447T>C;p.Cys1483Arg [MIM 601231]); two in *PIK3CA* (NM_006218.2:c.1624G>A;p.Glu542Lys and NM_006218.2:c.1633G>A;p.Glu545Lys [MIM 171834]); and one, in *TSC1* (NM_000368.4:c.298C>T;p.Gln100* [MIM 605284]). Additional candidates in novel genes are in the process of being confirmed. The confirmed mutations identified thus far further support the role for altered regulation of the mTOR pathway as a pivotal pathogenic mechanism for many MCD subtypes.

2041W

Whole exome sequencing reveals known and novel genomic variants in a cohort of intracranial vertebra-basilar artery dissecting aneurysm (IVAD). *S. Zhao[1,2,3], K. Wang[4], Y. Zhou[1,2,5], X. Song[6], J. Liu[1,2,7], W. Chen[1,2,3], Z. Wu[4,8,9], X. Yang[4], P. Liu[6], N. Wu[1,2,3].* 1) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, China; 2) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, China; 3) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China; 4) Department of Interventional Neuroradiology, Beijing Neurosurgical Institute and Beijing Tiantan Hospital, Capital Medical University, Beijing, China; 5) Department of Internal Medicine, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA; 7) Department of Breast Surgical Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; 8) Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, Beijing, China; 9) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730,China.

   INTRODUCTION:;Intracranial vertebra-basilar dissecting aneurysm (IVAD) can lead to life-threatening consequences if not diagnosed timely and susceptibility is linked to genetic risk factors. In this study, we aim to identify genetic variants contributing to the etiology of IVAD in sporadic patients. **METHODS:;**We enrolled a cohort of cases with definite diagnoses of IVAD, and performed whole exome sequencing for all the participants (cases with some of the parents). Variant interpretation and prioritization and new disease gene discovery were performed. We identified variants contributing to IVAD in known and novel disease genes in sporadic trios and singletons. We performed phenotypic characterizations for the cases with a possible molecular diagnosis. **RESULTS:;**Exome sequencing data were analyzed for the 44 participants consisting of 8 trios and 36 singletons. We identified five reported disease-causing variants (*COL3A1* and *FBN1*), two *de novo* variants (*VPS52* and *CDK18*) and two loss-of-function variants (*MYH9* and *LYL1*) as highly pathogenic changes. Subjects carrying these variants are considered possibly molecularly solved (7 of 44). In addition, eight novel variants in VAD-related genes and two biallelic variants in new disease genes for VAD were considered to be possibly contributing to the phenotype, but of unknown significance (8 of 44). An absence of hypertension and the symptom of headache were identified as distinctive phenotypic characteristics for patients carrying a genetic diagnosis of IVAD. **CONCLUSIONS:;**The next generation genomic analysis approach enabled detection of disease associated variants in known and novel IVAD related genes. A mutational burden in IVAD sporadic cases revealed the complexity of the disease. Genetic screening is suggested for IVAD patients with the symptom of headache.

2042T

Polygenic analysis of persistent cisplatin-induced peripheral neuropathy implicates immune-mediated processes. *O. El Charif[1], H. Wheeler[2], B. Mapes[1], E. Gamazon[3], S. Ardeshit-Rouhani-Fard[4], P. Monahan[4], D. Feldman[5], R. Hamilton[6], D. Vaughn[7], C. Beard[8], C. Fung[9], J. Kim[10], C. Kollmannsberger[11], S. Fossa[12], P. Dinh[4], T. Mushiroda[13], M. Kubo[13], L. Einhorn[4], N. Cox[3], L. Travis[4], M.E. Dolan[1], The Platinum Study Group.* 1) The University of Chicago Medicine-Hematology/Oncology, Chicago, IL; 2) Loyola University Chicago-Dept of Biology, Computer Science, Chicago, IL; 3) Vanderbilt University-Division of Genetic Medicine, Nashville, TN; 4) Indiana University-Dept of Medical Oncology, Indianapolis, IN; 5) Memorial Sloan-Kettering Cancer Center-Dept of Medical Oncology, New York, NY; 6) Princess Margaret Cancer Centre-Dept of Surgical Oncology, Toronto, ON; 7) University of Pennsylvania-Dept of Medicine, Philadelphia, PA; 8) Dana-Farber Cancer Institute-Dept of Radiation Oncology, Boston, MA; 9) University of Rochester Medical Center-J.P Wilmot Cancer Institute, Rochester, NY; 10) The University of Texas MD Anderson Cancer Center-Dept of Genitourinary Medical Oncology, Houston, TX; 11) University of British Columbia-Division of Medical Oncology, Vancouver, BC; 12) Oslo University Hospital Radiumhospital-Dept of Oncology, Oslo, Norway; 13) RIKEN Center for Integrative Medical Science, Yokohama, Japan.

Cisplatin-induced peripheral neuropathy (CisIPN) is a debilitating, often irreversible outcome of platinum-based chemotherapy. No approved therapy exists and mechanisms remain elusive. We aimed to uncover the polygenic architecture of CisIPN using GTEx expression quantitative trait loci (eQTL)-based analyses from a GWAS of CisIPN in 680 genetically European testicular cancer survivors (TCS). **Methods**: TCS completed validated questionnaires to define a binary case-control CisIPN phenotype. Genotyping, quality control, and GWAS were performed. The enrichment of GTEx cis-eQTLs ($p < 10^{-6}$ in any tissue) was assessed by permutation resampling. Briefly, GWAS was performed with randomized phenotype labels adjusting for covariates (age, 10 genotype principal components) 500 times. The number of GTEx eQTLs with $p < 5 \times 10^{-3}$ in the shuffled-phenotype GWAS was computed for each permutation and the null empirical distribution of eQTL counts was used to approximate a normal distribution. A Z-score was computed from the observed overlap of GTEx eQTLs in the non-permuted GWAS to assess the significance. Protein-protein interactions (PPI) of protein-coding gene targets of enriched eQTLs and functional pathway enrichment were evaluated with KEGG using one-tailed Fisher's exact tests. **Results**: Out of 3485 genotyped SNPs with $p < 0.005$ in the CisIPN GWAS, 725 were eQTLs for 498 gene targets, indicating significant eQTL enrichment ($p = 0.0048$). We reassessed enrichment of eQTLs in each of the 44 tissues at the same threshold and another more stringent threshold (GWAS $p < 0.005$ and 0.001 respectively) to ensure robustness. Seven of 44 tissues displayed eQTL enrichment ($p < 0.05$ at both GWAS p-value thresholds): tibial nerve, tibial artery, subcutaneous adipose tissue, thyroid, pituitary, hypothalamus, and the nucleus accumbens. Protein-coding gene targets (349 of 498) provided the input for PPI, which displayed significant interaction ($p = 2 \times 10^{-4}$), with 149/349 proteins forming a single network. KEGG pathway analysis implicated major immune functions, including autoimmunity pathways such as autoimmune thyroid disease (FDR $q = 1 \times 10^{-4}$), type I diabetes ($q = 2 \times 10^{-4}$) and rheumatoid arthritis ($q = 4 \times 10^{-4}$). **Conclusion**: Our analysis implicates regulatory regions of a 149-protein network in CisIPN. This network involves proteins functioning in immune processes, e.g., TLR2, TLR4 and MHC. Deeper analysis and functional studies are needed to validate our findings.

2043F

Hirschsprung's disease and the related genes in Taiwan. *W. Yang[1,2], P.L. Chen[2].* 1) Chang Gung Memorial Hospital, Taoyuan, Taiwan; 2) National Taiwan University.

Background Hirschsprung's disease (HSCR) is a congenital disease when the enteric ganglion cells were absent in the variable lengths of gastrointestinal tract. Multiple genes have been found related to the disease, such as, *RET, GDNF, EDNRB, SOX10,* etc. Although *RET* is suspected to be the main gene related to HSCR, the phenotypic variability and incomplete penetrance observed in HSCR might suggest the involvement of modifier genes. Thus, a more precise correlation among these genes in HSCR should be studied. Patients and Methods Patients proved to be HSCR by pathology will were included after obtain their written consent under the ethical guidance of the Institutional Review Board of Chung Gang Memorial Hospital, so as their parents, siblings, and/or off springs. Total 30 families and 150 participants will be included. Patients with syndromic abnormal (such as Down syndrome) will be excluded. Peripheral blood is drawn for 20cc in all participants if possible, for children younger than 1-year- old, the amount is 10cc. After review previous articles thoroughly, there are 31 genes might be related to HSCR. A next-generation sequence (NGS) panel was designed accordingly. The DNA samples of all the patients will be sent to do NGS for detection of possible genetic mutations. Further Sanger's sequencing will be done for the mutations detected by NGS in both the patient and the family members to confirm the mutation to be de Novo or inherited. Results There are 15 families with 17 patients were included for NGS. Mutations were found in 13 patients among 7 different genes. The detection rate of a possibly pathogenic mutation is 76%, which was much higher than previous reported. Frameshift in *RET, L1CAM* and a stop coden in *NRG1* were only found in long segment HSCR patients, which may imply that an obvious loss of function in this 3 genes were related to more severe disease. Besides, a SNP in *RET* (rs1800860) was found in 100% of our patients. For most of the patient, the relation between phenotype and the genotype is unclear, and our case number is too small to made solid conclusion in penetrance. Conclusion NGS is a powerful tool to work up for a multigenic disease, such as HSCR. An obvious loss of function in *RET, L1CAM* and *NRG1* could result in long segment HSCR, and larger case number is needed to make more solid conclusions.

SNP	Gene	BP	P value	Resp.
rs10779261_C	*USH2A*	216595306	0.004	-
rs6443266_A	*TTLL3*	9841989	0.004	-
rs17844444_A	*PCDHB6*	140532165	0.001	-
rs3733694_G	*PCDHB7*	140558528	0.001	-
rs17096961_A	*PCDHB7*	140559849	0.003	-
rs1982151_A	*RMI1*	86617265	0.004	-
rs10882386_A	*PLCE1*	95790669	0.005	+
rs1531394_A	*ANO3*	26353643	0.008	-
rs116903927_G	*KDM2A*	67022766	0.009	-
rs2230433_C	*ITGAL*	30518041	0.003	+

2044W

Leveraging large-scale exome sequencing data from >5,000 individuals to elucidate the genetic influences of amyotrophic lateral sclerosis. *S.M. Farhan[1,2], M.J. Daly[1,2], B.M. Neale[1,2].* 1) Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School. Boston, MA, USA; 2) Stanley Centre for Psychiatric Research, Broad Institute of MIT and Harvard. Boston, MA, USA.

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease characterized by degeneration of motor neurons, leading to progressive weakening of limb, bulbar, and respiratory muscles and is ultimately fatal, typically within 3-5 years of onset. Although ALS is classified as a rare disease, affecting 1-2 per 100,000 individuals, it is the most common motor neuron disease in adults and has a lifetime risk of 1 in 350/500 for men and women, respectively. There are two forms of ALS: familial (FALS) and sporadic (SALS), which differ in their etiology but are clinically indistinguishable. FALS accounts for only 5-10% of all cases however, the genes that have been identified through FALS pedigrees can explain up to 15% of SALS cases. Therefore, the application of genetic and genomic approaches in ALS has the potential to identify novel disease loci. The FALS and the ALSGEN Consortia are composed of international researchers who are willing to contribute sequencing data of ALS cases with the unifying goal of unveiling the complete ALS genetic landscape. We have aggregated sequencing data from >5000 cases and have integrated methods for quality control, statistical analysis, and variant discovery, while using >10,000 healthy individuals as controls. By processing the data jointly using Hail, a new open-source genetics pipeline developed in the Neale lab, we can improve the statistical power to potentially discover novel genetic risk factors as well as eliminate confounding biases. To the best of our knowledge, our study is the largest ALS exome case-control analysis pursued to date. In addition to our own statistical analyses, we plan to replicate the results of published reports to corroborate or refute previous associations. Upon novel genetic discoveries, we plan to study the defective gene product in stem cells from ALS patients included in our dataset. Our efforts in collecting the largest ALS exome data set will be valuable in identifying novel disease loci, which can potentially translate into clinically relevant targets.

2045T

Whole genome sequencing and rare variant analysis in essential tremor families. *Z. Odgerel[1], N. Hernandez[2], J. Park[2], R. Ottman[3,4,5,6], ED. Louis[2], LN. Clark[1,7].* 1) Department of Pathology and Cell Biology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA; 2) Department of Neurology, Yale School of Medicine, Yale University, New Haven, CT 06510, USA; 3) G.H Sergievsky Center, Columbia University, New York, NY 10032, USA; 4) Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA; 5) Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032, USA; 6) Division of Epidemiology, New York State Psychiatric Institute, New York, NY 10032, USA; 7) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA.

Essential Tremor, one of the most common movement disorders, is genetically heterogeneous. The genetic etiology of ET remains largely unexplained. Whole genome sequencing (WGS) is likely to be of value in furthering our understanding of a large proportion of ET with complex disease inheritance patterns and enrichment of functional rare coding and non-coding variants may explain the heritability of ET. We performed WGS analysis of eight ET families (n=40 individuals) enrolled in the family study of Essential Tremor (FASET) at Columbia University. The analysis included filtering on WGS data based on allele frequency in population databases, rare variant burden analysis using the Mixed-Model Kernel Based Adaptive Cluster (KBAC) test and prioritization of candidate genes identified within families using gene set enrichment analysis, network analysis and phenolyzer. Genomic DNA was extracted from peripheral blood cells. Capturing was performed using the Illumina TruSeq DNA PCR-Free Library Preparation kit (Illuminia, San Diego, CA.), paired end sequencing was performed at >30X coverage per sample on the Illumina HiSeq X ten instrument (Illumina, San Diego, CA). Sequence alignment to the human reference genome (UCSC hg19) was performed using the Burrows-Wheeler Aligner (BWA) algorithm, and variant calling was performed using the genome analysis toolkit (GATK; Broad Institute). QC was performed according to GATK best practices. Rare variants with MAF<0.01 and MAF<0.005 were included in the study. Burden analysis was performed using mixed model KBAC implemented in the Golden Helix SVS & Variation Suite. Variants were annotated using CADD, SIFT, Polyphen2, LRT, Mutation Taster, FATHMM, PROVEAN, MetaSVM and MetaLR. Overall, we identified a total of 167 genes with deleterious or damaging variants that co-segregated with ET within families. Gene set enrichment and network analyses of these genes identified significant enrichment of genes that encode proteins with a role in pathways or biological processes previously implicated in the pathophysiology of neurological and neurodegenerative disease. Phenolyzer prioritized candidate genes within families. This work has identified a number of candidate genes and pathways for ET in families that can now be prioritized for functional studies to further our understanding of the pathophysiology of ET using cellular and animal models.

2046F

Novel loss-of-function mutation of α-tocopherol transfer protein leads to vitamin E deficiency in a family. *E. Smith[1], H.C. Yu[1], J. Appis[1], A. Kuusela[1], R. Sainger[1], S. Bureeva[1], L. Guo[2], T.J. Jönsson[2], B.M. Whittman[2], P. Brar[1], R. Martin[1], B. Perkins[1], C.T. Caskey[1,3].* 1) Human Longevity, San Diego, CA; 2) Metabolon, Morrisville, NC; 3) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

TTPA is a gene involved in the binding and transport of vitamin E (α-tocopherol), which helps protect cells and tissues throughout the body from free radicals. Homozygous mutations in this gene have been reported to cause autosomal recessive ataxia with vitamin E deficiency. The condition has variable penetrance and expressivity. The pathophysiology of the condition is due to the accumulation of free radicals in neurons (Ulatowski L, et al. 2014). Currently, no data exists pertaining to a heterozygous clinical phenotype and individuals who are heterozygotes are assumed to be unaffected carriers. At the Health Nucleus (HN) In our studies, individuals have whole genome sequencing, microbiome sequencing, metabolomics, and clinical testing, such as echocardiogram, neuro-cognitive assessment, and full body MRI. The information collected from both genetics and clinical data is used to make health assessments for these individuals. Each person receives genetic counseling when receiving genetics results. A family consisting of a father, a mother, and two children visited our facility. One parent and both children were found to have a novel heterozygous nonsense variant in *TTPA*. Metabolomics data from this family was further analyzed after one of the sons had an abnormal vitamin E level which was surprising to the clinical team. The other family members were analyzed and were also found to have an abnormal of low vitamin E. The three family members demonstrated vitamin E levels ranging from -3.56 to -6.93 standard deviations from the mean. The other parent, who did not have the *TTPA* variant, had a normal vitamin E level. Family history review at the time of initial visit and then at the time of results disclosure, there was no family history of ataxia or balance problems reported. In the HN clinical testing, there was no gait or balance changes identified in exam. This novel variant leads to a premature stop codon at the N-terminus of TTPA protein, which is presumably associated with a more profound phenotype when found in the homozygous state. While initially counseled that they were unaffected carriers, a second discussion with the family members, we explained the value in increasing daily intake of vitamin E. This family demonstrates the value of using metabolomics and functional diagnostics in conjunction with genetics. In addition, it shows how there may be subclinical disease in the carriers of metabolic conditions previously believed to be unaffected.

2047W

Exome sequence analysis identifies novel loci associated with carpal tunnel syndrome in DiscovEHR study cohort. *S. Krishna Murthy[1], M. Malik[2], D. Carey[1], S. Friedenberg[2], R. Metpally[1] on behalf of the DiscovEHR collaboration.* 1) Department of Molecular and Functional Genomics, Geisinger Health System, Danville, PA; 2) Department of Neurology, Geisinger Health System, Danville, PA, USA.

Carpal tunnel syndrome (CTS) is one of the most common entrapment neuropathies of the hand due to compression of the median nerve at the wrist causing significant work related disability. The prevalence in the general adult population ranges from 3 to 6 percent [Atroshi et al 1999]. Symptoms include numbness, tingling, pain and weakness of hand muscles. The management includes conservative treatment with splints and pain medications but in severe cases requiring surgery. **Aim:** The aim of this retrospective study is to make use of available electronic health records (EHR) to understand the molecular genetic basis of CTS. This will help us to understand what genes/loci are associated with CTS, leading to the development of new treatments. **Method**: We analyzed both whole exome sequencing and corresponding imputed Human Omni Express Exome beadchip data from 51,289 individuals in the DiscovEHR cohort as described in Dewey et al 2016. Participants were phenotyped as cases by mining the EHR data with at least 3 independent calls of ICD9 codes for CTS, resulting in 3185 cases [6.2 % prevalence; 1072 are males, 2113 are females]. Mixed linear model based association analysis (MLMA) was performed on the whole exome sequences using GCTA-v12.6. Genome wide association study (GWAS) was also performed using PLINK-v1.9 on the imputed chip data. **Results**: We identified a common non-coding variant in *TPMTP1* [rs9959961: P $=8e^{-8}$; *18q21* locus] with near genome wide significance from the GWAS analysis and MLMA revealed two near exome-wide significant variants in *TSTD2* [rs41274254: Pvalue:$6e^{-7}$; *9q22.23* Locus], MTMR*11* [rs11205303: Pvalue:$9.4e^{-7}$; *1q21.2* locus]. **Conclusion**: This is the first large scale, retrospective, association study of CTS using EHR; Understanding genes associated with CTS might elucidate the underlying causative mechanisms. Further study is warranted to understand the specific molecular determinants of CTS risk and links to environmental exposures.

2048T

Genome-wide association study reveals candidate susceptibility loci for idiopathic hypersomnia. *K. Tanida[1], T. Miyagawa[1,2], M. Shimada[1,2], H. Toyoda[1], SS. Khor[1], K. Mishima[3], Y. Honda[4], M. Honda[2,4], K. Tokunaga[1].* 1) Human genetics, Graduate school of Medicine, The University of Tokyo, Tokyo, Japan; 2) Sleep Disorders Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 3) Department of Psychophysiology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan; 4) Japan Somnology Center, Neuropsychiatric Research Institute, Tokyo, Japan.

Idiopathic hypersomnia (IHS) is a rare sleep disorder characterized by excessive daytime sleepiness, great difficulty in awakening, and prolonged sleep time. The exact prevalence is unknown, but it has been estimated to be between 0.002 and 0.010%. Several familial studies have suggested that IHS is associated with genetic components. Narcolepsy is known to be strongly associated with a human leukocyte antigen (HLA) allele, specifically the *HLA-DQB1*06:02*. However, there is no association between IHS and any particular HLA alleles. To identify susceptibility genes associated with IHS, we conducted a genome-wide association study (GWAS) in 120 Japanese patients with IHS and 1,791 healthy individuals. This study represents the first GWAS of IHS. Genome-wide SNP genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 6.0. SNPs used in the case-control association analysis satisfied the following quality control criteria: (1) call rate ≥ 97%, (2) minor allele frequency ≥ 5% and (3) Hardy-Weinberg equilibrium $P ≥ 0.1\%$. The genomic inflation factor (lambda) of the GWAS was 1.01, suggesting that population stratification has negligible impact on this analysis. Consequently, we found 80 SNPs which reached suggestive significance level ($P < 10^{-4}$). After imputation of 1000 Genomes reference panels, 53 genotyped SNPs and 43 imputed SNPs were selected for a replication study (200 Japanese patients with IHS and 2,000 healthy individuals). We will conduct a multiple SNP genotyping assay using DigiTag2 for the candidate 93 SNPs selected in the GWAS stage. We are going to report the results of the replication study in this conference.

2049F

Genome- and phenome-wide association analyses uncover MET as a susceptibility gene of cerebral palsy. *J.J. Connolly[1], J. Li[1], Y. Guo[1], X. Chang[1], D. Abrams[1], F. Mentch[1], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Dept. of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Pulmonary Medicine, Children's Hospital of Philadelphia, Philadelphia, PA.

Cerebral Palsy (CP) is a heterogeneous group of early-onset disorders that impair motor function, with severe adverse effects on movement and posture. Epidemiological studies suggest that genetic components make an important contribution to CP etiology, in addition to major environmental and maternal risk factors. We performed a genome-wide association study (GWAS) of CP and further meta-analyzed with two independent replication cohorts. In this first GWAS of CP, we identified a genome-wide significant locus at chr7q31.2 harboring the MET Proto-Oncogene, Receptor Tyrosine Kinase (*MET*) gene (rs714180, P=5.0x10^{-10}). The GWAS signal correlates with *MET* gene expression and promoter methylation. A phenome-wide association study of SNP rs714180 further confirms its phenome-wide significant association with infantile CP. Exome-sequencing of a subset of severe CP cases revealed an initiator codon variant enriched in CP cases emphasizing *MET* as a potentially important susceptibility gene for CP.

2050W

RNAseq gene expression profiling of CD4+ and CD8+ T cells from multiple sclerosis patients and healthy controls. *S.D. Bos[1,2], I.S. Brorson[2], L. Barcellos[3], T. Berge[2,4], H.F. Harbo[1,2].* 1) Department of Neurology, University of Oslo, Oslo, OSLO, Norway; 2) Oslo University Hospital, Oslo, OSLO, Norway; 3) Division of Epidemiology, School of Public Health, California Institute for Quantitative Biosciences, University of California Berkeley, CA 94720-3220, USA; 4) Department of Mechanical, Electronical and Chemical Engineering, Oslo and Akershus University College of Applied Science, Oslo, OSLO, Norway.

Introduction: Multiple sclerosis (MS) is a complex disease characterized by an autoimmune attack against components of the central nervous system. Genetic variations, environmental factors and interaction between these contribute to an individuals' risk for MS. Close to 200 genetic associations have been identified in increasingly large genome-wide studies. More than one-third of these genetic associations can be linked to immune-regulatory genes, corroborating the clinical observations that the immune system plays a major role in disease etiology. Here we present the gene expression profiles of CD4+ T cells from untreated MS patients and age and gender matched healthy controls. **Materials and methods:** Blood was drawn from 20 MS patient and 20 controls, followed by peripheral blood mononuclear cell (PBMC) isolation by density gradient centrifugation. CD4+ T cells were purified from the PBMCs by negative selection. Flow cytometry was performed to confirm the purity of CD4+ T cells (>95%). RNA was extracted and submitted for sequencing using a 100-bp paired-end protocol. The resulting sequencing files were processed using the programs "kallisto" and "sleuth". Multiple Discriminant Analysis (MDA) analyses were used to identify potential outlier samples and these were excluded from further analyses. **Results:** 19 MS patient samples and 20 control samples were forwarded to analysis. Hierarchical clustering did not identify clear separation of MS patients and controls indicating there are no large-scale differences between CD4+ T cells from these groups. Individual gene differences were observed at a nominal significance level, however; when multiple testing was taking into consideration there was no significant difference between any genes. **Discussion and Conclusion:** At a global level we did not identify differences between MS patients and controls. Although the genes with most differential expression showed promising candidate genes with respect to the immune system, none reached a significant difference after correction for multiple testing. These patients were either newly diagnosed or have experienced a benign disease course. Global or individual gene expression differences may be more pronounced for severely affected patients. Using pathway analysis we may be able to identify pathways that exhibit several smaller changes in their gene expression, thereby providing insights in disease-related pathways.

2051T

***Gba1* haploinsufficiency in a Parkinson mouse impacts longevity and symptom severity independent of SNCA aggregate.** N. Tayebi[1], L. Parisiadou[2], B. Berhe[1], A. Gonzalez[1], R. Tamargo[1], E. Maniwang[1], H. Fujiwara[5], R. Grey[1], S. Hassan[1], Y. Blech-Hermoni[1], C. Makariou-Pikis[2], M. Brooks[3], E. Ginns[4], D. Ory[5], B. Giasson[3], E. Sidransky[1].* 1) NIH/NHGRI, .Bethesda MD; 2) Feinberg School of medicine, Northwestern Univ, Chicago, IL; 3) Dept. of Neuroscience, Univ of Florida, Gainesville, FL; 4) Lysosomal research program, Univ. of Massachusetts, Worcester, MA; 5) Dept. of Medicine, Washington Univ. St. Louis, MO.

Mutations in the glucocerebrosidase gene, *GBA1,* are a leading risk factor for Parkinson disease and associated Lewy body disorders. Many *GBA1* mutation carriers, especially those with severe or null *GBA1* alleles, have earlier and more progressive parkinsonism. To model the effect of partial glucocerebrosidase deficiency on neurological progression *in vivo*, mice with a human A53T α-synuclein (*SNCA^A53T*) transgene were crossed with heterozygous null *gba* mice (*gba^+/-*). Survival analysis of 84 mice showed that in *gba^+/-*//*SNCA^A53T* hemizygotes and homozygotes, the symptom onset was significantly earlier than in *gba^+/+*//*SNCA^A53T* mice (p-values 0.023-0.0030), with exacerbated disease progression (p-value <0.0001). Over-expression of *SNCA^A53T* resulted in reduced glucocerebrosidase levels, regardless of genotype or symptoms (p-value <0.05-0.001), recapitulating the relationship between SNCA and glucocerebrosidase in patients with parkinsonism. *gba* haploinsufficiency did not alter the abundance or distribution of SNCA pathology based on quantification of SNCA aggregates in 17 symptomatic mice, nor levels of monomeric SNCA. However, both *gba^+/-*//*SNCA^A53* mice and patients with synucleinopathies carrying null *GBA1* mutations had lower levels of insoluble high molecular weight SNCA than those without mutant glucocerebrosidase. This model indicates that the effect of *gba* deficiency on symptoms is not directly related to SNCA aggregation, providing a valuable resource to identify modifiers, pathways and possible moonlighting roles of glucocerebrosidase in Parkinson pathogenesis.

2052F

Effect of genetic variants associated with uric acid on multiple sclerosis: A Mendelian randomization study. *M.D. Montierth[1], J.C. Denny[2], M.F. Davis[1].* 1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Multiple sclerosis (MS) is a severe disease of the central nervous system that affects approximately 400,000 Americans. Despite the prevalence and severity of MS, its cause and risk factors are still poorly understood. Over 100 genetic risk factors have been identified, including the *HLA-DRB1*15:01* allele and other non-HLA loci. Studies have also indicated that oxidative stress plays a key role in the pathogenesis of MS. Uric acid, an antioxidant, has been suggested as having a protective affect against MS. However, previous studies of uric acid in MS are affected by many confounding factors, and as such it is still difficult to draw conclusions with regard to this association. In this study, we compared uric acid lab values between patients with MS and control patients in a large population of electronic medical records to better understand this difference. Additionally, Mendelian randomization studies can be used to lend power to epidemiological studies by using genetic variants as instrumental variables, and thus avoiding potential biases of observational studies. We performed a Mendelian randomization analysis to investigate the link between MS and uric acid. MS cases and controls were selected from the Vanderbilt University Medical Center's Synthetic Derivative (SD) database. The SD is composed of de-identified medical records set aside specifically for research purposes and includes more than 2 million individuals. Using lab values from 3803 individuals (500 cases, 3303 controls), we determined that uric acid levels were significantly lower in patients with MS than in the control population (p=1.64E-13). From the SD database, 1,625 cases and 6,187 controls were genotyped. In these individuals, we identified 37 SNPs associated with uric acid, and used them to generate a weighted genetic risk score (WGRS) for each individual. We then regressed MS case status on the uric acid WGRS. We also created an MS WGRS using 110 SNPs associated with MS, and regressed uric acid lab values on the MS WGRS.

2053W

A longitudinal metabolome-wide association study on beta amyloid in adults with increased risk for Alzheimer's disease. *B.F. Darst[1,2], M.J.P. Rush[1,3,4], P.D. Hutchins[1,3,4], T. Reddy[5], J.D. Russell[1,4,5], R. Gangnon[1,2,6], R.L. Koscik[1,7], S. Asthana[1,7,8,9], S.C. Johnson[1,7,8,9], K.J. Hogan[1,7,10], J.J. Coon[1,3,4,5,11], C.D. Engelman[1,2,7,9].* 1) University of Wisconsin, Madison, WI; 2) Department of Population Health Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI; 3) Department of Chemistry, University of Wisconsin, Madison, WI; 4) Genome Center of Wisconsin, University of Wisconsin, Madison, WI; 5) Morgridge Institute for Research, Madison, WI; 6) Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI; 7) Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI; 8) Geriatric Research Education and Clinical Center, Wm. S. Middleton Memorial VA Hospital, Madison, WI; 9) Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI; 10) Department of Anesthesiology, University of Wisconsin School of Medicine and Public Health, Madison, WI; 11) Department of Biomolecular Chemistry, University of Wisconsin, Madison, WI.

Several cross sectional metabolomics studies of Alzheimer's disease (AD) have recently been reported, but a longitudinal examination of metabolomic profiles prior to AD diagnosis is crucial to establish early biomarker trajectories that influence the disease. Beta-amyloid (Aβ) is one of the strongest predictors of AD and alterations to this protein are believed to predate the diagnosis of AD by decades. We investigated the longitudinal influence of metabolomics on Aβ using plasma samples from an ongoing longitudinal cohort of participants who were asymptomatic at enrollment and enriched with a parental history of AD. Metabolomic profiles were quantified with mass spectrometry on 152 individuals on up to four visits each. These individuals also had up to two visits with cerebral spinal fluid Aβ measures. Due to the fewer number of observations with Aβ, Aβ measures were averaged across visits, while metabolomics were used longitudinally, and a total of 305 observations used in the current analysis. A metabolome-wide association study (MWAS) was performed using linear mixed effects regression models to assess the metabolites' main effects on outcomes Aβ_{42} and Aβ_{42}/Aβ_{40}. Due to the heterogeneity of AD mechanisms that can differ by an individual's *APOE* genotype, we also tested for interactions between the metabolites and an *APOE* genetic risk score for both outcomes. Of the 87 metabolites tested, preliminary analyses identified urea and sugar metabolites as some of the top main effects for both Aβ_{42} and Aβ_{42}/Aβ_{40}. The *APOE* risk score was found to interact with beta alanine and oxoproline for both Aβ outcomes. The implications of these initial longitudinal results could further explain the biological relevance of previously implicated pathways to AD. Ongoing analyses within this cohort and an independent replication cohort are being performed to investigate these relationships.

2054T

Interplay of genetic risk at *SNCA* locus and dysbiosis of gut microbiome in Parkinson's disease. *Z.D. Wallen[1], C.P. Zabetian[2], R. Knight[3,4,5], S.A. Factor[6], E. Molho[7], H. Payami[1,8].* 1) Departments of Neurology and Genetics, University of Alabama - Birmingham, Birmingham, AL; 2) Veterans Affairs Puget Sound Health Care System and Department of Neurology, University of Washington, Seattle, WA; 3) Department of Pediatrics, University of California San Diego, La Jolla, CA; 4) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA; 5) Center for Microbiome Innovation, University of California San Diego, La Jolla, CA; 6) Department of Neurology, Emory University School of Medicine, Atlanta, GA; 7) Department of Neurology, Albany Medical College, Albany, NY; 8) Center for Genomic Medicine, HudsonAlpha Institute for Biotechnology, Huntsville, AL.

   Idiopathic Parkinson's disease (PD) is thought to result from gene-environment interaction. GWAS has identified 28 susceptibility variants, the strongest of which is rs356219 at 3' of *SNCA* (PMID:25064009). Recently, PD has been linked to the gut microbiome (PMID: 28195358). There is extensive evidence showing two-way interactions between the host genome and the microbiome, but the relationship between PD-associated genotypes and dysbiosis of the gut microbiome in PD has not been studied. Here, we studied the gut microbiome and *SNCA* rs356219 genotype in 214 PD cases and 122 healthy controls. *SNCA* was genotyped using DNA from whole blood. For the gut microbiome, DNA was extracted from stool, 16S rRNA amplicon sequencing was done using Illumina, and sequence variants were inferred using DADA2 R-package. First we tested the global composition of the microbiome as a function of PD, sex, age, *SNCA* genotype and *SNCA*xPD interaction (all included in the statistical model). Dissimilarities in the microbiomes were calculated using Canberra distances, and tested using permutational multivariate analysis of variance (PERMANOVA, with 10,000 permutations). PD (P<0.0001), sex (P=0.0002), age (P=0.0004), and *SNCA*xPD interaction (P=0.006) were significantly associated with the microbiome, *SNCA* genotype was marginal (P=0.09). Next, we sought to identify individual taxa whose abundance differed significantly as a function of the interaction, using the same statistical model (taxa~ PD+sex+age+*SNCA*+*SNCA*xPD), and employing generalized linear model with negative binomial distributions and zero inflation, and FDR correction for multiple testing. A significant *SNCA*xPD interaction was detected for *Corynebacteriaceae_Corynebacterium* (FDR Q= 0.001). *Corynebacterium* occur naturally and are often innocuous, but some members can be opportunistic pathogens in immune compromised individuals. In conclusion, we confirmed the known associations of PD, sex and age with the gut microbiome, and found new evidence for interaction between *SNCA* genotype and disease on the microbiome that is independent of the disease effect. The taxon that is altered suggests the immune system may be involved in the interactive effect of *SNCA* and PD in the gut.

2055F

Somatic copy number gains of alpha-synuclein (*SNCA*) in synucleinopathies (Parkinson's disease and multiple system atrophy). *C. Proukakis[1], K. Mokretar[1], D. Pease[1], A. Soenmez[1], A. Ejaz[1], T. Lashley[1], S. Gentleman[3], H. Houlden[1], J. Taanman[1], J. Holton[1], A. Schapira[1], E. Nacheva[2].* 1) Institute of Neurology, University College London, London, England, United Kingdom; 2) Academic Haematology, University College London, London, England, United Kingdom; 3) Department of Medicine, Imperial College London, United Kingdom.

   Parkinson's disease (PD), the second commonest neurodegenerative disorder, is usually sporadic, although rare inherited mutations, including CNVs, in *SNCA* and other genes can lead to it. The alpha-synuclein protein has a key role in pathogenesis. We have hypothesized a role of somatic *SNCA* mutations in sporadic synucleinopathies. The evidence for mosaicism in healthy and diseased brain is increasing rapidly. Somatic copy number (CN) gains of *APP* have been reported in Alzheimer's brain. We aimed to detect *SNCA* gains in post mortem brains from PD and multiple system atrophy (MSA), another synucleinopathy, and controls. We selected sporadic cases with relatively young onset and short disease duration, and obtained substantia nigra (SN) and other brain regions where possible. We analysed DNA by droplet digital PCR (ddPCR) for *SNCA* CN, and customized array CGH for all PD genes. We performed preliminary FISH using *SNCA* BAC probes. We then stained sections with Agilent SureFISH custom-designed probes for *SNCA* and multiple references, counting dopaminergic neurons (DN) in the SN with visible neuromelanin positivity and other cells separately in a blinded fashion, and identifying frontal neurons by NeuN staining. We excluded high level mosaicism by aCGH and / or ddPCR in at least the SN in 37 PD and two MSA cases, and 12 controls. BAC FISH probes showed occasional possible *SNCA* gains. We validated *SNCA* SureFISH probes on brain and fibroblasts with known CNVs, and analysed sections from 40 PD, 25 control, and 5 MSA cases, counting 142 DN and 306 other nuclei on average. PD cases were more likely than controls to have any DN *SNCA* gain (77.5% v 40%, p=0.004), or any cell with gain (95% v 68%, p=0.005). The % of DN gains was significantly higher in PD than controls (0.78% v 0.45%; p=0.017). There was a negative correlation of % DN gains with onset age (r=-0.39, p=0.013), but not with disease duration, or age of death in PD or controls. All MSA cases had gains, and the highest levels of gains in DN were in two of these (2.76%, 2.48%). We performed selective validation with different control probes after dye-swapping. Control probes showed minimal gains (<0.1% in DN). We found occasional *SNCA* gains in frontal neurons of 4 PD cases studied. We present evidence of somatic gains of *SNCA* in brain, commoner in DN of the SN in PD than controls, negatively correlated with onset age, and possibly commonest in MSA. Somatic *SNCA* gains may be a novel risk factor for synucleinopathies.

2056W

Protective effect of smoking and caffeine on Parkinson's disease: A gene-environment study. *V. Altmann[1], AF. Schumacher-Schuh[2], M. Medeiros[2], SM. Callegari-Jacques[3], CRM. Rieder[2], MH. Hutz[1].* 1) Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Porto Alegre, Brazil; 2) Serviço de Neurologia, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 3) Statistics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Parkinson's disease (PD) is the second most common neurodegenerative disease. Epidemiological studies have indicated several risk factors for PD. Among them, smoking and caffeine consumption have presented strong evidence of an inverse association with PD. Gene-environment interactions as risk modifiers might be important for PD susceptibility. Therefore, the objective of this study was to investigate the role of ABCB1, ADORA2A and NOS1 genes in interaction with smoking and caffeine consumption on PD risk. A total of 184 PD patients and 162 controls were recruited at Hospital de Clínicas de Porto Alegre, Brazil. Smoking was considered as never or ever (at least 100 cigarettes) used. Caffeine use included coffee and yerba mate (Ilex paraguariensis) and was evaluated as every day vs. occasionally or never used. *ABCB1* haplotype (rs2032582, rs1128508 and rs1045642), *ADORA2A* diplotypes (rs2298383 and rs3761422) and NOS1 polymorphism (rs478597) were genotyped by allelic discrimination assays. Interaction results were assessed by binary logistic regression, controlling for age, gender and ancestry. Cigarette smoking and caffeine use were protective factors for PD (respectively, OR 0.42; 95% CI 0.27-0.65; P < 0.001 and OR 0.57; 95%CI 0.36-0.88; P = 0.014). *ABCB1* haplotype T-non G-T was associated with PD in interaction with smoking ($P_{interaction}$ = 0.016). Smokers with T-non G-T haplotype had higher protection to this disease (OR 0.35; 95% CI 0.16-0.75) compared to those with C-G-C haplotype. *NOS1* CC genotype in interaction with caffeine was associated with PD as protective factor (OR 0.20; 95% CI 0.09-0.44; $P_{interaction}$< 0.001). *ADORA2A* diplotypes were not associated with PD in interaction with smoking. In conclusion, *ABCB1* and *NOS1* in interaction with smoking and caffeine seem to influence PD susceptibility in our population. PD risk seems to result of complex interplay and composite effects of both genetic and environment risk factors.

2057T

ONDRISeq: Genetic diagnosis of neurodegenerative disease patients using targeted next-generation sequencing. *A.A. Dilliott[1,2], S.M.K. Farhan[3], A.D. McIntyre[1], L. Racacho[4,5], J.F. Robinson[1], M. Ghani[6], C. Sato[6], R. Swartz[7], D. Sahlas[8,9], J. Mandzia[10,11], D. Dowlatshahi[12], G. Saposnik[13,14], L. Casaubon[14,15], A. Hassan[16], M.J. Strong[1,10], M. Masellis[7,14], P. St. George-Hyslop[6], D.E. Bulman[4,5], E. Rogaeva[6], R.A. Hegele[1,2], the ONDRI Investigators.* 1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 3) Analytic and Translational Genetics Unit, Center for Genomic Medicine, Harvard Medical School, Massachusetts General Hospital, Stanley Centre for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, Massachusetts, United States; 4) Faculty of Medicine, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, Canada; 5) CHEO Research Institute, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada; 6) Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada; 7) Sunnybrook Health Sciences Centre, University of Toronto, Toronto, Ontario, Canada; 8) Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; 9) Central South Regional Stroke Centre, Hamilton General Hospital, Hamilton Health Sciences, Hamilton, Ontario, Canada; 10) Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada; 11) Southwestern Ontario Stroke Network, London Health Sciences Centre, London, Ontario, Canada; 12) Neuroscience Program, Ottawa Hospital Research Institute; Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada; 13) Division of Neurology, St. Michael's Hospital, Toronto, Ontario, Canada; 14) Division of Neurology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada; 15) Division of Neurology and Stroke Program, Toronto Western Hospital, Toronto, Ontario, Canada; 16) Stroke and Neurology Service, Thunder Bay Regional Health Science Centre; Neurology/Internal Medicine, Clinical Division, Northern Ontario School of Medicine, Thunder Bay, Ontario, Canada.

As the population ages, an ever-increasing prevalence of individuals suffers with neurodegenerative disease, with over 50 million people estimated to be affected worldwide in 2015. The Ontario Neurodegenerative Disease Research Initiative (ONDRI) is a provincial-wide, multi-year, observational cohort study characterizing five neurodegenerative diseases: 1) Alzheimer's disease and mild cognitive impairment (AD/MCI); 2) amyotrophic lateral sclerosis (ALS); 3) frontotemporal dementia (FTD); 4) Parkinson's disease (PD); and 5) vascular cognitive impairment (VCI). The ONDRI genomic subgroup's objective is to elucidate the often discounted, yet extremely important genetic landscape of these neurodegenerative disorders. For this investigation, a custom next-generation targeted resequencing panel, ONDRISeq, was designed to target 80 genes previously associated with familial forms of the five neurodegenerative diseases of interest. DNA samples from 520 unrelated ONDRI participants were sequenced with ONDRISeq and sequencing data were analyzed with a custom bioinformatics workflow to call and annotate variants. Manual curation of the variants is being performed based on their likelihood to be disease causing. Variants of interest are classified based on their pathogenicity using the 2015 American College of Medical Genetics Standards and Guidelines. To date, rare variant analysis has been completed for the 161 VCI participants, with 62.1% of individuals harbouring at least one rare candidate variant of probable clinical significance. Further, one variant of interest was identified in 37.9%, two variants in 18.6%, and three or more variants in 5.6% of participants. This rare variant burden is higher than to be expected from an unaffected population, which will be further confirmed by sequencing of control samples with the ONDRISeq panel. The rare variant burden will also be determined for the remaining 359 enrolled individuals. These pilot data suggest that our custom next-generation sequencing workflow is a robust tool for genetic evaluation of patients with neurodegenerative disease. This can be important for many individuals who are facing diagnostic odysseys, and can allow for the next stages of treatment or life-planning. Additionally, future multivariate regression analyses with the phenotypic data obtained from other ONDRI technological platforms will allow for the identification of new genetic markers correlated with disease progression and outcomes.

2058F

Genome-wide CNV analysis identifies candidate causal genes for Parkinson's disease in the lysosomal pathway. *P.S. Eis[1], B. Schüle[2], S. Kim[2], E. Hatchwell[1].* 1) Population Bio, Inc., Melville NY; 2) The Parkinson's Institute and Clinical Center, Sunnyvale, CA.

Parkinson's Disease (PD) is the second most common neurodegenerative disorder and affects ~1% of the population >60 years of age, but early-onset forms are also found. Genetics were not thought to play a role, but from 1997 to 2004 five genes (*SNCA, PARK2, PARK7/DJ-1, PINK1, LRRK2*) were found to harbor rare mutations that caused autosomal dominant (AD) or autosomal recessive (AR) PD. Additional PD genes have since been found but, to date, only 5-10% of patients are known to have causal mutations in one of these genes [PMID 28332488]. Interestingly, the gene *GBA* (glucocerebrosidase), which causes the AR lysosomal storage disorder (LSD) Gaucher disease (MIM 606463), was found to increase the risk of PD at least 5-fold in heterozygous carriers. A similar neuropathological presentation is found in PD patients carrying pathogenic variants in *SNCA, LRRK2,* and *GBA,* namely the presence of alpha-synuclein-positive Lewy bodies and neurites and neuronal cell loss in the substantia nigra. In <1% of patients, aberrant aggregation of alpha-synuclein protein is caused by pathogenic single nucleotide variants (SNVs) or copy number variants (CNVs), but in a majority of patients, alpha-synuclein pathology involves a variety of cellular and molecular mechanisms, such as mitochondrial dysfunction, oxidative stress, neuroinflammation, and lysosome/autophagy defects. With the aim of discovering additional genes that contribute to PD, we performed genome-wide CNV analysis using array-based comparative genomic hybridization (aCGH) on a PD cohort (n = 555, European ancestry, familial and sporadic cases). Interpretation of benign vs. potentially pathogenic CNVs was performed using our in-house CNV database and algorithms. Given the rapid advances on *GBA* as a drug target for development of disease-modifying PD therapies, we initially analyzed the aCGH data for CNVs impacting lysosomal pathway genes. Loss-of-function (LOF), exonic CNVs (size range 6-242 Kb) were found in 5 PD patients, which were absent in our control CNV data (1,000 subjects, European ancestry). These 5 LOF CNVs impact 5 different genes (*CERK, GALNS, PSAP, SUMF1, SCARB2*), 4 of which are known to cause an AR LSD: mucopolysaccharidosis IVA (MIM 612222), combined SAP deficiency (MIM 176801), multiple sulfatase deficiency (MIM 607939), and epilepsy, progressive myoclonic 4 (MIM 602257). Like *GBA,* we hypothesize that carriers of heterozygous pathogenic variants in these 5 lysosomal genes have an increased risk for PD.

2059W

Genetic modifiers modulating the age of onset of amyotrophic lateral sclerosis caused by expanded GGGGCC repeats. *H. Kim[1], J. Lim[1], H. Bao[1], B. Jiao[1], J.D. Glass[2], T. Wingo[2], P. Jin[1].* 1) Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Neurology, School of Medicine, Emory University, Atlanta, GA.

Amyotrophic lateral sclerosis (ALS) is a fatal neurological disorder characterized by progressive muscular atrophy and respiratory failure, leading to death within 2-5 years after the initiation of clinical symptoms. Unfortunately, no curable treatment has been developed yet except the management of symptoms. Through studies with familial and sporadic cases, genetic factors were proposed as the major contributor of ALS pathology; however, the frequency of mutations is very low among ALS patients, and all the known ALS associated mutations cannot account for phenotypic variability in terms of age of onset, type of motor neuron involvement, and disease duration. Given rapid progression of neurodegeneration once clinical presentation appears, the delay of age of onset would be a key therapeutic point. To identify novel genetic modifiers modulating the age of onset in ALS, we focused on the patients with the pathogenic expansion of GGGGCC (G_4C_2) repeats in the first intron of C9orf72, which is the most frequent mutation in both familial (~40%) and sporadic cases (~25%). We have performed whole-genome sequencing of both early-onset (younger than 40-year old) and late-onset (later than 70-year old) ALS patients with expanded G_4C_2 repeat followed by functional testing using a previously developed fly model as well as target resequencing of additional cohort. Through these analyses, we have identified two genetic modifiers of the age of onset in ALS caused by the expanded G_4C_2 repeats, DLG2 (also known as PSD-93) and CELF5. Notably, the analysis of published RNA sequencing data showed that DLG2 expression was inversely correlated with age of onset while CELF5 was positively correlated, regardless the presence of expanded G_4C_2 repeat. Taken together, our analyses demonstrated the effectiveness of sequencing-based genomics combined with functional genomics in identifying the genetic factors that could contribute to complex human diseases, such as ALS.

2060T

CGG interruptions alter protein properties and increase disease penetrance in SCA8. *B.A. Perez[1-3], M. Banez-Coronel[1,2], J. Estabrook[1,2], H. Olafson[1,4], S.H. Subramony[1-3], T. Reid[5], J. Richardson[1,4], T. Ashizawa[1,4], G. Xia[6,7], K. Santostefano[6,7], N. Terada[5], A. Berglund[1,2], E. Wang[1-4], L.W. Ranum[1-3].* 1) Center for NeuroGenetics, University of Florida, Gainesville, FL; 2) Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL; 3) Genetics Institute, University of Florida, Gainesville, FL; 4) Department of Neurology, University of Florida, Gainesville, FL; 5) Department of Biochemistry, University of Florida, Gainesville, FL; 6) Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL; 7) Center for Cellular Reprogramming, University of Florida, Gainesville, FL.

Spinocerebellar ataxia type 8 (SCA8) is a dominantly inherited neurodegenerative disorder caused by a bidirectionally transcribed CTG•CAG expansion mutation. Both RNA and protein gain-of-function (GOF) mechanisms are thought to contribute to disease. Expanded CUG transcripts sequester MBNL proteins within RNA foci (RNA GOF) and ATG-initiated polyGln and polyAla and polySer repeat associated non-ATG (RAN) proteins accumulate in SCA8 brains. Although the SCA8 mutation is dominantly inherited, the reduced penetrance of the disease often makes it appear as a sporadic disorder because there is no family history of the disease. Among the 75 SCA8 families we have collected, 77% presented as sporadic cases, 16% as recessive families and only 7% showed a dominant family history. Sequencing of the expansion mutations shows that CCG•CGG interruptions within the CTG•CAG repeat tract were found at a higher frequency in families with multiple affected members (8/11) versus families with a single affected member (3/29) (p= 0.0003). Transfection experiments show SCA8 expansions with CGG interruptions in the polyglutamine reading frame change their electrophoretic, aggregation and toxicity properties. To further investigate molecular differences that could alter SCA8 disease penetrance, we developed patient-derived induced pluripotent stem cell (iPSC) lines from affected and unaffected SCA8 expansion carriers. Since *ATXN8* mRNA was not detected in iPSCs and SCA8 primarily affects the CNS, we differentiated the iPSCs into neuronal precursor cells (NPCs). We found that the size of the expanded allele was stable in fibroblasts, iPSCs, and NPCs. Transcriptome analyses of control and affected NPCs detected 150 differentially expressed genes. Genes typically associated with SCA8 (e.g. *ATXN8OS*, *KLHL1*, and *MBNL1, 2,* and *3*) showed no significant differences in expression. Asymptomatic expansion carrier NPCs are being analyzed, with initial data from one asyptomatic line showing a transcriptome profile more similar to control than affected samples. Ongoing studies aim to assess RNA-, and protein-GOF differences between symptomatic and asymptomatic groups and the molecular effects of CCG•CGG interruptions.

2061F

The NINDS human genetics DNA and cell line repository: A publicly available biospecimen collection for neurological disease research. *A. Resch[1], S. Heil[1], A. Green[1], J. Santana[1], A. Amberson[1], R. Zhang[2].* 1) Coriell Institute for Medical Research, Camden, NJ, USA; 2) National Institute for Neurological Disorders and Stroke-NIH, Bethesda, MD, USA.

The National Institute of Neurological Disorders and Stroke (NINDS) Human Genetics Resource Center, also known as the NINDS Human Genetics DNA and Cell Line Repository, is a public resource containing DNA and lymphoblastoid cell lines with corresponding de-identified clinical and phenotypic data. NINDS Repository samples cover a diverse population of neurological patients and neurologically normal controls. Since its establishment, specimens from over 44,000 individuals with cerebrovascular diseases, Parkinsonism, motor neuron diseases, epilepsy, Tourette syndrome, Dystonia, and neurologically-normal controls have been submitted to the repository and more than 34,500 unique samples have been successfully banked and can be accessed through an online catalog (http://catalog.coriell.org/NINDS). The burden of neurological disorders is a serious health concern and presents a massive challenge to healthcare systems. The causes of neurological disorders are often multifactorial, involving interactions of many genetic and environmental factors. In neurodegenerative disorders clear pathogenic mechanisms are not completely understood and the lack of reliable genetic markers for early diagnosis and intervention present great challenges for disease mitigation. Studies aimed at identifying pathogenic mechanisms will accelerate discovery of disease-specific genetic markers, the development of tools for monitoring disease onset and progression, and implementation of new therapies. Samples from the NINDS Repository have been used in GWAS, SNP, CNV, and sequencing studies and the results of these studies have been deposited in the Database of Genotypes and Phenotypes (dbGaP), a NIH/NLM-sponsored repository for restricted-access data from studies investigating the interaction of genotype and phenotype (http://www.ncbi.nlm.nih.gov/gap). The NINDS Repository regularly monitors new dbGaP studies to identify those that include genomic information about NINDS Repository samples. Cross-referencing of the NINDS repository and dbGaP databases allows NINDS Repository users to access additional clinical data for thousands of samples. By establishing a centralized, publicly accessible resource of biospecimens and associated clinical data, the NINDS Repository facilitates discovery of new genetic and molecular signatures related to neurological disorders.

2062W

**Burden analysis of ALS-gene variants in patients with and without
C9orf72 expansion.** *J.P. Ross[1,2], C.S. Leblond[1,2], D. Spiegelman[2,3], A. Genge[2,3],
N. Dupré[4], J.P. Bouchard[4], P. Corcia[5], W. Camu[6], P.A. Dion[2,3], G.A. Rouleau[1,2,3].*
1) Department of Human Genetics, McGill University, Montréal, QC, Canada;
2) Montreal Neurological Institute, McGill University, Montréal, QC, Canada; 3)
Department of Neurology and Neurosurgery, McGill University, Montréal, QC,
Canada; 4) Division of Neurology, CHU de Québec, and Faculty of Medicine,
Université Laval, Québec, QC, Canada; 5) Centre SLA, CHRU de Tours,
Tours, France; 6) Centre Hospitalier Universitaire Gui de Chauliac et Institut
National de la Santé et de la Recherche Médicale U1051, Montpellier University 1, Montpellier, France.

   Background: Over 30 genes are associated with ALS. While there exist
several genes with penetrant variants, such as *SOD1*, *FUS*, *TARDBP*, and the
C9orf72 hexanucleotide repeat expansion, variants in most ALS genes typically have a low or unknown level of penetrance. One low penetrance variant
may not directly cause the disease. However, the "Oligogenic Hypothesis,"
that several concurrently inherited variants in certain genes are sufficient to
cause ALS, could explain the disease in patients without a single known causal mutation. **Objectives:** We aimed to test whether patients without a known
causal mutation carry significantly more variants in ALS-associated genes
than patients with the *C9orf72* expansion. Further, we aimed to test whether
types of patients will cluster by underlying genetic variance in these genes.
Methods: We included 119 ALS patients carrying the *C9orf72* repeat expansion, 120 ALS patients with no known causal mutation, and 143 unaffected
controls. All individuals were recruited from clinics in Québec and France.
The *C9orf72* repeat expansion was tested using Repeat-Primed PCR and
genotyping assays. The protein-coding regions of 28 ALS-associated genes
were targeted by a custom Molecular Inversion Probe Sequencing panel.
Mutation burden was assessed using the 'Burden' method in the R package
"SKAT." Clustering was performed by Principal Component Analysis (PCA)
using the smartpca module in EIGENSOFT 6.1.4. **Results:** None of the genes
tested in the current study showed significantly increased burden between
ALS patients with and without the *C9orf72* repeat expansion, nor between
unknown-cause ALS patients and healthy controls. A significant difference in
the total number of variants in ALS-associated genes was not observed. Further, PCA clustering analysis did not suggest a difference in genetic variance
between *C9orf72*-positive patients and unknown cause patients. **Discussion:**
The Oligogenic Hypothesis could help to explain why certain genes have been
statistically associated with ALS, but that single mutations are often insufficient
to cause the disease. However, because we did not observe a significantly
increased burden of mutations within the ALS genes, it is unlikely that the
disease is caused by carrying multiple mutations in these genes at once. It
may be that there remain several undiscovered ALS genes that explain the
disease in these unknown-cause patients. **Conclusion:** Our evidence does
not support the Oligogenic Hypothesis in ALS.

2063T

**Identification of candidate amyotrophic lateral sclerosis risk loci using
pedigree based analyses of next-generation sequencing data.** *K.L. Russell[1], J.M. Downie[1], S.B. Gibson[2], K.P. Figueroa[2], M.B. Bromberg[2], S.M. Pulst[2],
L.B. Jorde[1].* 1) Department of Human Genetics, University of Utah School of
Medicine, 15 North 2030 East Rm 5100, Salt Lake City, UT 84112; 2) Department of Neurology, University of Utah School of Medicine, 175 North Medical
Drive East, Salt Lake City, UT 84132.

   Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease
that is characterized by the death of both upper and lower motor neurons.
This neuronal degeneration leads to loss of voluntary muscle movement and
eventual death 3-5 years after diagnosis. The molecular mechanisms that
cause neuronal demise in ALS are not well understood. ALS can be classified
into familial and sporadic forms. Familial ALS (FALS), which accounts for
10% of ALS cases, is defined by an index case that has a first- or second-degree relative that is also affected. Sporadic ALS (SALS) accounts for the
remaining 90% of cases and has a heritability estimate of ~0.60. However,
only 17% of SALS cases have a pathogenic variant in a known ALS-causing
gene, suggesting that additional causal genes remain to be discovered. To
investigate the potential role of *de novo* mutations in SALS causation, we used
ALS family trios to detect potential ALS risk variants present in probands, but
absent in both unaffected parents. Two trios were selected from a large cohort
of ALS cases on whom whole-genome sequencing (WGS) was carried out
at an average coverage of 60x. BWA and GATK (v3.3) were used to perform
read alignment and variant calling. To identify candidate ALS risk genes, the
pedigree-Variant Annotation Analysis and Search Tool (pVAAST) software
was used. pVAAST first calculates logarithm of odds (LOD) scores to estimate
linkage between the identified variants and disease. It then combines LOD
scores with VAAST scores, which identify genes that are more burdened with
deleterious, rare (MAF<0.01) variants in case genomes compared to 96 WGS
healthy controls. After removal of false positives, the pVAAST analysis identified *ZNF587* as the top candidate risk gene. Next, Phenotype Driven Variant
Ontological Re-ranking (Phevor) software was used to re-rank the pVAAST
output by utilizing medical ontology databases. After removing variants that
were common in control genomes, Phevor ranked *MTMR2*, a causative gene
for Charcot-Marie-Tooth disease, and *ZNF587* as the top two candidate loci.
The two probands had different, nonsynonymous variants in *ZNF587* (variants
rs77577775 and rs144186084). *ZNF587* encodes zinc finger protein 587,
thought to be involved in transcriptional regulation. Interestingly, knockdown
of *ZNF587* in human glioblastoma cells was shown to decrease cell growth.
These results indicate that ALS trio analysis can help to identify novel candidate ALS risk loci.

2064F

Common variants at 5q33.1 predispose to migraine in African-American children. *X. Chang[1], R. Pellegrino[1], J. Garifallou[1], M. March[1], J. Li[1], Y. Liu[1], P. Sleiman[1,2], H. Hakonarson[1,2].* 1) The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA.

Genome-wide association studies (GWAS) have identified multiple susceptibility loci for migraine in European adults. However, no large-scale genetic studies have been performed in children or African Americans with migraine. Here, we conducted a GWAS of 380 African-American children and 2,129 ancestry-matched controls to identify variants associated with migraine risk. We then attempted to replicate our primary analysis in an independent cohort of 233 patients and 4,038 non-migraine control subjects. We uncovered common variants at 5q33.1 associated with migraine risk in African-American children (rs1946225, $P = 8.28 \times 10^{-8}$). The association was further validated in an independent study ($P = 0.015$) for an overall meta-analysis P-value of 9.55×10^{-9}. Further, eQTL analysis of the GTEx data show the genotypes of rs1946225 were strongly correlated with the mRNA expression levels of *NMUR2* at 5q33.1. *NMUR2* encodes a G protein-coupled receptor of NMU. NMU, a highly conserved neuropeptide, participates in diverse physiological processes of the central nervous system (CNS). Our studies provide new insights into the genetic basis of childhood migraine and allow for precision therapeutic development strategies targeting migraine patients of AA ancestry.

2065W

A novel mutation in *INF2* gene: Expanding the genetic spectrum of Charcot-Marie-Tooth disease and glomerulopathy. *P. Gupta, A. Gomez.* Genetics, St. Josephs Regional Medical Cen, Paterson, NJ.

Introduction: Segmental glomerulosclerosis is a histological form of kidney injury characterized by focal and segmental areas of glomerular sclerosis and tubulointerstitial fibrosis (FSGS). Primary forms of FSGS usually occur as isolated proteinuria in children and adults, and may evolve to end-stage renal failure (ESRD). Charcot-Marie-Tooth neuropathy (CMT) is a heterogenous group of inherited disorders affecting peripheral neurons . Typical clinical features of CMT include progressive muscle weakness and distal limb atrophy, reduced tendon reflexes, and deformities in hands and feet. In CMT patients, an increased prevalence of glomerulopathies, mostly FSGS, has been documented. The estimated prevalence of FSGS in CMT is 1 in 400 CMT patients, compared to 1 in 1000,000 in the general population. However, it has been unclear whether these two processes represent in these individuals one disorder or if they are two separate diseases. Currently, mutations in several genes, highly expressed in podocytes, are known to cause familial FSGS. *INF2* mutations account for 12-17% of autosomal dominant (AD) cases of FSGS. Apart from podocytes, *INF2* is strongly expressed in Schwann cell cytoplasm. In recent years about 20 case reports have described a FSGS-CMT phenotype associated with *INF2* mutations. We present a 47 year old male with ESRD and CMT who has a novel mutation in *INF2*. Case Report: Our patient is a 47 year old male who presented to the genetics clinic for evaluation of CMT and possible recurrence risk in his children. He reports having an abnormal gait and foot drop when he was 11 years of age. CMT was confirmed through EMG studies and further neurological assessment. A routine urinalysis performed at 17 years of age revealed proteinuria and subsequent renal biopsy confirmed FSGS. He has received 3 renal transplants to date. Family history is significant for an 8 year old daughter who has frequent falls. Genetic testing revealed a variant of unknown significance (VOUS) in the *INF2* gene c.148T>G, which is a novel missense mutation with uncertain impact on protein function. Discussion: As we continue to learn more about the CMT-FSGS phenotype it is important to perform baseline urinalysis to assess renal function in order to make an early diagnosis especially in the event of a novel mutation in an affected family member.

2066T

No rare deleterious variants from *STK32B*, *PPARGC1A*, *CTNNA3* are associated with essential tremor. *G. Houle[1,2], A. Ambalavanan[1,2], J.F. Schmouth[1,3], C.S. Leblond[1,2], D. Spiegelman[1], S.B. Laurent[1], C.V. Bourassa[1], C. Grayson[4], M. Panisset[5], S. Chouinard[5], N. Dupré[6,7], C. Vilariño-Güell[8], A. Rajput[9], S.L. Girard[10], P.A. Dion[1,2,3], G.A. Rouleau[1,2,3].* 1) Montreal Neurological Institute and Hospital, Montreal, QC, Canada; 2) Department of Human Genetics, McGill University, Montreal, QC, Canada; 3) Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada; 4) Xenon® Pharmaceuticals Inc, 3650 Gilmore Way, Burnaby, BC V5G 4W8, Canada; 5) Centre Hospitalier Universitaire de Montréal (CHUM)-Notre-Dame, André Barbeau Movement Disorders Unit, Montreal, Quebec, Canada; 6) Department of Medicine, Faculty of Medicine, Laval University, Quebec City, QC, Canada; 7) Département des Sciences Neurologiques, CHU de Québec (Enfant-Jésus), Quebec City, QC, Canada; 8) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 9) Division of Neurology, Saskatchewan Movement Disorders Program, University of Saskatchewan, Saskatoon Health Region, Saskatoon, SK, Canada; 10) Département des Sciences Fondamentales, Université du Québec à Chicoutimi, Saguenay, Canada.

Introduction: Essential tremor (ET) is a common neurological disorder characterized by postural, kinetic and intention tremors of a body part. ET is known to cluster in families and it is believed to be influenced by several genetic and environmental factors. Previous studies have suggested genetic variants and loci associated with the condition which provides valuable insights into the complexity of the genetic architecture of the disorder. Despite these recent advances, causative variants yet explain only a small proportion ET cases. Recently, a two-stage genome-wide association study (GWAS) that included 2,807 cases and 6,441 controls of European descent recently identified associations between essential tremor (ET) and specific intronic variants within three genes: a serine/threonine kinase (*STK32B*), a transcriptional coactivator (*PPARGC1A*) and a cell-adhesion molecule (*CTNNA3*). The present study aims to establish if coding variants from these genes might be associated with ET. **Methods and Results:** To further assess the role of *STK32B, PPARGC1A, CTNNA3* as ET predisposing factors, we first screened the coding regions of these genes for rare variants. Looking at exome and whole genome sequencing data, no rare deleterious variant were found to co-segregate with the disease in 14 autosomal dominant multiplex ET families. Next we used a targeted massive parallel sequencing approach to examine the protein-coding region of these genes in 269 ET cases and 287 control individuals. Thirty-four variants were identified but none had a significantly different allelic distribution between cases and controls. Finally, we performed a gene-based, variance-component test (SKAT-O) to assess the cumulative impact of rare variants. No difference in variant distribution for any of the genes was observed between the ET cases and the controls. **Conclusion:** No difference emerged regarding the distributions of individual variants (or gene) between cases and controls. Thus no rare mRNA altering variant further validated one of these genes as a risk factor for ET. The recent GWAS offers promising avenues but the genetic heterogeneity of ET is nonetheless challenging for the validation of risk factors, ultimately larger cohorts of cases should help to overcome this task.

2067F

Whole-genome sequencing in primary progressive multiple sclerosis uncovers mutations in genes for inherited leukodystrophies and other MS phenocopies. *X. Jia[1], L. Madireddy[1], S.J. Caillier[1], A. Santaniello[1], F. Martinelli-Boneschi[2], B.C. Cree[1], S.L Hauser[1], J. Oksenberg[1], S.E. Baranzini[1].* 1) Department of Neurology, University of California San Francisco, San Francisco, CA; 2) Department of Neurology, San Raffaele Scientific Institute, Milan, Italy.

Primary progressive multiple sclerosis (PPMS) is characterized by accumulation of neurological disability without relapses at disease outset and poor response to immunotherapies used for relapsing-remitting multiple sclerosis (RRMS). Population studies demonstrate a significant concordance in disease course (PPMS vs RRMS) between affected siblings, implicating genetics in PPMS pathogenesis. Furthermore, genetic mapping studies do not distinguish between RRMS and PPMS, suggesting a common susceptibility pattern. However, whether genetic modifiers influencing a primary neural susceptibility or the presence of relapses exists is still unclear. We performed whole-genome sequencing in 49 PPMS patients of European ancestry and obtained 81 publicly-available controls sequenced on the same platform. We identified 17 rare pathogenic variants in PPMS patients (but not in controls) in genes affecting Mendelian neurologic disorders, and attempted replication in 411 PPMS and 460 RRMS patients. Five top candidates were subjected to a second round of replication in an additional 335 PPMS, 2589 RRMS patients, and 1000 healthy controls. Four of the top 5 replicated variants were reported in diseases that share clinical features with MS, were found at a higher frequency in this PPMS cohort compared to large public cohorts, and had evidence of clinical impact in patients carrying the mutation. These included *KIF5A*p.Ala361Val in Spastic Paraplegia 10 (RR = 23), *MFN2* p.Ala716Thr in Charcot-Marie-Tooth Type 2A (RR = 5.1), *MLC1* p.Pro92Ser in Megalencephalic Leukodystrophy with Subcortical Cysts (RR = 1.8), and *REEP1* c.606+43G>T in Spastic Paraplegia 31 (RR = 1.6). Lastly, we tested the hypothesis that PPMS patients are enriched for mutations in genes affecting inherited leukodystrophies, and screened the discovery cohort for mutations in 48 genes that cause progressive demyelination in the central nervous system. Interrogation of 48 leukodystrophy genes revealed a 1.75-fold enrichment of rare potentially deleterious variants in 49 PPMS patients compared to 81 controls (T-test p = 0.012, simulation normal distribution p = 8.4 x 10^{-6}). This study found that mutations in genetic disorders that share some features with multiple sclerosis appear to contribute to disability in PPMS patients.

2068W

Prioritizing Parkinson's disease genes using population-scale transcriptomic data. *G.T. Wong[1], Y.I. Li[2], T. Raj[1].* 1) Ronald M. Loeb Center for Alzheimer's Disease, Departments of Neuroscience, Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics, Stanford University, Stanford, CA.

Genome-wide association studies have identified 26 genetic loci associated with late-onset Parkinson's Disease (PD) but identifying putative causal genes in PD remains particularly challenging. To address this, we leveraged large-scale transcriptomic data from 7,088 genotyped individuals across 61 tissues including 1,028 brains (461 prefrontal cortex from an aging cohort), peripheral blood, monocytes, T-cells, neutrophils, and 44 tissues from GTEx including 10 brain regions to study the genetic etiology of PD and to identify putative causal genes. Unlike Alzheimer's disease (AD), we found a strong enrichment of PD heritability near genes expressed in the brain, and to a lesser extent near genes with myeloid expression. Using a transcriptome-wide association study, we identified 86 significant gene-disease associations at FDR 5%, 58 of are in novel PD loci with suggestive levels of GWAS association signal. We found 41 significant gene associations (e.g. *LRRK2*, *BST1*, etc.) in peripheral monocytes, of which two are shared genetic factors (*CD33* and *PILRA/B*) with AD. This points toward an overlap in the innate immune-mediated mechanisms that contribute to neurodegeneration. Interestingly, a large number of the associations were genes whose differential splicing in prefrontal cortex is associated with PD. These include known genes, e.g. the *MAPT* and *SNCA* loci which are explained by exon 3 inclusion and exon 5 exclusion, respectively. We also identified several novel genes (*MTOR*, *CLASP2*, *CAMLG*, *GALC*), whose splicing are associated with PD. A PD-associated variant (rs2076655) at the *MTOR* locus is associated with an increase in a minor *MTOR* isoform that extends the 5' end of an exon. Genes identified in our study are more likely to interact physically with PD genes and belong to the same or related pathways including lysosomal and autophagy function, which corroborates existing work on PD etiology. Overall, this study provides a strong foundation for further mechanistic studies that will elucidate the molecular drivers of PD. Thus, this work represents a significant step towards understanding the genetic basis of PD.

2069T

Genetic analysis of SNCA gene polymorphisms in Parkinson's disease in an Iranian population. *M. Rahimi[1], J. Jamshidi[2], A. Tafakhori[3], H. Darvish[1].* 1) Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, CA; 2) Noncommunicable Diseases Research Center, Fasa University of Medical Sciences, Fasa, Iran; 3) Department of Neurology, School of Medicine, Imam Khomeini Hospital AND Iranian Center of Neurological Research, Tehran University of Medical Sciences, Tehran, Iran.

Objectives: Parkinson's disease (PD) is a complex disorder influenced by genetic and environmental factors. One of several genes indicated to be important in the etiology of PD is *SNCA*. Here we aimed to investigate the association of rs2301134, rs2301135, rs356221 and rs11931074 polymorphisms located in *SNCA* with PD. **Methods:** A case-control study was designed using 500 sporadic PD patients and 500 healthy controls. DNA was extracted from peripheral blood of all subjects and *SNCA* variations were genotyped using the PCR-RFLP method. **Results:** Significant associations were found for the two promoter polymorphisms of the *SNCA* gene, rs2301134 and rs2301135 (p-value=0.009, OR=0.79 95%CI: 0.66-0.94 and p-value=0.001, OR=1.33 95%CI: 1.12-1.59 respectively for differences in allele frequencies). Genotype frequencies were also significantly different in case and control groups for rs11931074 polymorphism located in 3'UTR region of the gene(p-value=0.036). **Discussion:** Our study indicates the possible effect of *SNCA* variations in the etiology of PD in the Iranian population. Further studies in different populations and also functional studies of the polymorphisms are suggested for more certainty in future studies.

2070F

RNA-sequencing reveals novel immunological pathways in neuromyelitis optica. *I. Adrianto[1], J.A. Ice[1], G. Kumar[1], J.L. Quinn[1], A. Rasmussen[1], R.H. Scofield[1], G. Pardo[1], K.L. Sivils[1,2], R.C. Axtell[1], C.J. Lessard[1,2].* 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Neuromyelitis optica (NMO) is a rare devastating neuro-autoimmune disorder affecting primarily the optic nerve and spinal cord that has limited options for therapeutic intervention. NMO was initially characterized as a variant of relapsing-remitting multiple sclerosis (RRMS) but conventional RRMS therapies are not effective in NMO patients. There is now increasing evidence NMO shares auto-antibodies with Sjögren's syndrome (SS), an autoimmune disease of the exocrine glands. Currently, there are no studies that compare the molecular similarities and differences in patients with NMO, RRMS, progressive MS (PMS), and SS. In this study, we compared the transcriptional profiles of NMO, RRMS, PMS, and SS using RNA-sequencing. We obtained whole blood from treatment naïve patients (21 NMO, 24 RRMS, 23 PMS, and 27 SS anti-Ro-positive, 30 SS anti-Ro-negative) and 27 healthy controls. Whole blood RNA was isolated using the NuGEN Encore kit. Sequencing was performed using the Illumina HiSeq 2000 and 3000. Raw FASTQ files were aligned to the human genome using HISAT2. The read counts per transcript were generated using featureCounts. Differentially expressed (DE) transcripts were determined using DESeq2 with a false discovery rate q-value of less than 0.05 and fold change (FC) of >2 or <0.5. We found 1751 upregulated and 437 downregulated transcripts when comparing NMO to healthy controls. Although a large number of DE transcripts were protein-coding elements, ~40 of DE transcripts were non-coding RNAs. The majority of NMO and SS anti-Ro positive patients had elevated interferon-inducible genes and antibody isotype gene IGHA1 compared to SS anti-Ro negative, RRMS, PMS and healthy controls. In addition, pathway analyses of DE transcripts indicated the enrichment of those transcripts in synaptic transmission and nervous system development pathways, neurological disorders, and brain tissues. Our RNA profile analyses demonstrate that NMO and SS anti-Ro positive patients have many molecular similarities and both are significantly different than RRMS. Specifically, NMO and SS anti-Ro positive patients have elevated interferon signatures as well as humoral IgA responses. We are currently undertaking studies to further our understanding of the cellular and molecular similarities and differences between NMO, MS, and SS. These preliminary findings illuminate that potential treatment strategies that are under development for SS could be repurposed for use in NMO.

2071W

Decoding GWAS discoveries of neurodegenerative diseases: Gene expression changes in single neurons. *L. Tagliafierro[1,2], K. Bonawitz[1,2], J. Barrera[1,2], OC. Glenn[1,2], K. Dai[1,2], O. Chiba-Falek[1,2].* 1) Neurology, Duke University Medical Center, Durham, NC; 2) 2Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC.

In the post genome-wide-association-studies (GWAS) era we are shifting gears toward translation of genetic disease loci to molecular mechanisms of pathogenesis. Large multi-center GWAS have found associations between extensive lists of genomic loci and dementia spectrum disorders, and candidate genes were inferred by the proximity to the associated-SNP. However, the precise target genes within the associated genomic regions and the causal variants effecting disease-risk are yet to be uncovered. Determining neuronal-specific alterations in gene expression profiles among different dementia-related neuropathologies and compared to cognitively healthy controls will advance the understanding of the target genes within the associated loci that contribute to the genetic etiologies of these diseases. Here we focused on late-onset Alzheimer's disease (LOAD), mild cognitive impairment (MCI) due to AD, and dementia with Lewy body (DLB). We developed a method to quantify cell-type specific gene expression levels. Archive human frozen brain tissues were used to prepare slides for rapid immunostaining, and single-neurons were isolated by Laser Capture Microdissection (LCM). Following RNA extraction, gene expression was determined digitally using nCounter Single Cell gene expression assay (NanoString). Using this method, we have analyzed the expression profiles of genes within LOAD and DLB-risk regions in neurons isolated from cortical brain tissues of LOAD, MCI, DLB and normal controls. We detected common and distinct trends of differential expression for several genes. In the neurons, *APOE*- and *TOMM40*-mRNAs were higher in all neurodegenerative pathologies compared to control. *SNCA*-mRNA levels were elevated in DLB neurons, and the levels of its alternative spliced variant (*SNCA*112) showed a further increase. Several other GWAS-genes also exhibited expression variations. We then evaluated the association between expression changes and disease associated variants. For example, Rep1 genotypes were associated with *SNCA*-mRNA levels in the isolated neurons. In conclusion, we present a state-of-the-art method for detection of neuronal-specific gene expression changes using banked frozen human brain samples. Our results showed disease-associated changes in the expression of critical genes, also in early stages such as MCI. We suggest that the expression regulation of key genes possibly has a causative role in the etiology of dementia-related diseases in aging.

2072T

The generation of iPSC-derived astrocytes from patients with Gaucher disease with and without Parkinsonism provide a model to study Parkinson pathogenesis. *B. McMahon, E. Aflaki, E. Sidransky.* National Institute of Health, Bethesda, MD., U.S.A.

Parkinson disease (PD) is caused by the disruption of dopaminergic neuro-transmission in the basal ganglia, associated with a severe reduction in the number of dopaminergic neurons in the substantia nigra pars compacta and the formation of cytoplasmic inclusions called Lewy bodies. Currently, the role of glial cells such as astrocytes in PD pathogenesis is unknown. Astrocytes provide critical support for neuronal function by providing antioxidant protection, clearing glutamate, developing and/or maintaining the blood-brain barrier, and releasing gliotransmitters and cytokines. Mutations in GBA1, the gene encoding glucocerebrosidase (GCase) which is mutated in the lysosomal disorder Gaucher disease (GD), are the most common genetic risk factor for PD. We generated astrocytes from iPSCs derived from fibroblasts of five patients with type 1 GD (GD1), two with and three without parkinsonism (GD1-PD), including two siblings discordant for Parkinson disease. The patients were previously sequenced to establish their GBA1 genotype. The astrocytes exhibited decreased GCase activity and stored glycolipids when compared to astrocytes from a healthy control. Moreover, the astrocytes from patients with GD and PD had a decreased intracellular calcium response to ATP, an important molecule in glial-neuronal signaling. These cells therefore represent an excellent model for exploring the potential role of astrocytes in the pathogenesis of GBA1-associated PD.

2073F

Selective activation of caspase family of genes in multiple sclerosis patients inducing neuronal apoptosis. *Y. Kattimani, A. Veerappa.* Department of Studies in Genetics and Genomics, University of Mysor, Mysuru, India.

The risk of developing MS (Multiple Sclerosis) is strongly influenced by genetic predisposition. However, the exact mechanisms of MS initiation, its development, and progression are still elusive. It is imperative to understand the genetic complexity of multiple sclerosis by investigating the expression level of transcripts in MS cases with the age-matched controls. Whole transcriptome RNA-seq datasets from peripheral blood monocyte of 5 Multiple sclerosis cases and 3 age-matched controls were obtained from Binder et al (2016). whose original study did not report on the differential expression of genes between MS patients and controls. Therefore, we obtained the raw data files and these were aligned to a reference genome using STAR Align pipeline. Gene counts and transcripts counts were called and quantified using GSA and ANOVA pipelines. We identified 8948 genes to be up-regulated 2-34 fold, while 2727 genes were found down-regulated with fold change up to -9.3. Clustering of enriched Gene Ontology terms uncovered functional clusters relevant to neuron apoptotic pathway (positive regulation of neuron apoptotic process, Enrichment Score = 4.81 & P-value=0.00194; neuron apoptotic process= 5.51, P-value=0.038). Caspase family of genes *CASP2*, *CASP9*, *CASP6*, *CASP10*, and *CYCS* were found up-regulated 2-4 fold and were found activating intrinsic apoptotic signaling pathway, and endopeptidase activity pathways. *BCL2L1*, another pro-apoptotic regulator was found expressed across the cases, with P-value=0.000123, FDR =0.000996 with a fold change >2.17 indicating the constitutive activation of pro-apoptotic genes in cells undergoing death. The most prominent caveat in understanding MS is the primary cause that provokes autoimmunity against the CNS. The presence of not one but a number of different factors might trigger the demyelination leading to the activation of these caspase pathways. The initiation of autoimmunity might be the consequence of an efficient presentation of CNS antigens to T cells by DC in lymphoid organs. After initial activation, T cells migrate into the CNS, where they are reactivated. As a consequence the caspase dependent necrosis occurs. The implications of these findings will be presented.

2074W

Common genetic variation contributes to cognitive performance in Russian elderly population. *O. Makeeva[1,2], A. Marusin[1], A. Bocharova[1], K. Vagaitseva[1], V. Markova[2], L. Minaycheva[1,2], I. Zhukova[3], N. Zhukova[3], V. Stepanov[1]. 1) Institute of Medical Genetics, Tomsk, Russian Federation; 2) Nebbiolo Centre for Clinical Trials, Tomsk, Russian Federation; 3) Siberian State Medical University, Tomsk, Russian Federation.*

Background: Maintenance of cognitive abilities in older ages as well as the late onset Alzheimer's disease are to the big extent genetically determined. A set of 62 single nucleotide variants previously reported to be linked to cognitive endophenotypes, psychiatric conditions, or late onset Alzheimer's disease was tested in respect to Montreal Cognitive Assessment (MoCA) total score and timed executive function test (Trials B) in 710 older subjects from a population-based registry from Tomsk, Russia. **Methods:** Volunteers (n=2032) were identified through a local centralized medical care system and were invited for neurocognitive assessment as a part of Alzheimer's disease prevention program. Self-reported information on individual's medical and family history, demographic information, level of education, and profession was collected. Genetic variants in 45 genes previously reported to be associated with cognitive phenotypes, psychiatric disorders and Alzheimer's disease in multiple GWAS or meta-analysis studies were genotyped using MALDI-TOF mass spectrometry. **Results:** Mean age in the study group was 72±5 years (from 56 to 91), 77.1% of the sample were female. Age (r=-0.338,p<0.001) and education (r=+0.422,p<0.001) significantly influenced MoCA total score; male and female subjects performed similarly on the test. MoCA scores corrected for covariates were subjected to the analysis of variance in order to estimate the contribution of studied SNPs to individual variability. Nine variants were associated with the variance of MoCA score. The largest contribution that cumulatively accounted for 8.6% of the MoCA total score variance was found for *APOE-TOMM40-PVRL2* gene cluster on chromosome 19q13.32. The effect of single SNPs in this cluster varied from 0.7% to 1.8%. Other significant signals were detected for variants at *CSMD1, SNX29, FBXO40,* and *TENM4* loci. Executive function was measured as the time for the Trials B execution in subjects who succeed to complete the test (n=665). Association was found with *TOMM40, PVRL, NKAPL, STK24* variants. **Conclusions:** The study confirms the role of structural variability in *APOE-TOMM40-PVRL2* locus and points several other loci associated with the cognitive performance in Russian elderly population. *(The work was supported by the Russian Science Foundation grant #16-14-00020)..*

2075T

Identifying the genetic underpinnings of social withdrawal. *N. Roth Mota[1,2], J. Bralten[1], K. Klemann[1], T.E. Galesloot[3], L.A.L.M. Kiemeney[3], G. Poelmans[1], B. Franke[1,2]. 1) Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Psychiatry, Radboud University Medical Center, Nijmegen, Netherlands; 3) Department of Health Evidence, Radboud University Medical Center, Nijmegen, Netherlands.*

Social withdrawal is characterized by a tendency to avoid social contacts and activities, and to prefer being alone rather than with others. While it is a continuously distributed trait in the population, high levels of social withdrawal represent a common early manifestation of multiple psychiatric disorders - such as Schizophrenia (SCZ) and Major Depressive Disorder (MDD) - and neurological diseases, such as Alzheimer's disease (AD). Given the complex nature of this trait, multiple risk factors - e.g. genetic, environmental, and psychological factors - are expected to influence it. However, the underlying biological basis of social withdrawal is still to be identified. Here, we aim to investigate the genetic underpinnings of social withdrawal trait in the general population. In this respect, based on questionnaire data from a Dutch population cohort (the Nijmegen Biomedical Study) of 2988 adults (46.5% male; mean age=59), we constructed a proxy measure of social withdrawal. This measure is based on 4 items from the Extraversion subscale of the Eysenck's Personality Questionnaire (EPQ) regarding social interaction avoidance, and on 3 items from the Autism Spectrum Quotient (AQ) scale about (lack of) social skills. For the EPQ items, respondents indicated if the statement applied to them or not; for AQ items, they indicated how much it applied to them on a 4-point Likert scale and answers were then dichotomized; for each respondent, a total score of social withdrawal was calculated by adding the scores of the 7 items (range 0-7). Genome-wide association analysis was performed to test the effect of individual genetic variants on the total score for our social withdrawal measure. This was followed by gene-based analyses to determine the joint effect of variants in each gene. Although no single variant reached genome-wide significance, gene-based analyses revealed the association of one gene - *GIT1* -with social withdrawal (P=2.30E-06), which survived correction for the 17.876 genes tested. The implication of *GIT1* gene, located on chromosome 17, on social withdrawal is an interesting and promising finding since *GIT1* is known to play an important role in neurite outgrowth and synapse formation. Additional analyses are being conducted in order to further investigate the implication of *GIT1* on social withdrawal and also to test the potential of genetic factors previously associated to SCZ, MDD and AD to predict social withdrawal measure in the general population.

2076F

Genetic investigation of restricted and repetitive traits in autism. *M.L. Cuccaro[1,2], S. Luzi[1,2], E.R. Martin[1,2], H.N. Cukier[1,2,3], A.J. Griswold[1,2,4], D.D. Dykxhoorn[1,2], M.A. Pericak-Vance[1,2].* 1) University of Miami School of Medicine, Miami, FL; 2) John P. Hussman Institute for Human Genomics; 3) Department of Neurology; 4) Department of Pathology.

Restricted and repetitive behaviors (RRBs) are a defining feature of Autism Spectrum Disorder (ASD). Two RRB subdomains, repetitive sensory motor behaviors (RSMB) and insistence on sameness (IS), have yielded suggestive association in stratified analyses in prior genetic studies of ASD. We hypothesize that genetic variants modify the expression of RRBs in individuals with ASD and that this genetic variation will be associated with RSMB and IS scores. Using the Autism Diagnostic Interview-Revised (ADI-R), a semi-structured interview for ASD administered to a knowledgeable informant, we calculated RSMB and IS scores for 1118 ASD participants from the Hussman Institute for Human Genomics and the Simons Simplex Collection. All individuals had DNA sequence data available from a 17Mb custom capture covering 681 genes within regions identified by GWAS of ASD. Gene-based and single-variant tests for association with IS and RSMB as quantitative traits were conducted using SKAT-O. Combinations of synonymous, non-synonymous, missense, stop, loss-of-function and splice variants were investigated in different hypothesis tests. A Bonferroni correction for the number of genes tested was used as a significance threshold for each hypothesis with an experiment-wise significance level of 0.05. Gene-based tests revealed different genes in association for the respective traits although none passed our Bonferroni corrected significance level. For the IS trait, two zinc finger genes were most significant when all exonic variants were included (*ZNF397*, p=2.24E-03, *ZSCAN30*, p=2.63E-03) and when only missense variants were examined (*ZNF397*, p=2.73E-03, *ZSCAN30*, p=3.16E-03). For the RSMB trait, we observed a convergence on the gene *PTPRT* for analyses of both damaging (p=1.49E-04) and missense (p=1.63E-04) variants. *PTPRT* (*protein tyrosine phosphatase, receptor type, T*) is a gene that is highly expressed in the developing and adult CNS and is involved in both CNS signal transduction and cellular adhesion. These results are in line with prior studies showing IS and RSMB as distinct type of RRBs. Our work shows that these different RRBs are potentially mediated by different genes. Our results suggest the value of identifying genes associated with phenotypic traits that are a part of ASD as an alternative to using the broad ASD diagnostic phenotype.

2077W

MAPK3 identified as candidate gene influencing schizophrenia and BMI in the 16p11.2 CNV region. *L. Davis[1], Y. Ji[1], J. Sutcliffe[1], E. Cook[2], H. Im[3], A. Barbeira[3], B. Li[1].* 1) Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Psychiatry, University of Illinois, Chicago, IL; 3) Section of Genetic Medicine, Department of Medicine, University of Chicago.

Deletions of the 16p11.2 chromosomal region have been reproducibly associated with autism spectrum disorders (ASDs), intellectual disability, and obesity, while duplications are associated with ASDs, ID, microcephaly, schizophrenia, and low body mass index. The 600 kb region between break points 4 and 5 has been most consistently associated with neurodevelopmental phenotypes, however, recent studies have shown the transcriptional consequences of copy number variants (CNVs) in this region extend beyond the boundaries of the CNV itself (Blumenthal et al., 2014). Thus, we sought to narrow the list of contributory candidate genes in the 16p11.2 chromosomal region. We hypothesize that altered expression of risk genes in the region contributes to disease predisposition and used MetaXcan (Barbeira et al., 2017), a method which utilizes genetic and gene expression data from GTEx and the Depression Genes Network to develop prediction models of gene expression levels in 45 human tissues. The genetically regulated component of gene expression is then "imputed" in an independent sample of cases and controls ascertained for genome-wide association studies (GWAS). MetaXcan allows the use of summary statistics for model testing bypassing the need for individual level genetic data and facilitating the use of very large sample sizes collected through consortium efforts. We tested the association between the predicted expression of 179 RefSeq genes (annotated to the 16p11.2 chromosomal region) across 45 tissues, and 101 publicly available GWAS phenotypes. A total of 32 genes demonstrated expression-phenotype associations that exceeded a Bonferroni correction for the total number of tests conducted (p=3.2E-06). Consistent with expectation, predicted increased expression of MAPK3 demonstrated a significant association with schizophrenia based on a GTEx whole blood model (Z = 6.51, p = 7.48E-11), a DGN whole blood model (Z = 4.68, p = 2.75E-06), and an EBV transformed lymphocyte model (Z = 4.91, p = 8.97E-07), while predicted expression demonstrated significant associations with decreased hip circumference based on a GTEx whole blood model (Z = -5.27, p = 1.36E-7) and a DGN whole blood model (Z = -5.08, p = 3.59E-07), and with waist circumference based on a GTEX whole blood model (Z = -4.78, p = 1.73E-06). Our findings suggest that MAPK3 plays a role in both obesity and schizophrenia phenotypes and highlight this gene as a potential risk gene influencing the 16p11.2 phenotypes.

2078T

Inherited mutations in Human Accelerated Regions (HARs) are associated abnormal social and cognitive behavior. *R.N. Doan[1], T. Shin[1], B.I. Bae[1], M. Nieto[2], B. Cubelos[3], S. Al-Saad[4], N.M. Mukaddes[5], C.A. Walsh[1,6,7], The Homozygosity Mapping Consortium for Autism.* 1) Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Departamento de Bıologıa Molecular, Centro de Bıologıa Molecular 'Severo Ochoa', Universidad Autonoma de Madrid, UAM-CSIC, Nicolas Cabrera, 1, Madrid 28049, Spain; 3) Department of Molecular and Cellular Biology, Centro Nacional de Biotecnologıa, CNB-CSIC, Darwin 3, Campus de Cantoblanco, Madrid, 28049 Spain; 4) Kuwait Center for Autism, Kuwait City 73455, Kuwait; 5) Istanbul Faculty of Medicine, Department of Child Psychiatry, Istanbul University, Istanbul 34452, Turkey; 6) Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA; 7) Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA.

Comparative analyses have identified genomic regions potentially involved in human evolution, but do not directly assess function. Human accelerated regions (HARs) represent conserved genomic loci with elevated divergence in humans compared to other primate and non-primate species. At least some HARs are thought to contribute to neurodevelopmental functions underlying the unique social and behavioral traits of humans. If some HARs regulate human-specific social and behavioral traits, then mutations would likely impact cognitive and social disorders. Recent studies support the role of HARs in complex neurodevelopmental functions and associated disorders, but the contribution of individual rare HAR mutations has yet to be fully characterized. Strikingly, we find an important contribution for inherited HAR mutations in disease manifestation through large-scale whole genome and targeted "HAR-ome" sequencing in individuals with abnormal social and cognitive development. Through a combined approach of chromatin interaction sequencing, massively parallel reporter assays (MPRA), and transgenic mice, we identified disease-linked, inherited HAR mutations in active regulatory elements for *CUX1, PTBP2, GPC4, CDKL5*, and other genes implicated in neural function, autism spectrum disorder, or both. Our data provide genetic evidence that specific HARs are essential for normal development, consistent with suggestions that their evolutionary changes may have altered social and/or cognitive behavior.

2079F

CNV meta-analysis of major neurodevelopmental and neuropsychiatric disorders. *J.T. Glessner, J. Li, M. March, C. Kim, K. Thomas, C. Kao, P.M.A. Sleiman, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia PA.

Neurodevelopmental and neuropsychiatric disorders represent a wide spectrum of heterogeneous yet inter-related disease conditions. The overlapping clinical presentations of these diseases suggest a shared genetic etiology. In a meta-analysis of copy number variation (CNV) analysis in five cohorts of neuropsychiatric disorders, 14 genes containing CNVs with exonic overlap reached genome-wide significance threshold of meta P-value $< 2 \times 10^{-6}$; and another 41 genes had meta P-value $< 5 \times 10^{-4}$. Twenty of these genes demonstrated abundant brain expression during infancy and among them and their direct interactors we observed significant enrichment in CNV burden involving ErbB, neurotrophin and MAPK signaling pathway genes. We observed significant associations of genes in several loci previously reported to be associated with neuropsychiatric disorders. In addition, we identified and validated novel significant associations of *DOCK8/KANK1* duplications (Meta P-value=7.5×10^{-7}) across all cohorts. Taken together, in our large scale meta-analysis of CNVs across multiple neurodevelopmental/psychiatric diseases, we uncovered significant associations of structural variants in multiple genes shared by five major neurodevelopmental/psychiatric diseases suggesting common etiology of these clinically distinct neurodevelopmental conditions.

2080W

Identification of novel variants in autism spectrum disorder using whole-exome trio sequencing. *R.S. Harripaul[1,2], N. Vasli[4], A. Rodrigues[1], A. Rabia[5], S. Mahmood[5], A. Heidari[6], L.A. Ayadhi[7], B. Bozorgmehr[1], R. Sasanfar[8], M. Ayub[9], J.B. Vincent[1,2,3].* 1) Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Sciences, Toronto, Ontario, Canada; 3) Department of Psychiatry, Toronto,Ontario, Canada; 4) SickKids Research Institute, Toronto,Ontario, Canada; 5) University of Health Sciences, Lahore, Punjab, Pakistan; 6) Qazvin University, Qazvin, Qazvin Province,Iran; 7) King Saud University, Ryadh, Saudi Arabia; 8) University of Massachusetts Memorial Hospital, Worcester, MA,USA; 9) Department of Psychiatry, Queen's University, Kingston, Ontario, Canada.

Autism Spectrum Disorder (ASD) affects 1% of the population and signs of ASD occur before the age of three years and are often comorbid with other psychiatric disorders such as Epilepsy and Intellectual Disability (ID). ASD is believed to have a high degree of genetic heterogeneity with many rare variants contributing to the disorder. Thus far, most ASD research has focused on *de novo* and dominant variants because the emphasis has been placed on small Western families where recessive variants are hard to identify. This has led to a substantial knowledge gap about the effects of recessive variants on diseases such as ASD. The aim of this study is to assess the role recessive variants play in ASD and neurodevelopmental disorders. We used Whole Exome Sequencing (WES) for 43 trios so far. In addition, we have DNA from 257 Iranian, 115 Pakistani, and 20 Saudi ASD trios that will be analyzed over the next few months. The analysis pipeline consists of genome alignment using SNAP and the Genome Analysis Toolkit GlobalHaplotyper (GATK) implemented on a high-performance Hadoop infrastructure. Variant calls are annotated using recently released databases and pathogenicity scores including M-CAP, Intervar, Sift, Polyphen2, MutationTaster and comparing against 123 000 exomes in gnoMAD and 2497 exomes in the GME Server. We have identified *de novo* variants in known ID or ASD genes *ZNF292*, *MYT1L*, *CHD8*, *DYRK1A*, and novel genes such as *ANXA4*; also an X-linked variant in known ID gene *THOC2*. Biallelic mutations have been found in the known ID/ASD gene *CC2D1A*, and new candidate gene *FAXDC2*. The *FAXDC2* is a homozygous truncating mutation that has not been associated with ASD or ID previously. *FAXDC2* is a fatty acid hydroxylase domain containing protein with no known function or previous associations with disease. *ANXA4* is part of the annexin family of genes previously associated with ID/ASD. It is a calcium dependent phospholipid binding protein and is expressed in the brain. Mouse models of *ANX4* indicate behavioral and neurological impairments when knocked out. We are in the process of analyzing 300 trios the results of which will be available in next few months. Once the function of these mutations and the associated clinical features is characterized, these genes can be utilized for molecular diagnosis, better clinical management and in the long-term may act as potential targets for personalized therapeutics.

2081T

Discovery of multifaceted genomic features as a promising approach to novel autism risk gene identification. *Y. Ji, Q. Wang, R. Chen, B. Li.* Vanderbilt Genetics Institute, Light Hall, 2215 Garland Ave. Nashville, TN.

Autism spectrum disorders (ASD) is a common and highly heritable neurodevelopmental disease. Understanding the underlying risk genes of ASD will deepen our knowledge of brain function and fuel therapeutic development for more appropriate treatment. GWAS studies only identified 5 GWS loci so far. De novo mutations (DNM) in ASD cases have been used effectively to identify risk genes due to their larger effect size and pathway convergence. However relying on recurrence of loss of function DNMs in the same gene only helped us find 10s of the estimated ~1000 ASD risk genes due to the rarity of DNMs. A big challenge is how to sift true risk genes from the massive amount of DNMs discovered in cohorts of ASD probands. We hypothesize that ASD risk genes have distinct genomics characteristics, and in this study we aim to systematically discover such genomic features and leverage these multi-dimensional predictive features in an integrative framework to identify ASD risk genes from DNMs not simply rely on recurrence. Specifically, we considered 3 categories of features: function based (e.g. disease-related pathway), sequence based (e.g. loss of function intolerant, conservation) and tissue-specificity based (e.g. genes of brain development). Using the 65 curated high-confident ASD genes (Sanders SJ, et al. Neuron 87-6), we identified a handful of genomic features that are characteristic of ASD risk genes compared to the genome background. In particular, features like target genes of RNA-binding protein FMRP (p=2.2e-16, OR=17.22, fisher-exact one-sided test) are highly enriched for ASD genes, pLi (loss of function intolerant) scores (p=6.9e-12, t-test) distinguish clearly ASD genes from the rest. In total we have found 14 features that reflect the characteristics of ASD genes. We reasoned that integrating the collection of distinguishing genomic features as a comprehensive source of evidence provides greater power to identify ASD risk genes. To test that, we used an independent set of ASD genes from SFARI database and found significant difference between ASD genes and random selected genes from genome background in the "integration score" of 14 features (p=5.67e-20). We have developed a Bayesian framework that enables the integration of the discovered predictive features as well as the de novo discovery of ASD relevant pathways, and are employing this new approach to systematically identify ASD risk genes from public DNMs generated by large ASD consortia. .

2082F

Gene expression profiling predicts clinical outcome in autism spectrum disorder: Confirmation of potential biomarkers and initial characterization of clinically homogeneous subgroups. *F. Macciardi[1,3], G. Guffanti[2], R. Miramontes[3], R. Padilla[3], D. Babek[4], H. Bennett[5], C. Eckman[6], E. Eckman[6], K. Selvaggi Fadden[5], T. Gleeson[5], M. Rabin[4], C. Sabatino[4], C. Gergati[1], M. Mapstone[3], G. Cossu[7], H. Federoff[3], M. Fiandaca[3], R. Kurlan[8].* 1) Dept of Psychiatry, University of California, Irvine, CA; 2) McLean Hospital, Department of Psychiatry, Harvard Medical School; 3) Dept of Neurology, University of California, Irvine, CA; 4) Atlantic Neuroscience Institute, Summit, NJ; 5) Department of Child Neurology and Developmental Pediatrics, Goryeb Children Hospital, Morristown, NJ; 6) Biomedical Research Institute of New Jersey (BRINJ), Cedar Knolls, NJ; 7) Centro di Foniatria, Padova, Italy; 8) Center for Neurodevelopmental and Neurological Health, Voorhees, NJ.

We present the initial results from the first reported blood whole transcriptome expression profiling (RNA sequencing) in 20 patients with Autism Spectrum Disorder (ASD) and 19 age- and gender-matched normally developing children, as part of a larger exploratory study of gene expression in childhood ASD. We first identified 152 differentially expressed transcripts that are significant at a nominal p-value < 0.05, of which one third overlap with genes already found associated with ASD from the SFARI (Simon Foundation Autism Research Initiative) database, while the remaining two thirds present potentially new findings. We also confirmed the differential expression of a subset of transcripts (N=10), originally reported by Glatt and colleagues (2012) using RNA microarrays from peripheral blood. Using these 10 transcripts in a disease-predictive model, we can distinguish between our cases and controls with high sensitivity and specificity. This finding in an independent sample represents the first confirmation of potential peripheral biomarkers indicative of an increased risk of developing ASD. Since the current diagnosis of ASD is still based on a clinical phenotype that only becomes discernable when an affected child is age 3 years or older, the availability of early actionable biomarkers indicative of ASD provides clinical relevance in avoiding delay in possible therapeutic interventions. Additionally, an exploratory analysis of co-expressed gene networks allowed us to untangle the heterogeneity of the ASD phenotype into at least two homogeneous subgroups characterized by contrasting clinical features, including the presence or absence of severe language impairment, IQ scores, and certain motor traits. .

2083W

Using the DGRP to identify gene networks associated with autism-like behaviors. *L.T. Reiter[1], P. Cavich[1], B. May[1], D. Johnson[2], K.A. Hope[1,3].* 1) Dept Neurology, Univ Tennessee HSC, Memphis, TN; 2) Molecular Bioinformatics Core, Univ Tennessee HSC, Memphis, TN; 3) IPBS Program, Univ Tennessee HSC, Memphis, TN.

Appropriate social behaviors are essential for reproduction and survival. Defects in social interaction and communication in humans can cause autism spectrum disorder (ASD). The basic neurogenetics controlling complex ASD-like behaviors in humans is still largely unknown. Here we used an unbiased quantitative trait locus (QTL) approach in *Drosophila melanogaster* to identify gene interactions that influence communication during mating (receptive and expressive), repetitive behaviors and social interactions in flies. The Drosophila Genetic Reference Panel (DGRP) has been used to analyze a range of naturally occurring behaviors including sleep, appetite, olfaction, and aggression. The DGRP represent >4.8M single nucleotide polymorphisms (SNPs) and >2.9M non-SNP variants mirroring the variation for each of 205 flies in a wild type population. Genetic variation within each line is fixed, while variation across lines represents the naturally occurring variation in the original population. This is the most powerful system available for behavioral genetics among model organisms. Using genome wide association tools designed for the DGRP, we identified genes associated with ASD-like behaviors across 40 DGRP lines. Social communication during mating was used as a proxy for expressive and receptive communication, grooming behavior as a proxy for repetitive behavior and physical spacing as a measure of social interaction among flies. Measures used for genome wide association were time to latency (expressive/receptive communication), average time grooming and average distance between flies (social space). We identified at least one gene, *sfl* (*sulfateless*), associated with all three ASD-like behaviors. *sfl* encodes a heparan sulfate N-deacetylase known to be essential for synaptic vesicle endocytosis. We show that knock down for *sfl* levels in neurons causes both decreased grooming and increased social spacing defects. We have also identified QTLs in 130 genes affecting ASD-like behaviors including homologous of *NLGN2* and *CNTNAP2*, but have only tested ~20% of the fully sequenced DGRP lines. Optimizing resolution and statistical power of these associations will require assessment of all 205 lines for all ASD-like behaviors. Our end goal is to identify new gene-gene relationships or even variants in known ASD genes that are associated with these behaviors in humans.

2084T

Epigenetic dysregulation of *DYRK1A* may have a role in ASD development in a discordant monozygotic twin pair. *C. Sjaarda[1,2], L. Santavy[1,2], A. McNaughton[1,2], M. Hudson[1,2], A. Guerin[3], M. Ayub[2], X. Liu[1,2].* 1) Queen's Genomics Laboratory at Ongwanada, Ongwanada Resource Center, Kingston, Ontario, Canada; 2) Department of Psychiatry, Queen's University, Kingston, Ontario, Canada; 3) Department of Pediatrics, Queen's University, Kingston, Ontario, Canada.

Autism spectrum disorder (ASD) is a complex and highly heritable neurodevelopmental condition with a poorly understood etiology. Genetic sequence variation, aberrant gene expression, and dysregulation of gene expression have all been implicated in ASD, though few studies have examined the combined effect of all three genetic factors. Using an integrative genomics approach, we investigate potential genetic and epigenetic mechanisms underlying the ASD phenotype in a family with a discordant monozygotic twin pair and a non-twin sibling with ASD. Whole exome sequencing of both subjects with ASD identified potentially significant variants. Transcriptome sequencing identified differentially expressed genes between monozygotic twins that are enriched for pathways relevant to ASD including MAPK signaling. Most importantly, expression of *DYRK1A*, a kinase involved in neurodevelopment and previously implicated in ASD, was significantly reduced in both individuals with ASD relative to unaffected family members. MicroRNA sequencing identified 46 differentially expressed microRNA between monozygotic twins including two microRNAs, miR-150-3p and miR-369-3p, that are significantly overexpressed in the twin with ASD and predicted to target the *DYRK1A* gene transcript. These results may indicate a role for aberrant *DYRK1A* expression in development of ASD in the affected twin and sibling and suggests that one possible mechanism of *DYRK1A* dysregulation is by differential microRNA expression.

2085F

Epigenetic factors and gene-environment interactions in autism: Prenatal maternal stress and the SERT gene. *Z. Talebizadeh[1], A. Shah[1], J. Noel-MacDonnell[1], P. Hecht[2], D. Beversdorf[2].* 1) Children's Mercy and University of Missouri-Kansas City School of Medicine, Kansas City, MO; 2) University of Missouri, Columbia, MO.

There is insufficient knowledge of the potential environmental factors or gene X environment interactions (GxE) in autism. Maternal stress exposure may be important in autism. By surveying for history/timing of prenatal psychosocial stressors, we found autism mothers to have a higher incidence of stressors compared to controls, which was subsequently confirmed by larger epidemiological studies. To explain why prenatal stressors might result in autism in some cases but not others, we began to explore GxE. The serotonin transporter (SERT) gene is well studied for its role in stress reactivity. The most widely studied variation is an insertion/deletion within the promoter region, resulting in long (L) or short (S) alleles. The S-allele has been linked to autism in some but not all studies. These contradictory findings could, in part, be explained by the presence of a gene/stress interaction or epigenetic factors. There is also emerging evidence that microRNAs may play a regulatory role in the serotonergic pathway and prenatal stress, particularly, in response to chronic stress through interaction with serotonergic genes. In the present study, we examined microRNAs in blood samples (n=34) from mothers of children with autism, with known pregnancy stress history. The samples were divided into 5 groups based on SERT genotypes (LL, LS, and SS) and prenatal stress level (High and Low). Five-way ANOVA showed differential expression (DE) of 119 microRNAs (P<0.05), 90 (76%) of which showed a similar pattern of expression in High vs Low stress groups (stress-dependent microRNAs). Intriguingly, two of them, miR-1224-5p and miR-331-3p, were recently reported by our group to exhibit stress-dependent expression in rodent brain samples from embryos exposed to prenatal stress. Another stress-dependent microRNA, miR-145-5p, has been reported in association with maternal stress. To assess the role of SERT genotype, we conducted a 3-way ANOVA on groups exposed to a high level of prenatal stress. This analysis showed a smaller number of DE microRNAs (n=20), 5 (25%) of which were among the stress-dependent microRNAs. These 5 microRNAs may be candidates for stress X SERT genotype interactions. One of them, miR-663a, has been previously reported to be DE in response to Fluoxetine (SSRI drug). We have also performed exome analysis and will correlate the microRNA expression with exome data to detect other potential serotonergic genes that may play a role in this GxE model.

2086W

Assembling the effects of genetic and environment risk factors in autism spectrum disorder using patient-derived neurons. *K. Tammimies[1], M. Becker[1], M. Arora[2], S. Stamouli[1], M. Kele[3], S. Bölte[1], A. Falk[3].* 1) Women's and Children's Health, Center of Neurodevelopmental Disorders (KIND), Karolinska Institutet, Stockholm, Sweden; 2) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, USA; 3) Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden.

Autism spectrum disorder (ASD) is a common neurodevelopmental disorder affecting 1-2% of the population. Hundreds of genes and multiple environmental risk factors have been linked to ASD susceptibility. Using a highly informative twin cohort of more than 150 twin pairs, enriched for monozygotic (MZ) pairs discordant for ASD, we are studying genetic and environmental risk factors underlying the disorder. No discordant genetic variants have been identified in the discordant MZ twins in saliva- or blood-derived genomic DNA. However, we have identified shared rare copy number and sequence level variants in genes and loci earlier implicated in ASD. Additionally, we have demonstrated a dysregulation of essential and toxic metals, including zinc and lead, during the prenatal and early postnatal periods (Arora *et al.* Nat Commun. 2017 Jun 1;8:15493). We are generating induced pluripotent stem cells (iPSCs) from the same twins to simultaneously model the identified genetic and environmental risks during neuronal development. The iPSCs are derived based on earlier reported protocol (Uhlin *et al.* Stem Cell Res. 2017 Jan;18:22-25) and the cells are further reprogrammed into neuroepithelial stem cells (NES) before the neuronal differentiation. We have selected twin pairs for further characterization based on the identified genetic risk factors. These include ASD-discordant MZ twin pairs with 0.95 Mb duplication of chr16q14.1 and another pair with a splice site mutation affecting an X-linked synaptic protein *CASK*. We will mimic the environmental risk conditions such as the prenatal zinc deficiency using modified cell culture media which we have optimized using chelating agents or by the addition of toxic metals. The effect of the risk conditions is measured using single cell transcriptomics during different time points of neuronal differentiation. Furthermore, we are testing different chelating drugs such as Clioquinol in our system. We will validate the molecular pathways affected by both the genetic variants and environmental exposures using additional cellular assays, i.e. measuring cellular energy levels and synaptogenesis. With our approach, we aim to understand better the mechanisms underlying ASD, identify novel drug targets and putative therapeutic strategies for subgroups with ASD.

2087T

Complete gene knockouts in autism spectrum disorder. *T.W. Yu[1], R.N. Doan[1], E.T. Lim[1], S. De Rubeis[2], C. Betancur[3], D.J. Cutler[4], A.G. Chiocchetti[5], L.M. Overman[6], S. Goetze[1], C.M. Freitag[5], M.J. Daly[7], C.A. Walsh[1], J.D. Buxbaum[2], Autism Sequencing Consortium.* 1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts, USA, 02115; 2) Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 3) INSERM, U1130, Paris 75005, France CNRS, UMR 8246, Paris 75005, France Sorbonne Universités, UPMC Univ Paris 6, Institut de Biologie Paris Seine, Neuroscience Paris Seine, Paris 75005, France; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA; 5) Autism Research and Intervention Center of Excellence, University Hospital Frankfurt, Goethe University, 60528 Frankfurt, Germany; 6) Human Developmental Biology Resource, Institute of Genetic Medicine, Newcastle University, International Centre for Life, Newcastle upon Tyne, NE1 3BZ; 7) Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA.

Autism spectrum disorder (ASD) exhibits complex genetic architecture with well demonstrated contributions of both common and rare, especially *de novo*, variants. Biallelic mutations provide another powerful, but comparatively underexplored, avenue for illuminating disease-relevant biological processes. We report the largest analysis to date of biallelic mutation in ASD using whole exome sequencing data from >8,000 individuals in the ASC. We confirm an excess of biallelic, autosomal loss-of-function (LOF) mutations in ASD, and extend these observations to biallelic damaging missense variation as well. The observed excess is especially striking in females (10% of affected females), with patterns consistent with a female protective (or male susceptibility) effect. We also find that ~60% of genes subject to biallelic LOF in the ASC are also knocked out in ExAC. From this observation we develop gene and allele filters based on ExAC and demonstrate that their application cuts observed rates of biallelic mutation by half, but leaves the ascertainment differential between cases and controls intact, thus indicating that the excess of biallelic mutation in ASD is driven by genes under recessive constraint. We identify several patients diagnosed with ASD due to biallelic knockout of known and/or emerging recessive neurodevelopmental genes (*CA2, DDHD1, NSUN2, PAH, RARB, ROGDI, SLC1A1, USH2A*). We also identify ASD-specific, biallelic gene knockouts in genes that affect neurobiological mechanisms not previously implicated by *de novo* mutations, including *FEV*, a transcription factor required for development and function of serotonergic circuitry, a system long suspected to be dysregulated in ASD. Our data refine the contributions of recessive mutation to ASD and suggest paths for using rare recessive mutations to illuminate novel biological pathways underlying ASD.

2088F

Chronic psychosocial stress in mice alters brain myelination in a genetic background-dependent manner. *I. Hovatta[1], M. Laine[1], K. Trontti[1], Z. Misiewicz[1], N. Kulesskaya[1], S. Saarnio[1], E. Jokitalo[2], D. Greco[2], I. Balcells[1], E. Sokolowska[1].* 1) Department of Biosciences, University of Helsinki, Helsinki, Finland; 2) Institute of Biotechnology, University of Helsinki, Helsinki, Finland.

Chronic psychosocial stress is a well-established risk factor for anxiety disorders and major depression. Mechanisms by which chronic stress impacts susceptibility and resilience to psychiatric disorders are largely unknown. Animal models allow administration of specific stressors in a controlled manner in inbred genetic backgrounds to reveal gene-environment interactions. To identify brain gene expression changes taking place after chronic psychosocial stress, we used 10-day chronic social defeat model in mice, and carried out RNA-seq from medial prefrontal cortex (mPFC) and ventral hippocampus (vHPC) of C57BL/6 (B6; a non-anxious inbred strain) and DBA/2 (D2; an anxious inbred strain) mice. The two mouse strains showed a distinct behavioral response to stress, as measured by the social avoidance test carried out after social defeat. Sixty % of the B6 mice but only 10% of D2 mice were resilient to stress, the remainder being susceptible. We discovered that the expression levels of several myelination-related genes differed substantially due to chronic stress in both brain regions, and 'myelination' pathway was over-represented in gene-set enrichment analysis. To establish whether and how myelin thickness and structure are altered after stress, we carried out transmission electron microscopy. We found that the B6 stress resilient mice had significantly thicker myelin of small diameter mPFC axons compared to the B6 control mice. In the vHPC, susceptible B6 mice had thinner myelin compared to B6 control mice. In the D2 strain, susceptible mice had thicker myelin than resilient mice within the mPFC and resilient mice had thinner myelin of small diameter axons compared to the controls in the vHPC. Our findings suggest significant white matter plasticity in response to chronic psychosocial stress in mice. Such differences have previously been observed in response to early-life stress or social isolation in adult mice. Our results extend these previous findings to psychosocial stress, and demonstrate that the pattern of myelination differences is dependent on the genetic background and varies across brain regions.

2089W

Analysis of the genetics and heritability of a shared endophenotype in ADHD and FASD. *J. Kapalanga[1,2,5], T. Trudeau[3], B. Laufer[4].* 1) Pediatrics, Western University, London, ON, Canada; 2) Pediatrics/Genetics, Grey Bruce Health Services, Owen Sound, ON, Canada; 3) Biology, Western University, London, ON, Canada; 4) Schulich School of Medicine, Western University, London, ON, Canada; 5) Summerside Medical Center, Summerside, PEI, Canada.

ADHD and FASD share an behavioral endophenotype characterized by impaired executive function, emotional regulation, cognition, learning and social adaptation. Genetic factors play a significant role in both FASD and ADHD, and in ADHD families the shared endophenotype is heritable. While FASD is causally associated with prenatal alcohol exposure (PAE) and includes severe forms known as fetal alcohol syndrome (FAS) characterized by distinctive physical features, PAE and distinctive physical features are not associated with ADHD. However, individuals with undiagnosed ADHD are at risk for substance and alcohol abuse during the reproductive years. Children with ADHD often have mothers who have ADHD and children with FASD often have mothers with undiagnosed ADHD. This raises the question of whether inheritance of genetic factors and not PAE is the cause of the shared endophenotype in FASD. It also raises the question of whether or not FAS is in fact a separate PAE – induced disorder that does not share etiology with FASD and ADHD. We hypothesize that ADHD and FASD share a heritable endophenotype that is associated with specific etiologic factors. This study is i) to determine whether or not ADHD and FASD share an endophenotype, ii) to determine whether the endophenotype is heritable, and iii) to determine if mothers of FASD patients have undiagnosed ADHD. **Methods:** Children aged 3 to 17, with FASD and/or ADHD and their consenting mothers were recruited after ethics approval. Patient interviews and chart analysis were conducted to extract clinical information focused on behavioral features, and maternal childhood history of ADHD and alcohol consumption. **Results:** Data analysis identified 184 ADHD and 166 FASD children, and 84 ADHD and 74 FASD mothers. Of these 91% ADHD and 93% FASD children had mothers with the endophenotype; and 85% of ADHD and 88% of FASD mothers had children with the endophenotype, respectively. Further, for the ADHD and FASD children 80% and 77% of the mothers with the endophenotype, respectively had at least 2 affected children. **Conclusion:** Children with ADHD and FASD share a certain heritable behavioral endophenotype that is also observed in their corresponding mothers. Genetics and not PAE is the explanation for the shared endophenotype in children diagnosed as FASD. This implies that ADHD and FASD are the same disorder and are caused by the same etiologic factors. FAS is a separate disorder.

2090T

Genetic polymorphism and gene-environment interactions of dopamine receptor genes and nicotine dependence in the population of the Northwest Indian region. *J. Kaur, R. Kaur.* Department of Human Genetics, Punjabi University, Patiala, Punjab, India.

    Purpose: Nicotine Dependence (ND) like many other addictive behaviors has a genetic component. Furthermore, D1 and D2 like dopamine receptor genes involved in dopaminergic reward pathways of the brain are considered excellent candidates for contributing to the genetic components of these addictive behaviors. In the present study we aimed to ascertain the association, if any, between DRD1, ANKK1 *TaqIA*, DRD2 *TaqIB*, DRD2 *TaqID* and DRD3 gene SNPs with nicotine dependence by case-control. **Methods**: Genomic DNA was extracted using standard salting out method and genotyped using PCR-RFLP method in ND cases (n=173) and healthy controls (n=188) inhabiting North-West region of India. The degree of nicotine dependence was ascertained by FTND, HSI and HWRSS (six reasons for smoking). Data were analyzed using MedCalc and SPSS 16.0, Haploview program, version 4.2. ,SHEsis and Multifactor Dimensionality Reduction (MDR) Analysis. **Results**: A total of 361 subjects (173 nicotine dependent cases with a mean age 35.37±14.29 years and 188 healthy age/ethnicity matched controls with a mean age of 35.79±13.37 years) participated in the study. Both the groups were found to be in Hardy- Weinberg equilibrium. On applying contingent chi-square test, DRD2 *TaqID* showed significant differences between cases and controls at the allelic level (χ^2= 20.2, d.f=1, 0.000<p<0.05) and DRD3 showed significant differences both at the genotypic (χ^2= 5.84, d.f=2, 0.054<p<0.05) and allelic level (χ^2= 5.63, d.f=1, 0.018<p<0.05). However, after applying the stringent Bonferroni correction the differences remained significant only at the allelic level for DRD3 and DRD2 *TaqID*. *DRD3* gene showed a 2 fold risk of ND (OR 2.11, 95% CI 1.118-3.985, 0.021<p<0.05) with rare genotype *(Gly/Gly)* compared to predominant *(Ser/Ser)* genotypes. D′ value suggested a strong linkage between the three sites ANKK1 *TaqIA*, DRD2 *TaqIB* and DRD2 *TaqID.* Gene-environment interplay unveiled synergistic interactions between DRD1-ANKK1 *TaqIA (*gene-gene)*, DRD3-Pack years smoked (Gene-environment), NAR-DRD1 (Gene-environment), DRD2 *TaqID*-NAR (Gene-environment), DRD2 *TaqIB*-NAR (Gene-environment) and DRD1-Pleasure (Gene-environment) for nicotine dependence risk prediction in ND cases and controls. **Conclusion**: Our results suggest a higher risk of nicotine dependence in individuals carrying minor *T* allele of *DRD3* and *D1D1* genotype of DRD2*TaqID* in their lifetime than others.

2091F

Whole-genome sequencing to identify risk loci for nighttime eating in American Indians. *C. Koroglu[1], M.E. Gluck[2], S. Vortruba[2], J. Krakoff[2], M. Traurig[1], P. Piaggi[1], P. Chen[1], L.J. Baier[1].* 1) Diabetes Molecular Genetics Section, Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, AZ; 2) Obesity and Diabetes Clinical Research Section, Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, AZ.

    Night eating syndrome (NES) is characterized by increased food intake in the evening and night time, with or without nocturnal awakenings to eat. Familial aggregation has been shown for NES, suggesting that this syndrome has genetic determinants. Individuals with NES are prone to weight gain and the disorder can complicate diabetes management; therefore, our study focused on identifying genetic determinants of nighttime eating (NE) behavior in American Indians who are at high risk for obesity and type 2 diabetes. We performed whole-genome sequencing on 23 individuals with NE and 27 controls, where all individuals were >50% American Indian. These individuals had been clinically characterized over a 3-day period as inpatients in our Clinical Research Center, and NE status was determined based on the occurrence of food intake on any of the 3 nights between 11 pm and 5 am, recorded by a computerized vending machine. Among sequence data from these individuals, 6.2 million variants of high quality were detected. As expected with this small sample size, no variant achieved genome-wide significance for NE in a case/control analysis. Therefore, the top signals for NE (396 variants with a P < 1 x 10^{-4}) were further analysed in 6,789 American Indians with longitudinal data on BMI. Among these top signals for NE, only rs2268953 in *AKAP6* also associated with maximum lifetime BMI (P = 5.1 x 10^{-5}, adjusted for age, sex, birth year, admixture and the principal components 1-5), where the risk allele for NE (G allele) associated with a higher BMI. *AKAP6* is most highly expressed in the brain, but it is also expressed in cardiac and skeletal muscle. Its protein product is involved in anchoring PKA to the nuclear membrane or sarcoplasmic reticulum. Interestingly, *AKAP6* has previously been reported as a risk locus for anorexia nervosa in a genome-wide association, which supports a potential role of *AKAP6* in determining eating behaviour. We are currently pursuing functional studies of AKAP6, but are also looking at variation in other candidate loci (e.g. genes that have a role in regulation of leptin, ghrelin, melatonin, and cortisol levels) which may have only achieved nominal statistical evidence for association with NE in this small sample size.

2092W

RNAseq transcriptome study of schizophrenia in the MGS African American sample. *A. Sanders[1,2], J. Duan[1,2], E. Drigalenko[3], H. Göring[4], P. Gejman[1,2], Molecular Genetics of Schizophrenia (MGS) Collaboration.* 1) Res Inst, NorthShore Univ HealthSystem, Evanston, IL; 2) Dept of Psychiatry and Behavioral Neuroscience, Univ of Chicago, Chicago, IL; 3) Dept of Genetics, Texas Biomedical Res Inst, San Antonio, TX; 4) South Texas Diabetes and Obesity Inst, Univ of Texas Rio Grande Valley Sch of Med, San Antonio, TX.

   GWAS on predominantly European ancestry (EA) samples have implicated over 100 loci as being associated at genome-wide significant levels with risk for schizophrenia (SZ). Some of these risk loci are shared across ancestral groups, including within variably admixed African American (AA) samples. Regulation of mRNA expression may be involved in the etiology for some of these loci, thus suggesting utility of a transcriptomic profiling approach. In both of two previous large transcriptomic profiling studies on lymphoblastoid cell lines (LCLs) from non-overlapping EA SZ case-control samples, using arrays (N=714; Sanders et al., 2013) and RNAseq (N=1,189; Sanders et al., 2017), we found genes differentially expressed by affection status to be enriched for immune-related genes, consistent with hypothesized immune contributions to SZ risk. The single largest reported SZ GWAS on an AA sample is on the Molecular Genetics of SZ (MGS) dataset (Shi et al., 2009), from which we derived the current transcriptomic (RNAseq) sample. After various sample quality control steps including identity verification ensuring concordance between known sex (i.e., dosages of X and Y chromosomes) and RNAseq expression levels of sex-dimorphic expressed genes on chromosomes X and Y, and RNAseq-derived genotype concordance versus previous GWAS genotypes, we moved samples with at least 8M read depth from 378 SZ cases and 386 controls to analyses. To adjust for known or potentially confounding effects, we collected data on measured (EBV load, growth rate, and energy status), RNAseq batch, ancestry (top two genotypic principal components from GWAS), and other demographic (sex, age) covariates to include in the regression analysis to identify genes differentially expressed by affection status. We will present and discuss results in several main areas: (1) degree of replication of AA transcriptomic results for previous EA findings, (2) pathway and network analyses of differentially expressed genes, (3) gene set enrichment analyses, especially for immune related genes, (4) differences compared to findings for previous EA transcriptomic work, and (5) combined analyses of the current AA and previous EA transcriptomic datasets. This work was supported NIH grant R01MH098059 using samples from the MGS collaboration.

2093T

Minds, genes, and machines: Performance on online cognitive assessments is correlated with individual characteristics in 23andMe customers. *O.V. Sazonova[1], L. Germine[2], C. Wilson[1], P. Fontanillas[1], C.E. Carey[3], 23andMe Research Team.* 1) 23andMe, Mountain View, CA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Department of Psychological and Brain Sciences, Washington University, St. Louis, MO.

   Personality and cognitive traits can have a profound impact on individual quality of life. Together with environmental factors, a significant body of literature points to a genetic component of individual variation in human behavior. Specifically, cognitive traits such as risk-taking tendency, impulsivity, and processing speed have shown heritability in twin studies. Uncovering novel causal relationships between cognition and the combination of genetics, lifestyle, and demographics can deepen our understanding of the mechanisms by which cognitive processes manifest. 23andMe is a consumer genetics company with a database of over 2 million genotyped customers, the majority of whom consent to participate in 23andMe research. 23andMe has recently deployed a series of online cognitive assessments designed to capture phenotypic variation in cognitive traits. In the present study, we mined our database for associations with performance on the Balloon Analog Risk Task ("BART") and the Digit Symbol Substitution Test ("DSST"). Higher BART scores have been associated with neuropsychiatric disorders that involve risk-seeking behavior. The DSST is a test of processing speed, attention and cognitive efficiency and is broadly sensitive to brain dysfunction. Ongoing data collection has resulted in over 10,000 consented research participants for each of the BART and the DSST. We modeled scores as continuous and discrete phenotypes (adjusting for extreme outliers) and explored genetic and phenotypic associations in our database. Preliminary results suggest a genetic component to performance on both tasks, as well as associations between BART scores and phenotypes pertaining to education and "Big Five" personality traits. As independent validation, we also compared performance on the BART with a survey of risk-taking, where 23andMe research participants rated themselves as "adventurous" on a 5 point scale. This study represents an ongoing effort to combine genetics and self-reported phenotypic data with online cognitive assessments. A better understanding of the mechanisms that influence cognitive traits may guide the development of more effective treatments for neuropathic conditions such as impulse control disorder or bipolar disorder. As a novel paradigm in crowdsourcing biomedical discoveries, our research model holds great promise for uncovering causal factors that contribute to the continuum of human cognition.

2094F

Network analysis of gene polymorphisms in GABA, dopamine, cannabinoid, mu-opioid and alcohol metabolism pathways with alcohol dependence in scheduled class (SC) population of Punjab (Northwest India). *R. Sharma, AJS. Bhanwer.* Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

The role of GABA, dopamine, cannabinoid, mu-Opioid and alcohol metabolism pathways in alcohol dependence (AD) are evident from animal models and human studies. Aim of this study was to investigate association between genes in the different pathways of AD. This case-control study was designed to investigate the association of SNPs in *ADH2* (rs1229984)*, ALDH2* (rs671), *CNR1* (rs806368), *DRD2* (rs1800497, rs1079597, rs1799732), *DRD4* (rs916455), *GABRA2* (rs567926, rs279858) and *OPRM1* (rs1799971) genes with AD in the SC population. Male subjects comprising of 300 AD cases and 300 age, ethnicity matched normal controls were enrolled in this study. The G allele of *CNR1* (rs806368) variant conferred 1.5 fold risk towards AD [p=0.001, OR=1.58 (1.2-2.08)]. Of the three polymorphisms analyzed in *DRD2* gene, significant risk to AD was posed by minor alleles of rs1800497 (T allele) [p=1.5×10^{-5}, OR=1.8 (1.38-2.35)] and rs1799732 (D allele) [p=0.0003, OR=1.67 (1.26-2.22)]. The minor alleles (C allele) of both variants of *GABRA2* (rs567926, rs279858) posed 1.4 fold risk for alcohol abuse. The G allele of *OPRM1* rs1799971 was associated with increased risk for AD [p=1.96×10^{-4}, OR=1.82 (1.33-2.51)]. Both *ADH2* (rs1229984) and *ALDH2* (rs671) were observed to be monomorphic for G allele. Haplotype analysis of *DRD2* gene created two novel combinations: TGC providing risk for AD [p=7.9×10^{-4}, OR=2.12 (1.35-3.31)] while CAA playing protective role against AD [p=5.1×10^{-4}, OR=0.51 (0.35-0.75)]. TT combination of *GABRA2* provided protection from AD [p=2.8×10^{-4}, OR=0.6 (0.46-0.79)]. MDR analysis detected 3 loci genetic model comprising of *CNR1* (rs806368), *DRD2* (rs1800497) and *GABRA2* (rs567926) variants to increase the susceptibility of AD by 5 fold [p<0.0001, OR=5.77 (3.82-8.72)]. Prediction of RNA secondary structure from the base sequence using minimum free energy (MFE) based on thermodynamic approach was also outlined which revealed that ancestral type alleles to be structurally more stable. This analysis further evaluated the impact of studied variants on the RNA structure. Further, network analysis was conducted to unravel the remaining interrelating genes among the studied pathways. In conclusion, combinatorial effect of genes in different studied pathways conferred the major genetic risk for AD and can serve as potential biomarkers for determining the genetic predisposition towards AD in the SC population of Punjab.

2095W

Understanding remission on venlafaxine in late-life depression: A genome-wide approach. *V.S. Marshe[1,2], M. Maciukiewicz[2], S. Rej[3,4], A.K. Tiwari[2,4], E. Sibille[2,5], D. Blumberger[4], J.F. Karp[6,7], E.J. Lenze[8], C.F. Reynolds III[6], J.L. Kennedy[1,2,4], B.H. Mulsant[1,2,4], D.J. Mueller[1,2,4].* 1) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 2) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 5) Department of Pharmacology, University of Toronto, Toronto, ON, Canada; 6) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA; 7) VA Pittsburgh Health System, Geriatric Research Education and Clinical Center, Pittsburgh, PA, USA; 8) Healthy Mind Lab, Department of Psychiatry, Washington University, St. Louis, MO, USA.

Background: In geriatric depression, antidepressant treatment response is often slow and incomplete; new biomarkers for antidepressant response could lead to precision medicine and uncover new mechanisms of illness. Therefore, we conducted a genome-wide association study (GWAS) in older adults treated with venlafaxine. **Methods:** Three-hundred and fifty participants (>60 years) of mixed ethnic ancestry, diagnosed with major depression (MADRS > 15), were treated with venlafaxine (~12 weeks). Individuals were genotyped using the Illumina PsychArray, and genetic data was imputed to 7.5 million variants per individual. Associations with remission status (MADRS < 10) and change in MADRS score were conducted using logistic/linear regressions, adjusted for ancestry, sex, recruitment site, length in treatment, depressive episode duration, and baseline depressive severity. We also conducted pathway enrichment analysis using DEPICT. **Results:** No variants were significant at the genome-wide level. Our top hit variant, in a schizophrenia susceptibility gene, showed an association with MADRS score change. This gene has been implicated in post-ketamine treatment down regulation of GABA and glutamate levels in the rat prefrontal cortex and hippocampus, resulting in an antidepressant effect. We also observed a suggestive association between remission status and a genetic variant upstream of neuro-specific gene expressed in the hypothalamus and amygdala, suggesting a role in synaptic neurotransmitter signalling. Our top hit pathways included processes in the metabolic pathway of amyloid precursor protein. **Conclusions:** Our findings suggest novel gene variant associations with measures of venlafaxine remission in older adults, as well as, interesting neuro-relevant genetic pathways. A new, larger study of geriatric depression treatment will confirm these novel findings.

2096T

Abnormal expression of sonic hedgehog as a biomarker and therapeutic target for depression and suicide in bipolar disorder. *M. Galdzicka[1], J. Karlgren[2], E.I. Ginns[2].* 1) Pathology, UMass Memorial Medical Center, Worcester, MA; 2) Psychiatry, UMass Memorial Medical Center, Worcester, MA.

   Purpose: Bipolar affective disorder (BPAD) is a global public health problem. Compelling clinical-epidemiologic evidence supports a significant genetic predisposition. Our GWAS for BPAD protective genes in the Old Order Amish (OOA) identified genomic loci containing Shh antagonists (EVC and Hhip). No Amish individual with Ellis-van Creveld (EvC) dwarfism in these bipolar families has had bipolar disorder. The molecular change causing EvC disrupts Shh signaling. There is no cure, and many patients with depression and risk for suicide find that their bipolar symptoms are only partially managed with available treatments. More effective treatment for BPAD thus remains an enormous unmet need. A critical barrier to uncovering causes of BPAD is lack of biological markers for BPAD. The goal of this study is to determine the extent that sonic hedgehog expression is a biomarker and potential therapeutic target for depression and suicide in bipolar disorder.   Methods: RNA was extracted from fibroblasts of three normal and three BP1 affected related OOA family members using Qiagen RNeasy Mini Kit. Residual genomic contamination in samples was removed by the on-column DNase digestion step. Integrity of each RNA sample was verified on the Agilent® Bioanalyzer. The cDNA synthesis was performed from two micrograms of total RNA using the RT2 First Strand Kit. Real Time PCR was performed using ABI7500 Fast instrument according to the RT2 RNA QC PCR Array protocol using Human Hedgehog Signaling Pathway array (Qiagen). The array contains primers for 84 key genes in this pathway.   Summary of Results: Quantitative PCR using the Qiagen human Shh signaling pathway array (PAHS-078Z) indicated a greater than 600 fold elevated expression of Shh mRNA in fibroblasts from Amish BP1 compared to Amish controls.   Conclusion: Our research has benefited from a rare, informative "experiment of nature" of families where EvC individuals at high risk for BPAD do not get BP1. Ketamine, lithium and SSRIs act as antagonists of Wnt and Shh signaling, mimicking the EvC mutation BPAD protective effect. We expect that results from this study will increase our understanding of altered brain signaling in bipolar disorder, and lead to more personalized molecular diagnosis. Our discovery suggests that sonic hedgehog signaling can be modulated to help manage bipolar symptoms in adolescents and adults by repurposing FDA approved drugs or those already in clinical trials for other medical conditions.

2097F

A multi-omics analysis towards understanding of the polygenicity in schizophrenia. *P. Jia[1], X. Chen[2,3], W. Xie[4], Z. Zhao[1,5].* 1) School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Psychology, University of Nevada Las Vegas, Las Vegas, NV; 3) Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV; 4) Department of Electrical Engineering & Computer Science, Vanderbilt University, Nashville, TN; 5) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX.

   Numerous high throughput omics studies have been conducted in schizophrenia, providing an accumulated catalogue of susceptible variants and genes. The results from these studies, however, are highly heterogeneous and complex. The variants and genes nominated by genetic studies often have limited overlap with those from transcriptomic and epigenetic studies, and vice versa. There is thus a pressing need for integrative analysis to unify the different types of data and provide a convergent view of candidate genes in schizophrenia. In this study, we collected a comprehensive, multidimensional dataset, including 7475 brain-expressed genes. The data hosted genome-wide association evidence in genetics (e.g., genome-wide association studies, linkage studies, copy number variations, *de novo* mutations), epigenetics (methylation), transcriptomes, and literature mining. We developed a method named Mega-analysis of Odds Ratio (MegaOR) to prioritize candidate genes. Application of MegaOR in the multidimensional data resulted in consensus sets of candidate genes associated with schizophrenia (SZgenes), which were enriched with condensed evidence from multi-dimensions. We proved that our SZgenes had highly tissue-specific expression in brain and nerve and had intensive interactions significantly higher than randomly expected. Furthermore, we found that our SZgenes were involved in the human brain development by showing strong spatiotemporal expression patterns, which were replicated in independent brain expression data sets. Finally, we found the SZgenes were enriched in critical functional gene sets involved in neuronal activities, ligand gated ion signaling, and Fragile X mental retardation protein targets. In summary, these results assessed rich association evidence to demonstrate a converged set of schizophrenia candidate genes, whose characteristics provided insights into the pathophysiology underlying schizophrenia.

2098W

Genetics of schizophrenia in Cooperative Studies Program #572. *N. Sun[1,2], Q. Lu[2], Y. Hu[2], B. Li[2], Q. Chen[1,2], M. Aslan[1,3], K. Radhakrishnan[1], K.H. Cheung[1,4], Y. Li[1,5], N. Rajeevan[1,5], F. Sayward[1,5], S. Pyarajan[6,7], J.M. Gaziano[6,7], J. Concato[1,3], H. Zhao[1,2], P. Harvey[8,9] on behalf of the Cooperative Studies Program (CSP) #572 and VA Million Veteran Program (MVP).* 1) VA Clinical Epidemiology Research Center (CERC), VA Connecticut Healthcare System, West Haven, CT; 2) Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut; 3) Department of Medicine, Yale University School of Medicine, New Haven, Connecticut; 4) Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT; 5) Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT; 6) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA; 7) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 8) Research Service, Miami VA Healthcare System, Miami, FL; 9) Department of Psychiatry, University of Miami School of Medicine.

Background: Given cognitive impairments and disability associated with schizophrenia (SZ) and bipolar disorder (BP), considerable interest has arisen in identifying determinants of the diseases and their features. These psychiatric conditions are known to have a substantial hereditary component to their etiology, including possible overlap in genetic contributions. Various loci have been implicated to date, based on results from large international efforts focusing on genetic determinants of SZ. **Design and Methods:** Veterans Affairs (VA) Cooperative Studies Program (CSP) Study #572 recruited and evaluated cognitive and functional capacity (FC) measures among BP and SZ patients. In conjunction with the VA Million Veteran Program (MVP), and using a customized Affymetrix Axiom genotyping array for samples from 3,388 SZ #572 patients and 195,843 MVP controls, we conducted genomewide association analyses for European Americans (1,133 cases and 146,842 controls) and African Americans (1,647 cases and 29,502 controls), separately. **Results:** One SNP in the CHD gene region (2.9E-08) had genome-wide statistical significance in the EA samples; no statistically significant results were found for the AA samples. We applied the summary statistics-based version of PrediXcan to calculate gene-level test statistics across 44 tissues and did not identify any associated genes. Using LD score regression and tissue and cell type-specific annotations (GenoSkyline-Plus), 12 out of 66 tissues/cell types showed significant enrichments in functionally annotated regions at the 0.05 significance level, with brain_inferior_temporal_lobe being most significant (0.005). Based on the comparison between summary statistics from this study and those from published GWAS of 55 diseases or traits, statistical significance was determined for genetic correlation between this current SZ GWAS and published GWASs on SZ (7.8E-55), BP (1.5E-14), and major depression (6.0E-07). In addition, statistically significant correlations were identified between CSP#572 SZ GWAS and CSP#572 BP GWAS results (1.0E-4). **Conclusion:** Although only one region showed statistically significant results in this GWAS, a strong genetic correlation was found between the current results and previously published GWASs on SZ, BP, and major depression. The lack of statistical signals is likely due to a relatively small sample size. Meta-analysis of our results with other studies may lead to novel genes associated with SZ. .

2099T

A genome-wide association study reveals a novel locus significantly associated with ADHD in African Americans and replicating in ADHD cases of European ancestry. *B. Almoguera[1], L. Vazquez[1], J. Connolly[1], T. Lingren[2], B. Namjou-Khales[2], B.C. Cobb[2], J.B. Harley[2,3], F. Mentch[1], M. March[1], P. Sleiman[1], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Center for Autoimmune Genomics and Etiology, Department of Pediatrics, University of Cincinnati, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA; 3) US Department of Veterans Affairs Medical Center, Cincinnati, Ohio USA.

Introduction: Attention-deficit/hyperactivity disorder (ADHD) is the most prevalent neurobiological disorder among children, with an incidence of 6-7% that has remained stable for decades. ADHD is highly heritable, and genetic factors may account for 75%-90% of etiology. Although several genome-wide association studies (GWASs) have been performed, no single locus has yet been reported surpassing the genome-wide level of significance. **Methods**: We performed a GWAS study where ADHD cases and controls were identified with a phenotyping algorithm that mines electronic health record (EHR) data for pertinent diagnostic (ICD9 codes) and medication information. The EHR-algorithm was validated by manual chart review of 50 predicted cases and 50 controls at the Center for Applied Genomics (CAG) and Cincinnati Children's Hospital Medical Center (CCHMC). Positive predictive values (PPVs) were calculated. Imputation was performed on the Michigan Imputation Server against release 1.1 of the Haplotype Research Consortium (HRC) reference panel. Samples were separated by ancestry into African American (AA) and European American (EA) based on the principal component analysis, which was performed with Eigenstrat 3.0. SNPTEST was used for the GWAS of the EA and AA samples, and meta-analysis was performed using METAL. **Results**: The algorithm was validated by manual chart review internally at CAG and yielded positive predictive values (PPVs) of 96% and 86% for cases and controls, respectively. Secondary validation was performed at CCHMC and the PPVs were 89% and 91%, for cases and controls. A total of 11,322 ADHD cases (N=1,652) and controls (N=9,670) were included in the analysis: 4,369 AA (N=714 cases) and 6,953 EA (N=938 cases). A GWAS of the AA sample yielded a significant association of rs2105158 in the intronic region of *SMYD3* (p=9.63x10^{-9}), which replicated in the EA sample (p=0.018). Two ethnic-specific associations, including rs114359002 at chr 12 in the region between the genes *ANO6* and *PLEKHA8P1* (p=5.88x10^{-9}) and rs189771980, downstream of *EFEMP1* (p=4.27x10^{-8}) at chr 2 were also identified in the AA cases. No significant results were uncovered in the EA sample. **Conclusions**: We have identified a novel locus at *SMYD3* that is significantly associated with ADHD in AA children and replicating in cases of EA ancestry. While further characterization of the locus is underway, the present study provides the first report of GWAS variants significantly associated with ADHD.

2100F

Using genetic diversity from East Asia to improve the biological insight into schizophrenia. *H. Huang[1,2], M. Lam[3], C. Chen[1,2], A. Martin[1,2], Z. Li[4], S. Ripke[1,2,5], M. O'Donovan[6], M. Daly[1,2], PGC Schizophrenia Asia Initiative.* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA, USA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge MA, USA; 3) Institute of Mental Health, Singapore; 4) Bio-X Institutes, Shanghai Jiaotong University, China; 5) Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Germany; 6) Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, UK.

Schizophrenia is a disabling psychiatric disorder affecting about 1% of the world population. While the biological mechanisms of schizophrenia remain largely elusive, over 100 genomic loci have been associated with schizophrenia through genome-wide association studies (GWAS). However, to date, the majority of schizophrenia genetic studies are derived from the European population, severely limiting our understanding of the schizophrenia biology. Here we present a large-scale schizophrenia genetic study in the Asian population, comprising of 13 cohorts from Singapore, Hong Kong, Taiwan, mainland China, Japan, and Indonesia. After quality control, there are 13,760 cases and 16,883 controls. We found 12 loci significantly (p < 5e-8) associated with schizophrenia, including four previously unreported loci. One of the new loci implicates *CACNA2D2*, a calcium channel auxiliary subunit associated with early infantile epileptic encephalopathy. European studies, with 3 times more samples, failed to find this association because it is extremely rare (0.7% minor allele frequency, vs. 45% in Asians). We found that genetic effects derived from the Asians are consistent with those derived from the Caucasians, indicating that schizophrenia etiology is shared across major world populations. More specifically, across the two populations, all known schizophrenia associations have the same direction of effect, and effect sizes show no significant heterogeneity. Given this observation we conducted a fixed-effect meta-analysis across the Asian and Caucasian populations, which identified over 40 new loci associated with schizophrenia. While the genetic effects of causal variants are consistent across populations, effects of non-causal variants in linkage disequilibrium with them can vary across populations owing to differences in recombination history and drift. A novel fine-mapping algorithm leveraging this observation significantly improved the association resolution. Using a published fine-mapping method (Huang et al., Nature 2017) and a Caucasian-only dataset (Ripke et al. Nature 2014), we mapped 11 schizophrenia associations to credible sets each has ≤ 5 variants. The number increased to 18 associations when we used the novel fine-mapping algorithm on a combined European and Asian dataset. This improved resolution allows us to better isolate the causal alleles and facilitate the functional interpretation of schizophrenia associations.

2101W

A missense variant in *PER2* is associated with delayed sleep phase disorder. *T. Miyagawa[1,2], M. Shimada[1,2], A. Hida[3], Y. Honda[4], K. Mishima[3], K. Tokunaga[2], M. Honda[1].* 1) Sleep Disorders Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Department of Psychophysiology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan; 4) Japan Somnology Center, Neuropsychiatric Research Institute, Tokyo, Japan.

Delayed sleep phase disorder (DSPD) is a circadian rhythm sleep disorder, and is characterized by an inability to fall asleep until very late at night and awaken at a socially acceptable morning time. The pathogenesis and genetic background of DSPD are poorly understood. Recently, several large scale GWASs of chronotype have reported genetic variants associated with variation in chronotype. The associated variants were located in or near genes with a known role in circadian rhythms. Therefore, this study was performed to identify variants associated with DSPD from known circadian genes. Genes related to circadian rhythms were selected as candidates using previous studies and pathway databases such as KEGG. We focused on infrequent and rare missense variants, which are expected to have larger effects on the phenotype. We utilized data obtained from whole-exome and whole-genome sequencing databases, such as ExAC and Integrative Japanese Genome Variation Database. Candidate variants were extracted by integrating these sequencing data and in silico assessment. DNA samples from 236 patients with DSPD and 1,436 healthy controls were genotyped to examine whether the candidate variants are associated with DSPD. As a result, a missense variant (p.Val-1205Met) in *PER2* showed a significant association with DSPD (minor allele frequency (MAF) of 2.5% in cases and 1.1% in controls, $P = 0.026$, odds ratio = 2.32). In addition, MAF of the variant in 569 patients with central disorders of hypersomnolence other than narcolepsy was also significantly higher than in 1,079 healthy controls (MAF of 1.9% in cases and 1.0% in controls, $P = 0.037$, odds ratio = 1.91). MAFs of the variant in European and African were less than 0.1%. PER2 is noted for its major role in circadian rhythms. PER2 forms a heterodimer with CRY, and the heterodimer plays an important role in the regulation of the circadian rhythm. The p.Val1205Met substitution was located in the PER2 CRY-binding domain. The substitution in *PER2* could be a potential genetic marker for circadian rhythms and sleep phenotypes.

2102T

Schizophrenia and adult height show an inverse polygenic correlation within specific functional domains of the genome. *A.P.S. Ori[1], L.M. Olde Loohuis[1], R.A. Ophoff[1,2,3].* 1) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California, USA; 3) Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands.

Many epidemiological studies have reported on an inverse relationship between schizophrenia (SCZ) and adult human height (H), two highly heritable traits. Genome-wide analysis has not provided evidence of a genetic correlation between the traits, however. To address the shared genetic architecture more systematically, we performed a genetic covariance analysis using advanced statistical methods and annotation-stratified domains across the genome. We applied GNOVA (GeNetic cOVariance Analyzer), a novel statistical framework that estimates annotation-stratified genetic covariance using GWAS summary statistics (Qiongshi et al., Biorxiv 2017). Annotation-stratified analyses can identify genetic overlap that would otherwise be missed across the whole genome. We used the most recent published GWAS summary statistics for both SCZ and H and estimated LD from 1000 Genome reference panel. We analyzed their genetic covariance (ρ) stratified by minor allele frequency (MAF) bins, the predicted functional genome, and transcriptomic and epigenomic annotations from a broad range of tissues and cell types from the GTEx and Roadmap Epigenomics consortia. Across the whole genome, we identified a significant inverse genetic relationship ($\rho = -0.0036$, r = -0.03, P = 0.03), which was not detectable using the LD score regression method (LDSC; r= 0.00, P= 0.91). We found that the covariance is concentrated in variants of lower allele frequencies (MAF bin1: 5-18%, $\rho = -0.0044$, P=0.003) that are predicted to be functional ($\rho = -0.0043$, P=0.001). Across 53 tissues of GTEx, using the subset of lower frequency variants, we observe the SCZ and H covariance in 13 tissues (P< 9.4e-4). These include 6 brain tissues, but also spleen, whole blood, and lung. Across 220 chromatin profiles, we find significant covariance in 11 profiles (P < 2.3e-4) that are mainly brain- and blood-specific cell types and tissues. In summary, we are the first to report a significant inverse genetic relationship between SCZ and H. We show that this genetic overlap is small, but concentrated in variants of lower allele frequencies and specific functional domains of the genome. Current work involves Mendelian randomization analysis in a large SCZ cohort with anthropometric data available to further dissect their pleiotropic relationship. Our work confirms observations of decades of epidemiological studies and provides an integrative framework to investigate the shared genetic architecture between two complex traits.

2103F

Copy number variation analysis of psychiatric traits in a large population-based sample of youth. *M. Zarrei[1,2], D. Merico[1,3], W. Engchuan[1,2], J. MacDonald[1], J. Wei[1], E. Higginbotham[1,2], G. Pellecchia[1], M. Faheem[1,2], A. Chan[1,2], B. Xiao[2], L. Strug[1,2,4], C. Marshall[1], C. Burton[5], J. Crosbie[5], P. Arnold[6,7], R. Schachar[5], S. Scherer[1,2,8].* 1) The Centre for Applied Genomics (TCAG), The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, the Hospital for Sick Children, Toronto, Canada; 3) Deep Genomics Inc., Toronto, Ontario, Canada; 4) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 5) Neurosciences and Mental Health, Psychiatry, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Mathison Centre for Mental Health Research and Education, Hotchkiss Brain Institute, University of Calgary, Alberta, Canada; 7) Departments of Psychiatry and Medical Genetics, Cumming school of Medicine, University of Calgary, Alberta, Canada; 8) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

'Spit for Science' is a sample of over 17,000 youth (6-17 years of age) seen at a local science museum with psychiatric phenotypes collected and DNA from saliva. Quantitative data on Attention Deficit Hyperactivity Disorder (ADHD) traits using the Strengths and Weaknesses of ADHD and Normal Behavior (SWAN) scale, Obsessive Compulsive Disorder (OCD) traits using the Toronto Obsessive-Compulsive scale and response inhibition using the Stop-Signal task (an endophenotype for ADHD) was collected. We genotyped unrelated Caucasians (n=5,636) using Illumina HumanCoreExome beadchips. Here, copy number variations (CNVs) were called from 4,815 Caucasian samples using three algorithms (PennCNV, QuantiSNP, iPattern). We defined a 'stringent' set of rare CNV (<1% population frequency) being >10kb in size and detected by a minimum of two algorithms, and performed global burden and functional gene-set burden analysis, modelled using linear regression against different traits. Overall, we identified 9,490 rare CNVs across the dataset (1.97/individual). From this, we identified sex-chromosome and autosomal aneuploidies (e.g. 45,XO, 47,XXY, 47,XXX, 47,XYY, 48,XXXY, and 47,XX,+21) in 12 (0.25%) of individuals. In 59 (1.2%) different participants, we identified CNVs affecting 23 genomic loci previously implicated in neurodevelopmental conditions (e.g. 1q21.1, 15q11.2, 16p12.1, 17p12, 22q11.2). Large CNVs affecting known psychiatric genes such *ASTN2*, *CNTN4*, *MACROD2*, *NRXN1*, and others, were also identified. ADHD traits were associated with increased CNV deletions (*p*=0.015). Although response inhibition was also associated with deletions with a borderline significance (*p*=0.072), there was an increased burden of duplication for this trait (*p*=0.004). We found no global burden of duplications or deletions for OCD traits. ADHD traits showed significant enrichment for deletions in genes associated with mouse neurophenotypes (*p*=0.01) and with gene (*p*=0.003) and protein expression (*p*=0.007) specifically in the human brain. OCD traits showed enrichment for duplications in genes specifically expressed in the brain (*p*=0.003) while response inhibition was enriched for duplications or deletions in genes specifically expressed in the brain (*p*=0.007).

2104W

Genome-wide association study of cognitive flexibility assessed by Wisconsin Card Sorting Tests. *H. Zhang[1,2], H. Zhou[3,6], T. Lencz[7], L. Farrer[2], H.R. Kranzler[8], J. Gelernter[3,4,5,6].* 1) Department of Psychiatry, Boston University School of Medicine, Boston, MA; 2) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA; 3) Department of Psychiatry, Yale University School of Medicine, New Haven, CT; 4) Department of Genetics, Yale University School of Medicine, New Haven, CT; 5) Department of Neurobiology, Yale University School of Medicine, New Haven, CT; 6) VA Connecticut Healthcare System, West Haven, CT; 7) Department of Psychiatry, Hofstra Northwell School of Medicine, Hempstead, NY; 8) Department of Psychiatry, University of Pennsylvania Perelman School of Medicine and VISN4 MIRECC, Crescenz VAMC, Philadelphia, PA.

Cognitive flexibility is a critical component of executive function strongly influenced by genetic factors. We conducted a genome-wide association study of cognitive flexibility (as measured by perseverative errors on the Wisconsin Card Sorting Test) in two sets of African American (AA) and European American (EA) samples (Yale-Penn 1: 1,411 AAs and 949 EAs, genotyped by Illumina's HumanOmni1-Quad v1.0 microarray; Yale-Penn 2: 1,178 AAs and 1,335 EAs, genotyped by Illumina's HumanCore Exome BeadChip). The R package GWAF was used to test the association of both genotyped and imputed SNPs with cognitive flexibility. In AAs, two close and linked common SNPs (rs7165213 and rs35633795) in the downstream region of noncoding gene *LOC101927286* on chromosome 15 showed genome-wide significant associations ($P<5\times10^{-8}$) with cognitive flexibility. In EAs, no GWAS associations were observed. Enriched gene sets by DEPICT analysis of top SNPs ($P_{met-analysis}<10^{-5}$) included signalosome and ubiquitin specific peptidase 9, X-linked (USP9X) subnetwork in AAs as well as abnormal frontal and occipital bone morphology in EAs. We also performed polygenic risk score (PRS) analysis to examine the genetic correlation of cognition-proxy phenotypes (general cognitive function, education attainment, childhood intelligence, and infant head circumference) and cognitive flexibility in EAs. The PRS derived from general cognitive function-associated SNPs was significantly associated with cognitive flexibility. Non-genetic factors (age, education, sex, and tobacco recency) exerted significant effects on cognitive flexibility. Our study demonstrated that both genetic and non-genetic factors impact cognitive flexibility, and variants in genes involved in protein degradation and brain development may contribute to cognitive decline.

2105T

Integrating multi-omics data to boost the translation of GWAS to biology and therapeutics for schizophrenia. *Q. Wang, R. Chen, Q. Wei, H. Yang, Y. Ji, B.S. Li.* Vanderbilt Univerisy, NASHVILLE, TN.

Genome-wide association studies (GWAS) have identified numerous variants associated with schizophrenia (SCZ). However, how to translate GWAS findings to biological mechanism, especially clinical applicability, for example the development of new therapeutics, remains challenging. Few drugs are now available for SCZ, and the drug development for SCZ, or generally psychiatric disorders, is largely stalled. The therapeutic impasse is largely due to the lack of the understanding of the biology of the diseases and therefore the lack of novel, accurate drug targets. One reason is that GWAS hits, in most cases, refer to genomic loci, not specific genes. It is thus difficult to identify the risk genes driving the significant associations without further scrutiny. It is increasingly clear that drugs that target genes with genetics evidence for a particular disease are more effective, making it extremely important to systematically identify risk genes for SCZ and other psychiatric disorders. Here we proposed a Gibbs sampling algorithm in a Bayesian framework which is capable of integrating multi-omics data, to accurately infer risk genes through which GWAS SNPs exert their functions by a sampling strategy. The proposed method is very flexible to leverage data from different relevant sources, and cumulatively orchestrate them to calculate the sampling probability for risk gene prediction. By applying the framework to the 108 SCZ loci, we predicted a set of high-confident risk genes (HRGs) that are supported by multiple lines of evidence: 1) the predicted HRGs are consistent with the leading pathophysiological hypotheses of SCZ; 2) HRGs are enriched in the *de novo* mutations (DNMs) identified in SCZ and other diseases like autism and intellectual disability, confirming the functional convergence between regulatory and coding variants, and the genetic sharing across neurodevelopmental disorders; 3) HRGs are highly expressed in brain tissues; 4) knockout of HRGs shows central nervous system (CNS) relevant phenotypes in mouse; 5) HRGs are more druggable, suggesting the potential of repositioning drugs of other diseases for the treatment of SCZ. Collectively, these results strongly demonstrated the utility of multi-omics data in interpreting GWAS and our framework is a powerful tool to translate GWAS into biology and therapeutics.

2106F

Neurodevelopmental copy number variants and clinical risk: A pediatric record population study. *K. Ahn[1], F. Mentch[2], S. An[3], C. Kao[2], H. Hakonnarson[2], J. Rapoport[1].* 1) Child Psychiatry Branch, National Inst Mental Health, NIH, Bethesda, MD; 2) Center for Applied Genomics, Children's Hospital at Philadelphia, Philadelphia, PA; 3) Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

Background: Chromosomal copy number variants (CNVs) associated with schizophrenia also carry etiological risk for neurodevelopmental disorders, including autism and intellectual deficiency. It is not known however what is their risk for any disorders. Here we applied well-replicated 'neurodevelopmental' CNVs (NRXN1 del, MYT1L dup, 15q13.3 del/dup, 16p11.2 del/dup, and 22q11.2 del/dup) and interrogated their frequency and overall clinical penetrance for broad categories of disorder in a general pediatric population. **Methods:** Eight selected CNVs were screened within a population sample of 60,644 pediatric patients. CNV carriers were examined with respect to pre-scored digitized pediatric records and compared to records for a 5:1 non-carrier control group matched for each individual CNV. Prevalence for 12 clinical categories were estimated. **Results:** Seven of these CNVs (NRXN1 dep, MYT1L dup, 22q11 dup, 16p11 del/dup, and 15q11.3 del/dup) were detected in the pediatric population at expected rates based on large control population studies. Having any CNV predicted an increase for two of the 12 disease categories after Bonferroni correction: Nervous system (p=1.43X10^{-5}) and mental/neurodevelopmental disorder (p=2.9X10^{-10}), and were associated with greater cost of medical care. Individual CNVs were examined in relation to their matched non-carriers for previously described associated disorders. As anticipated, 15q13.3 del was associated with mental/neurodevelopmental disorders, and 16p11 del was associated with congenital deficits, CNS disorders and mental/neurodevelopmental disorders. 16p11 del carriers showed more circulatory disorders than their non-carriers. Finally, in addition to association with mental/neurodevelopmental and congenital defects, a significant association was found between 22q11 dup and pediatric gastrointestinal reflux disorder (GERD) (p=0.0037) which was more likely to be present in the presence of developmental delay. **Conclusions:** These findings extend previous clinically based studies providing a high prediction of risk for 16p11.2 del, 15q13.3 del and 22q11.2 dup**.** A broader concept of overall clinical penetrance is important for understanding the pathophysiology of these disorders and may inform genetic counseling. To our knowledge, this is the first report of a CNV association (22q11 dup) for pediatric GERD, which may represent a delay in vagal nerve maturation, and first report of a 16p11 del associated with circulatory disorders.

2107W

Potential role of rare variants in the genetics of tardive dyskinesia. *A. Alkelai[1], L. Greenbaum[2,3], D. Goldstein[1], B. Lerer[4].* 1) Institute for Genomic Medicine, Columbia University Medical School, Columbia University Medical Center, New York, NY, USA; 2) The Danek Gertner institute of Human Genetics, Chaim Sheba medical center, Tel Hashomer, Ramat Gan, Israel; 3) The Joseph Sagol Neuroscience Center (JSNC), Chaim Sheba medical center, Tel Hashomer, Ramat Gan, Israel; 4) Biological Psychiatry Laboratory, Department of Psychiatry, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.

Tardive dyskinesia (TD) is a chronic, irreversible side effect of antipsychotic treatment. TD occurrence is influenced by both clinical and demographic variables, as well as genetic factors. Contribution of common variants to TD susceptibility has been investigated in recent years, including by the genome-wide association study approach, but results are inconsistent. In order to discover the involvement of rare variants in this phenotype, we implemented the whole exome sequencing method. We performed whole exome sequencing in 20 Ashkenazi and 20 Non-Ashkenazi Jewish schizophrenia patients with severe TD, and 18 Ashkenazi and 27 Non-Ashkenazi patients without any manifestation of TD (total AIMS score of zero) despite more than 10 years of exposure to antipsychotics. We performed a case-control, exome-wide collapsing analysis including 40 patients with severe TD and 45 patients without any manifestation of TD. We also performed an additional analysis including 40 patients with severe TD and 880 unrelated healthy Jewish controls. Comparing genetic variation across 18,668 protein-coding genes we found a suggestive case-enrichment of rare loss of function and deleterious non-synonymous variants in genes that seem to be relevant to this disorder. Follow-up case-control study in a larger Jewish TD sample is needed to confirm these initial results.

2108T

Exome sequencing study of bipolar disorder in a genetically isolated population. *L. Hou[1], L. Kassem[1], TG. Schulze[2], F. Lopes[1], L. Willie[1], N. Akula[1], D. Chen[1], M. Crawford[3], AR. Shuldiner[4], FJ. McMahon[1].* 1) Human Genetics Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA; 2) Ludwig Maximillians University, Munich, Germany; 3) University of Kansas, Lawrence, Kansas, USA; 4) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, New York, USA.

Bipolar disorder is a highly heritable, mental illness that affects about 2% of the US population. Genome-wide association studies (GWAS) have found many common, low-risk variants, but genetic heterogeneity has hindered the identification of high-risk alleles. We are pursuing an exome sequencing study of 1000 cases and 2500 unaffected relatives ascertained from Amish and Mennonite (Anabaptist) communities, where genetic isolation, large families, and low rates of drug and alcohol abuse may facilitate the identification of high-risk alleles. Here we present a preliminary analysis of the first 460 exomes in order to sample the genetic architecture and plan future analyses. Capture was performed with the IDT xGen Exome Research Panel v1.0, followed by sequencing by Illumina Hi-Seq. Raw reads were mapped to reference genome build 38 using BWA followed by variant calling with GATK. After sample and variant level QC, 227K single nucleotide and 8.8K indel variants were identified, of which 64K (27%) were novel. Variants that met the following criteria qualified for further analysis: shared by at least 2 cases; functional annotation as stopgain, stoploss, frameshift, or nonsynonymous in an ENSEMBL gene; <10% missing calls in cases and unaffected relatives; maximum frequency <1% (or missing) in 1000Genomes, ExAC, and esp6500 reference populations; called in at least one of two Anabaptist reference samples, with frequency <1% (or missing) in 1000 Anabaptists sequenced by the TopMED Consortium. After filtering, there were 3,789 variants in 3,100 unique genes. No variant was shared by more than 6% of cases, and there was no evidence of excess homozygosity. 110 genes carried three or more qualifying variants each, including *ANK3*, *SYNE1*, and *TENM4 (ODZ4)*. Of these genes, 29 had a >90% probability of intolerance to loss of function (lof) mutations by LoFtool. As a set, these 29 genes were enriched for ion channels and transporters by GeneMANIA. The results suggest 3 interim conclusions: 1) there is considerable genetic heterogeneity among people with bipolar disorder, even in Anabaptist communities; 2) genes with multiple otherwise rare, functional variants shared by cases include some genes previously implicated in bipolar disorder by GWAS; 3) the most lof-intolerant of these genes comprise a non-random set enriched for biological processes thought to contribute to mental illness. Future work will focus on family-based burden tests in the full sample.

2109F

Maternal antenatal depression and child socio-emotional outcomes: Investigating intervening child genetic risk for ADHD and biological pathways. *L.M. Chen[1,2,3], A-A. Bouvette-Turcot[1,2,3], E. Garg[2,3], T.T.T. Nguyen[2,3], P.P. Silveira[1,2,3], J. Diorio[2,3], M.S. Kobor[4], K.J. O'Donnell[1,2,3], M.J. Meaney[1,2,3,5], MAVAN Team.* 1) McGill University, Montreal, Quebec, Canada; 2) Douglas Mental Health University Institute, Montreal, Quebec, Canada; 3) Ludmer Centre for Neuroinformatics and Mental Health, Montreal, Quebec, Canada; 4) University of British Columbia, Vancouver, British Columbia, Canada; 5) Singapore Institute for Clinical Sciences, Singapore.

Background Maternal antenatal depression is associated with offspring socio-emotional problems. Yet, individual differences are observed, which are likely due to genetic factors. **Aim** We sought to identify genes and biological pathways that moderate the relation between maternal antenatal depressive symptoms (MADS) and child socio-emotional outcomes. **Methods** Our sample included 190 mother-child dyads. MADS was measured with the Center for Epidemiologic Studies – Depression Scale. Child socio-emotional outcomes were mother-reported at 5 years with the Child Behavior Checklist. We constructed genomic profile risk scores from the children's SNPs genome-wide that account for polygenic risk for attention deficit/hyperactivity disorder ($GPRS_{ADHD}$) based on risk information from the genome-wide associations with ADHD by the Psychiatric Genomics Consortium and applied them in our gene-environment interaction (GxE) model. SNPs constituting the best-fit $GPRS_{ADHD}$ in the GxE model were used in a genome-wide by environment association analysis follow-up. We used MetaCore to examine the enriched gene ontology in a subset of SNPs in the GxE model with p-values < .01. **Results** $GPRS_{ADHD}$ moderated the relation between MADS and child internalizing problems ($t(186)=3.22$, $p<.05$). A positive association between MADS and internalizing trait was found only in children with high $GPRS_{ADHD}$. The top significant genes underlying this moderating effect of the $GPRS_{ADHD}$ were enriched in processes such as "neuron development" ($p=1.56e-7$) and "synapse organization" ($p=5.58e-7$), in molecular functions such as "inositol 1,4,5-trisphosphate-sensitive calcium-release channel activity" ($p=3.66e-4$) and "retinoic acid receptor binding" ($p=4.49e-4$), and in pathway networks such as "synaptic contact" ($p=1.12e-4$), "synaptogenesis" ($p=2.90e-4$), "axonal guidance" ($p=1.11e-3$), "attractive and repulsive receptors" ($p=3.80e-3$), and "transmission of nerve impulse" ($p=1.60e-3$). **Conclusion** The enriched biological pathways are implicated in axon development and synaptic functions, vital in fetal brain development. Some genome-wide significant genes include *TNFRSF19*, which is essential in fetal neurodevelopment, and *CACNA1C*, which is a shared risk factor in major psychiatric disorders and is important in hippocampal neurogenesis. These findings suggest that specific genes involved in the biological framework for antenatal neurodevelopment can confer sensitivity or resilience to the influences of MADS on mental health.

2110W

Genome-editing of the *RERE* super-enhancer alters expression of genes in independent schizophrenia GWAS regions. *C. Barr[1,2], Y. Feng[1,2], A. Dineen[1,2], K. Wigg[1,2], A. Sarkar[1,2].* 1) Genetics & Development, Krembil Research Institute, UHN, Toronto, ON, Canada; 2) Program in Neurosciences and Mental Health, Hospital for Sick Children, Toronto, ON, Canada.

Background: The majority of associated markers for complex genetic traits reside in gene regulatory regions, particularly enhancers and super-enhancers. Enhancers can reside megabases from the gene they regulate (target gene) and their targets are often not the nearest gene. Thus, the assumption that the gene nearest a GWAS-significant marker will be the risk gene will in many cases be incorrect. Methods: To identify the target genes of enhancers with GWAS significant markers, we analyzed Capture-HiC data selecting enhancers for functional studies using CRISPR/Cas9 in human neural precursor cells (hNPCs). The impact on expression was measured by RT-qPCR and RNA-seq in the edited versus the mock-transfected cells. Results: We selected the super-enhancer spanning the 3' end of the *RERE* gene for study, the site of 12 GWAS significant SNPs for schizophrenia. Capture-HiC data indicate interactions of the super-enhancer with *RERE* (co-repressor/co-activator involved in retinoic acid signaling), *PARK7* (Parkinsons 7, protects neurons from oxidative stress) and *PER3* (Period 3). These 3 genes are transcription co-regulators or transcription factors. Using CRISPR/Cas9, we deleted a 2kb region of the super-enhancer in hNPCs and analyzed the transcriptome by RNA-seq. We identified 107 genes that were differentially expressed, including 14 regulated by retinoic acid. Importantly, 3 of these are located in independent GWAS regions for schizophrenia. Conclusions: Capture-HiC provides important new leads in pinpointing the target genes of enhancer-mediated regulation emanating from the GWAS findings and functional studies confirm altered expression of interacting genes. We are currently differentiating the CRISPR/Cas9 edited NPCs to neurons to examine the impact of the edits on differentiation and neuronal phenotypes. The finding of altered expression of genes in independent GWAS regions is an important new lead in understanding the regulation of schizophrenia risk genes.

2111T

Identification of ADHD risk genes in extended pedigrees by combining linkage analysis and whole-exome sequencing. *M. Klein[1], J. Corominas[1], T. Zayats[2], O. Rivero[3], G. Ziegler[3], M. Pauper[1], K. Neveling[1], G. Poelmans[1,4], C. Jansch[3], J. Geissler[5], H. Weber[6,7], A. Reif[7], A. Arias Vasquez[1,8,9], T.E. Galesloot[10], L.A.L.M. Kiemeney[10], J.K. Buitelaar[9], J.A. Ramos-Quiroga[11,12,13,14], B. Cormand[15,16,17,18], M. Ribases[11,13,19], K. Hveem[20,21], M. Gabrielsen[20], P. Hoffmann[22,23,24,25], S. Cichon[24,25,26], J. Haavik[2,27], S. Johansson[28,29], C.P. Jacob[6], M. Romanos[5], B. Franke[1,9], K.P. Lesch[3,30].* 1) Radboud university medical center, Donders Institute for Brain, Cognition and Behaviour, Department of Human Genetics, Nijmegen, The Netherlands; 2) K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway; 3) Division of Molecular Psychiatry, Clinical Research Unit on Disorders of Neurodevelopment and Cognition, Center of Mental Health, University of Würzburg, Würzburg, Germany; 4) Radboud university medical center, Donders Institute for Brain, Cognition and Behaviour, Department of Molecular Animal Physiology, Nijmegen, The Netherlands; 5) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital of Würzburg, Würzburg, Germany; 6) Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University of Würzburg, Würzburg, Germany; 7) Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt am Main, Germany; 8) Radboud university medical center, Donders Institute for Brain, Cognition and Behaviour, Department of Psychiatry, Nijmegen, The Netherlands; 9) Radboud university medical center, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, The Netherlands; 10) Radboud university medical center, Radboud Institute for Health Sciences, Department for Health Evidence, Nijmegen, The Netherlands; 11) Biomedical Network Research Center on Mental Health (CIBERSAM), Institute of Salud Carlos III, Spain; 12) Department of Psychiatry and Forensic Medicine, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain; 13) Department of Psychiatry, University Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain; 14) Psychiatric Genetics Unit, University Hospital, Vall d'Hebron, Barcelona, Spain; 15) Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Catalonia, Spain; 16) Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain; 17) Instituto de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Institute of Salud Carlos III, Spain; 18) Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, Catalonia, Spain; 19) Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain; 20) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 21) HUNT Research Centre, Department of Public Health, Norwegian University of Science and Technology, Levanger, Norway; 22) Institute of Human Genetics, University of Bonn, Bonn, Germany; 23) Department of Genomics, Life&Brain Center, University of Bonn, Bonn, Germany; 24) Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland; 25) Department of Biomedicine, University of Basel, Basel, Switzerland; 26) Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany; 27) Division of Psychiatry, Haukeland University Hospital, Bergen, Norway; 28) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 29) K.G. Jebsen Centre for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway; 30) Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, I.M. Sechenov First Moscow State Medical University, Moscow, Russia; Department of Translational Neuroscience, School of Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands.

Attention-Deficit/Hyperactivity Disorder (ADHD) is a common neurodevelopmental disorder with a complex genetic background, hampering identification of underlying genetic risk factors. We hypothesized that combining linkage analysis and whole exome sequencing (WES) in multi-generation pedigrees with multiple affected individuals can point towards novel ADHD genes. Three families with multiple ADHD-affected members (N_{total}=70) and apparent dominant inheritance pattern were included in this study. Genotyping was performed in 37 family members, and WES was additionally carried out in 10 of those. Linkage analysis was performed using multi-point analysis in Superlink Online SNP 1.1. From prioritized linkage regions with a LOD score ≥2, a total of 24 genes harboring rare variants were selected. Those genes were taken forward and were jointly analyzed in gene-set analyses of exome-chip

data using the MAGMA software in an independent sample of patients with persistent ADHD and healthy controls (N=9 365). All 24 genes together, and particularly the gene-set from one of the three families (12 genes), were significantly associated with persistent ADHD in this sample. Among the latter, gene-wide analysis for the *AAED1* gene reached significance. A rare variant (rs151326868) within *AAED1* segregated with ADHD in one of the families. The analytic strategy followed here is an effective approach for identifying novel ADHD risk genes. Additionally, this study suggests that both rare and more frequent variants in multiple genes act together in contributing to ADHD risk, even in individual multi-case families.

2112F

Limited contribution of rare, noncoding variation to autism spectrum disorder from sequencing of 2,076 genomes in quartet families. *D.M. Werling[1], H. Brand[2,3,4], J-Y. An[1], M.R. Stone[2], J.T. Glessner[2,3,4], L. Zhu[5], R.L. Collins[2,3,6], S. Dong[1], R.M. Layer[7,8], E.C. Markenscoff-Papadimitriou[1], A. Farrell[7,8], G.B. Schwartz[1], X. He[9], A.R. Kriegstein[10], J.L. Rubenstein[1], N. Sestan[11], S.A. McCarroll[4,12], B.M. Neale[4,13,14], H. Coon[15,16], A.J. Willsey[1,17], J.D. Buxbaum[18,19,20,21], M.J. Daly[4,13,14], M.W. State[1], A. Quinlan[7,8,16], G.T. Marth[7,8], K. Roeder[22], B. Devlin[23], S.J. Sanders[1], M.E. Talkowski[2,3,4,24].* 1) Department of Psychiatry, University of California, San Francisco, San Francisco, CA; 2) Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Department of Neurology, Harvard Medical School, Boston, MA; 4) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA; 5) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 6) Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, MA; 7) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; 8) USTAR Center for Genetic Discovery, University of Utah School of Medicine, Salt Lake City, UT; 9) Department of Human Genetics, University of Chicago, Chicago, IL; 10) Department of Neuroscience, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA; 11) Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT; 12) Department of Genetics Harvard Medical School, Boston, MA; 13) Analytical and Translational Genetics Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 14) Department of Medicine, Harvard Medical School, Boston, MA; 15) Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT; 16) Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT; 17) Institute of Neurodegeneration, University of California, San Francisco, San Francisco, CA; 18) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY; 19) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 20) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 21) Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 22) Departments of Statistics and Computational Biology, Carnegie Mellon University, Pittsburgh, PA; 23) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 24) Departments of Pathology and Psychiatry, Massachusetts General Hospital, Boston, MA.

Introduction: Genomic studies in autism spectrum disorder (ASD) have largely focused on *de novo* mutations that disrupt protein-coding sequence. Given the size of the noncoding genome and its regulatory role in gene function, it is likely that rare, noncoding variation represents an important component of the genetic architecture of ASD. Characterization of noncoding mutations may also highlight the cell types and critical developmental stages involved in ASD. However, there is no obvious noncoding equivalent to protein-truncating variants, thus necessitating an unbiased and statistically rigorous association strategy for noncoding variation that parallels standards set by common variant analyses. **Methods:** Using whole genome sequencing (WGS) of 2,076 individuals from 519 quartets, we evaluated the contribution of genomic variation across the spectrum of variant size and frequency. The results of 12 variant discovery algorithms were integrated to identify *de novo* SNVs, indels, and structural variants (SVs); cross-site validation exceeded 93% for all variant classes. Variants were annotated using an extensive series of noncoding functional annotations and gene sets of plausible relevance to ASD, resulting in 51,801 combinations of annotation categories. ASD association within each category was assessed using a binomial test to compare variant counts in cases and controls in a Category-Wide Association Study (CWAS). To account for multiple testing, correlations of p-values were assessed between the 51,801 categories from 20,000 sets of simulated variants. Eigenvalue decomposition estimated that 4,211 effective tests explained 99% of the variation. **Results:** A small excess of *de novo* noncoding variants was observed in cases (1.02-fold), which was not significant after correcting for paternal age. While we identified several variants that were likely causal for an individual, no annotation category was significant after correcting for 4,211 tests in the CWAS and the lowest p-values were observed in coding regions, including missense variation and SVs. Similarly, no category of rare inherited variants demonstrated parental transmission bias or ASD association. **Conclusion:** There is no clear category of rare noncoding variation with equivalent impact on ASD risk as large SVs or protein-disrupting mutations in evolutionarily constrained genes. We estimate substantially larger sample sizes will be required to quantify this risk, likely in excess of 5,000 cases.

2113W

Genetics of bipolar disorder in Cooperative Studies Program #572. *M. Aslan[1,2], N. Sun[1,3], Q. Lu[3], Y. Hu[3], B. Li[3], Q. Chen[1,3], K. Radhakrishnan[1], K.H. Cheung[1,4], Y. Li[1,5], N. Rajeevan[1,5], F. Sayward[1,5], S. Pyarajan[6,7], J.M. Gaziano[6,7], J. Concato[1,2], H. Zhao[1,3], P. Harvey[8,9] on behalf of Dept of Veterans Affairs Cooperative Studies Program (#572) & Million Veteran Program.* 1) Cooperative Studies Program, VA Clinical Epidemiology Research Center (CERC), VA Connecticut Hlthcare System, West Haven, CT; 2) Department of Medicine, Yale University School of Medicine, New Haven, CT; 3) Department of Biostatistics, Yale University School of Public Health, New Haven CT; 4) Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT; 5) Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT; 6) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA; 7) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 8) Research Service, Miami VA Healthcare System; 9) Department of Psychiatry, University of Miami School of Medicine.

Background: Given the prominence of cognitive impairments and disability associated with schizophrenia (SZ) and bipolar disorder (BP), substantial interest has arisen in identifying determinants of the diseases and their features. Both of these psychiatric conditions have long been shown to have a substantial genetic component to their etiology, including possible overlap in genetic contributions. From large international efforts to identify genomic associations for the etiology of BP, various loci have been implicated to date. **Design and Methods:** Veterans Affairs (VA) Cooperative Studies Program (CSP) Study #572 recruited and evaluated cognitive and functional capacity (FC) measures among BP and SZ patients. In conjunction with the VA Million Veteran Program (MVP), and using a customized Affymetrix Axiom genotyping array for samples from 4,663 BP #572 patients and 195,843 MVP controls, we conducted genomewide association analyses (GWASs) for European Americans (2,908 cases and 146,842 controls) and African Americans (1,030 cases and 29,502 controls), separately. **Results:** No SNPs were identified meeting genome-wide statistical significance in either the EA or AA samples. We applied the summary statistics-based version of PrediXcan to calculate gene-level test statistics across 44 tissues and did not identify any associated genes. Based on the comparison between summary statistics from this study, and those from published GWAS of 55 diseases and traits, a statistically significant genetic correlation was found between this current BP GWAS and 9 published GWASs, including SZ (3.8E-35), BP (4.5E-20), smoking (7.1E-09), and depression (1.1E-08). In addition, a statistically significant correlation (0.4554) was observed between CSP#572 SZ GWAS and CSP#572 BP GWAS results (p-value=1.02E-04). **Conclusion:** Although no region was detected showing statistically significant results from the CSP#572 BP GWAS, a strong genetic correlation was evident between the current results and previously published GWASs on SZ, BP, and depression, suggesting many shared factors with weak to moderate effects on disease risk. The lack of statistical signals is likely due to the relatively small number of case patients with the disorder. Meta-analysis of our results with other studies may lead to novel genes associated with BP.

2114T

Genome-wide associations with generalised anxiety disorder symptoms and self-reported anxiety disorder: An investigation of unique and shared genomic contributions. *K.L. Purves, J.R.I. Coleman, G. Breen, E.C Eley.* SGDP, King's College London, London, United Kingdom.

Generalised Anxiety Disorder (GAD) is a pervasive condition characterised by states of consistently high anxiety across multiple situations. GAD is highly comorbid with other Anxiety Disorders (AD's) with approximately 60% of GAD patients being additionally diagnosed with one of these disorders (Wittchen et al. 2011). In addition to phenotypic overlap, GAD shows significant genetic overlap with other AD's (Waszczuk et al. 2014). There is evidence that traits central to GAD (such as heightened anxiety sensitivity) are associated with specific genetic variants (Davies et al. 2015). A meta-analysis of genome-wide association studies of AD's identified loci associated with anxiety case status, and with a common anxiety factor score (Otowa et al. 2016). Thus, there is evidence that core components of GAD symptom spectrum are associated with variation in specific genetic loci, and that AD's share common genetic factors. To date, no studies have investigated GAD-specific genome-wide associations, nor the overlap between these and other AD's. The current study aims to analyse the association between GAD and genome-wide single nucleotide polymorphisms (SNP's), and investigate the extent to which these these associations are unique to GAD, or are shared across AD's. Analyses were undertaken in approximately 138,811 individuals from the UK Biobank. Three approaches to defining anxiety phenotypes were adopted. (1) Adjusted raw scores on the Generalised Anxiety Disorder-7 (GAD-7) were used for all participants, resulting in a continuous quantitative GAD-7 phenotype, (2) GAD case status was established using responses to items from the Composite International Diagnostic Interview – short form (CIDI - SF), and (3) AD case status was established through self-report of any life time diagnosis of any AD. Associations between adjusted GAD-7 total scores and genome wide SNP's were analysed using linear regression. Associations between GAD and AD case status and genome-wide SNP's were analysed using logistic regression. SNP heritability was estimated for all phenotypes. The correlation between SNP heritability of GAD and non-specific AD diagnoses is investigated.

2115F

Parent-of-origin and maternal effects in attention deficit hyperactivity disorder. *D. Smajlagic[1,6,8], S. Connoly[2], H. Hakonarson[3,4], I. Waldman[5], J. Elia[7], E. Heron[2], J. Haavik[8], S. Johansson[1,6,8], T. Zayats[8].* 1) KG Jebsen Center for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway; 2) Neuropsychiatric Genetics Research Group, Department of Psychiatry, Trinity College Dublin, Trinity Centre for Health Sciences, Dublin, Ireland; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 4) Department of Pediatrics, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA, USA; 5) Emory University, Atlanta, GA, USA; 6) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 7) Department of Child and Adolescent Psychiatry, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 8) KG Jebsen Center for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway.

Attention Deficit Hyperactivity Disorder (ADHD) is one of the most common and most heritable childhood-onset neuropsychiatric disorders, characterized by multifaceted genetics. To date, genetic studies of ADHD focused on additive effects only, explaining just a fraction of its heritability. Thus, we aimed at examining parent of origin effects (POE) together with maternal and additive effects, providing novel insight into the complex genetic architecture of ADHD. We compiled parent-offspring data collected through the Psychiatric Genomics Consortium and the Norwegian Mother and Child Cohort, consisting of 2060 trios and 328 duos. Additional parent-offspring data is being added. ADHD was diagnosed based on DSM-IV, ICD-10 and child behavior checklist. POE, maternal and additive genetic effects are being evaluated using multinomial modelling implemented in EMIM software. We explored our signals in the light of known imprinted genes (POE) and the largest ADHD genome-wide association study (GWAS) in children (N=17666). Gene based analyses are being performed using MAGMA software. Heritability estimates and genetic correlations of the examined effects are being calculated using LD score regression. Our preliminary results indicate the presence of non-additive genetic effects in the development of ADHD. Our preliminary strongest imprinting signal is located in *CALD1* locus (rs11980823, effect=0.77, SE=0.14, P=1.21E-07). This gene also revealed strong association signal in the previously reported large-scale childhood ADHD GWAS (rs79846815, P=2.03E-06). *CALD1* plays essential role in nerve regeneration, a function previously implicated in a number of neuropsychiatric disorders. Our preliminary gene-based analyses of the known imprinted genes revealed strong association with *TP73* locus (P=0.0034), encoding a transcription factor implicated in disorders of nervous system (e.g. neuroblastoma). Additional hits were noted in the non-coding RNA genes, adding to the recent observations in neuropsychiatric genetics of gene regulation playing a pivotal role in the development of disorders of mental health. In conclusion, this is the first and the largest genome-wide parent-offspring study in ADHD, exploring its non-additive genetic effects by detecting and distinguishing between POE (imprinting), maternal and child effects. As we increase our sample size, we will provide estimates of such effects as well as those of their heritability and genetic correlations.

2116W

GWAS to drug: PTPRD as a drug target for addictions, RLS and neurofibrillary neurodegenerations. *G.R. Uhl[1,2,3], M.J. Martinez[1,2], J. Drgonova[4], A. Suliman[5], K.C. Rice[5].* 1) New Mexico VA Healthcare System, Albuquerque NM; 2) Biomedical Research Institute of New Mexico, Albuquerque NM; 3) University of New Mexico, Albuquerque NM; 4) Center for Scientific Review, Rockville MD; 5) Chemical Biology Research Branch, NIDA, Rockville MD.

We and others have identified PTPRD in GWAS studies of addiction/quit success phenotyes, RLS and neurofibrillary neuropathologies. We have reported a) 70% differences in brain PTPRD mRNA levels in individuals with common PTPRD haplotypes; b) lack of baseline behavioral abnormalities in heterozygous PTPRD knockouts that nearly mimic these common human differences; c) striking reductions in cocaine reward in heterozygous knockout mice (assessed by conditioned place preference); d) sleep onset difficulties in PTPRD knockouts that resemble those in RLS. There are no reported PTPRD ligands. We now report synthesis of a 7-butoxy illudalic acid analog (7-BIA) based on its reported ability to inhibit activity of phosphatases from related rPTPs, PTPRS and PTPRS. 7-BIA inhibits phosphatase activity of recombinant PTPRD and PTPRS fusion proteins with 3 and 30 μM EC_{50} values, respectively, The time course of action suggests pseudoirreversible inhibition, as postulated for activities at PTPRF. Acute and repeated intraperitoneal (ip) administration to C57Bl mice resulted in no apparent compound related toxicity by behavioral, gross pathologic or histopathologic criteria. In initial results, 7-BIA pretreatments prior to cocaine conditioning sessions resulted in significant decreases in subsequent preference for places previously paired with effects of 10 mg/kg cocaine doses, supporting reduced cocaine reward from 7-BIA. These results support progress from GWAS to animal model to compounds that provide paths to lead compounds for pharmaceutical development for common, disabling disorders that lack effective pharmacotherapeutics.

2117T

Genome-wide association study of comorbid alcohol dependence and major depression. *H. Zhou[1], R. Polimanti[1], BZ. Yang[1,2], Q. Wang[3], S. Han[4,5], R. Sherva[6], YZ. Nunez[1,2], H. Zhao[3,7,8,9], L. Farrer[6,10,11,12,13], H. Kranzler[14,15], J. Gelernter[1,2,8,16].* 1) Dept of Psychiatry, Yale University, New Haven, CT; 2) VA CT Healthcare Center, West Haven, CT; 3) Program in CBB, Yale University, New Haven, CT; 4) Dept of Psychiatry, University of Iowa, Iowa City, IA; 5) Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA; 6) Dept of Medicine (Biomedical Genetics), Boston University, Boston, MA; 7) Dept of Biostatistics, Yale University, New Haven, CT; 8) Dept of Genetics, Yale University, New Haven, CT; 9) VA Cooperative Studies Program Coordinating Center, West Haven, CT; 10) Dept of Neurology, Boston University, Boston, MA; 11) Dept of Ophthalmology, Boston University, Boston, MA; 12) Dept of Genetics and Genomics, Boston University, Boston, MA; 13) Dept of Epidemiology and Biostatistics, Boston University, Boston, MA; 14) Dept of Psychiatry, University of Pennsylvania, Philadelphia, PA; 15) VISN 4 MIRECC, Crescenz VA Medical Center, Philadelphia, PA; 16) Dept of Neurobiology, Yale University, New Haven, CT.

Importance Alcohol dependence (AD) and major depression (MD) are leading causes of disability, and they often co-occur. Genetic epidemiologic data have shown that they share, at least to some extent, a common genetic etiology. The molecular nature of this shared genetic basis is poorly understood. **Objective** To detect genetic risk variants for comorbid AD and MD and to determine whether polygenic risk alleles are shared with neuropsychiatric traits or subcortical brain volumes. **Design, setting and participants** This genome-wide association study (GWAS) analyzed criterion counts of comorbid AD and MD in the Yale-Penn sample of subjects recruited for drug or alcohol dependence studies including African Americans (AA) and European Americans (EA). After excluding subjects who were never exposed to alcohol or had missing information for any diagnostic criterion, we performed GWAS on two samples ("Yale-Penn-1" and "Yale-Penn-2") totaling 4,653 AAs and 3,169 EAs. AAs and EAs were analyzed separately. We then tested whether polygenic risk scores (PRS) derived from potentially related traits (using data from SS-GAC, PGC and ENIGMA consortia) in EAs could be used to predict comorbid AD and MD. **Main Outcome(s) and Measure(s)** Comorbid criterion counts (ranging from 0-14) comprised of *DSM-IV* AD (7 criteria) and MD (9 criteria, scaled to 7). **Results** Under a linear regression model adjusted for age, sex and 10 PCs, rs139438618 at the *SEMA3A* locus was significantly associated with AD-MD comorbidity in AAs in the Yale-Penn-1 sample ($P=2.76\times10^{-8}$). In the Yale-Penn-2 sample, the association was also significant ($P=2.06\times10^{-4}$). Meta-analysis of the two samples yielded $P=2.41\times10^{-11}$. There was no significant association identified in EAs. PRS analyses showed that individuals with a higher level of neuroticism ($P=1.03\times10^{-4}$) or depressive symptoms ($P=1.80\times10^{-3}$) and a lower level of subjective well-being ($P=4.19\times10^{-4}$) and educational attainment ($P=5.13\times10^{-4}$) were at higher risk of AD+MD; while larger intracranial ($P=2.18\times10^{-4}$) and smaller putamen volumes ($P=1.22\times10^{-3}$) predicted higher risks of AD+MD. **Conclusions and Relevance** *SEMA3A* variation is significantly associated with comorbid AD and MD in AAs. Using PRS analyses, we identified pleiotropy with neuropsychiatric traits and brain volumes. Further studies are warranted to understand the biological and genetic mechanisms of this comorbidity, which could facilitate medications development and other treatments for comorbid AD and MD.

2118F

Expanding the neurological and skeletal phenotypes of individuals with *de novo KMT2A* mutations. *A.J.S. Chan[1,2], R.K.C. Yuen[1,2,3], M. Uddin[1,4], M. Zarrei[1], S. Walker[1], N. Hoang[5], I. Drmic[5], C. Cytrynbaum[2,3,6], K. Drew[1], T. Higginbotham[1,2], R. Schachar[7], M. Woodbury-Smith[8], R. Weksberg[3,6,9,10], S.W. Scherer[1,2,3,11].* 1) University of Toronto, Toronto, Ontario, CanadaThe Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Mohammed Bin Rashid University of Medicine and Health Sciences, College of Medicine, Dubai, United Arab Emirates; 5) Autism Research Unit, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Division of Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Psychiatry and Research Institute, The Hospital for Sick Children, University of Toronto, Toronto, Canada; 8) Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada; 9) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 10) Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada; 11) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

De novo loss of function (LoF) mutations in *KMT2A* have been associated with Wiedemann-Steiner Syndrome (WSS), which is characterized by hypertrichosis cubiti, dysmorphic facial features, short stature, developmental delay, and intellectual disability. Also, *de novo* LoFs in *KMT2A* have been identified in some individuals with neurodevelopmental disorders (NDDs), including autism spectrum disorder (ASD). *KMT2A* encodes a histone methyltransferase that is required for neurogenesis, early development and hematopoiesis. Here, we present the clinical and genomic characterization of four unrelated individuals with ASD, attention-deficit and hyperactivity disorder (ADHD) or WSS. By combining our four cases with a meta-analysis of *KMT2A* mutations, we found that *de novo* LoF mutations in *KMT2A* were significantly enriched in NDD individuals compared to controls (43/13,431 vs. 0/2,299, p=0.0011), suggesting *de novo* LoF mutations are highly penetrant for NDDs. By examining the phenotypes of WSS and NDD individuals with *de novo* LoF or missense variants in *KMT2A, s*eizures/epilepsy, speech and language delay/difficulties, anxiety, ADHD, obsessive compulsive symptoms, tics, oppositional defiant disorder, and Klippel-Feil Syndrome are neurodevelopmental and skeletal phenotypes that were not previously reported in WSS patients. These findings expand the neurological and skeletal phenotypes associated with *KMT2A* and suggest that these features and other NDDs are under-diagnosed in individuals with *KMT2A* mutations.

2119W

Polygenic burden analysis of longitudinal clusters of psychopathological features in a cross-diagnostic group of individuals with severe mental illness. *E.C. Schulte[1,2], I. Kondofersky[3], M. Budde[2], K. Adorjan[1,2], F. Aldinger[1,2], H. Anderson-Schmidt[4], T. Andlauer[5], K. Gade[4], U. Heilbronner[2], J. Kalman[1,2], S. Papiol[2], F.J. Theis[3], P. Falkai[1], N.S. Mueller[3], T.G. Schulze[2].* 1) Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Munich, Germany; 2) Institute for Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Munich, Germany; 3) Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany; 4) Department of Psychiatry and Psychotherapy, Medical Center Georg August University, Göttingen, Germany; 5) Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany.

Bipolar disorder (BD), schizophrenia (SZ) and schizoaffective disorder (SZA) are complex genetic disorders with largely polygenic architecture and severe and overlapping psychiatric symptoms. Stratification of cases into homogeneous subgroups across diagnoses using both psychometric and genetic information could identify individuals with higher risk for severe illness. Examined at four time points over 18-months, a subset of 198 participants (46.9±12.4 yrs; 46% female) with DSM-IV diagnoses of SZ, SZA or BD from an ongoing longitudinal cohort study (www.kfo241.de) were genotyped on Illumina's Infinium PsychArray and imputed using the 1000genomes. 67 variables from the Positive and Negative Syndrome Scale (PANSS), the Inventory of Depressive Symptoms (IDS) and the Young Mania Rating Scale (YMRS) entered cluster analyses. Longitudinal trajectories derived from abstract data dimensions computed by factor analysis for mixed data (FAMD) were used for clustering. SZ-polygenic risk scores (PRS) based on the Psychiatric Genetics Consortium 2 SZ results were tested for cluster association at 11 thresholds. Two clusters were identified in the first two dimensions: (A) individuals with continuously low scores on PANSS and IDS (70.7%) and (B) individuals with consistently high scores on PANSS and IDS (29.3%). Clusters differed significantly with regard to Global Assessment of Functioning (higher in (A); FDR-adjusted p-value=2.23x10-10), while there were no significant differences regarding sex, age, diagnosis, center, age at onset, family history, duration of illness, or association with the SZ-PRS. In this preliminary data set, longitudinal clustering identified cross-diagnostic homogeneous subgroups. Surprisingly, more severe psychopathological features were not associated with increased genetic risk burden. .

2120T

The regulatory landscape of genetic variants associated with psychiatric disorders and neurodegenerative diseases. *A. Amlie-Wolf[1,2], L. Qu[2], E.E. Mlynarski[2], C.D. Brown[1,2,3], G.D. Schellenberg[1,2,3], L.S. Wang[1,2,3].* 1) Genomics and Computational Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

   Genome-wide association studies (GWAS) have identified thousands of mostly noncoding genetic variants associated with phenotypes, but due to linkage disequilibrium (LD), these variants tag many potentially causal variants. To identify the causal variants and tissue-specific regulatory mechanisms underlying GWAS signals, we propose the INFERNO (INFERring the molecular mechanisms of NOncoding genetic variants) pipeline. INFERNO builds LD blocks from GWAS variants using 1,000 Genomes Project data. Each linked variant is then overlapped with transcription factor binding sites (TFBSs) for 332 TFs and active enhancer annotations across 239 tissues and cell types from FANTOM5 and Roadmap Epigenomics. GWAS summary statistics are used in a Bayesian model to identify colocalized GWAS and eQTL signals across 44 GTEx tissues. Tissues and cell types are grouped into 32 tissue categories to allow data integration and provide unbiased identification of the affected tissue contexts. Background sampling is performed to quantify the enrichment of enhancer overlaps in each tissue category. To characterize the landscape of regulatory genetic variation across brain-related traits, we applied INFERNO to variants associated with neurodegenerative diseases (Alzheimer's disease (AD), amyotrophic lateral sclerosis (ALS), frontotemporal dementia, corticobasal degeneration (CBD)) and psychiatric disorders (schizophrenia (SCZ), bipolar disorder, attention deficit hyperactivity disorder, major depressive disorder, and autism). We identified significant enrichments of enhancer overlaps in the brain category in CBD, the brain and nervous categories in ALS, 18 categories including the blood and brain categories in AD, and 11 categories including blood, brain, and stem cell categories in SCZ. The blood category includes immune-related cell types, providing evidence of an immune system role in AD and SCZ etiology. INFERNO identified 1,266 strongly colocalized GWAS-eQTL signals across these phenotypes, including 403 supported by variants overlapping concordant tissue enhancers with TFBS overlap and/or a high probability of underlying the colocalization signal. One AD signal has been validated by luciferase assays, with more experiments underway, and we continue to analyze the strongly colocalized signals. In summary, INFERNO provides a comprehensive and principled tool for integrating hundreds of diverse functional genomics datasets to characterize the regulatory effects of noncoding variants.

2121F

Common genetic variation contributes to risk of severe developmental disorders. *M. Niemi, H. Martin, J. Barrett, The Deciphering Developmental Disorders Study.* Wellcome Trust Sanger Institute, Hinxton, United Kingdom.

   Recent studies have conclusively implicated both rare and common genetic variation in common neurodevelopmental conditions, such as autism and schizophrenia. For rare developmental disorders, rare variants are presumed to account for the disease phenotype, and thus the contribution of common genetic variation to disease risk has not been widely studied. The Deciphering Developmental Disorders (DDD) study has recruited 13,500 children with severe developmental disorders without a previous genetic diagnosis from around the UK and Ireland. We exome sequenced the children and their parents to search for rare or *de novo* variants of strong effect, because these individuals are presumed to have Mendelian genetic disorders. To investigate the contribution of common genetic variants (MAF>0.05) to developmental disorders, we genotyped the subset of DDD patients of European ancestry whose disorder included an abnormality of nervous system physiology or morphology. Our genome-wide association analysis of 7,274 DDD patients and 13,087 population-matched controls and 728 trios found no SNPs with genome-wide significance ($p < 5 \times 10^{-8}$). This is unsurprising given the diversity of the phenotypes in the DDD cohort. However, using LD score regression, we estimated the observed-scale SNP heritability to be 0.07 (SE=0.03), which implies there are subtle polygenic effects contributing to risk of being in the neurodevelopmental DDD cohort ('neuro-DD risk'). We further found that this neuro-DD risk is significantly negatively genetically correlated with years of schooling, which is a proxy for cognitive performance ($r_g = -0.62$, $p = 6.6 \times 10^{-5}$). We also found evidence for a positive genetic correlation with schizophrenia ($rg = 0.30$, $p = 0.003$) and obesity ($rg = 0.29$, $p = 0.009$). We constructed a polygenic risk score (PRS) of neuro-DD risk, and found that DDD patients with a diagnostic exome mutation do not differ in their PRS from as yet undiagnosed patients, which suggests that polygenic effects act additively with Mendelian variants. We also find that DDD patients with autism spectrum disorder (ASD) have a higher polygenic risk for ASD ($p = 0.001$) and educational attainment ($p = 0.0005$) than other patients. Altogether these data suggest there is a polygenic component to developmental disorders in at least a subset of patients, and that there may be shared common genetic architecture among a range of neurodevelopmental disorders.

2122W

Link genetic variation to schizophrenia through cognitive and brain anatomical phenotypes. *Z. Liu[1,2], P.C. Sham[1,2], M. Li[1,2,3].* 1) Centre for Genomic Sciences, The University of Hong Kong; 2) Department of Psychiatry, The University of Hong Kong; 3) Precision Medical Genetics Lab, Zhongshan School of Medicine, Sun Yat-sen University.

Background: Schizophrenia (SCZ) is an important cause of mental disability. Although several studies have suggested associations between SCZ and its possible endo-phenotypes, the power to discover the causality is limited. It has been predicted that deficits of cognitive phenotypes may lie upstream of the liability for SCZ and brain volume changes may lie downstream of SCZ liability (Touloupoulou, T *et al*. *Mol Psychiatry*. 2015). Thus, we hypothesize that the causation of different phenotypes to SCZ show different directions in this pathway. **Methods:** In this study, we collected GWAS summary data for cognitive, brain anatomical phenotypes and SCZ. Then, we used Mendelian randomization (MR) to detect the causal effects between cognitive/anatomical traits and SCZ in a gene-centered method. For each gene region, we selected LD-pruned single nucleotide polymorphisms (SNPs) with a P-value of less than 1e-6 as instrumental variables (IVs) for each exposure trait. We further filtered these SNPs by integrating eQTL data from brain tissues. Then the corresponding effects of selected SNPs in the outcome trait were included to perform MR analysis. Next, we ran sensitivity tests (including heterogeneity test, pleiotropy test and directionality test *etc*) to keep the valid MR results. For the remaining causal relations, we investigated whether the genes, which was used for IVs selection in the MR analysis, were also responsible for the exposure/outcome traits by performing transcriptome-wide association study using summary data from GWAS and eQTL studies. **Results:** We found that causality exists between cognitive/anatomical traits and SCZ. For example, using SNPs from *ALDH2* gene whose mutation was reported to confer susceptibility to SCZ (Zheng F *et al. Cereb Cortex.* 2017), one cognitive trait (reaction time) showed significant causation on SCZ (P = 0.0004); using SNPs from *SRR* gene which was reported as the age-related oxidative stress selectively target within the hippocampus (Billard JM. *Eur J Neurosci.* 2013), SCZ showed significant causation on anatomical traits (Hippocampus, P=0.0007). **Conclusions:** Although the association of cognitive/anatomical phenotypes with SCZ has been observed in recent GWASs, our preliminary study shows that some of them are causes and some are consequences. More efforts are required to validate these causal relations between SCZ and cognitive/anatomical traits and investigate the underlying molecular basis.

2123T

DNA banking and genetic analysis of adverse drug reactions in the New Zealand healthcare setting. *M.A. Kennedy[1], S.D.S. Maggo[1], J. Foulds[2], S. Luty[2], E.W. Chua[3], S. Cree[1], K. Ton[1], Y. Liau[1], A.L. Miller[1], P. Chin[4], K. Lehnert[5], M. Doogue[4].* 1) Dept Pathology, University of Otago, Christchurch, Christchurch, New Zealand; 2) Department of Psychological Medicine University of Otago, Christchurch, New Zealand; 3) Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia; 4) Department of Clinical Pharmacology, University of Otago Christchurch and Christchurch Hospital, New Zealand; 5) Biological Sciences, Faculty of Science, University of Auckland, New Zealand.

Adverse drug reactions (ADRs) cause considerable morbidity and mortality, consume substantial healthcare resources and compromise therapeutic relationships. It is now recognised that many ADRs have significant underlying genetic factors which may be useful for risk prediction and prevention. Understanding the range of genetic variants that contribute to ADR risk is important as clinical genome sequencing becomes affordable and more widely used. One of the limiting factors in defining genetic contributors to ADRs is the availability of consented DNA samples from clinically well-defined cases. We have established a programme called UDRUGS (Understanding Drug Reactions Using Genomic Sequencing) to allow informed consent, DNA collection, storage and analysis for various adverse drug reactions in the New Zealand healthcare setting. In the pilot stages of this study we recruited~80 participants with ADRs from 12 general drug classes, through clinician referrals and specific clinical studies. We have carried out candidate gene analyses (for relevant pharmacogenes such as CYP2D6 and CYP2C19) in 25 cases, particularly those with ADRs to antidepressants, and applied whole exome sequencing to 20 cases. We consent participants for return of findings, and report these back via the referring physicians. Our hospital (Christchurch, New Zealand) recently implemented electronic prescribing and administration software (MedChart®) such that every drug prescription and every drug administration is recorded electronically and patients' ADRs are recorded. These data facilitate efficient identification of potential cases, and this software is being progressively implemented throughout New Zealand hospitals. We are now aiming to expand the UDRUGS DNA Bank and research studies on genetic factors contributing to ADRs using these electronic tools. Initial target ADRs include severe hyponatraemia due to a range of commonly used drugs, severe ADRs due to proton pump inhibitors (PPIs), and ADRs to antidepressant drugs. As appropriate, we will carry out candidate gene or whole exome/genome sequencing to identify variants that may be relevant to each ADR, and contribute samples to genome-wide association studies. The establishment of this ADR biobank provides a resource to support pharmacogenomic research specific to the New Zealand population, and will enable contribution to international collaborative research in this area.

2124F

Association of *HTR3C*, *HTR3D*, *HTR5A*, and *HTR6* gene polymorphisms with treatment response to risperidone in Chinese schizophrenia patients. *S. Qin, W. Zhou, L. He.* Bio-X Institutes, Shanghai Jiaotong University, Shanghai, Shanghai, China.

Previous assessments of the pharmacological profile of antipsychotics have shown that the serotonin receptor family is involved in their pharmacodynamics. The current study was designed to evaluate the effect of serotonin receptor polymorphisms on risperidone response in Han Chinese patients with schizophrenia. Two independent sets of Han Chinese in-patients with schizophrenia (N = 201) were followed for a 4-week risperidone monotherapy treatment. Psychopathological symptoms were assessed using the Positive and Negative Syndrome Scale (PANSS) and response was defined as a reduction of 50% or more of PANSS scores. Fourteen single nucleotide polymorphisms (SNPs) selected from the *HTR3C*, *HTR3D*, *HTR5A*, and *HTR6* genes were genotyped. A significant association between rs1467257 and the rs939334–rs1467257–rs6792482 haplotype in the *HTR3D* gene and risperidone efficacy was detected (for rs1467257, allele: odds ratio [OR] = 2.51, 95% confidence interval [CI] = 1.26–5.02, P=0.008; genotype: P=0.03; for the haplotype, P=0.035, OR = 2.299, 95% CI = 1.046–5.056). Furthermore, the rs6801865 polymorphism in the *HTR3C* gene significantly influenced total (P=0.018) and negative (P=0.032) PANSS score improvement during risperidone treatment. These findings suggests that variants in *HTR3C* and *HTR3D* may act as biomarkers to predict risperidone response in Han Chinese populations. The current study has clinical significance in evaluating the response to risperidone.

2125W

Risperidone-induced multi-dimensional phenotypic alteration in first-episode drug-naive schizophrenia patients: A longitudinal study of DNA methylation and neurophysiological phenotyping. *C. Chen[1], Y. Xia[1], M. Hu[2], X. Zong[2], S. Jiang[3], Z. Li[1], Y. He[2], J. Bishop[3], C. Liu[3], X. Chen[2].* 1) State key lab of medical genetics, Central South University, Changsha, China; 2) Department of Psychiatry, Second Xiangya Hospital, Central South University, Changsha, China; 3) Department of Psychiatry, University of Illinois at Chicago, Chicago, IL.

Schizophrenia (SCZ) is a complex neuropsychiatric disorder and whether the biological alterations are due to chronic illness or to antipsychotic treatment effects remains to be elucidated. In this study, we performed intensive phenome-wide assessments for 38 Han Chinese drug-naïve, first-episode SCZ patients (FESPs) at baseline and after eight weeks of risperidone monotherapy and for 38 matched healthy controls. DNA methylation (DNAm) of peripheral blood was measured using Illumina 450K BeadChip. We calculated brain anatomical network using diffusion tensor imaging and spontaneous brain activity using functional magnetic resonance imaging. Cognitive function and FESPs' clinical states were also evaluated. We first conducted the epigenome-wide association study for FESPs and controls, next combined differential DNAm and co-methylation connectivity to explore significant connectivity and DNAm changes during treatment. Finally, we linked DNAm changes with all other phenotypes and constructed phenome networks. Three independent studies were used as validation. We found 8326 CpGs were differentially methylated between FESPs and controls, 9428 CpGs between post-treatment patients and controls, and 5904 CpGs in pre- versus post-treatment patients (P<0.05). Based on 5904 CpGs we detected 200 differentially methylated regions mapped to 134 Genes. These genes enriched for calcium signaling pathway and long-term depression (adjusted P<0.05). We found 586 CpGs showed differential DNAm between FESP and controls also showed differential between pre- and post-treatment, 583 of them were normalized, and 392 of these changes correlated with PANSS-positive symptoms and cognitive function as well as the anatomical brain network and spontaneous brain activity alterations. In the co-methylation networks, we extracted the CpGs with significant connectivity changes, which mapped to 831 unique Genes enriched for negative regulation of GTPase activity. Finally, using CpG and brain phenotypes network, we observed the most connected CpG site, cg00478599, correlated with the white matter connectivity in amygdala, ACG and PCG, and the white matter connectivity in these regions have correlated with spontaneous brain activity in different brain regions. Overall, we found risperidone not only normalized DNAm but also linked to network connectivity changes in patient blood. Our complex web analysis depicts relationships among multi-dimensional variables induced by treatment.

2126T

Genetic and functional analysis of the GRIN2C gene as a candidate gene of schizophrenia. *M. Cheng, S. Tsai, H. Tsai, S. Hsu.* Department of Psychiatry, Yuli Branch, Taipei Veterans General Hospital, Hualien County, Taiwan.

Schizophrenia is a chronic debilitating mental disorder with a high genetic component in its etiology. Hypofunction of the *N*-methyl-*D*-aspartate (NMDA) receptor-mediated signal transduction has been implicated in the pathophysiology of schizophrenia. The *GRIN2C* gene encodes a subunit of the NMDA receptor, which is a subtype of ionotropic glutamate receptor involves synaptic function, learning, memory, and neurodevelopmental disease. The study aimed to investigate the involvement of the *GRIN2C* gene in the genetic susceptibility to schizophrenia. We resequenced the exon regions of the *GRIN2C* gene in 1000 subjects, including schizophrenic patients and normal control, both being Han Chinese from Taiwan. We identified 11 missense mutations, including seven schizophrenia-specific mutations. Bioinformatic analysis showed some of these mutations were damaging or pathological to the protein function. Also, we detected one schizophrenia-specific nonsense mutation (c.3355C>T), which results in a premature stop codon, and in a truncated nonfunctional the GRIN2C protein. Taken together, multiple rare missense and nonsense mutations in the *GRIN2C* gene contribute to the pathophysiology of schizophrenia in some patients, but functional gene assays are needed to verify their relevance to the pathogenesis of schizophrenia. Our study supports common disease and rare variants hypothesis for schizophrenia.

2127F

Major depressive disorder and nausea and vomiting during pregnancy. Shared genetic factors? *L. Colodro Conde[1], B. Couvy-Duchesne[2], P. Lind[1], J. Painter[1], M. Wright[2], G. Montgomery[2], D. Nyholt[3], S.E. Medland[1].* 1) QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia; 2) University of Queensland, Brisbane, QLD, Australia; 3) Queensland University of Technolog, Brisbane, QLD, Australia.

Nausea and vomiting in pregnancy (NVP) affects ~70% of pregnant women to different degrees. Around 14% of pregnant women experience severe NVP and this progresses to hyperemesis gravidarum (HG) in around 1-3%. Women who experience severe NVP and HG have higher rates of major depressive disorder (MDD) prior to the index pregnancy. We therefore hypothesize that there is a genetic correlation between NVP/HG and depression. To test this hypothesis, we examined the extent to which individual differences in the occurrence of NVP, HG and the severity of NVP can be predicted by a polygenic risk score (PRS) derived from the most recent Psychiatry Genomics Consortium MDD analyses (PGC-MDD2, leave QIMR out). Our preliminary analyses using data on NVP were obtained from two studies undertaken at the QIMR Berghofer (Australia), (N= 1,440 unrelated women). The phenotypes were: 1) presence of NVP for 7 days or more, 1) degree of severity of NVP (5 point scale), and 3) presence of severe NVP, characterized by disruption of daily routine, prescription of medication, loss of weight, and intravenous feeding (proxy for HG). Polygenic risk scores were calculated using the PLINK profile score method for clumped SNPs (thresholds of p ≤ 0.5*10-8, 0.1*10-5, 0.001, 0.01, 0.05, 0.1, 0.5 or 1). Logistic and linear regressions on the profile scores were performed, controlling for ancestry, age and age2 at survey time, cohort, and if it was a twin pregnancy. Among the women from our sample, in at least one of their pregnancies, 53.1% had symptoms of NVP for 7 or more days and 18.7% had severe symptoms. After fitting covariates, the MDD PRSs (independent of the time at which any depressive symptoms manifest) predicted the probability of presenting with NVP, the degree of severity (all PRS under threshold <0.001), and very severe NVP or HG (all PRS under threshold 0.05) at a nominally significant level. However, only MDD-PRS< 0.1 predicting the degree of severity of NVP survived multiple correction testing. Further genotyping of an HG case-control cohort is currently underway and we expect the increase in power will increase the support for these findings and allow the calculation of genetic correlations using LD score regression. We will also present an update to the GWAS meta analysis for NVP.

2128W

Whole genome sequence association analysis of tobacco use in the Trans-Omics for Precision Medicine Whole Genome Sequencing Program (TOPMed). *G. Datta[1,2], R. Wedow[1,22], M. Liu[1,2], Y. Jiang[13], S. David[10], L. Emery[3], S. Gharib[5], D. Glahn[6,21], M. Hall[11], J. He[7], J. Hokanson[8], S.J. Hwang[9], A. Justice[12], C. Laurie[3], D. Levy[9], D. Liu[4], L. Martin[14], A. Pandit[15], E. Schmidt[15], R. Reed[16], A. Shadyab[17], W. White[18], L. Yanek[19], K. Young[12], W. Zhao[20], G. Abecasis[15], K. North[12,23], S. Vrieze[1,2], GWAS and Sequencing Consortium of Alcohol & Nicotine, Trans-Omics for Precision Medicine Consortia.* 1) Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO; 2) Department of Psychology, University of Minnesota, Minneapolis, MN; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Department of Public Health Sciences, Institute for Personalized Medicine, the Pennsylvania State University College of Medicine, Hershey, PA; 5) Computational Medicine Core at Center for Lung Biology, Division of Pulmonary & Critical Care Medicine, University of Washington, Seattle, WA; 6) Department of Psychiatry, Yale University School of Medicine, New Haven, CT; 7) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 8) Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO; 9) Population Sciences Branch, Division of Intramural Research, NHLBI, NIH, Bethesda MD; 10) Department of Medicine, Stanford University, Stanford, CA; 11) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 12) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 13) College of Medicine, the Pennsylvania State University College of Medicine, Hershey, PA; 14) Division of Cardiology, School of Medicine, George Washington University, Washington, DC; 15) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 16) Division of Pulmonary and Critical Care Medicine, University of Maryland School of Medicine, Baltimore, MD; 17) Division of Epidemiology, Department of Family Medicine and Public Health, University of California, San Diego School of Medicine, San Diego, CA; 18) Undergraduate Training and Education Center - UTEC, Tougaloo College, Jackson, MS; 19) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 20) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 21) Olin Neuropsychiatric Research Center, Hartford, CT; 22) Department of Sociology, University of Colorado Boulder, Boulder, CO; 23) Carolina Center for Genome Sciences, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC.

Genome-wide array-based analyses of tobacco use have identified many associated common variants, but the vast majority of the heritability remains unexplained. Rare variants likely contribute to risk for smoking and other addictive behaviors, and cost-effective whole genome sequencing is now making genome-wide analyses of rare variants possible. The present study reports rare variant association results for tobacco use from TOPMed, which has at present called genotypes from 9,536 30x whole genome sequences in 8 distinct studies with smoking phenotypes. We conducted a genome-wide association study (GWAS) of all discovered variants with minor allele count > 10 for four phenotypes: regular vs never smoker (n=9,536;n variants=34,115,958), current vs former smoker (n=4,601;n variants=27,243,212), age of initiation of regular smoking (n=4,467;n variants=27,141,662) and cigarettes smoked per day (n=4,590;n variants=27,547,744). The GWAS yielded no significant results after Bonferroni correction per phenotype, although there was a signal suggested for the known variant rs16969968 in *CHRNA5* (p=7.718e-05). Rare variant burden tests of highly deleterious variants also yielded no significant results. Finally, we used the TOPMed sequences to fine map 182 recently discovered smoking-related loci--defined as 1MB regions around top associated variants--from array-based GWAS meta-analyses across all four smoking phenotypes (N ranging from 256,658 to 991,257 depending on the phenotype). Of the 182 most significantly associated common variants within these 182 loci, two were significant after Bonferroni correction for 182 tests (rs56113850 in *CYP2A6*, p=7.3e-5 and rs112878080 in *CHRNA3*, p=2.6e-5), and 121 of the 182 variants showed the same direction of effect in TOPMed and GSCAN (p=5.1e-6). Forward selection fine mapping tests in TOPMed within the 182 loci discovered one additional conditionally independent variant in an intron of *ASXL1* (rs148986225, MAF=0.003, p=1.7e-7) associated with cigarettes smoked per day after Bonferroni correction per phenotype. We expect to present updated results on a much larger data freeze approaching 50,000 total individuals, greatly increasing power and precision of all analyses.

2129T

Genetics of cognitive function in schizophrenia and bipolar disorder. *P.D. Harvey[1,2], N. Sun[3,4], Q. Lu[4], B. Li[4], Q. Chen[3,4], M. Aslan[3,5], K. Radhakrishnan[3], K.H. Cheung[3,6], Y. Li[3,7], N. Rajeevan[3,7], F. Sayward[3,7], S. Pyarajan[8,9], J.M. Gaziano[8,9], J. Concato[3,5], H. Zhao[3,4], U.S. Department of Veterans Affairs Cooperative Studies Program (#572) and Million Veteran Program.* 1) Miami VA Healthcare System, Miami, FL; 2) University of Miami, Miami, FL; 3) VA Clinical Epidemiology Research Center (CERC), VA Connecticut Healthcare System, West Haven, CT; 4) Department of Biostatistics, Yale University School of Public Health, New Haven CT; 5) Department of Medicine, Yale University School of Medicine, New Haven, CT; 6) Department of Emergency Medicine, Yale University School of Medicine, New Haven, CT; 7) Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT; 8) Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA; 9) Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Background: Cognitive impairment is a frequent and serious problem in patients with various forms of severe mental illnesses, including schizophrenia (SZ) and bipolar disorder (BP). Cognitive performance in patients with bipolar disorder or schizophrenia has been found to be very similar in underlying structure, and recent research has suggested genomic linkage to several different cognitive phenotypes. In addition, recent evidence from the Cognitive Genomics Consortium (COGENT) study, focusing on a pooled population of more than 35,000 healthy European ancestry patients from 24 studies, reported an estimated SNP heritability of 21.5%. The current study examined the genomic correlates of cognitive and functional skills performance in 8,051 U.S. veterans with either SZ or BP. **Design and Methods:** Veterans Affairs (VA) Cooperative Studies Program (CSP) Study #572 evaluated cognitive and functional capacity (FC) measures among BPI and SZ patients, using a subset of tests from the MATRICS Consensus cognitive battery (MCCB) and the UCSD Performance-based skills assessment-brief version (UPSA-B). In conjunction with the VA Million Veteran Program, and using a customized Affymetrix Axiom genotyping array for samples from 3,388 SZ and 4,663 BP CSP #572 patients (N=8,051 total), we conducted genome-wide association analyses for European Americans (3,952) and African Americans (2,598) separately. **Results:** There are no SNPs with genome-wide statistically significant results for either EA or AA samples. We also applied MetaXcan to calculate gene-level test statistics across 44 tissues and did not identify any associated genes. However, based on the comparison between summary statistics from this study and those from published GWAS of 55 diseases/traits, there is statistically significant evidence of genetic correlation between cognitive function and schizophrenia (1.7E-09), but not bipolar illness. **Conclusion:** No SNPs with strong evidence of association with cognitive performance were identified in this study, but genetic correlation analysis between this study and published GWAS on schizophrenia from other studies suggests shared factors between cognitive functions and schizophrenia. Later analyses will perform targeted examination of the linkage of previously identified genes and SNPs to overall performance and specific cognitive phenotypes.

2130F

Transcriptional signatures of childhood onset schizophrenia in hiP-SC-derived NPCs and neurons are concordant with signatures from post mortem adult brains. *G.E. Hoffman[1,4], B.J. Hartley[3,5], E. Flaherty[2,5], I. Ladran[3,5], P. Gochman[6], D. Ruderfer[1,4], J. Rapoport[6], P. Sklar[1,3,4,5], K.J. Brennand[1,2,3,4].* 1) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 4) Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 5) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 6) Childhood Psychiatry Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA.

Large-scale genome wide association studies (GWAS) revealed that schizophrenia (SZ) risk reflects both rare and common variants. Whereas highly penetrant rare variants have proven well-suited to human induced pluripotent stem cell (hiPSC)-based models, the power of hiPSC-based studies to resolve the much smaller effects of common variants within the size of cohorts that can be realistically assembled remains uncertain. Childhood-onset-SZ (COS) patients have both a higher rate of rare SZ-associated copy number variations (CNVs), and also stronger common SZ polygenic risk scores. We developed a case/control hiPSC cohort (12 individuals with COS and 10 controls) with which to conduct a comprehensive evaluation of global gene expression in hiPSC-derived neural progenitor cells (NPCs) and neurons. Although we identified many sources of variation across our 94 RNAseq samples, this was partially addressed by establishing a rigorous series of bioinformatic practices to reduce and assess the variance inherent in this approach. We report that the donor-specific signal is enriched for post-mortem brain expression quantitative trait loci (eQTLs), but also that there is a significant correlation between the case/control differential expression observed in our hiPSC-derived COS neurons and the CommonMind Consortium post-mortem study. While the issue of small sample size is shared between post-mortem and hiPSC-based studies, and may be exacerbated in hiPSC-based experiments through the variability that arises as a result of the reprogramming and neural differentiation processes, we find that the common variation effects predicted in GWAS and detected in much larger post-mortem analyses are reflected in hiPSC-neurons. We predict a growing convergence of common variation findings between hiPSC and post-mortem studies as both approaches expand to larger cohort sizes.

2131W

Copy number variation in Thai individuals with schizophrenia and schizoaffective disorder. *N. Jinawath[1,2], W. Thammachote[1], S. Promsri[1], N. Jantaratnotai[3], D. Wattanasirichaigoon[4], R. Kongsakon[5].* 1) Program in Translational Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Bangkok, Thailand; 2) Integrative Computational Bioscience Center, Mahidol University Salaya Campus, Nakhon Pathom, Thailand; 3) Department of Pharmacology, Faculty of Science, Mahidol University, Thailand; 4) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 5) Department of Psychiatry, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand.

Schizophrenia is a chronic and severe neuropsychiatric disorder presented in approximately 1% of the general population. Patients with schizophrenia may lose touch with reality, exhibit emotional and behavioral disruptions, and impaired thinking and memory. Schizoaffective disorder is a rarer entity with more pronounced mood disorder symptoms. These mental disorders are multifactorial illnesses and share some genetic etiologies including copy number variations (CNVs) with other neuropsychiatric and neurodevelopmental disorders such as bipolar disorder, intellectual disability, and autism spectrum disorder (ASD). The aims of this study were to identify novel schizophrenia and schizoaffective disorder-associated CNVs and to evaluate the diagnostic yield of chromosomal microarray (CMA) in Thai patients with these disorders. Illumina CytoSNP-850K BeadChip was used to identify CNVs in the 67 patients who met the DSM-IV-TR criteria for schizophrenia (46 patients) and schizoaffective disorder (21 patients). Multiple large curated CNV databases including Thai CNV database were utilized for interpretation of clinical significance. We detected pathogenic CNVs and variants of uncertain significance (VOUS)-likely pathogenic CNVs in 5 of 67 patients (7.5%). Of the five cases, four had known schizophrenia-associated alterations including 15q13.3 microduplication, *NRXN1* exon deletion, 16p13.11 microdeletion, and triple X syndrome. Another case had 10q26 microdeletion syndrome, a syndrome with no previous link to schizophrenia. Interestingly, we observed a significantly high frequency of 15q13.3 duplications involving *CHRNA7* in Thai schizoaffective patients (2 of 21 patients, 9.5%), while none was seen in the 46 schizophrenia patients and 114 ASD patients previously studied. This duplication is known to exhibit incomplete penetrance and variable expressivity. Its frequency in the 3,017 general Thai individuals is 0.36%. In conclusion, this is the first SNP-based microarray study in a cohort of Thai patients with schizophrenia and schizoaffective disorder. Our findings provide supportive evidence that CMA should be considered as a first tier genetic analysis for patients with schizophrenia. Additionally, CNV analysis using ancestry-matched controls in underrepresented population is a valuable tool for identification of novel disease-associated alterations and phenotypes, thus expanding both the genetic and phenotypic spectrum of the disorders.

2132T

Sex-stratified analysis of obsessive-compulsive disorder reveals minor differences in genetic architecture. *E.A. Khramtsova[1,2], L.K. Davis[3], B.E. Stranger[1,2,4], P.G.C. OCD Workgroup[5].* 1) Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, IL; 2) Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL; 3) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 4) Center for Data Intensive Science, The University of Chicago, Chicago, IL; 5) Psychiatric Genomics Consortium Obsessive-Compulsive Disorder Workgroup.

Obsessive-compulsive disorder (OCD), demonstrates sexual dimorphism in age of onset and clinical presentation, suggesting a possible sex difference in genetic architecture. We present the first genome-wide assessment of sex-specific genetic architecture of OCD on currently the largest OCD cohort available from the Psychiatric Genomics Consortium (N=9870 after QC, 1:1.4 male/female ratio). First, we used heritability (h^2) analysis to test for evidence of variable liability threshold for OCD between sexes and to assess the proportion of overall OCD h^2 explained by the X chromosome. There were no significant differences between sexes in OCD h^2. The X chromosome contributed 6.7% to total h^2 which is not statistically different from expectation. The genetic correlation between sexes is high: GCTA GREML (1.0, se=0.27) and LDSC (1.04, se= 0.51, p=0.0405). Second, we performed a sex-stratified meta-GWAS to identify specific autosomal and sex chromosome risk factors with different effects in sexes. There were no genome-wide significant associations in either sex. Third, we assessed whether heterogeneous OCD risk alleles and top ($p<10^{-3}$) sex-specific genome-wide associations are involved in gene regulation to elucidate the biological mechanisms by which those variants may impact dimorphism. On a genome-wide level, heterogenous SNPs (including and excluding SNPs in the HLA region) were strongly enriched for immune expression quantitative trait loci ($p<0.001$); top female associations were significantly enriched for brain eQTLs (p=0.003), while top male associations were significantly enriched for immune eQTLs (p=0.003). Using Disease Association Protein-Protein Link Evaluator we identified 10 genes implicated by heterogeneous OCD risk brain eQTLs (FDR<0.05) that showed significantly more network connectivity than expected by chance. Among them are two solute carrier family transporters: SLC8A2 (sodium/calcium exchanger) which is differentially expressed between males and females in the cerebellum (q=0.027), and SLC25A1 (calcium-binding mitochondrial protein carrier), a polymorphism in which has been associated with autism and global cerebral hypomyelination. In conclusion, we identified minor sex differences in genetic architecture of OCD, and although the sex-stratified sample size is likely too small to identify variants with small effect, we have developed a pipeline for sex-stratified genetic analysis which will be applied when a larger OCD cohort is available.

2133F

The transcriptional consequences of bipolar disorder polygenic risk and medication use. *C.E. Krebs[1], A.P.S. Ori[2], L.M. Olde Loohuis[2], M.P. Boks[3], R.S. Kahn[3], R.A. Ophoff[1,2,3,4].* 1) Department of Human Genetics, UCLA, Los Angeles, CA; 2) Center for Neurobehavioral Genetics, UCLA, Los Angeles, CA; 3) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Department of Psychiatry and Biobehavioral Sciences, UCLA, CA.

Bipolar disorder (BPD) is a debilitating psychiatric disorder that is highly heritable and polygenic. Genome-wide association studies of BPD have yielded a number of associated loci, yet little is known about the disorder's etiology. We sought to characterize the functional transcriptomic impact of BPD polygenic risk and disease state in a large cohort from The Netherlands. We collected genotyping and RNA sequencing data from peripheral whole blood as well as extensive phenotypic information, including medication use, from 240 BPD cases and 240 controls. We aligned the RNA-seq reads with TopHat2, quantified gene counts using HTSeq, and normalized, voom transformed, and corrected for known contributing technical and biological covariates using limma. Of note, medication use and specifically the current use of lithium, which is the most commonly prescribed mood stabilizer for BPD patients, contributed significantly to variation in gene expression. With these corrected gene counts we performed several analyses to assess the transcriptional consequences of BPD genetic risk and disease state in our cohort. Using limma we identified five differentially expressed genes (DEGs) at FDR < 0.05 between cases and controls, 641 DEGs between cases currently taking lithium and cases not currently taking lithium, and 2,631 DEGs between cases currently taking lithium and controls. Altogether these DEGs had small effect sizes (log2 fold change mean = 0.13, max = 1.33) in line with the complex nature of BPD. Interestingly, among the five case-control DEGs is *DOCK3*, which exhibits reduced expression in cases and is well characterized as inducing axonal outgrowth in the central nervous system. The lithium DEGs were enriched for biological terms related to apoptosis and cell death, which are known cellular consequences of lithium treatment. Gene-set analysis using MAGMA showed no enrichment of BPD-associated genetic risk in the DEGs we identified. We discovered BPD-specific eQTLs using Matrix-eQTL after additionally correcting for hidden covariates using PEER, but none remained significant after FDR correction. Overall our analyses point to a lack of transcriptomic consequences of BPD polygenic risk in blood but they do reveal a strong effect of medication use.

2134W

Rare heterozygous mutation in glutamate receptor gene segregating in a schizophrenia family. *P. Kukshal[1], J. John[1], U. Bhattacharya[1], T. Bhatia[2], V.L. Nimgaonkar[3], S.N. Despande[2], B.K. Thelma[1].* 1) Department of Genetics, University of Delhi, South Campus, New Delhi, Delhi, India; 2) Department of Psychiatry, Dr. RML Hospital, New Delhi, India; 3) Departments of Psychiatry and Human Genetics, Western Psychiatric Institute and Clinic, University of Pittsburgh , Pittsburgh, USA.

Objectives: Schizophrenia (SZ) is a common complex disorder affecting 1% of the population worldwide and gene-environment interactions are known to be the etiological factors. With the advent of Next Generation Sequencing (NGS) tools, and the discovery of rare variants and *de novo* mutations, the paradigm has shifted to common disease rare variant. Recently a large number of *de novo* mutations have been identified in pathways of brain development. It is hypothesised that many neural circuits are important in schizophrenia and subject to spontaneous genetic mutations. Employing NGS tools on familial cases of schizophrenia would facilitate identification of rare putative causal variants in known or novel candidate genes. In this study we analysed a schizophrenia family with multiple member affected SZ family by Whole Exome Sequencing (WES) approaches. Methods: An extended family with three affected members and a probable autosomal dominant mode of inheritance was recruited from Dr R.M.L Hospital, New Delhi, following institutional ethical committee clearance and informed consent. Diagnosis for SZ was confirmed using Hindi version of Diagnostic Interview for Genetic Studies and Family Interview for Genetic Studies. WES of three affected and one unaffected member were captured using Agilent V5+UTR capture kit and sequenced on an Ilumina Hiseq 2000 platform. WES processing, variant calling and variant annotations were performed using bioinformatic tools and software. The variants were prioritised based on the likely mode of inheritance, functional significance and rarity. **Results:** Of the several variants found most promising was a rare heterozygous missense variant in exon 3 (rs528711346) in *GRIN3B* (Glutamate Ionotropic Receptor NMDA Type Subunit 3B). *In silico* analysis predicted the variant to be damaging. **Conclusion:** Common SNPs in this gene has been previously reported to be associated with SZ. However, rare putative causal variants identified in this gene and segregating with SZ phenotype in the family reiterates the role of this gene in disease etiology.

2135T

Large meta-analysis of Scandinavian exome sequencing studies of schizophrenia. *F. Lescai[1,2,3], F.K. Satterstrom[4,5], T.D. Als[1,2,3], J. Grove[1,2,3,8], J.B. Maller[4,5], J. Grauholm[2,9], C. Stevens[4,5], M. Mattheisen[1,2,3], R. Walters[4,5], J.I. Goldstein[4,5], P.F. Sullivan[6,7], S.A. McCarroll[4,5], D.M. Hougaard[2,9], T.M. Werge[2,10,11], B.M. Neale[4,5,12], M.J. Daly[4,5,12], A.D. Børglum[1,2,3], iPSYCH-Broad Consortium.* 1) Biomedicine, Aarhus University, Aarhus, Denmark; 2) iPSYCH, The Lundbeck Initiative for Integrative Psychiatric Research, Denmark; 3) iSEQ, Centre for Integrative Sequencing, Aarhus, Denmark; 4) Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 6) Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA; 7) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 8) BiRC, Bioinformatics Research Centre, Aarhus University, Denmark; 9) Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 10) Mental Health Centre Sct. Hans, Institute for Biological Psychiatry, Capital Region of Denmark, Roskilde, Denmark; 11) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 12) Department of Medicine, Harvard Medical School, Boston, MA, USA.

The current data indicated that the allelic architecture of complex traits is caused by many common and rare variants. Large GWAS meta-analyses have been successfully applied in the search for common variants affecting the risk of developing psychiatric disorders, in particular schizophrenia and major depressive disorder. However, these studies are designed to examine only the common variant proportion of the genomic landscape whereas high-throughput sequencing allows direct examination of both common and rare alleles. Schizophrenia is one of the most disabling mental disorders; it has a lifetime prevalence of around 1% and a heritability estimated as high as 80%. Although around one third of the heritability can be attributed to common variants, prior studies indicate that rare variants also contribute to its etiology. A collaboration between iPSYCH (The Lundbeck Initiative for Integrative Psychiatric Research, Denmark) and the Broad Institute (in Cambridge, MA, USA) is conducting large scale exome sequencing of major psychiatric disorders, including schizophrenia. DNA is extracted from dried blood spots (DBS) stored at the Danish Neonatal Screening Biobank. The samples can be matched to a wide range of phenotypic information available in Danish registries. We present here the analysis of a large collection of schizophrenia individuals recruited in clinical centers (1,866) and from DBS samples (3,710): a total of 5,576 cases and 7,468 controls from Denmark. We follow up in a similarly sized sample from Sweden with 4,969 cases and 6,245 controls. The combined analysis represents the largest exome sequencing study in schizophrenia to date. We have found evidence showing a higher per-person rate of disruptive ultra-rare loss-of-function variants located in constrained genes among schizophrenia cases compared to controls (combined p-value 1.6e-07). Meta-analysis SKAT of rare and ultra-rare variants at gene level has been performed. The results from burden and enrichment analysis, as well as ranking of mutational load, will be discussed in the larger perspective of findings from both GWAS and other exome analyses in schizophrenia.

2136F

Increased predicted *C4A* expression is associated with cognitive deficit in both schizophrenia and Alzheimer's disease. *N.S. McCarthy[1,2,3], S.M. Laws[3,4,5], T. Porter[3,4], S.C. Burnham[6], E.K. Moses[1], A. Jablensky[2].* 1) The Centre for the Genetic Origins of Health the Disease, School of Biomedical Sciences, University of Western Australia and Curtin University, MRF Building, 50 Murray St, Perth, WA 6000, Australia; 2) The Centre for Clinical Research in Neuropsychiatry, School of Psychiatry and Clinical Neurosciences, University of Western Australia, MRF Building, 50 Murray St, Perth, WA 6000, Australia; 3) Cooperative Research Centre for Mental Health, Carlton, South, Victoria, Australia; 4) Collaborative Genomics Group, Centre of Excellence for Alzheimer's Disease Research & Care, School of Medical and Health Sciences, Edith Cowan University, 270 Joondalup Dr, Joondalup WA 6027; 5) School of Biomedical Sciences, Faculty of Health Sciences, Curtin Health Innovation Research Institute, Curtin University; 6) CSIRO Health and Biosecurity, Parkville, Victoria.

Recently, Sekar *et al.* (Nature, 2016) reported that the top GWAS hit for schizophrenia in the major histocompatibility complex (MHC) locus implicated structural variants in the *C4* gene which are associated with *C4A* expression and may affect synaptic pruning. Synapse loss is associated with cognitive decline in a number of disorders, including Alzheimer's disease and schizophrenia. We hypothesised that *C4A* expression predicted from *C4* structural alleles is associated with cognition. We tested this in the Western Australian Family Study of Schizophrenia (WAFSS) (n=770, including 402 schizophrenia cases and 368 controls), and the Australian Imaging, Biomarkers and Lifestyle Study (AIBL) (n=1,112, including 211 Alzheimer's disease cases, 133 with mild cognitive impairment and 768 controls). Sekar *et al.* showed that *C4* structural alleles could be imputed from GWAS data with an accuracy of ~70%. We used 222 reference haplotypes with *C4* structural alleles measured by Sekar *et al.* to impute *C4* alleles in all individuals using Beagle. *C4A* expression was predicted for each individual using the regression coefficients calculated by Sekar *et al.* Composite measures of cognition were available for both WAFSS and AIBL (Hallmeyer *et al.*, Am J Hum Genet. 2005; Burnham *et al.*, J Alzheimers Dis. 2015). Linear regression was used to assess their relationship with predicted *C4A* expression. *P* values for these relationships were calculated using linear mixed models to account for the relatedness between some WAFSS participants, and repeated measurements in the AIBL dataset. All analyses were adjusted for case status, age, sex, and education. Although mean *C4A* expression was not significantly different between schizophrenia cases and controls in the WAFSS (P=0.7), cases were overrepresented in those with the highest *C4A* expression levels. There was a small but consistent association between increased cognitive deficit and increased levels of *C4A* in both the WAFSS and the AIBL samples (*P*<0.05) which was particularly pronounced in cases. These data suggest that *C4* structural variants previously shown to be associated with schizophrenia may specifically implicate the cognitive deficit which is a core feature of schizophrenia. Further, our finding that a similar relationship between *C4A* levels and cognition exists in patients with schizophrenia and Alzheimer's disease suggests that the locus may be pleiotropic and have implications for a range of neuropsychiatric disorders.

2137W

BBS1 M390R/M390R mice have impaired anxiety-like behavior. *T. Pak[1,2], C. Carter[1], S. Gupta[3], J. Wemmie[3], S. Huang[1,2], N. Nuangchamnong[4], C. Searby[1], P. Abbott[3], H. Stevens[3], V. Sheffield[1].* 1) Pediatrics, University of Iowa, Iowa City, IA; 2) Medical Scientist Training Program, University of Iowa, Iowa City, IA; 3) Psychiatry, University of Iowa, Iowa City, IA; 4) Obstetrics and Gynecology, University of Iowa, Iowa City, IA.

Bardet Biedl Syndrome (BBS) is a clinically and genetically heterogeneous autosomal recessive ciliopathy with core features that include polydactyly, obesity, retinopathy, and nephropathy. There is also increased prevalence of psychiatric illness in BBS patients. The connection between BBS and psychiatric illness is not well understood. We studied the behavioral phenotypes in mice with the most common human BBS mutation, BBS1[M390R/M390R]. Open field testing, carbon dioxide evoked freezing, elevated plus maze evaluation, and marble burying testing examine anxiety-like behavior. Anxiety-like behavior is expressed by set patterns of defensive behavior such as apprehensions, sustained arousal, and vigilance. For the open field testing and carbon dioxide evoked freezing, there was no difference between BBS1[M390R/M390R] and wild type mice. For elevated plus maze evaluation, BBS1[M390R/M390R] mice showed less inhibitive anxiety than wild type mice. This was seen by the increased duration the BBS mutant mice stayed in the open arm (p=0.027). For the marble burying test, we found that BBS1[M390R/M390R] mice show an increased obsessive anxiety phenotype. This was accounted for by the increased number of marbles buried by the BBS mutant mice (p=0.015). To understand the impaired anxiety-like behavior in our BBS mutant mice, we studied the lateral amygdala (LA). The LA receives sensory inputs about potential threats. We assessed neuronal processing of the LA by testing its synaptic plasticity and strength using electrophysiology. The synaptic plasticity is impaired in the LA of BBS mutant mice as seen by the increased paired pulse ratio of the lateral amygdala (p=0.030). In addition, the synaptic strength of the LA is lower in the BBS mutant mice as seen by the decreased frequency and amplitude of the miniature excitatory postsynaptic currents (p=0.003, p<0.001). Impaired neuronal processing of potential threats by the lateral amygdala aligns with the impaired anxiety-like behavior in our BBS1[M390R/M390R] mice.

2138T

Initial results from the meta-analysis of the whole-exomes of 20,000 schizophrenia cases and 45,000 controls. *T. Singh[1,2,3], B. Neale[1,2,3], M.J. Daly[1,2,3] on behalf of Schizophrenia Exome Sequencing Meta-Analysis Consortium.* 1) Analytic and Translational Genetics Unit (ATGU), Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Schizophrenia is a debilitating psychiatric disorder that affects nearly 1% of the general population. Schizophrenia has been demonstrated to have a substantial genetic component, with common and rare variants contributing to individual risk. While over a hundred common risk loci have been implicated for schizophrenia, rare variant analyses from sequencing studies have yet to reliably implicate individual genes, presumably owing to power limitations. Here, we present the Schizophrenia Sequencing Meta-Analysis Consortium (SEMAC), a large multi-site collaboration to aggregate, generate, and analyze high-throughput sequencing data of schizophrenia to advance gene discovery. To date, we have sequenced and processed the whole exomes of over 20,000 schizophrenia cases and 45,000 matched controls using a standardized variant-calling and QC pipeline based on GATK and Hail, yielding one of the largest sequencing data sets of a complex trait to date. The included studies actively recruited from diverse global populations, and sampled individuals of European, Latin American, East Asian, Ashkenazi Jewish, and African American ancestry. Because our exomes were sequenced with various capture technologies over a span of seven years, we developed a novel pipeline to estimate exon-by-exon coverage of all sequencing batches in our data set, and incorporated this information during the quality control and analysis steps. After adjusting for local differences in coverage, we observed comparable quality metrics between different cohorts and sequencing waves, which we then took forward for analysis.   We successfully replicated findings from earlier sequencing studies of schizophrenia, including a substantial enrichment of ultra-rare damaging variants in genes with a near-complete depletion of protein-truncating variants. We also recapitulated a significant burden of disruptive variants in genes and gene sets previously associated with other neurodevelopmental disorders. We present the first gene-based burden results from our exome meta-analysis, and extend the sample using an independent set of 4,000 whole-genome sequenced cases and 7,000 controls. Finally, we present a new browser for exome association results that allows for easy viewing of identified variants and gene-based results. In summary, we introduce the largest multi-center effort to aggregate sequencing data of a psychiatric trait, and the initial results from the harmonization and analysis of over 65,000 exomes.

2139F

Enrichment of treatable metabolic disease gene variants in a large cohort of schizophrenia, bipolar and major depressive disorder patients. *V. Sriretnakumar[1,2], R. Harripaul[4,5], K. Mittal[1], J. Vincent[4,5], J. Kennedy[1,3,4], J. So[1,2,3,4,6].* 1) Neurogenetics Lab, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada; 2) Laboratory Medicine and Pathobiology (LMP), University of Toronto, Toronto, Ontario, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 4) Institute of Medical Sciences, University of Toronto, Toronto, Ontario, Canada; 5) Molecular Neuropsychiatry and Development Laboratory, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada; 6) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, Ontario, Canada.

Genetic diseases are individually rare but collectively common. Many genetic conditions can mimic mental health disorders, with psychiatric symptoms that are difficult to treat with regular medications. Treatment of the underlying genetic disease can cure the associated psychiatric symptoms or help regular medications work better. Discovery of rare genetic diseases in psychiatric patients would reveal specific treatment options, and give information about the chances of other family members being affected. In this study, we test the hypothesis that psychiatric populations are enriched for pathogenic variants associated with selected treatable genetic disorders. Using targeted exome sequencing, we screened 2,046 schizophrenia, bipolar and major depressive disorder patients for variants in genes associated with Niemann-Pick disease type C (NPC), Wilson disease (WD), homocystinuria (HOM), and acute intermittent porphyria (AIP). Our study is the first to explore the prevalence of NPC, WD, HOM and AIP gene variants in well-defined psychiatric populations. We found carrier rates of 0.88%, 0.88% and 0.20% for NPC, WD and HOM, respectively. Furthermore, an AIP affected rate of 0.20% was observed across the entire psychiatric cohort, a 200X enrichment in comparison to what is expected in the general population. Discovering genetic diseases in psychiatric patients will shift how health care is delivered to these vulnerable patients by addressing underlying conditions rather than masking symptoms with medications and will especially help patients who don't respond to regular medications. This will lead to significant cost savings to the health care system.

2140W

Integrated analysis supports *ATXN1* as a schizophrenia risk gene. *B. Su, JW. Liu.* Kunming Inst Zoology, Chinese Academy Sci, Kunming, Yunnan, China.

Protein-protein interaction (PPI) is informative in identifying hidden disease risk genes that tend to interact with known risk genes usually working together in the same disease module. With the use of an integrated approach combining PPI information with pathway and expression analysis as well as genome-wide association study (GWAS), we intended to find new risk genes for schizophrenia (SCZ). We showed that *ATXN1* was the only direct PPI partner of the know SCZ risk gene *ZNF804A*, and it also had direct PPIs with other 18 known SCZ risk genes. *ATXN1* serves as one of the hub genes in the PPI network containing many known SCZ risk genes, and this network is significantly enriched for the MAPK signaling pathway. Further gene expression analysis indicated that *ATXN1* is highly expressed in prefrontal cortex, and SCZ patients had significantly decreased expression compared with healthy controls. Finally, the published GWAS data supports an association of *ATXN1* with SCZ as well as other psychiatric disorders though not reaching genome-wide significance. These convergent evidences support *ATXN1* as a promising risk gene for SCZ, and the integrated approach serves as a useful tool for dissecting the genetic basis of psychiatric disorders.

2141T

The role of miRNAs in 22q11.2 deletion syndrome. *A.K. Victor[1,2], M.E. Fuller[2], S. Zakharenko[5], L.T. Reiter[2,3,4].* 1) IPBS Program, University of Tennessee Health Science Center, Memphis, TN; 2) Neurology, University of Tennessee Health Science Center, Memphis, TN; 3) Pediatrics, University of Tennessee Health Science Center, Memphis, TN; 4) Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN; 5) Department of Developmental Neurobiology, St. Jude's Children's Research Hospital, Memphis, TN.

22q11.2 deletion syndrome (22q11DS) has a variable clinical phenotype, which includes mild dysmorphic features, learning difficulties, hypernasal speech, congenital abnormalities and an increased risk for developing psychiatric disorders. Effected individuals have an ~30 times higher risk of developing schizophrenia than the general population. Using neuronal cultures derived from dental pulp stem cells of affected individuals and neurotypical controls we performed both miRNA and mRNA expression studies to look for gene dysregulation associated with schizophrenia (SCZ). We identified several mRNAs and miRNAs that may be involved in the 22q11DS. We identified 16 miRNAs significantly differentially expressed between 22qDS and control which putatively target three known SCZ associated genes: *DISC1*, *SYN2* and *HTR2A*. We hypothesize that when these miRNAs are overexpressed in 22q11DS neurons, transcript levels decrease for these SCZ associated genes. Here we overexpressed three *DISC1* targeting miRNAs into HEK293T cells to validate the interactions of the suspect miRNAs with the *DISC1* transcript. Pilot qRT-PCR experiments have shown that at least one of the predicted miRNAs decreases expression of *DISC1*. These experiments are the first steps in identifying gene expression and regulatory miRNA changes in 22qDS neurons that may contribute to the SCZ phenotype in 22qDS.

2142F

Identifying a shared regulatory background for neurodevelopmental disorders through meta-analysis of genomewide association studies.
Z. Yang[1], T. Tsetsos[1,2], P. Paschou[1]. 1) Dept of Biological Sciences, Purdue University, West Lafayette, IN, US; 2) Dept of Molecular Biology and Genetics, Democritus University of Thrace, Xanthi, Greece.

Gilles de la Tourette Syndrome (TS) is a neuropsychiatric disorder characterized by motor and vocal tics and high comorbidities rate with other neurodevelopmental disorders including Attention Deficit/Hyperactivity Disorder (ADHD), Obsessive-Compulsive Disorder (OCD), and Autism Spectrum Disorder (ASD). The high level of comorbidity rates among these disorders indicates a possible shared genetic basis, which has not yet been fully established. We set out to perform a cross-disorder meta-analysis across all four phenotypes based on published Genome-wide Association Studies (GWAS) on a total of 11394 cases across all disorders and 13,897 controls (ASD: 5,305 cases, ADHD: 2,787 cases, OCD: 1,889 cases, TS: 1,413 cases). We analysed 1,492 single-nucleotide polymorphism (SNPs) which we found as overlap among these datasets. We undertook a meta-analysis as implemented via the software METAL under a fixed-effects analysis model. According to the inverse variance based analytical strategy, the effect size estimates of each input dataset was weighed by the inverse of their corresponding standard errors. For heterogeneity analysis, we used Cochran's Q-test and I2 test. As a result, two SNPs, rs4575830 (p = 3.05e-08) and rs4306813 (p = 4.03e-08), reached the genome-wide significant threshold. Both SNPs belong to a piece of intron variant on chromosome 3 and they are in linkage disequilibrium with each other. They reside on a non-coding RNA gene LOC101928135, which is associated with a set of enhancers including brain-enriched protein coding gene ARPP21, and other non-coding RNA genes including MIR128-2. Besides these, 16 additional SNPs were proven significant over Bonferroni correction. To our knowledge this is the first study trying to reveal the common genetic foundation for TS, ADHD, OCD and ASD. The regulation effect of the implicated gene sheds light into a possible shared underlying neurobiology for the studied disorders.

2143W

A study of subthreshold hallucinatory experiences and thier relationship to genetic liability for schizophrenia.
H. Young[1], M. Liu[1], R. Wedow[1], A. Pandit[2], E.M. Schmidt[2], G.J.M. Zajac[2], W. Li[2], K. Brieger[2], G.R. Abecasis[2], S. Vrieze[1]. 1) Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI,, USA.

One hallmark symptom of schizophrenia is the presence of hallucinations. Consistent with the liability threshold model of schizophrenia, clinically subthreshold hallucinatory symptoms also occur in non-psychotic healthy individuals in the general population. Over 100 common variants have been associated with schizophrenia to date, but symptom specificity of these associations and associated neurobehavioral mechanisms are unknown and require study. The aim of the present work is to assess if subthreshold hallucinatory experiences and schizophrenia share genetic liability, and whether any known schizophrenia-associated locus influences hallucinations. In Genes for Good, an online study of health and behavior, 28,221 participants were administered the Launay-Slade Hallucination Scale (LSHS), a measure of predisposition to hallucinations. Of these individuals, 3,028 were genotyped on a genome-wide array and imputed to the Haplotype Reference Consortium panel. In addition to a genome-wide association study, we conducted a targeted genetic association study of 124 known risk variants associated with schizophrenia were tested for association with subthreshold hallucinatory experiences. Genetic correlations between schizophrenia and subthreshold hallucinations were tested with polygenic risk scores and LD score regression. In the dataset of 3,028 genotyped individuals, no reliable associations with hallucinatory experiences were found that reached genome-wide significance. None of the 124 previously established significantly associated schizophrenia risk variants were found to be significant after correcting for 124 tests. A polygenic risk score based on genetic vulnerability to schizophrenia with 80% power to detect a correlation of 0.05 was unable to successfully predict hallucinatory experiences using scores from the LSHS (*p* > 0.05). LD score regression results will also be presented. All results will be presented on the current full dataset of 28,221 individuals, of whom ~8,000 will be genotyped. No substantial evidence was found to support the association between genetic liability for subclinical hallucinatory experiences and schizophrenia. However, more data from Genes for Good participants has become available and the above analyses are currently underway for a sample of 5,655 and growing, providing much greater power to rule out the existence of genetic correlations and shared associations between schizophrenia and subthreshold hallucinations.

2144T

Genetics of attention deficity hyperactivity disorder dimensions.
T. Zayats[1], C. Burton[2], S. Medland[3], B. Xiao[2], J. Crosbie[2], P. Arnold[2,7], R. Schachar[2], S. Johansson[4], B. Neale[5,6], J. Haavik[1]. 1) K.G.Jebsen Center for Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway; 2) SickKids Research Institute, University of Toronto, Toronto, Canada; 3) Queensland Institute of Medical Research, Brisbane, Australia; 4) K.G.Jebsen Center for Neuropsychiatric Disorders, Department of Clinical Science, University of Bergen, Bergen, Norway; 5) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 6) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 7) Mathison Centre for Mental Health Research & Education, University of Calgary.

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common and heritable childhood onset psychiatric conditions. The clinical phenotype is multidimensional with main symptoms being inattention and/or hyperactivity/impulsivity. As ADHD persists into adulthood (in up to ~80% of cases), some symptoms become more pronounced than others. Furthermore, the various aggregates of dimensions are associated with different clinical features (e.g. co-morbidities) and treatment outcomes of ADHD. So far, the biology behind such symptomatology of ADHD is far from understood. In this study, we aimed to explore the genetics of ADHD dimensions in children and their correlation with neuropsychiatric phenotypes in adulthood. The childhood sample consisted of 11,784 children from (1) Australia (N=4,000), (2) Canada (N=4,339) and (3) Norway (The Norwegian Mother Child Cohort (MoBa), N=3,445). ADHD dimensions were measured by the Strengths and Weaknesses of ADHD and Normal Behavior (SWAN) rating scale in Australia and Canada, and by Swanson, Nolan and Pelham, Teacher and Parent Rating Scale (SNAP) in Norway. The genetic similarity of SWAN and SNAP measures was assessed by LD regression. We also examined how genetic risk for ADHD symptoms correlated with neuropsychiatric phenotypes in adulthood, including schizophrenia, major depression disorder, bipolar disorder, Alzheimer's disorder, subjective well-being, neuroticism and educational attainment. Overall, our preliminary results (N=3,445, SNAP data) suggest that the genetics of hyperactivity/impulsivity dimension in childhood show higher correlation with genetics of neuropsychiatric phenotypes in adulthood than that of childhood inattention, with strong contrasts between genetic correlations of hyperactivity/impulsivity and inattention with schizophrenia (h2=-0.115 and 0.0037), depressive symptoms (h2=0.483 and 0.278), subjective well-being (h2=-0.564 and -0.369), anxiety (h2=0.307 and 0.188) and Alzheimer's disease (h2=0.280 and 0.076). We will present the results from the full sample(N=11,784). Further exploration of how ADHD dimensions in childhood relate/lead to neuropsychiatric pathology in adulthood will aid our understanding of biological processes underlying ADHD symptomatology and, potentially, lead to better diagnosis, treatment and prevention options for ADHD.

2145F

Gene-based meta-analysis of GWAS in African American and European ancestry populations implicates novel genes for PTSD. *H. Zhang, J. Jung, S.P. Chou.* NIH/NIAAA, Rockville, MD.

Post-traumatic stress disorder (PTSD) is a prevalent, serious mental health condition. Genetic factors account for up to 30-40% of the risk for PTSD. Previous genome-wide association studies (GWAS) mainly focus on individual genetic variant. This approach limits the statistical power to detect the aggregated effects of a set of SNPs within truly susceptibility genes. By using publicly available summary data, gene-based tests for association were performed on the largest GWAS of PTSD (summary results downloaded from Psychiatric Genomics Consortium), including 2,479 cases and 6,744 controls in the seven African American (AA) cohorts and 2,424 cases and 7,113 controls in the nine European ancestry (EA) cohorts. The GCTA-fastBAT (a fast and flexible set-Based Association Test) program was employed to conduct gene-based association analysis using GWAS summary data in each ethnic group separately, followed by meta-analysis on gene-based association results across ethnic groups using Fisher's method. A total of 17 PTSD-associated genes (P < 10^{-3}) were identified by gene-based association method in the AA discovery cohort. Among them, 2 genes (SPPL2C and RHCG) passed a marginally significant threshold (P ≤ 0.06) in the EA replication cohort and produced significant outcomes (P < 10^{-3}) in meta-analysis. Furthermore, 7 additional genes (ISOC1, MIR4633, VAMP5, MIR4320, P2RY2, RNF181, TMEM150A) passed the significant threshold in both AA and EA cohorts (discovery P ≤ 0.01, replication P ≤ 0.05, meta-analysis P≤ 0.001) and were suggested to be related to PTSD susceptibility. For example, MIR4320 gene (microRNA 4320) on 18q21.1 was associated with PTSD (discovery P = 7.7×10^{-3}, replication P = 5.7×10^{-3}, meta-analysis P = 4.9×10^{-4}). Our gene-based meta-analysis of GWAS identified several novel genes underlying susceptibility to PTSD. Our results demonstrated the importance and power of the gene-based GWAS methods in identifying and further prioritizing loci of interest for complex traits or common diseases.

2146W

Centrality pattern of susceptibility genes to complex disorders in functional specific protein-protein interaction sub-networks. *T. Zhang[1], N.L. Saccone[3], J.P. Rice[2].* 1) Dept. of Epidemiology and Biostatistics,School of Public Health, Xi'an Jiaotong University, Xi'an, Shaanxi, China; 2) Department of Pyshiatry, Washington University in Saint Louis, Saint Louis, MO, USA; 3) Department of Genetics, Washington University in Saint Louis, Saint Louis, MO, USA.

The biological process of human beings is regulated through complex molecular networks. One of the most important features of network is that the effect that is caused by blocking one pathway within a network can be bypassed through some other "back door pathways". This indicates that, if a gene loses its function because of a mutation occurs within it, and if it does not locate at the central position of a gene network, it may have little impact to the biological process due to the bypassing effects. On the other hand, if this gene has a relatively high centrality in the gene network, the loss of its function may block several pathways simultaneously and therefore no bypassing pathway can be used to supplement its loss of function. In this study, we will present that what the centrality distribution pattern would be in function specific sub-networks for several sets of susceptibility genes contributing to complex disorders. Gene expression data of human brain tissue recorded in the Human Protein Atlas were extracted and utilized to construct a series of brain function specific sub-networks. Susceptible genes from three categories of complex disorders, including neurodegenerative disorder, psychiatric disorder and non-brain related disorder, were extracted from the GWAS catalogue. We identified a significant enrichment pattern of high centrality of susceptibility genes contributing to neurodegenerative and psychiatric disorders in these sub-networks. Our findings indicate that susceptibility genes of complex disorder might have higher centralities in functional specific sub-networks. Our study provide a primary evaluation of the significance of centrality measures calculated from function specific sub-networks as indicators of the biological functions of relevant genes.

2147T

Rare human knockouts in consanguineous pedigrees aggregated with schizophrenia and bipolar disorder compared to matched healthy population controls. *Q. He[1], A. Dionne-Laporte[2], D. Spiegelman[2], R. Joober[3], G.A. Rouleau[2], L. Xiong[1,4].* 1) Centre de recherche, Institut universitaire en santé mentale de Montréal, Université de Montréal, Canada; 2) Montreal Neurological Institute and Hospital, McGill University, Montreal, Canada; 3) Douglas Mental Health University Institute, McGill University, Montreal, Canada; 4) Département de psychiatrie, Faculté de médecine, Université de Montréal, Montréal, Canada.

Introduction: Recent publications reported that highly consanguineous populations like Pakistan would result in offspring carrying certain homozygous loss-of-function (homLOF) variants without major disease phenotypes. Isolated populations, such as Icelandic and Finnish population with recent bottlenecks, could also carry such rare complete human knockouts without severe outcomes. However, study of unrelated Swedish schizophrenia cases and controls reported an elevated burden of gene-disruptive and protein-damaging ultra-rare variants (variants unique to individuals) in affected individuals. **Samples and methods:** 123 schizophrenia and bipolar disorder affected individuals and 107 unaffected family members from 10 large consanguineous pedigrees underwent whole-exome sequencing(WES). WES data of 3222 healthy British Pakistani-heritage adults and 96 Punjabi Pakistanis samples from 1000 Genomes Project were used as external matched population controls. Unified variant-calling algorithms, consistent variant quality control and aligned gene annotation pipeline were applied to these datasets. **Results:** 1) In total, 3225 high-quality LOFs were identified from the 230 exomes of 10 pedigrees (20.5% as singletons), each individual carries on average 5.5±2.6 rare homozygotes (rhomLOFs) and 53.2±13.0 rare LoF heterozygotes; however, the affected individuals tend to have slightly more rare homozygotes (5.9±2.6) compared to unaffected individuals (5.1±2.4); 2) No perfect segregation of the rhomLOFs (shared by all affected but not any unaffected) are running in each pedigree; 3) 104 selected genes with rhomLOFs have only 0-23 overlaps with reported rhomLoF genes from the ExAC samples, the healthy British Pakistanis, and the Pakistanis in Myocardial Infarction Study, the Icelandic and Finnish populations (total of 44 overlapped genes collectively); 4) Only 6.7% of 104 genes are intolerant to LOFs, similar to the proportion of constraint genes in other human knockout reports from isolated and inbred populations. **Conclusion:** The rare homozygous knockout variants seem to be mainly population specific and the ones found in our consanguineous pedigrees with psychiatric disease show a complicated profile, without satisfying the classical recessive model. More comprehensive and in-depth comparisons with the unaffected family members and matched population controls are under way, as well as comparing to the LOF genes from other cohorts.

2148F

Catalogue of 1 billion candidate ultra rare variants across 11670 Han Chinese individuals. *S. Mangul, E. Eskin, J. Flint.* University of California, Los Angeles, Los Angeles, CA, 90025, USA.

Currently, we have relatively complete catalogs of common genetic variants and assessments of their contribution to complex traits. However, we lack similar knowledge for ultra-rare variants (URVs). Next-generation sequencing has the potential to generate the required information, but the cost currently is too high to collect data from sufficiently large sample sizes. Here we describe the use of a new method (RAREV) to capture URVs from low-coverage sequencing data. Standard methods of variant calling typically discard singleton variants supported by a single read, since the error rates are assumed to be too high for this information to be useful. We show that it is possible to obtain reliable information about sequence variants even from a single read. To test RAREV, we applied the method to 500 individuals from the 1000 Genomes Project who were subject to both whole exome (WES) and whole genome (WGS) sequencing. To obtain a gold standard and evaluate the performance of RAREV on low coverage WGS data, we used alleles with a coverage of greater than 10 reads per allele from WES. Direct detection of variants based on pileup data, including URVs supported by a single read, resulted in a false positive rate of 1%. We then applied RAREV, which includes a multi-step filtering procedure to exclude spurious signals due to error introduced by sequencing machine or mismapping. RAREV was able to reduce the false positive rate down to 0.0001%. We applied RAREV, to 11670 samples from the CONVERGE Consortium, which were sequenced at a median coverage of 1.7X. RAREV was able to detect 1 billion variants, including 30 million tri-allelic variants (i.e., sites with two different alternative alleles). We identified 480 million singletons variants, including variants supported by a single read. The vast majority of variants fall into intronic and intergenic regions. Intronic includes 47% of the variant and intergenic 32% of the variant. While the positions of URVs across the genome are in line with current views, the distribution of alleles is non-random, with a higher than expected proportion lying within gene intronic regions. We identify 2540 genes with low rates of variants (LRV) and compare those genes to LRV genes reported in ExAC. We then explore the use of our variants to understand the contribution of URVs to common complex diseases, such as major depression disorder, and we identify specific sets of genes that are enriched in deleterious mutations (LOF).

2149W

Association between TNF-α G-308A polymorphism and depression: A meta-analysis. *T. Kim[1], H.C. Jeong[2], S.N. Kim[3], K. Shin[2].* 1) Department of Clinical Pharmacology, Konkuk University Medical Center, Seoul, Republic of Korea; 2) College of Pharmacy, Research Institute of Pharmaceutical Sciences, Kyungpook National University, Daegu, Republic of Korea; 3) Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea.

Although a number of studies have investigated the association between a single nucleotide polymorphism (SNP) TNF-α G-308A and depression, the relationship between TNF-α G-308A and depression remains controversial. To clarify the role of TNF-α G-308A polymorphism on the risk of depression, we performed a meta-analysis Related studies were searched from PubMed, Medline, Embase and Scopus (up to April 18, 2017). The odds ratios (ORs) and 95% confidence intervals (CIs) were estimated in the models of allele comparison (G vs. A), homozygote comparison (GG vs. AA), dominant (GG + GA vs. AA), and recessive (GG vs. GA+AA) to estimate the strength of the associations. A total of 10 case-control studies were included in this meta-analysis. Overall, no significant association between TNF-α G-308A and depression was found (G vs. A: OR [95% CI]=1.09 [0.92, 1.29]; GG vs. AA: 1.24 [0.71, 2.15]; GG+GA vs. AA: 1.22 [0.70, 2.13]; GG vs. GA+AA: 1.01 [0.76, 1.35]). However, in Asian subgroup analyses, there was a trend that the frequency of depression was higher in genotype AA than in other genotypes (GG vs. AA: 2.34 [0.98, 5.57], *P*=0.06; GG+GA vs. AA: 2.29 [0.96, 5.49], *P*=0.06). In conclusion, we found that the TNF-α G-308A polymorphism is not associated with susceptibility to depression. However, the result of Asian subgroup analysis suggested potential ethnic difference in the effect of TNF-α G-308A polymorphism on the susceptibility of depression.

2150T

Antidepressant effectiveness study in major depressive disorder in STAR*D patients. *W. Guo[1], L. Hou[1], A.J. Rush[2], F. McMahon[1], Y.Y. Shugart[1].* 1) National Institutes of Health, Bethesda, MD; 2) Duke – National University of Singapore, Singapore.

Sequenced treatment alternatives to relieve depression (STAR*D) Study aimed to assess the effectiveness of depression treatments in 4,041 outpatients with major depression. In the STAR*D data, four levels of the study were designed to assess the effectiveness of depression medication in patients with major depressive disorder. Our GWAS study will focus on the level 1, in which participants were given the All participants began at Level 1, and were treated with the antidepressant citalopram (Celexa), a selective serotonin reuptake inhibitor (SSRI) from 2 to 14 weeks. In this study, QIDS-SR$_{16}$ was used as the primary measure to report Celexa's efficacy. To check the imputation quality, the imputed genotypes of STAR*D data were compared with the original genotyped genotypes or imputed genotypes of STARD-Omni1S data to evaluate the quality of imputation. About 80% of the genotyped genotypes showed Pearson correlation between the imputed genotypes of STAR*D data and the original genotyped genotypes of STARD-Omni1S data. The genome-wide significant SNP was rs17071498 (P =1.49×10^{-8}) which was obtained in GWAS of QIDS-SR$_{16}$ on exit week and linked to ARF4P4 gene (r^2= 0.9 and distance=28kb). The ARF4P4 gene has an association with Nerve tissue in GTEx samples. .

2151F

Analysis of leukocyte telomere length in children and adolescents at risk of developing mental disorders. *G. Xavier[1,2], L.M. Spindola[1,2,3], V.K. Ota[1,2,3], P.K. Maurya[9], P.F. Tempaku[4], C.M. Carvalho[1,2,3], P.N. Moretti[1,2], M.L. Santoro[1,2,3], L.B Rizzo[3], M.L. Levandowski[6], P. Pan[2,3], L.A. Rohde[6], A. Gadelha[2,3], D.R. Mazzotti[8], J.R. Sato[7], R. Grassi-Oliveira[6], E.C Miguel[5], G.A. Salum[6], R.A. Bressan[2,3], E. Brietzke[3], S.I. Belangero[1,2,3].* 1) Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, São Paulo - SP, Brazil; 2) Interdisciplinary Laboratory of Clinical Neurosciences, Universidade Federal de São Paulo, São Paulo - SP, Brazil; 3) Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, São Paulo - SP, Brazil; 4) Department of Psychobiology, Universidade Federal de São Paulo, São Paulo, São Paulo - SP, Brazil; 5) Department of Psychiatry - Faculty of Medicine, Universidade de São Paulo, São Paulo, São Paulo - SP, Brazil; 6) Department of Psychiatry, Universidade Federal do Rio Grande do SUL, Porto Alegre - RS, Brazil; 7) Center of Mathematics, Computation and Cognition, Universidade Federal do ABC, São Paulo, São Paulo - SP, Brazil; 8) Laboratory Center for Applied Genomics of Children's Hospital of Philadelphia, United States; 9) Amity Institute of Biotechnology - Amity University, Noida - India.

Psychiatric disorders are one of the main causes of youth disability. Maltreatment suffered during childhood might be an important risk factor for the development of these disorders. Furthermore, there are studies showing a strong relationship between telomere length (TL) and child maltreatment (CM). We investigated whether there is a relationship between TL and psychiatric diagnoses, psychopathology and CM in childhood and adolescence. We accessed biological data of 559 children from a large prospective community school-6based study in Brazil, the High Risk Cohort (HRC) Study for Psychiatric Disorders in Childhood, from 6 to 14 years old, wherein 165 were children and adolescents with a psychiatric disorder (according to DSM-IV) and 394 controls (children with no psychiatric disorder). We also had access to the children psychopathology scales (CBCL) and to CM data. TL was accessed from blood and measured by a multiplex quantitative polymerase chain reaction (q-PCR). The technique consists in determining the relative ratio (T/S) between the telomere region copy number (T) and a single copy gene (S), using a relative standard curve. We used the albumin gene (*ALB*) as the single copy gene. After using TL as dependent variable in generalized linear models, we found no significant association was observed between TL measurement and psychiatric disorder, comparing cases and controls. There was also no correlation between TL and CM (i.e. physical neglect, physical, sexual or emotional abuse) or psychiatric symptoms. On the other hand, considering the interaction effect between gender and CM and correlation with TL, boys who experienced more early life traumas seem to have a shorter TL. Other studies have shown a diminished telomere in adults who reported experiencing early life stress, as CM and related trauma. Although our result is in agreement with literature, we expected to find a shorter TL also among girls. The maintenance of TL in girls might be influenced by hormonal processes. Estrogen is thought to induce telomerase activity, providing a possible protection of telomeres from oxidative stress from early life adversity. The telomere shortening mechanism seems to biologically reflect the damages early life stress as CM may cause.

2152W

Preliminary analysis of whole genome sequences of simplex autism spectrum disorder. *M.B. Neu[1,2], D. Gray[1], M. Thompson[1], J. Lawlor[1], B. Weaver[1], S. Hiatt[1], J. Whittle[1], K. Bowling[1], C. Finnila[1], G. Cooper[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Alabama at Birmingham, Birmingham, AL.

While large-scale DNA sequencing has led to the discovery of many genes and genetic variants underlying Autism Spectrum Disorder (ASD) risk, many cases remain in which genetic causes have not been identified. To this end, we exploit whole genome sequencing (WGS) data of simplex ASD families from the Simons Simplex Collection (SSC) to identify novel genetic variation in ASD-affected individuals, who previously underwent exome sequencing, but no causal variation was detected. WGS from ~540 simplex ASD families from the SSC, each of which includes an affected proband, unaffected sibling, and both biological parents, are undergoing variant analysis to identify candidate diagnostic variants. In parallel, computational methods are being used to logically group variants according to biological pathway, known ASD genes/pathways, and mutation type, with an emphasis on annotations of non-coding functionality. Differences between affected probands and unaffected siblings are being assessed using CADD, a quantitative method for scoring variant impacts. Due to ASD's higher prevalence in males, we are also examining variants in sexually dimorphic pathways and assessing variant subsets transmitted from mothers to affected sons to explore a possible "female protective" mechanism for otherwise pathogenic single variants. Preliminary comparisons between ASD probands and their unaffected siblings reveal no statistically significant difference in burden or CADD score of de novo variants. Furthermore, maternally inherited variants grouped by gene do not have a statistically significant CADD score distribution difference between probands and their unaffected siblings. A more clear understanding of genetic variant contributions to ASD risk may serve to both increase clinical diagnostic effectiveness and identify ASD-associated genes, pathways, and molecular mechanisms. This data may also be useful for identifying potential future therapeutic targets.

2153T

Autism spectrum disorder in the Amish: Exome sequencing in a founder population unveils novel coding variants. *C.G. Tise[1], J.W. Kleinberger[1], M.A. Pavlovich[1], M.L. Daue[1], J.A. Perry[1], J.R. O'Connell[1], J.G. Reid[2], J.D. Overton[2], L.M. Yerges-Armstrong[1], A.R. Shuldiner[1,2], N.A. Zaghloul[1].* 1) Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Regeneron Genetics Center LLC, Regeneron Pharmaceuticals Inc., Tarrytown, NY.

Background: Autism spectrum disorder (ASD) is a heterogeneous disorder that affects 1 in 68 children in the U.S., occurring in all racial, ethnic, and socioeconomic groups, with males affected almost five-times more often than females. The heritability of ASD is estimated to be 50%, suggesting both genetic and environmental factors contribute significantly to the etiology of the disorder. To date, sequencing efforts to identify genetic determinants of ASD have focused largely on *de novo* copy number variants and loss-of-function variants in severely affected children. In this study, we evaluated exonic variants for association with quantitative measures of social impairment within the autism spectrum among adult subjects. Methods: The Social Responsiveness Scale™, Second Edition (SRS-2), Adult Self-Report Form by Western Psychological Services was administered to 203 generally healthy Amish adults (110 women, 93 men) who were participants in ongoing studies at the University of Maryland Amish Research Clinic. The SRS-2 is a 65-item, Likert-scale, objective measure of symptoms associated with ASD used to identify the presence and severity of social impairment within the autism spectrum, with higher scores correlating with increased severity (normal ≤59, mild 60-65, moderate 66-75, severe ≥76). Using exome sequencing data, an exome-wide association analysis (ExWAS) of SRS-2 score as a continuous trait was performed using Mixed Models for Analysis of Pedigree/Populations (MMAP), adjusting for age, sex, and relatedness. Results were limited to exonic variants in Hardy-Weinberg equilibrium (P≥0.05) with at least five heterozygotes. Results: SRS-2 scores ranged from 37-75 with a mean of 49.3 ± 6.1 (SD). Two variants, each associated with increased SRS-2 score, reached exome-wide significance ($P<1\times10^{-6}$): rs368577988 in *SR-related CTD associated factor 1* (*SCAF1* p.A262V; β=9.5, $P=4.6\times10^{-7}$) and rs752371351 in *Spermatogenesis associated 13* (*SPATA13* p.V540M; β=14.7, $P=8.8\times10^{-7}$). Both genes are highly expressed in brain tissue according to data from the Genotype-Tissue Expression (GTEx) project. Conclusion: We evaluated adults across the full spectrum of ASD-related behavioral traits, and through ExWAS, identified germ-line variants in several genes that may provide insight into disease mechanisms and identify potential drug targets. Extension and replication of these findings will be required to confirm and further refine these novel genetic signals.

2154F

Mid-childhood adaptive function in individuals with 22q11.2 deletion syndrome is associated with immune-deficiency, but not oral/palatal or cardiac phenotypes. *J.G. Baskin[1], O.Y. Ousley[2,3], K. Coleman[1,3], J.F. Cubells[1,2].* 1) Dept. of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Emory Autism Center, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA; 3) Marcus Autism Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA.

Common features of 22q11.2 deletion syndrome (22q11.2DS) include heart malformations, oral/palatal anomalies, and immune deficiency. These and other clinical features of 22q11.2DS vary widely among patients. We tested the hypothesis that the presence of at least one of these features would associate with poorer adaptive function in patients with the syndrome. While most studies suggest no association between cardiac malformations and adaptive function, no studies to date have examined oral/palatal or immune phenotypes with regard to adaptive function. This retrospective medical chart review used the Adaptive Behavior Assessment System, Second edition (ABAS-II) to evaluate adaptive function in 54 children, 6-12 years old, with 22q11.2DS (M = 9.0, SD= 2.2). The ABAS-II is a parental report questionnaire from which a general adaptive composite (GAC) score can be derived. Children were assessed for heart malformations, oral/palatal anomalies, and immune deficiency as well as other clinical features during visits to a specialty clinic for 22q11.2DS. During those visits, parents filled out the ABAS-II. We compared ABAS-II GAC scores among children with and without each 22q11.2DS-associated clinical feature. GAC scores followed a normal distribution that was shifted downward from the general population norms (M = 72.5, SD = 17.1). Based on the results of t-test analysis, we found that presence of oral/palatal anomalies or cardiac defects did not associate with GAC scores. However, children without immune deficits had significantly higher mean GAC scores (M = 86.2, SD = 17.7), compared to children with immune deficits (M = 71.1, SD = 16.6); $t(52)$ = 1.93, $p < .05$, one-tailed, suggesting that less immune impairment associates with higher adaptive function in children with 22q11.2DS. In conclusion, children with 22q11.2DS exhibit lower mean adaptive function compared to normative data, but the presence of cardiac or oral/palatal anomalies does not further impair functioning. However, children with more immune impairment show lower GAC scores than those without immune deficiencies. Our results suggest that all children with 22q11.2DS should receive periodic assessment of adaptive function to guide specific interventions. Those with documented immune deficits should receive even closer attention. These observations may be useful to parents receiving genetic counseling for children with 22q11.2DS.

2155W

PYROXD1 is responsible for cellular functions in myoblasts and homozygous missense mutation in PYROXD1 causes limb-girdle muscular dystrophy among patients from Saudi Arabian cohort. *M. Saha[1], H.M. Reddy[1], M. Salih[2], M. Jones[1], S. Mitsuhashi[3], K. Cho[1], E. Estrella[4], M. Lek[5,6], B.B. Cummings[5,6], E. Valkanas[5,6], D.G. MacArthur[5,6], L.M. Kunkel[4], C. Pacak[7], P.B. Kang[1,7,8].* 1) Department of Pediatrics, University of Florida, GAINESVILLE, FL; 2) Department of Pediatrics, Division of Neurology, King Saud University, Riyad, Saudi Arabia; 3) Department of Neurology, Boston Children's Hospital & Harvard Medical School, Boston, MA; 4) Division of Genetics & Genomics, Boston Children's Hospital and Harvard Medical School, Boston, MA; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 6) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA; 7) Department of Neurology and Department of Molecular Genetics and Microbiology, University of Florida College of Medicine, Gainesville, FL; 8) Genetics Institute and Myology Institute, University of Florida, Gainesville, FL.

Statement of purpose: Limb-girdle muscular dystrophy (LGMD) is a neuromuscular disorder characterized by progressive proximal muscle weakness, accompanied by classic histological findings on muscle biopsy. *PYROXD1*, a gene that was recently associated with myofibrillar myopathy, is located at chromosome 12p12.1 in humans and encodes a pyridine nucleotide-disulfide oxidoreductase domain-containing protein. The missense mutation p.Asn155Ser in *PYROXD1* causes reduced expression of the encoded protein in muscles. Not much is known about the influence of PYROXD1 in the cellular functions for muscular development. **Methods used**: Subcellular fractionation was confirmed by western blotting of subcellular fractions with respective antibodies. We used Dharmafect reagent to successfully transfect C2C12 myoblasts with *Pyroxd1* siRNA. RT-PCR was performed using Taqman probes. Proliferation assays, MTT assays, and Seahorse analysis were used to measure developmental and metabolic function. Desmin staining and measurement of myoblast fusion indices were performed as previously published. **Summary of results**: We recruited and analyzed 16 families with LGMD from Saudi Arabia using traditional approaches and whole exome sequencing. One family was found to have a mutation in *PYROXD1*, demonstrating that mutations in this gene may also cause LGMD. Subcellular fractionation and immunofluorescence analysis indicated that the localization of endogenous Pyroxd1 to be nucleo-cytoplasmic in myoblasts. *Pyroxd1* siRNA was transiently transfected into C2C12 myoblasts and successful knockdown was confirmed via RT-PCR and western blot analysis. *Pyroxd1* knockdown myoblasts demonstrated significant proliferation defects compared to control scrambled siRNA treated myoblasts, as shown in NF CyQuant assays. Impaired desmin staining and reduced myoblast fusion indices indicate that the process of differentiation is also impaired in *Pyroxd1* knockdown myoblasts. MTT assays suggested reduced metabolic activity in *Pyroxd1* knockdown myoblasts. More specifically, *Pyroxd1* knockdown myoblasts exhibited decreased oxygen consumption rates (OCRs), basal respiration, and spare respiratory capacity on Seahorse assays. **Conclusion**: These analyses reveal that *PYROXD1* mutations may cause LGMD as well as myofibrillar myopathy, and that Pyroxd1 deficiency leads to developmental and metabolic defects in mouse myoblasts.

2156T

Using electronic health records for the identification of novel genes associated with adverse drug reactions. *L. Milani[1], K. Krebs[1], T. Tasa[1,2], M. Kals[1], R. Mägi[1], T. Esko[1], J. Vilo[2], A. Metspalu[1].* 1) Estonian Genome Center, University of Tartu, Tartu, Tartu, Estonia; 2) Institute of Computer Science, University of Tartu, Tartu, Estonia.

Combining genomic information with electronic health records (EHRs) provides new opportunities for the interpretation of the role of genetic variation in different diseases and traits. Despite recent advances in pharmacogenomic research, interindividual variability in drug metabolism and sensitivity for drug toxicity persists as a major problem for drug development and treatment. To address this, we designed a study of adverse drug reactions using specific ICD-10 codes extracted from the EHRs of biobank participants, and tested for their association with rare and common genetic variants among individuals who had purchased prescription drugs according to records from the Estonian Health Insurance Fund. We sequenced the whole genomes of 2,240 participants of the Estonian biobank using PCR-free sample preparation and 30x coverage at the Broad Institute, and further included all variants imputed for 14,219 genotyped subjects in the biobank. In the Estonian population, we found 4.1 million previously undetected single nucleotide variants (SNVs). When focusing on 64 genes highly involved in drug response (pharmacogenes) and genetic variants predicted to have high impact on protein function in this large dataset, we detected 46 loss-of-function (LoF) and 656 non-synonymous SNVs of which more than 22% were novel and more than 80% were rare with minor allele frequencies (MAF) <0.5%. Each individual carried an average of 1 LoF and 34 non-synonymous variants in these genes, which all require further studies to confirm their effects on drug response phenotypes. When focusing on 64 genes important in drug response and by linking genetic data with the EHRs we were able to validate several previously documented genetic variants associated drug induced ADRs. By genome wide analysis we identified a novel gene, *CTNNA3*, to associate with ADRs among individuals treated with oxicams. As drug prescription and EHRs were available for the entire cohort of 52,000 biobank participants, we were able to replicate this finding in an extended cohort of the biobank. In conclusion, EHRs can serve as a reliable tool for explorative pharmacogenomic studies. Very rare variants, however, will require deeper phenotyping and further studies to confirm their effects on drug response.

2157F

The intersection of type 2 diabetes and cognitive impairment in Mexican Americans: Insights from the mitochondria. *T. Silzer[1,6], N. Phillips[1,3,4], J. Sun[1,3], R. Barber[1,3,4], L. Johnson[2,5], S. O'Bryant[2,4].* 1) Institute for Molecular Medicine, University of North Texas Health Science Center, Fort Worth, TX; 2) Institute for Healthy Aging, University of North Texas Health Science Center, Fort Worth, TX; 3) Center for Medical Genetics, University of North Texas Health Science Center, Fort Worth, TX; 4) Center for Aging and Neurodegenerative Disease Research, University of North Texas Health Science Center, Fort Worth, TX; 5) Center for Neuroscience Discovery, University of North Texas Health Science Center, Fort Worth, TX; 6) Graduate School of Biomedical Sciences, Genetics Program, University of North Texas Health Science Center, Fort Worth, TX.

Mitochondrial dysfunction and systemic stress response is common in many complex age-related diseases. Multiple environmental and genetic factors, including variants in both nuclear and mitochondrial DNA (mtDNA), may contribute to the cycle. The role of mitochondrial function in cognitive decline has not been studied in Mexican American populations. This group is of interest because of their rapidly growing population within the United States and because of the high prevalence of type 2 diabetes (T2D). T2D is strongly associated with accelerated cognitive decline, and Mexican American populations may be at higher risk for developing Alzheimer's disease (AD) due to the high prevalence of this comorbid condition. **Hypothesis**. Mitochondrial dysfunction and resulting systemic stress may provide a mechanistic link between T2D and cognitive decline. We investigate the intersection of cognitive impairment and T2D with respect to indicators of mitochondrial function (expression of key genes) and mitochondrially-sourced proinflammatory stressors (cell-free mtDNA, CFmtDNA). **Methods**. Peripheral blood samples (buffy coat and plasma fractions) were obtained from subjects of the Healthy Aging Brain among Latino Elders (HABLE) cohort. Individuals were partitioned into groups based on T2D diagnosis, and were matched across groups based on age, gender and cognitive status. Expression of key mitochondrial function genes was quantified from buffy coat RNA isolates (TRIzol™ Plus RNA Kit, ThermoFisher) using the Mitochondria RT2 Profiler qPCR array (Qiagen). Circulating cell-free mtDNA (CFmtDNA) was quantified within plasma using TaqMan®-based absolute quantification (Kavlick *et al.*, 2011) via real-time PCR performed on the 7500 Real-Time PCR system (Applied Biosystems); additional assays were conducted to further characterize CFmtDNA. **Results**. Preliminary analysis indicate group differences in CFmtDNA both with respect to T2D status and cognition; notably, a significant interaction was observed as individuals with T2D and cognitive impairment was responsible for the bulk of the effect between T2D+ and T2D- groups. Differential expression of mitochondrial related genes between T2D+ and T2D-individuals is anticipated as a result. **Conclusion**. CFmtDNA and mitochondrial dysfunction may be particularly important in key populations where inflammatory comorbities (e.g., T2D) can exacerbate age-related cognitive decline and predispose individuals to earlier onset of dementia.

2158W

Expanding the *RTN4IP1/OPA10* genotype-phenotype correlation: From isolated optic neuropathy to severe mitochondrial encephalopathy. *I. Barbosa[1], M. Charif[2], A. Nasca[3], K. Thompson[4], S. Gerber[5], C. Makowski[6], N. Mazaheri[7], C. Bris[2], D. Goudenège[2], A. Legati[3], R. Marrofian[8], M. Simpson[1], H. Galehdari[8], C. Deshpande[9], T. Haack[10,11,12], J-M. Rozet[5], R. Taylor[4], D. Ghezzi[3], P. Amati-Bonneau[2], G. Lenaers[2].* 1) Division of Genetics and Molecular Medicine, King's College London School of Medicine, London, UK; 2) MitoLab team, UMR CNRS 6015 - INSERM U1083, Institut MitoVasc, Angers University and Hospital, Angers, France; 3) Unit of Molecular Neurogenetics, Foundation IRCCS Institute of Neurology « Besta », 20126 Milan, Italy; 4) Wellcome Trust for Mitochondrial Research, Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK; 5) Laboratory of Genetics in Ophthalmology, INSERM UMR1163 - Institute of genetic diseases, Imagine. Paris, France; 6) Department of Paediatrics, Technische Universität München, Munich, Germany; 7) Dept. of Genetics, Shahid Chamran University of Ahvaz, Iran ,2 Narges Medical Genetics and Prenatal Diagnosis Laboratory,East Mihan Ave., Kianpars, Ahvaz, Iran; 8) University of Exeter Medical School, RILD Wellcome Wolfson Centre, Royal Devon & Exeter NHS Foundation Trust, Exeter, UK; 9) Clinical Genetics Department, Guys and St Thomas' NHS Foundation Trust, London, SE1 9RT, UK; 10) Institute of Human Genetics, Technische Universität München, Munich, Germany; 11) Institute of Human Genetics, Helmholtz Zentrum München, Munich, Germany; 12) Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tuebingen, Germany.

Neurological disorders with isolated symptoms or complex syndromes are relatively frequent among inherited mitochondrial diseases. Recently, we have shown that recessive pathogenic variants in *RTN4IP1* cause isolated and syndromic optic neuropathies (Angelbault et al. *Am. J. Hum Genet.* 2015 https://www.ncbi.nlm.nih.gov/pubmed/26593267). We have used targeted and whole exome sequencing to delineate the spectrum of clinical phenotypes associated with pathogenic variants in *RTN4IP1/OPA10*, encoding a mitochondrial quinone oxidoreductase. Here we present the clinical and molecular investigations performed on 12 individuals from 11 families with CNS presentations ranging from isolated optic atrophy to severe encephalopathy, thereby clarifying the molecular diagnosis for patients with severe CNS diseases. Recessive, pathogenic RTN4IP1 variants were identified in all patients. Six families, among them 5 consanguineous, presented with isolated optic atrophy in whom we identified previously-reported p.(Arg103His) or the novel p.(Ile362Phe), p.(Met43Ile) and p.(Tyr51Cys) amino acid substitutions (RTN4IP1 Genbank accession mumber: NM_032730.5). Five other families presented severe neurological syndromes with a common core of symptoms consisting of optic atrophy, seizures, intellectual disability, growth retardation and elevated lactates. Additional features were deafness, stridor, brain MRI abnormalities, and abnormal EEG patterns, eventually leading to death before the age of 3 years. In these patients, we identified rare homozygous and compound heterozygous pathogenic variants, associated with a loss of steady-state RTN4IP1 protein levels, a defect in mitochondrial complex I assembly and mild fragementation of the mitochondrial network. Our results confirm that a broad clinical spectrum of neurological features ranging from isolated optic atrophy to severe early-onset epileptic encephalopathies is associated with bi-allelic *RTN4IP1* variants, and should prompt screening of this gene in both syndromic neurological presentations and non-syndromic recessive optic neuropathies.

2159T

Novel mutation in the *MT-ND4* gene at low heteroplasmy level likely associated with a mild phenotype: A case report. *A. Gonzalez Garcia[1], A. Rajadhyaksha[2], P. Jayakar[2].* 1) Medical Education, Nicklaus Children's Hospital, Miami, FL; 2) Genetics and Metabolism, Nicklaus Children's Hospital, Miami, FL.

The patient is an 8-year-old female with normal development who presented with seizures, headaches, tremors and exercise intolerance since age 3. Her symptoms have been stable and not progressive. MRI brain was normal and genetic testing including metabolic workup, CMA and WES were negative. Family history is significant for mother having migraines with aura. A muscle biopsy done showed mild degree of ultrastructural mitochrondrial cytopathy with small fibers being affected; however mitochondrial enzymes in muscle were negative for respiratory chain enzyme deficiency. NGS on muscle showed the following: Heteroplasmic (4%) novel variant m.12054 G>A, on the *MT-ND4* gene causing an amino acid change from Arginine to Glycine at position 432 (p.R432G) and homoplasmic variant on *MT-ND2* m.4917 A>G (p.N150D) . Mitochrondrial sequencing in blood showed the same findings. Maternal testing for the variant in *MT-ND4* was negative in blood. The variant in the *MT-ND4* has not been described in literature. It is a semi-conservative amino acid substitution that may impact secondary protein structure as these residues differ in some properties. This position is conserved across species and In silico analysis predicts this is probably damaging to the protein. Based on these results, the variant was interpreted as a strong candidate for a disease causing mutation. Mutations in the *MT-ND4* have been reported to cause a variety of signs and symptoms including ataxia and seizures. Limited understanding of the phenomenon of heteroplasmy continues to make counseling on mitochondrial diseases challenging. The threshold for mutation load to produce a clinical phenotype or affect the mitochondrial respiratory chain is unknown. This novel heteroplasmic variant has not been identified as benign in the general population nor has it been described in literature as disease causing. Furthermore, this variant has not been noted in the mother probably due to lower than detection level in blood. However, it maybe present in tissues not tested. Although functional testing is required to further define the phenotype caused by this mutation we predict the p. R432G mutation in the *MT-ND4* gene to be pathogenic, likely disease causing.

2160F

A hypertension-associated mtDNA mutation alters the tertiary interaction and function of tRNALeu(UUR). *M. Guan, M. Zhou, M. Wang.* Genetics, Zhejiang University, Hangzhou, Zhejiang, China.

   Several mitochondrial tRNA mutations have been associated with hypertension. However, the pathophysiology of these tRNA mutations remains poorly understood. In this report, we identified the novel homoplasmic 3253T>C mutation in the mitochondrial tRNALeu(UUR) gene in one Han Chinese pedigree with maternally inherited hypertension. The m.3253T>C mutation affected a highly conserved uridine at position 22 at D-stem of tRNALeu(UUR). The m.3253T>C mutation introduced the G-C base-pairing (13G-22C) at the D-stem and formed a tertiary base pairing (22C-G46) with the G46 in the variable loop of tRNALeu(UUR). It was anticipated that the m.3253T>C mutation may alter the structure and function of tRNALeu(UUR). Using cybrid cell lines derived from the Chinese family, we demonstrated that the m.3253T>C mutation perturbed the conformation of tRNALeu(UUR), as suggested by faster electrophoretic mobility of mutated tRNA with respect to the wild-type molecule. Northern analysis revealed ~45% decrease in the steady-state level of tRNALeu(UUR) in the mutant cell lines carrying the 3253T>C mutation. Furthermore, ~35% reduction in aminoacylated efficiency of tRNALeu(UUR) was observed in mutant cells derived from this Chinese JBC C o n f i d e n t i a l hypertension linked mitochondrial tRNALeu(UUR) mutation 2 family. A failure in tRNALeu(UUR) metabolism impaired mitochondrial translation, especially for polypeptides with a high proportion of Leu(UUR) such as ND6. Altered mitochondrial translation caused the decreasing activities of complexes I and V. Furthermore, marked decreases in the levels of mitochondrial ATP and membrane potential were observed in mutant cells. These mitochondrial dysfunctions caused an increasing production of reactive oxygen species in the mutant cells. Our findings may provide new insights into the pathophysiology of maternally inherited hypertension.

2161W

Identification of genetic causes for age-related hearing loss. *S.H. Blanton[1,2,3], B.A. Schulte[4], J.R. Dubno[5], L.J. Matthews[5], S. Guo[2], X.Z. Liu[1,2,3].* 1) Department of Human Genetics, University of Miami School of Medicine, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Otolaryngology, University of Miami School of Medicine, Miami, FL; 4) Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC; 5) Department of Otolaryngology-Head and Neck Surgery Medical University of South Carolina, Charleston, SC.

   Age-related hearing loss (ARHL), also known as presbyacusis, affects more than 28 million Americans, 75% of whom are over age 55. ARHL is a consequence of accumulated environmental stresses and part of the intrinsic genetically controlled aging process with heritability estimates of ~35-55%. Although several studies have attempted to identify the genetic underpinnings of ARHL, little is known about the underlying genes involved in its development and progression. Studies of animal models, human temporal bones, and human subjects have suggested several subtypes of ARHL recognizable by different pathological alterations and auditory phenotypes. To understand the development and progression of ARHL, we are analyzing results from the Medical University of South Carolina (MUSC) longitudinal database of >1,500 individuals who have been followed for up to 28 years. To investigate the contribution of specific genetic variants, a cohort of 501 adults >55 years of age and of European ancestry was assembled (mean age 72.3 years; 53% female). Whole exome sequencing was performed using the Agilent SureSelect XT Human All Exon V5-UTRs library capture kit. Study participants were classified using Quadratic Discriminant Analysis of pure-tone thresholds into one of four audiometric phenotypes: Older-Normal, Metabolic, Sensory, and mixed Metabolic+Sensory. We took a candidate gene approach and examined the data for association between audiometric phenotypes and variants in 203 known/possible deafness genes. Gene based associations were tested using SKAT-O with sex and age as covariates. In the analysis of Sensory vs. Older-Normal, one gene, *TBX4*, was significant (p=0.000166), assuming a Bonferroni correction (p<0.00025); 11 genes had a nominal p<0.05. None of the genes were significant in the Metabolic vs Older-Normal analysis; the top gene, *CLPP*, had a p=0.0073 and an additional 5 genes had p<0.05. *KCNQ1* was the only gene with a p<0.05 in both groups (0.034 and 0.023). *KCNQ1* encodes a voltage-gated potassium channel protein and mutations in this gene cause Jervell and Lange-Nielsen syndrome, which is characterized by congenital deafness, prolongation of the QT interval, syncopal attacks due to ventricular arrhythmias, and a high risk of sudden death. These results suggest that variation in deafness genes, while not a primary cause of ARHL, may contribute to the genetic etiology as well as to specific subtypes.

2162T

Cochlear nerve deficiency presenting as auditory neuropathy spectrum disorder. *A. Pandya[1], A. O'Brien[1], P. Roush[2].* 1) Pediatric Genetics and Metabolism, University of North Carolina, Chapel Hill, NC; 2) Department of Otolaryngology, Head neck and Surgery, University of North Carolina, Chapel Hill, NC.

Cochlear nerve deficiency (CND) is a condition with an absent or hypoplastic cochlear nerve which may present with an auditory brainstem response (ABR) pattern consistent with Auditory Neuropathy Spectrum Disorder (ASND). ANSD is clinically heterogeneous, with both, congenital and acquired causes. It accounts for more than 10% of all children with congenital hearing loss. Infants with ANSD have an abnormal auditory nerve response, as indicated by absent or abnormal ABR and, evidence of normal outer hair cell function as measured by either presence of a cochlear microphonic in the ABR or present otoacoustic emissions. Multiple etiologies including prematurity, hyperbilirubinemia, viral infections, hypoxia, and single gene sequence variants in the OTOF, and pejvakin genes can result in ANSD. Buchman et al. (2006) first reported the association of CND in nine of 51 children with ANSD, demonstrating unilateral or bilateral CND. To date, we know very little about the prevalence or etiology of CND in children. We present one of the largest series of patients with CND in the USA, ascertained through the Otolaryngology and Audiology clinic at UNC-Chapel Hill, and describe the demographic and clinical characteristics including their hearing loss profile, auditory neuropathy pattern, MRI findings, and existing co-morbidities. After IRB approval, we reviewed records on 267 patients diagnosed with ANSD from 1999-2016. Of the 254 probands in our cohort who had information on the laterality of ANSD, the majority (67.7%) had a bilateral and 32.3% had a unilateral auditory neuropathy. Overall, 15% (n=40) children had CND identified, and of those with a unilateral auditory neuropathy, 34% had cochlear nerve hypoplasia/aplasia identified by temporal bone imaging studies. All of them had a "refer" on their newborn audiologic screen however the age of diagnosis of CND ranged from birth to six years. In contrast to bilateral ANSD, only a minority of children with CND were born preterm and/or had hyperbilirubinemia, which are risk factors for ANSD. Response to cochlear implantation is also suboptimal in children with CND. Elucidating the clinical presentation and etiology for CND is critical to provide families with an accurate and timely diagnosis and to offer optimal rehabilitation. Children with CND represent a unique and homogenous group amongst the extreme heterogeneity noted with ANSD, and therefore offer an opportunity to determine a unifying genetic etiology. .

2163F

Association of PMEL missense variants with ocular pigment dispersion and pigmentary glaucoma. *M.A. Walter[1], A.A Lohola-Chomiak[1], T. Footz[1], T-K. Nguyen-Phuoc[1], G.J. Neil[2], W.T. Allison[1,2], O.J. Lehmann[1,3].* 1) Dept Med Gen, Univ Alberta, Edmonton, AB, Canada; 2) Dept Biol Sci, Univ. Alberta, Edmonton,AB, Caanda; 3) Dept Ophthalmol, Univ Alberta, Edmonton, AB, Canada.

Pigmentary glaucoma (PG) is the most common secondary glaucoma making it a major cause blindness worldwide. Previous research into PG demonstrated a substantial heritable component, but to date no candidate genes have been identified. We here report that variants in the melanosome structural protein premelanosome protein (PMEL) are associated with pigment dispersion syndrome (PDS) and PG. Whole exome sequencing of members in two Mennonite kindreds affected by PDS/PG identified several candidate genes of which *PMEL* was selected for further functional characterization based on its well established role in ocular pigmentation and the ocular phenotypes observed in several Pmel animal models. Targeted sequencing of *PMEL* in 114 additional sporadic cases revealed three additional missense variants. Functional analysis of these four (p.N111S, p.A340V, p.G370D, and p.L389P) variants using an *in vitro* cell culture system revealed defects in proteolytic cleavage (1 of 4), glycosyl group maturation (2 of 4), and melanosome fibril formation (4 of 4). Together these structural deficits provide evidence that the PG-associated variants in *PMEL* cause functional defects impacting PMEL's normal ability to form amyloid fibrils. Intriguingly, these data may indicate these *PMEL* variants are gain-of function mutations that switch PMEL from a functional to a pathological amyloid. Disruption of the homolog *pmela* in zebrafish caused profound pigmentation defects in the eye further supporting PMEL's role in ocular pigmentation. While PMEL mutations have previously been associated with pigmentation and ocular defects in animal models this research represents the first report of human disease-causing coding variants in *PMEL*.

2164W

Gene-set enrichment analysis identifies pathways involved in tinnitus. *E. Fransen[1], A. Gilles[2], P. Van de Heyning[2], G. Van Camp[1].* 1) Center for Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) Department of Otorhinolaryngology, Head and Neck surgery, Antwerpe University Hospital, Edegem, Belgium.

Tinnitus, the perception of an auditory phantom sound in the form of ringing, buzzing, roaring or hissing in the absence of an external sound source, is perceived by approximately 15% of the population and 2.5% experiences a severely bothersome tinnitus. Previous studies suggested only a minor contribution of genes to this condition, with a heritability estimate of 14% in a Norwegian study and a small GWAS in which the variance explained by all SNPs in the GWAS was estimated to be only 3.2%. None of the SNPs in this latter GWAS reached the threshold for genome-wide significance (p< 5.0e-8), with the most significant SNPs, situated outside coding genes, reaching a p-value of 3.4e-7. Here we report the results of gene set enrichment analysis on the results of this GWAS using the program MAGENTA. This highlighted several statistically significant metabolic pathways, including pathways involved in oxidative stress, endoplasmatic reticulum (ER) stress and serotonin reception mediated signalling. These results are a promising basis for further research into the genetic basis of tinnitus, including GWAS with larger sample sizes and considering tinnitus subtypes for which a greater genetic contribution is likely.

2165T

New risk loci for primary open-angle glaucoma. *P. Gharahkhani[1], A.W. Hewitt[2], K.P. Burdon[2], J.N. Cooke Bailey[3], D.A. Mackey[2,4], J.L. Wiggs[5], J.E. Craig[6], S. MacGregor[1].* 1) Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) University of Tasmania, Hobart, Tasmania, Australia; 3) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH; 4) Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Australia; 5) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA; 6) Department of Ophthalmology, Flinders University, Adelaide, South Australia, Australia.

Primary open-angle glaucoma (POAG), characterised by irreversible optic nerve degeneration, is a major cause of blindness worldwide. Optic nerve parameters including cup area (the central area without nerve fibers), disc area (the total area of optic disc including cup area and the surrounding area containing axons of the retinal ganglion cells), and Vertical cup-disc ratio (the ratio of the cup area to the total diameter of the optic disc) are key measurements used to assess POAG progression. Raised intraocular pressure (IOP) is a known risk factor for POAG. Genome-wide association studies (GWASs) have identified several genetic risk loci associated with POAG, however, these risk loci together explain only a small proportion of the heritability. To identify additional risk loci, we performed a GWAS in 3,071 POAG cases and 6,750 unscreened controls. To increase our statistical power to identify new risk loci for POAG, we performed a meta-analysis of the GWAS summary results for POAG with publically available GWAS data for IOP and optic disc parameters. We identified and validated four novel genome-wide significant associations within or near *MYOF/CYP26A1*, *LINC02052/CRYGS*, *LMX1B*, and *LMO7*. We also used gene-based and pathway-based approaches to identify further genes and relevant cellular mechanisms involved in the development of POAG. We identified two additional new risk loci (*MIR3125* and *C9*) using gene-based tests, and two genetic pathways -"response to fluid shear stress" and "abnormal retina morphology"- in pathway-based tests. Interestingly, some of these new risk loci contribute to risk of other genetically-correlated eye diseases including myopia and age-related macular degeneration. Of further note, *LMX1B* is mutated in Nail-Patella Syndrome, characterised by nail, patella and elbow dysplasia, in which some patients develop POAG. Bioinformatics functional analyses suggested that these loci have functional relevance in the development of POAG. This study highlighted the potential of combining genetic data from genetically-correlated eye traits for the purpose of gene discovery and mapping. The novel risk loci, in addition to the previously known risk loci, will improve risk profiling for glaucoma, providing better opportunities to identify high-risk individuals. Fine-mapping and functional studies of the new risk loci will help better understand the aetiology of POAG, which may lead to improved prevention and intervention strategies.

2166F

Identifying genes that underlie eye disorders and vision loss using predicted gene expression. *J.B. Hirbo[1], E.R. Gamazon[1], P. Evans[1], R. Tao[2], M. Brantley[3], N.J. Cox[1].* 1) Genetic Medicine, Vanderbilt University School of Medicine, Nashville, TN; 2) Biostatistics, Vanderbilt University School of Medicine, Nashville, TN; 3) Vanderbilt Eye Institute, Vanderbilt University School of Medicine, Nashville, TN.

Eye disorders and vision loss pose a large public health and economic burden that is estimated at nearly 140 billion annually. One of the main limiting factors in developing effective therapies for eye disorders is lack of clear understanding of the molecular mechanism in disease etiologies. Genome-Wide Association Studies (GWAS) have successfully identified numerous variants associated with some common eye diseases like glaucoma, age dependent macular degeneration, cataract etc., but the biological mechanism underlying these associations are yet to be elucidated. Over 95% of these variants are noncoding that might be part of instructions that determine when, where, and how much of a gene is expressed. Thus, gene expression and how it correlates with disease can provide a powerful method to prioritize genes involved in the etiology of the eye phenotypes. We used two recently developed gene based method that provide framework for directly correlating imputed gene expression data with individual's genetic profile, PrediXcan and MetaXcan. The imputed gene expression is generated using a transcriptome reference panel. We used both genome wide data generated from biobank samples at Vanderbilt University Medical Center (BioVU) and publicly available data on eye diseases to determine which genes exhibit pattern of altered gene expressions in 5,100 eye disease cases relative to 6,000 unaffected controls using PrediXcan. We also analyzed summary statistics of published GWAS studies of eye diseases using MetaXcan. We identified novel genes that fall in regions with sub-genome wide significant SNPs in GWAS results. We also confirmed genes that were previously suggested to be causal in eye phenotypes because of being in vicinity to loci identified in GWAS. Most of the genes identified are also associated with complex neurological, vascular and metabolic disorders. Our results points to potential pathophysiology that underlie eye diseases.

2167W

Exome sequencing identifies susceptibility genes for chronic central serous chorioretinopathy. *R.L. Schellevis[1], M. Pauper[1,2], D. Zafeiropoulou[2], M.B. Breukink[1], C.B. Hoyng[1], C. Gilissen[2], C.J.F. Boon[3], A.I. den Hollander[1,2], E.K. de Jong[1].* 1) Dept. of Ophthalmology, Donders Institute of Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands; 2) Dept. of Genetics, Donders Institute of Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen,, the Netherlands; 3) Dept. of Ophthalmology, Leiden University Medical Center, Leiden, the Netherlands.

Chronic central serous chorioretinopathy (cCSC) is a rare multifactorial eye disease characterized by subretinal fluid accumulation that leads to vision loss. Clinically, cCSC is associated with stress and corticosteroid use, although it is unclear how this contributes to disease etiology. Strikingly, cCSC is more frequent in males (80%, prevalence 9.9:100,000). Only a limited number of genetic studies have thus far been performed for cCSC, identifying associations with both Single Nucleotide Polymorphisms (SNPs) and common Copy Number Variants (CNVs). However, familial occurrence of cCSC suggests a role for (rare) genetic variants in the disease susceptibility. We therefore performed gene association analysis on exome sequencing data of cCSC patients to elucidate the role of (rare) protein-altering variants in the disease. Exome sequencing was performed using the SureSelect all exon V4 capture library on 269 cCSC patients and 1,586 controls. Data was processed according to the Genome-Analysis-Toolkit (GATK) best practices (v3.4) and batch effects were minimized by filtering according to the recommendations of Carson, AR et al. 2014 prior to variant recalibration in GATK. Population stratification was assessed with a principal component analysis in PLINK (v1.90b4.1), and only subjects of European descent were retained. Cryptically related individuals were removed after identification with KING (v2.0). Variants were annotated with Anno(tation) (v. Sep2011), and protein altering variants were used for downstream analyses. Using the SKAT R-package (v1.2.1), Burden, SKAT and SKAT-O tests were performed. Variants were weighted based on minor allele frequency. Bonferroni correction was applied for 20,000 genes and $P \leq 2.5 \times 10^{-6}$ were deemed statistically significant. After filtering, the dataset contained 261 cCSC patients, 1325 controls and 189,346 protein-altering variants. Analyses corrected for gender and 2 principal components identified 1 significantly associated gene involved in DNA repair and actin dynamics ($P_{burden} = 1.72 \times 10^{-6}$; $P_{SKAT} = 1.33 \times 10^{-7}$; $P_{SKAT-O} = 4.24 \times 10^{-7}$) and 3 suggestive associations with $P < 1 \times 10^{-5}$ in at least one of the tests. With these results, we have confirmed that (rare) protein-altering variants play a role in the etiology of cCSC. We are currently performing additional functional studies to determine the effect of these variants on protein function and to elucidate the exact role of the identified genes in the biological mechanism of cCSC.

2168T

Optimizing accurate classification of electronic health record case control status for age-related macular degeneration in the Million Veteran Program. *S.K. Iyengar[1,2,3], C.W. Halladay[4], T.M. Hadi[3], R.P. Igo, Jr[1], P. Greenberg[5,6], D.C. Crawford[1,2,7], J.M. Sullivan[8], S.J. Fliesler[8], W.C. Wu[9,10], P.E. Konicki[11,12], N.S. Peachey[11,13,14] on behalf of the VA Million Veteran Program.* 1) Dept of Population and Quantitative Health Sciences, Case Western Reserve Univ, Cleveland, OH 44106, USA; 2) Dept of Genetics and Genome Sciences, Case Western Reserve Univ, Cleveland, OH 44106, USA; 3) Dept of Ophthalmology and Visual Sciences, Case Western Reserve University, Univ Hospitals Eye Institute, Cleveland, OH 44106, USA; 4) Center of Innovation in Long Term Services and Supports, Providence VAMC, Providence, RI, USA; 5) Section of Ophthalmology, Providence VA Medical Center, Providence, RI USA; 6) Division of Ophthalmology, Alpert Medical School, Brown University, Providence RI USA; 7) Institute for Computational Biology Case Western Reserve University Wolstein Research Building 2103 Cornell Road, Suite 2527 Cleveland, Ohio 44106, USA; 8) Research Service, VA Western NY Healthcare System, Buffalo, NY, USA; 9) Section of Cardiology, Medical Service, Providence VA Medical Center, Providence, RI USA; 10) Division of Cardiology, Dept. of Medicine, Alpert Medical School, Brown University, Providence RI USA; 11) Louis Stokes Cleveland VA Medical Center, Cleveland, OH, USA; 12) Dept of Psychiatry, Case Western Reserve Univ, Cleveland, OH 44106, USA; 13) Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH, USA; 14) Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA.

Age-related macular degeneration (AMD), a complex oligogenic disease, leads to vision loss and blindness. To foster identification of new genes and validate existing loci for complex diseases, the Million Veteran Program (MVP) was initiated in 2011. To date, >500,000 participants have been enrolled and genotyped with a custom Affymetrix Axiom Biobank array. To classify participants as definite/probable cases, controls or unknown for AMD, we developed algorithms utilizing structured [International Classification of Diseases, 9th and 10th Revisions, Clinical Modification (ICD-9-CM and ICD-10-CM) and Current Procedural Terminology (CPT) codes] and unstructured data (clinical free text) in the electronic health records (EHRs). Imaging data is not yet available in MVP, so our algorithms do not include fundus photography, optical coherence tomography, or other gold-standard tools that clearly define AMD. We developed test sets of N>50 cases and controls across each of three VA medical centers (Cleveland, Providence, Buffalo) and refined the algorithm via detailed chart and imaging reviews by retinal specialists. Of 123,606 participants, cases were ≥50 and controls ≥65 years old with comprehensive eye exams (CPT codes 92004 and 92014). Definite cases (N=26,255) were identified via two separate clinic visits with ICD-9-CM (36.43, 362.50, 362.51,362.52 or 362.57) or ICD-10-CM (H35.30, H35.31, H35.32) codes for AMD; probable cases (N=12,598) with a single code mention are poised for further evaluation with >50 text mining terms to detect AMD in clinical notes. Individuals with no AMD codes but with recent eye exams were designated as controls (N=84,753). Recently, the Cleveland VA validated 79 cases and 69 controls representing European Americans (85.9% and 75.0%), African Americans (AA; 11.5% and 25.0%), and Pacific Islanders (2.6% and 0.0%). Review of their EHRs showed ~14% and ~1% of cases vs. controls were misclassified by sole use of ICD-9-CM and ICD-10-CM codes; sample positive predictive value 86.08%, confidence interval 78.1-91.5%. Three of 11 misclassifications were AA, disproportionate to the number of AA cases observed herein. Additional cases and controls will be examined at Providence and Buffalo to improve the algorithm. In summary, our algorithm led to high misclassification rates for cases. Text mining will be used to improve the algorithm prior to finalizing case and control identity for downstream genome-wide association and rare variant studies.

2169F

Identification of rare sequence variants in genes involved in focal adhesion and Wnt signaling pathways in keratoconus human corneas. *J.A. Karolak[1,2], T. Gambin[3,4], M. Rydzanicz[5], P. Stawinski[5], P. Polakowski[6], R. Ploski[5], J.P. Szaflik[6], M. Gajecka[1,2].* 1) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland; 2) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland; 4) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; 5) Department of Medical Genetics, Medical University of Warsaw, Warsaw, Poland; 6) Department of Ophthalmology, Medical University of Warsaw, Warsaw, Poland.

Keratoconus (KTCN, OMIM 148300) is a protrusion and thinning of the cornea, resulting in loss of visual acuity. The etiology of KTCN condition remains unclear. To assess a potential genetic contribution to KTCN, corneal tissues derived from five Polish unrelated individuals with KTCN were examined using whole exome sequencing (WES). Variants were filtered in a step-wise manner to exclude variants with minor allele frequency greater than 0.001 in the public and our internal database, and to exclude variants predicted as neutral by prediction tools. Genes with potential causative variants were enriched in 14 pathways, including focal adhesion and Wnt signaling (c.1066C>A in *PTK6*, c.1063G>T in *WNT1*, c.4204A>C in *COL5A3*, and c. 4822C>T in *LRP6*). All patients have rare variant in at least one gene from focal adhesion and/or Wnt signaling pathways. These pathways play a role in corneal organization and were previously recognized as involved in KTCN. This first WES profiling of human KTCN corneas indicates that the accumulation of sequence variants in several genes from focal adhesion and Wnt signaling pathways might cause the phenotypic effect and supports the importance of previously described molecular pathways in KTCN etiology.Supported by National Science Centre in Poland, Grant 2013/10/M/NZ2/00283.

2170W

Updated carrier rates for deafness-inducing mutation c.35delG (*GJB2*) in Russia and common haplotypes associated with c.35delG in Siberia. *O. Posukh[1,2], M. Zytsar[1,2], N. Barashkov[3,4], O. Shubina-Olejnik[5], N. Danilenko[5], V. Danilchenko[1,2], A. Solovyev[3,4], I. Morozov[2,6], A. Bondar[6].* 1) Federal Research Center Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russian Federation; 2) Novosibirsk State University, Novosibirsk, Russian Federation; 3) Yakut Scientific Center of Complex Medical Problems, Yakutsk, Sakha Republic, Russian Federation; 4) M.K. Ammosov North-Eastern Federal University, Yakutsk, Sakha Republic, Russian Federation; 5) Institute of Genetics and Cytology, National Academy of Sciences, Minsk, Belarus; 6) Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russian Federation.

Mutations in gene *GJB2* (13q11-q12) are the major causes of deafness and their spectrum and prevalence are specific for various populations. Mutation c.35delG is more frequent in patients of European origin and absent or found with very low frequencies in Asian populations. There is generally accepted hypothesis that c.35delG is an "old" mutation that arose ~ 10000-14000 years ago from a common ancestor in Middle East or the Mediterranean and spread throughout Europe with Neolithic migrations. This hypothesis is supported by the presence of common STR- and SNP-haplotypes for c.35delG in individuals of European descent. Recently, we evaluated mutational spectrum and *GJB2*-contribution to deafness in patients of different ethnicities living in Altai, Tuva (Southern Siberia) and Yakutia (North-Eastern Siberia) and found predominance of c.35delG among Russian patients in contrast to indigenous Siberian peoples. Carrier frequency of c.35delG was found to be 4.1% in Russian residents of Novosibirsk city (Western Siberia). Our data supplement scarce information on the c.35delG prevalence in Siberia and generally among ethnically heterogeneous populations of Russia. Carrier frequency of c.35delG in Russians varies from 7.5% in north-western part of Russia to 2.5% in Yakutia averaging as 4-5% in central part of Russia. Therefore, Siberia can be considered as eastern "endpoint" of the c.35delG prevalence in Eurasia. Genotyping of intragenic rs3751385 and the STRs flanking *GJB2* (D13S141, D13S175, D13S1853) in homozygotes for c.35delG and in control individuals living in Siberia revealed two common specific haplotypes strongly associated with c.35delG: 126-T-105-202 (37.5%) and 124-T-105-202 (25.0%) (~ 316 kb). Moreover, comparative analysis revealed common conservative region between D13S141 and D13S175 (~ 125 kb) among the most frequent c.35delG-haplotypes in deaf patients living far from each other (Belarus, Volga-Ural region of Russia and Siberia). We roughly estimated the age of c.35delG in Siberia as ~ 8800-6800 years ago. The settlement of Siberia by the Russians began only in the late 16th century and contemporary Siberian population of Caucasian descent was formed as a result of multiple migrations from European part of Russia. Thus, roughly estimated age of the c.35delG in Siberia can rather reflect complex processes of early formation of population of Europe. Study was supported by Project #0324-2015-0004 and the RFBR grants #14-04-90010 Bel_a, #15-04-04860a.

2171T

Forty novel genetic loci associated with intraocular pressure in a large multi-ethnic genome-wide association study. *H. Choquet[1], K.K. Thai[1], J. Yin[1], T.J. Hoffmann[2,3], M.N. Kvale[2], Y. Banda[2], C. Schaefer[1], N. Risch[1,2,3], R. Melles[4], E. Jorgenson[1].* 1) Kaiser Permanente Northern California (KPNC), Division of Research, Oakland, CA; 2) Institute for Human Genetics, University of California San Francisco (UCSF), San Francisco, CA; 3) Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA; 4) KPNC, Department of Ophtalmology, Redwood City, CA.

Purpose: Intraocular pressure (IOP) is associated with glaucoma, a common eye condition which can lead to vision loss. Family studies indicate that IOP has a moderate genetic component, with heritability estimates ranging from 0.29 to 0.67. Previously published genetic association studies have reported 11 loci associated with IOP, suggesting that additional loci remain to be discovered. **Methods:** We first conducted a trans-ethnic genome-wide association study (GWAS) linking 373,967 IOP measurements to 73,107 individuals from four race/ethnicity groups (non-Hispanic white, Latino, East Asian, and African American) in the Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort. Confirmation of our findings was then conducted in an external sample consisting of 37,930 individuals of European and Asian descent. We also conducted genome-wide conditional analysis by including all lead single nucleotide polymorphisms (SNPs) identified in the GWAS analyses of IOP as covariates. Epistasis analysis of all pairs of lead SNPs was also performed. We also investigated a total of 13 SNPs associated with IOP at a genome-wide significance level ($P < 5 \times 10^{-8}$) in previous studies. Array heritability was estimated in the non-Hispanic white sample using GCTA v1.26.0, and the proportion of IOP variation explained by our newly-identified SNPs was estimated using a genetic risk score. **Results:** We identified 47 genome-wide significant IOP-associated loci, of which 40 were novel, that is, not previously-reported to be associated with IOP. Thirty novel IOP-associated loci (75%) replicated in an external sample. Conditional analysis did not identify any additional genome-wide significant signals for IOP, and epistasis analysis did not detect any significant epistatic interactions between IOP SNPs after Bonferroni correction. Our results also confirmed 92% of previously identified IOP-associated loci. Array-based heritability of IOP explained by all autosomal variants was estimated to be 24%, and our newly-identified SNPs increased the proportion of variance explained from 1.1% to 3.7%. **Conclusions:** The large number of new and replicated IOP-associated loci represents a major step forward to understanding the role of genetic variation in this important risk factor for glaucoma.

2172F

A transethnic genome-wide association study identifies five novel genetic loci associated with primary open angle glaucoma. *E. Jorgenson[1], K.K. Thai[1], J. Yin[1], T.J. Hoffmann[2,3], M.N. Kvale[2], Y. Banda[2], C. Schaefer[1], N. Risch[1,2,3], R. Melles[4], H. Choquet[1].* 1) Division of Research, Kaiser Permanente, Oakland, CA; 2) Institute for Human Genetics, University of California San Francisco, San Francisco; 3) Department of Epidemiology and Biostatistics, UCSF; 4) KPNC, Department of Ophthalmology, Redwood City, CA, USA.

Purpose: Glaucoma is the second leading cause of blindness worldwide. Genome-wide association studies have reported a number of loci associated with primary open angle glaucoma (POAG), the most common type of glaucoma, but the discovery of additional risk loci has been limited by available sample size. **Methods:** We conducted a transethnic meta-analysis of primary open angle glaucoma (POAG) in 5,235 cases and 65,213 controls, all adults from four race/ethnicity groups (non-Hispanic white, Latino, East Asian, and African American) in the Kaiser Permanente Northern California (KPNC) Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort. We identified POAG cases using KPNC longitudinal electronic health records as those GERA participants who had at least two diagnoses of primary open angle or normal tension glaucoma by a KPNC ophthalmologist on separate days and no diagnoses of any other type of glaucoma. Controls had a history of at least one vision examination at which an intraocular pressure measurement was taken and no diagnoses of any type of glaucoma, nor any diagnosis of ocular hypertension. We conducted a genome-wide association study of POAG, using genome-wide genotype data generated on Affymetrix Axiom arrays and imputed to the 1000 Genomes reference panel. Genetic association analyses with POAG were performed using logistic regression adjusted for age, sex, and ancestry principal components. **Results:** We identified 14 loci with at least one associated variant reaching genome-wide significance (P < 5.0 x 10[-8]), of which 5 were novel. We confirmed 9 POAG-loci previously reported, including *TMCO1*, *AFAP1*, *GMDS*, *CDKN2B-AS1*, *ABCA1*, *TMTC2*, *SIX1-SIX6*, *GAS7*, *TXNRD2*. We conducted replication analysis of the lead SNP in the novel loci using summary statistics from published GWAS studies. **Conclusions:** The identification of 5 new loci underlying primary open angle glaucoma represents an important step toward understanding glaucoma genetic etiology.

2173W

Trial to make the combined genotyping to detect high-risk individuals for cold medicine related Stevens-Johnson syndrome (CM-SJS) with severe ocular complications (SOC). *M. Ueta[1], K. Tokunaga[2], C. Inoue[1], H. Sawai[2], C. Sotozono[3], S. Kinoshita[1].* 1) Department of Frontier Medical Science and Technol, Kyoto Prefectural University of Medicine, kyoto, Japan; 2) Department of Human Genetics, Graduate School of Medicine, University of Tokyo; 3) Department of Ophthalmology, Kyoto Prefectural University of Medicine.

In future, it is possible that combined genotyping for the associated variants may help to predict the risk for CM-SJS/TEN in Japanese individuals. We previously reported that in Japanese individuals, *HLA-A*02:06* was strongly associated with CM-SJS/TEN. We also documented that *HLA-A*02:06* and T*LR3* polymorphisms exerted more than additive effects, and that *HLA-A*02:06* and *PTGER3* polymorphisms exhibited additive effects in CM-SJS/TEN with SOC. We recently reported that *HLA-A*02:06* and *Rec114* rs16957893 CG also exerts more than additive effects in CM-SJS/TEN with SOC (OR = 110, p = 4.45×10[-8]). These findings might suggest that there are many interactions between polymorphisms associated with CM-SJS/TEN with SOC, and these interactions might show high odds ratio. In this study, we tried to make a prediction system to detect high-risk individual for cold medicine related Stevens-Johnson syndrome. We selected the 81 SNPs which showed associations with p<10[-5] with CM-SJS/TEN with SOC in Japonica array. We also selected 19 HLA-A, B, C alleles which tended to be associated with CM-SJS/TEN with SOC. Using the 19 HLA class I alleles and the 81 SNPs of 117 cases and 628 controls, we made pairs of HLA-SNP, SNP-SNP, HLA-HLA, and calculate p values and odds ratio, respectively. We found about 70 pairs which showed p<10[-5] and OR>15, although maximum OR was 6.4 for HLA-A*02:06 in separate analysis. Thus, we could find many interaction between the polymorphisms, although our sample numbers were small. In future, it is possible that combined genotyping for the associated variants may help to predict the risk for CM-SJS/TEN with SOC in Japanese individuals.

2174T

Loss of ELOVL6, a fatty acid elongase, rescues ER stress-induced apoptosis in model of retinitis pigmentosa. *R.A.S. Palu, C.Y. Chow.* Human Genetics, University of Utah, Salt Lake City, UT.

An important goal of the Precision Medicine Initiative is to address the phenotypic heterogeneity that impedes diagnosis and treatment in both Mendelian and complex genetic diseases. Individuals with the same causative mutation may display vast differences in disease penetrance and expressivity. Cryptic genetic variation is a key contributor to this heterogeneity, but the underlying genetic architecture and modifiers are largely unknown. Understanding the roles of genetic modifiers in disease processes will enable the development of individualized therapeutic approaches. Recently developed model organism tools allow researchers to study the effects of natural genetic variation on disease phenotypes. One of these, the *Drosophila* Genetic Reference Panel (DGRP), is a collection of ~200 inbred strains, each representing a unique and fully sequenced wild-derived genome. In a previous study, we utilized the DGRP to study the effects of natural genetic variation on a model of ER stress-associated retinal degeneration. Overexpression of a mutant, misfolded rhodopsin protein ($Rh1^{G69D}$) induces ER stress in the developing eye, ultimately resulting in apoptosis and a small, degenerate eye in the adults. Degenerate eye phenotypes were incredibly variable among the 200 DGRP strains. Using genome-wide association methods, we identified 84 conserved candidate modifiers, many of which are associated with ER stress and apoptosis. One of these, *baldspot*, is the *Drosophila* orthologue of *ELOVL6*, an ER-associated fatty acid elongase. ELOVL6 activity has been linked to ER stress, insulin sensitivity, and obesity in mammals. We demonstrate that RNAi-mediated knockdown of *ELOVL6* alleviates degeneration in the $Rh1^{G69D}$ model of retinal degeneration. This is associated with a reduction in ER stress and apoptosis and appears to be linked to the fatty acid elongase activity of ELOVL6. Furthermore, loss of *ELOVL6* in other tissues also reduces degenerative phenotypes induced by ER stress, suggesting that ELOVL6 is not retinal-specific, but a general modifier of ER stress. We hypothesize that ELOVL6 acts by altering the fatty acid composition of the ER and the activity of IRE1. We will present evidence from cell culture experiments for this. Our findings suggest that ELOVL6 is a general modifier of ER stress that functions in a variety of disease contexts. ELOVL6 may be a possible therapeutic target not only in retinal degeneration, but also in many ER stress-associated diseases.

2175F

Evaluation of genetic polymorphisms in the determination of optic disc parameters and circumpapillary retinal nerve fiber layer thickness in normal individuals in a sample of the Brazilian population. *M. Atique-Tacla, C. Zangalli, B.B. Souza, D. Stancato, G. Ananina, V.P. Costa, M.B. Melo, J.P.C. Cabral.* Unicamp, Campinas, Campinas, Brazil.

Background: Glaucoma is a degenerative disease of the optic nerve considered the leading cause of irreversible blindness worldwide. Primary open angle glaucoma (POAG[MIM137760]) is the most common subtype in African and European descendants and is a complex heterogeneous disease with molecular mechanisms poorly understood. The study of endophenotypes in glaucoma offers the possibility of reducing its genetic complexity and includes optic nerve head (ONH) parameters, such as optic disc area (ODA) and neuroretinal rim area, and circumpapillary retinal nerve fiber layer thickness (RNFLT), among others. *SALL1* [602218], *ATOH7* [609875]*, CDC7/TGFRB3* [603311] and *CARD10* [607209] polymorphisms have already been associated to ODA. Anatomic differences in the ONH between normal individuals, specially when they belong to different ethnic groups, as well as a non standardized evaluation of ONH parameters might have led to less precise genetic associations in previous studies. **Objectives**: to evaluate the association of *SALL1, CDC7/TGFRB3, CARD10* and *ATOH7* polymorphisms with ODA, global neuroretinal rim, circumpapillary RNFL and axial length, in a cohort of 195 normal individuals (without glaucoma). **Methods**: The project was approved by the Research Ethics Committee of the Faculty of Medical Sciences of the State University of Campinas Genomic. DNA was amplified by polymerase chain reaction and the polymorphisms were analyzed by Sanger sequencing. ONH images were acquired by spectral domain optic coherence tomography - SD-OCT- Spectralis (Heidelberg Engineering, GmbH, Heidelberg, Germany). **Results**: To evaluate if each of the dependant variables could be associated to any of the four polymorphisms, a multiple linear regression analysis with a stepwise backward elimination was performed. *SALL1 rs1362756* genotypes G/C and G/G were significantly associated with ODA (p = 0.035 and p = 0.01, respectively) and with global neuroretinal rim area (p = 0.016 and p = 0.032, respectively). There were no significant associations with this SNP to other parameters analyzed (axial length and RNFL thickness). Other SNPs included in our study (*ATOH7 rs 1900004, CDC7/TGFRB3 rs1192415* and *CARD10 rs9607469*) did not present association with any of the parameters evaluated. **Conclusions**: *SALL1 rs1362756* SNP presented a significant association with optic disc area and with global rim area in a cohort of normal individuals.

2176W

Lineage-specific linkage analysis localizes novel rare variant-driven genomic loci for the glaucoma endophenotype of cup to disc ratio in a large extended pedigree from Nepal: The Jiri Eye Study. *M.P. Johnson[1], S.S. Thapa[2], T.D. Dyer[1], S. Laston[1], K.L. Anderson[3], B. Towne[4], J. Subedi[5], H.H.H. Göring[1], J. Blangero[1], S. Williams-Blangero[1].* 1) South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA; 2) Tilganga Institute of Ophthalmology, Kathmandu, Nepal; 3) Department of Ophthalmology, School of Medicine, UT Health San Antonio, San Antonio, TX, USA; 4) Department of Community Health, Boonshoft School of Medicine, Wright State University, Kettering, OH, USA; 5) Department of Sociology and Gerontology, College of Arts and Science, Miami University, Oxford, OH, USA.

The leading cause of irreversible blindness worldwide is glaucoma, a heterogeneous group of eye diseases that results in progressive damage to the optic nerve, loss of retinal ganglion cells, and specific visual field deficit. Glaucoma development and progression is governed by a complex interplay of genetic and environmental risk factors. Identifying genetic risk factors for a heterogeneous disease such as glaucoma is best achieved by dissecting the genetic architecture of its biological constituents (endophenotypes) that contribute to the disease process. Intraocular pressure (IOP), central corneal thickness (CCT), and optic nerve cupping as measured by cup to disc ratio (CDR) are well established endophenotypes for glaucoma. The Jiri Eye Study is an ongoing, family-based study (target sample size: 2,000) with the objective of identifying rare functional variants that influence ocular health and disease in the Jirel population of eastern Nepal. In pursuit of genomic loci (QTLs) that harbor rare functional variants influencing glaucoma endophenotype trait variation in the Jirels, we adopt a novel analytical strategy that is based on founder-specific lineages and is able to identify founders who segregate rare functional variants. The model employed is a variance component mixed model with a fixed effect for a given founder's identity-by-descent (IBD) probability vector that is extracted from the usual IBD matrix. In our current sample size (n = 1,047), IOP (h^2 = 0.29, p = 4.0E-07), CCT (h^2 = 0.65, p = 4.9E-28), and CDR (h^2 = 0.48, p = 3.9E-16) are all significantly heritable. Using existing genome-wide markers, we identify five suggestive (p-values from 0.001 to 0.004) CDR QTLs likely due to rare variation on chromosomes 1 (2 QTLs), 7, 9 and 12. The chromosome 1 loci harbor genes (*TMCO1*; ~0.28 Mb from our maximum signal) previously identified in glaucoma and IOP GWAS, and genes (*NENF*; ~0.70 Mb from our maximum signal) that encode proteins with known/hypothesized roles in the glaucomatous process; the chromosome 7 locus harbors genes (*TMEM248*; ~1.18 Mb from our maximum signal) previously identified in CCT GWAS; the chromosome 9 locus has been previously implicated in juvenile-onset primary open angle glaucoma (GLC1J locus); and the chromosome 12 locus shares some overlap with a previously detected CDR QTL. To our knowledge, the loci on chromosomes 1, 7, and 9 have not been implicated in CDR trait variation.

2177T

***ANGPT1* association with adult-onset primary open angle glaucoma.** *J.N. Cooke Bailey[1], P. Gharahkhani[2], S.W. Tompson[3], T. Souma[4,5], O.M. Siggs[6], T.L. Young[3], A.P. Tanna[7], L.R. Pasquale[9], S.E. Quaggin[4,5], S. MacGregor[2], J.E. Craig[6], J.L. Haines[1], F. Pasutto[8], J.L. Wiggs[9], NEIGHBORHOOD Consortium.* 1) Department of Epidemiology and Biostatistics, Case Western University, Cleveland, OH, USA; 2) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 3) Ophthalmology and Visual Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA; 4) Feinberg Cardiovascular Research Institute, Northwestern University, Chicago, Illinois, USA; 5) Division of Nephrology/Hypertension, Northwestern University, Chicago, Illinois, USA; 6) Department of Ophthalmology, Flinders University, Adelaide, South Australia, Australia; 7) Department of Ophthalmology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; 8) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 9) Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA.

Primary open-angle glaucoma (POAG), a major worldwide cause of blindness, is a genetically and clinically heterogeneous disease. POAG has a significant genetic component; however, the 15 genomic regions associated with POAG risk account for only a fraction of disease heritability, suggesting the existence of additional disease-related genes. Recently, *TIE2* (*TEK*) loss-of-function mutations have been identified in congenital glaucoma families with age of disease onset varying from birth to adulthood. The purpose of this study is to examine POAG association for SNPs located in the genomic regions that include *TIE2* and its ligand *ANGPT1* as well as a related receptor ligand pair *ANGPT2* and *TIE1*. We first accessed results from two previously reported POAG imputed genome-wide association study (GWAS) datasets: the NEIGHBORHOOD consortium (3,853 POAG cases, 33,480 controls) and the Australian ANZRAG consortium (1,155 POAG cases, 1,992 controls). We assessed 2,302 SNPs located within the genomic regions corresponding to *TIE1, TIE2, ANGPT1*, and *ANGPT2* in the NEIGHBORHOOD; SNPs with $P<1\times10^{-5}$ were carried forward to ANZRAG. We additionally accessed results from Illumina HumanExome chip data in the NEIGHBORHOOD (2836 cases, 3207 controls) as well as sequence data from 528 German POAG cases. We detected significant association ($P < 2.17 \times 10^{-5}$, correcting for multiple testing of 2,302 SNPs) at 19 SNPs located in the genomic region around *ANGPT1* (top SNP rs34597339[T], $P = 3.24\times10^{-6}$, OR = 0.81). Thirteen of these SNPs are located in a 15-kb promoter region (ENCODE) that includes an alternative transcription start site used by various cell types, including human umbilical vein endothelial cells (HUVEC). We did not observe significant association with POAG at SNPs located in the *TIE1, TIE2*, and *ANGPT2* genomic regions ($P>2.17\times10^{-5}$) in the NEIGHBORHOOD GWAS. Further evidence of replication was detected in three nonsynonymous variants in *ANGPT1* in three individuals with POAG with onset >60 years and high intraocular pressure. Prior studies have shown that ANGPT1 and TIE2 are necessary for development of Schlemm's canal and that loss-of-function mutations in *TIE2* cause glaucoma with variable age of onset. The results from our study highlight SNPs located in the *ANGPT1* genomic region that are associated with POAG, which suggests that modulation of ANGPT1-TIE2 signaling may also contribute to the development of adult-onset glaucoma.

2178F

Common variants in *KLHL2* and *C4orf50* are associated with poorer anti-VEGF treatment response in age-related macular degeneration. *O. Garcia Rodriguez[1], S.S. Pan[1], P. Whitehead-Gay[1], L.D. Adams[1], J.K. Welch[1], R.A. Laux[2], J.A. Fortun[3], M.A. Brantley, Jr.[4], J.L. Kovach[3], S.G. Schwartz[3], A. Agarwal[5], J.L. Haines[2], M.A. Pericak-Vance[1], W.K. Scott[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 3) Bascom Palmer Eye Institute, University of Miami. MIami, FL; 4) Ophthalmology & Visual Sciences, Vanderbilt University Medical Center, Nashville TN; 5) West Coast Retina, San Francisco, CA.

Age related macular degeneration (AMD), a leading cause of vision loss in older adults, has a complex etiology comprising genetic and environmental risk factors. In advanced stages, vision loss results from geographic atrophy or choroidal neovascularization (CNV). Response to anti-vascular endothelial growth factor (anti-VEGF) treatment for CNV is variable, with 10-30% of patients continuing to lose visual acuity (VA) after 6 months' treatment. Although a replicated association with the rs4910623 variant in *OR52B4* was recently reported in a genome wide association study (GWAS), there are few genetic factors reproducibly associated with treatment response. Therefore, we conducted a second GWAS in 90 individuals (107 eyes) with CNV followed for at least one year. Samples were genotyped using the Illumina CoreExome Array and imputed against the 1000Genomes reference panel; after quality control, 6.3 million SNPs with minor allele frequency > 0.05 were retained for analysis. Association of one-year change in visual acuity (VA, analyzed as logMAR) with SNP dosage was analyzed using linear mixed models, adjusting for age at exam, antibody type (ranibizumab, bevacizumab, or aflibercept), and sex, and allowing for correlations between eyes for individuals with bilateral CNV. Two regions on chromosome 4 were identified as significantly associated ($p<5\times10^{-8}$) with change in VA: 2 intronic variants in *KLHL2* and 2 variants in the 3'UTR of *C4orf50*. In both regions, the two SNPs were in complete linkage disequilibrium and the minor allele was associated with reduced VA after 1 year. Imputation quality at all 4 SNPs was good (R-sq > 0.75) and validation with TaqMan genotyping showed strong correlation with imputed dosages. Association was not detected at rs4910623 in *OR52B4* (nominal p>0.05) or at any of the 52 known AMD risk variants (Bonferroni-corrected p>0.05) possibly due to limited power in the relatively small sample. Our findings identified new loci specific to anti-VEGF treatment response in CNV, and suggest that mechanisms governing risk of AMD and response to treatment may be distinct.

2179W

Additive effects of genetic variants associated with primary open-angle glaucoma. *F. Mabuchi[1], N. Mabuchi[1], Y. Sakurada[1], S. Yoneyama[1], K. Kashiwagi[1], H. Iijima[1], Z. Yamagata[2], M. Takamoto[3], M. Aihara[4], T. Iwata[5], K. Kawase[6], Y. Shiga[7], K. Nishiguchi[8], T. Nakazawa[7,8,9], M. Ozaki[10], M. Araie[11] for the Japan Glaucoma Society Omics Group (JGS-OG).* 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Health Sciences, Univ Yamanashi, Chuo, Yamanashi, Japan; 3) Dept Ophthalmology, Tokyo Metropolitan Police Hospital, Nakano-ku, Tokyo, Japan; 4) Dept Ophthalmology, Graduate School of Medicine, Univ Tokyo, Bunkyo-ku, Tokyo, Japan; 5) Division of Molecular and Cellular Biology, National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Center, Meguro-ku, Tokyo, Japan; 6) Gifu University Hospital, Gifu, Gifu, Japan; 7) Dept Ophthalmology, Tohoku Univ Graduate School of Medicine, Sendai, Miyagi, Japan; 8) Dept Advanced Ophthalmic Medicine, Tohoku Univ Graduate School of Medicine, Sendai, Miyagi, Japan; 9) Dept of Retinal Disease Control, Tohoku Univ Graduate School of Medicine, Sendai, Miyagi, Japan; 10) Ozaki Eye Hospital, Hyuga, Miyazaki, Japan; 11) Kanto Central Hospital of the Mutual Aid Association of Public School Teachers, Setagaya-ku, Tokyo, Japan.

Purpose: Primary open-angle glaucoma (POAG) is clinically classified into high tension glaucoma (HTG), in which elevated intraocular pressure (IOP) is a major feature and normal tension glaucoma (NTG), in which the IOPs are consistently within the statistically normal population range. Genome-wide association studies identified several genetic variants associated with POAG, and these can be classified into two types of genetic variants. One is a genetic variant associated with IOP elevation (IOP-related genetic variant), while the other is a genetic variant associated with a vulnerability of optic nerve (non-IOP-related genetic variant). The present study was conducted to investigate the additive effects of genetic variants associated with POAG. Methods: Nine variants identified as IOP-related genetic variants, including rs1052990 (near gene: *CAV2*), rs11656696 (*GAS7*), rs59072263 (*GLCCI1/ICA1*), rs2472493 (*ABCA1*), rs58073046 (*ARHGEF12*), rs2286885 (*FAM125B*), rs6445055 (*FNDC3B*), rs8176743 (*ABO*), and rs747782 (*PTPRJ*), and 11 variants identified as non-IOP-related genetic variants, including rs3213787 (*SRBD1*), rs735860 (*ELOVL5*), rs1063192 (*CDKN2B*), rs1900004 (*ATOH7*), rs10483727 (*SIX6*), rs4619890 (*AFAP1*), rs11969985 (*GMDS*), rs3785176 (*PMM2*), rs1192415 (*TGFBR3*), rs2745572 (*FOXC1*), and rs35934224 (*TXNRD2*), on genome-wide association studies, were genotyped for 762 Japanese participants, including the patients with NTG (n = 261) and HTG (n = 255) and 246 control subjects. The total number of risk alleles of the 9 IOP-related, 11 non-IOP-related, and 20 IOP-related or non-IOP-related genetic variants were calculated for each participant as genetic risk scores (GRSs), and the association between these GRSs and POAG (NTG or HTG) was evaluated. Results: The GRS (9.1 ± 1.9, mean ± standard deviation) of 9 IOP-related variants in patients with HTG was significantly higher (P = 0.011) than that (8.7 ± 1.8) in control subjects. The GRSs of 11 non-IOP-related variants in patients with NTG (14.7 ± 1.9) and HTG (14.8 ± 1.7) were significantly higher (P < 0.0001) than that (14.0 ± 2.0) in control subjects. The GRSs of 20 IOP-related or non-IOP-related variants in patients with NTG (23.6 ± 2.6) and HTG (23.9 ± 2.5) were significantly higher (P < 0.0001) than that (22.7 ± 2.6) in control subjects. Conclusion: These data support that POAG is a polygenic disorder by additive effects of multi-locus genetic variants associated with POAG.

2180T

Genome-wide gene expression profiling in the retina of common marmosets exposed to hyperopic or myopic defocus reveals large-scale sign-of-defocus-specific changes in gene expression, pathway switching, and strong left-right eye yoking. *A. Tkatchenko[1,2], D. Troilo[3], A. Benavente-Perez[3], T. Tkatchenko[1].* 1) Dept of Ophthalmology, Columbia University, New York, NY; 2) Dept of Pathology and Cell Biology, Columbia University, New York, NY; 3) SUNY College of Optometry, New York, NY.

Previous studies demonstrated that the eye can compensate for both minus- and plus-lens defocus. However, the ability of the retina to recognize the sign of defocus has been a subject of much controversy. Comprehensive information about transcriptome changes underlying compensation for lens imposed defocus has also been lacking. The purpose of this study was to identify sign-of-defocus-specific transcriptome changes in the primate retina and elucidate retinal signaling pathways underlying lens compensation at the whole-genome level. In 4 groups of marmosets (3 animals each), -5D or +5D lenses were applied to the right eye for 10 days or 5 weeks. The left eye was fitted with a plano lens control. Following lens treatment, retinae were harvested and used to perform whole-genome gene expression profiling using RNA-seq. A total of 119 genes were differentially expressed between -5D-treated eyes and the contralateral plano control eyes after 10 days of treatment, 309 genes were differentially expressed between -5D eyes and control eyes after 5 weeks of treatment. In +5D eyes compared to controls, 79 genes were differentially expressed after 10 days, and 740 genes were differentially expressed after 5 weeks of treatment. Comparison of differential genes from the four experimental groups revealed that each experimental condition leads to differential expression of a unique set of genes with very little overlap between the datasets and a large degree of yoking between left and right eyes. Although no genes were differentially expressed in opposite directions between -5D and +5D lens treatment groups at 10 days, 13 genes were differentially expressed in opposite directions at 5 weeks. Pathway analysis revealed that there was almost complete transition from one set of pathways to another between 10 days of treatment and 5 weeks of treatment for both -5D and +5D experimental groups. Our data suggest that retina can distinguish between hyperopic and myopic defocus and responds to defocus of different sign by activation of largely distinct pathways. There is also, however, strong yoking between left and right eyes at the gene expression level. The response to both hyperopic and myopic defocus over time is characterized by signaling pathway switching.
.

2181F

Using GWAS data to identify copy number variants associated with orofacial clefts. *A.L. Petrin[1], L.J.J. Gowans[2], P.A. Mossey[3], W.L. Adeyemo[4], M.A. Eshete[5], T.D. Busch[6], R.J. Petrin[6], P. Donkor[2], A. Owais[7], M.L. Marazita[8], A.A. Adeyemo[9], J.C. Murray[6], A. Butali[10].* 1) Dows Research Institute, University of Iowa, Iowa City, Iowa; 2) Cleft Clinic, Komfo Anokye Teaching Hospital and Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; 3) Department of Orthodontics, University of Dundee, Dundee, UK; 4) Department of Oral and Maxillofacial Surgery, University of Lagos, Lagos, Nigeria; 5) Addis Ababa University, School of Public Health, Addis Ababa, Ethiopia; 6) Department of Pediatrics, University of Iowa, Iowa City, IA, USA; 7) Department of Pediatric Dentistry, University of Iowa, Iowa City, IA, USA; 8) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, U.S.A; 9) National Human Genomic Research Institute, Bethesda, Maryland, USA; 10) Department of Oral Pathology, Radiology and Medicine, University of Iowa, Iowa City, IA, USA.

Orofacial clefts (OFCs), including cleft lip and/or palate (CL/P) and cleft palate (CP), are among the most common birth defects in humans. Approximately 1 in 700 newborns are affected; however, the prevalence of OFCs differs widely across populations. Although genome-wide association studies have successfully identified at least 26 risk loci for OFCs, these loci only account for part of the heritability, leaving a gap to be filled by contributions of rare variants, copy number variants (CNVs), epigenetics or interactions. CNVs are segments of the genome that exist in different copy numbers in the population. Copy number variations (or CNVs) can encompass genes leading to dosage imbalances and therefore, contribute to an abnormal phenotype. Here, we report the ongoing genome wide survey of CNVs associated with OFCs using the data from the first GWAS for OFCs in the African population. A total of 3,178 participants were genotyped on the pre-release consortium version of Illumina Multi Ethnic Genotyping Array (MEGA), and imputation was done into the 1000 Genomes Phase 3 reference imputation panel using IMPUTE2. These included 814 cases of CLP, 205 cases of isolated CP, and 2,159 related and unrelated controls. Over 45million SNPs were imputed including the 2.2million SNPs in the Multi Ethnic Genotyping Array. We are using PennCNV combined with statistical analysis to identify potential CNVs associated with OFCs. These analyses integrate multiple sources of information to infer CNVs call for individual samples, in addition to using family information to generate family based calls and to determine the *de* novo or inherited origin of the CNV. The ascertainment of CNVs from SNP array data offer an excellent opportunity to identify common CNVs associated with common diseases and this has resulted in the discovery of both de novo and inherited CNVs. Our CNV analysis on GWAS data will help elucidate the molecular etiology of orofacial clefts, and will improve our understanding of how multiple forms of genetic variation act together to influence these craniofacial phenotypes.

2182W

Fluctuating dermatoglyphic asymmetry and familial recurrence of cleft lip/palate in a high-prevalence cluster of South America. *J. Ratowiecki[1], J.S. Lopez Camelo[1,2,3], F.M. Carvalho[3,4,5], C.A. Brandon[6,8], A.R. Vieira[6,7,8], M.L. Marazita[6,8,9], I.M. Orioli[3,5], E.E. Castilla[1,2,3,4], F.A. Poletta[1,2,3].* 1) ECLAMC (Latin-American Collaborative Study of Congenital Malformations at CEMIC-CONICET, Centro de Educación Médica e Investigaciones Clínicas, Buenos Aires, Argentina; 2) CONICET (National Research Council of Argentina), Argentina; 3) INAGEMP (Instituto Nacional de Genética Médica Populacional), Brazil; 4) ECLAMC at Laboratório de Epidemiologia de Malformações Congênitas, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; 5) ECLAMC at Laboratório de Malformações Congênitas, Departamento de Genética, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; 6) Departments of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 7) Pediatric Dentistry, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 8) Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 9) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA.

Non-syndromic orofacial clefts, notably cleft lip with or without palate (CL/P), are common congenital anomalies with complex and heterogeneous etiology. A high prevalence of CL/P was previously identified in the Argentine Patagonia, probably associated with Amerindian ancestry and low socioeconomic status. While it is known that there is a relationship between abnormal dermatoglyphics and birth defects, there are relatively few studies investigating dermatoglyphic asymmetry and familial risk of CL/P, and as far as we know, this is the first report in South America populations. The aim of this study was to compare levels of fluctuating dermatoglyphic asymmetry among non-syndromic CL/P probands with different degrees of family recurrence, versus unrelated controls, from a high-CL/P-prevalence cluster in South America. The sample included 168 individuals (84 CL/P cases and 84 unrelated controls) ascertained by ECLAMC (The Latin-American Collaborative Study of Congenital Malformations) hospitals in Patagonia. Dissimilarity score of fingerprint patterns was used as an indicator of bilateral fluctuating asymmetry. For each CL/P case we calculated the family history score (FHS) from pedigree data, using the observed risk in first-degree relatives, and the expected frequency according to family size and population risk prevalence. Individuals were then distributed in 3 groups: (1) Unrelated controls (without family history of CL/P), (2) CL/P cases with low FHS, and (3) CL/P cases with high FHS. Means and standard deviations were calculated for each FHS group and sex. Differences among groups were analyzed with two-way ANOVA including FHS group, sex, and an interaction term (FHS group*sex). We observed statistically significant differences between FHS groups (p = 0.025) with a higher asymmetry in cases with higher FHS; but not between sexes, nor evidence of interaction FHS group*sex. Our results suggests that higher FHS (higher genetic load of susceptibility) could influence the developmental instability in CL/P cases. Bilateral asymmetry in dermatoglyphic patterns could be a good phenotypic indicator to identify families where the developmental instability is playing a role in the causal model of CL/P, and therefore to obtain more homogenous groups of cases for linkage/association studies conducted to identify major genes for CL/P.

2183T

Robinow, Ter Haar, Teebi or a new syndrome? Complex genotype with distinctive craniofacial features. *F. Uysal[1], B. Turkgenc[2], SG. Temel[3].* 1) Department of Paediatrics, Uludag University, Bursa, Turkey; 2) University of Acibadem, Acibadem Genetic Diagnostic Center, Istanbul, Turkey; 3) Department of Histology & Embryology, Uludag University, Bursa, Turkey.

We report a 45 days old boy who initially presented with the history of antenally diagnosed ompholocele, growth retardation and craniofacial features in 2009. He was the only born from a non-consanguineous marriage at term. Main clinical findings included hypertelorism, prominent forehead, nevus flammateus, prominent eyes, megalocornea, lowset ears, broad nasal root, long philtrum, short frenulum, high palate, thinner upper lip, maloclusion of the teeth, gingival hypertrophy, micrognatia, 5th finger clinodactyly and rocker bottom feet. Echocardiograhy demonstrated LV hyper-trabeculation with myocardial thickness 6 mm, moderate tricuspite regurgitation, mild tricuspite valve displasia and LV non compaction. LV systolic functions were normal. Cranial and 3D calvarium BT demonstrated closure of metopic and sagittal sutures. The diameter of calvarium from anterior to posterior is minimally elevated (scaphocephaly?/dolicocephaly?) according to the lateral one; suggesting sagittal synostosis. Sequencing of possible genes responsible from the phenotype like *Wnt5a, SH3PXD2B, BRAFV600E, KRAS, MEK1* and *MEK2* genes revealed normal genotype. The whole genome chromosome SNP microarray analysis was also normal. Exome sequencing revealed several candidate gene variants homozygous c.1954C>T (p.Arg652Tryp) missense variant in *POTEE* gene, homozygous c.874G>A (p. Arg292Cys) missense variant in *PIK3C2B* gene, and compound heterozygote missense c.7687C>T (p.Gly2563Ser) and c.5212T>C (p.Ile1738Val) variantss in *ODZ4* gene. Seggregation analysis was confirmed with Sanger sequencing. Complex genotype will be discussed with the phenotypic features and pathway interactions of the candidate genes. Functional studies and RNA sequencing are ongoing projects to explain the complex genotype.

2184F

Clinical aspects associated with orofacial clefts in patients of smile operation in a Colombian population. *J. Martinez[1], I. Briceno[1], L. Arias[1], A. Zarate[1], A. Collins[2].* 1) Bioscience, Univ de La Sabana, Chia, Colombia, Colombia; 2) University of Southampton.

Objectives: To present descriptive epidemiology of Orofacial Clefts and to determine the association of syndromic forms with antenatal high-risk conditions, preterm birth, and comorbidities among nested-series of cases. **Methods:** A study of nested-series of cases was conducted. Frequencies of cleft type, associated congenital anomalies, syndromic, non-syndromic and multiple malformation forms, and distribution of Orofacial Clefts according to sex and affected-side were determined. Odds ratios were calculated as measures of association between syndromic forms and antenatal high-risk conditions, preterm birth and comorbidities. A total of three hundred and eleven patients with Orofacial Clefts were assessed in a 12-month period. **Results:** The most frequent type of Orofacial Clefts was cleft lip and palate, this type of cleft was more frequent in males, whereas cleft palate occurred more often in females. The most common cases occurred as non-syndromic forms. Aarskog-Scott syndrome showed the highest frequency amongst syndromic forms. Hypertensive disorders in pregnancy, developmental dysplasia of the hip, central nervous diseases and respiratory failure showed significant statistical associations (p <0.05) with syndromic forms. **Conclusions:** These data provide an epidemiological reference of Orofacial Clefts in Colombia. Novel associations between syndromic forms and clinical variables are determined. In order to investigate causality relationships between these variables further studies must be carried out.

2185W

Exploring the impact of sex-specific genetic effects on orofacial clefting. *J. Carlson[1], N. Nidey[2], J. Shaffer[1], A. Butali[2], C. Buxo[3], K. Christensen[4], F. Deleyiannis[5], J. Hecht[6], L. Field[7], L. Moreno[2], I. Orioli[8], E. Castilla[9], F. Poletta[9], C. Padilla[10], A. Vieira[1], S. Weinberg[1], G. Wehby[2], E. Feingold[1], J. Murray[2], M. Marazita[1], E. Leslie[1].* 1) University of Pittsburgh, Pittsburgh, PA 15261, USA; 2) University of Iowa, Iowa City, IA, 52246, USA; 3) University of Puerto Rico, San Juan, 00936, Puerto Rico; 4) University of Southern Denmark, Odense, DK-5230 Denmark; 5) University of Colorado School of Medicine, Denver, CO, 80045, USA; 6) McGovern Medical School University of Texas Health, Houston, TX,77030, USA; 7) University of British Columbia, Vancouver, V6H 3N1, Canada; 8) Federal University of Rio de Janeiro, Rio de Janeiro, 21941-617, Brazil; 9) Center for Medical Education and Clinical Research, Buenos Aires, 1431, Argentina; 10) University of the Philippines Manila, Manila, 1101, The Philippines.

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is the most common craniofacial birth defect in humans, affecting approximately 1 in 700 births worldwide. NSCL/P has complex etiology and heterogeneous phenotypic presentation including differences in prevalence by sex. Approximately twice as many males are affected with NSCL/P than females, and the source of this disparity is largely unknown. Interactions between genetic effects and sex effects have been hypothesized to help explain some of the differences seen in NSCL/P prevalence. To this end, we examined gene-by-sex interactions in a worldwide sample of 2142 NSCL/P cases and 1,700 controls recruited from 18 sites across 13 countries from North America, Central/South America, Asia, Europe and Africa. Joint tests of genetic and gene-by-sex effects were tested genome-wide (6,926,378 SNPs with MAF>0.05) using logistic regression assuming an additive genetic effect and adjusting for 18 principal components of ancestry. We further interrogated loci with suggestive results from the joint test (p < 1E-05) by examining the gene-by-sex effects from the same model. Suggestive results (p < 1E-05) for the joint test were observed in 133 loci. From these regions, one genome-wide significant gene-by-sex effect in the 10q21 locus (rs72804706; p=6.69E-09; OR=2.62[1.89, 3.62]) and 16 suggestive gene-by-sex effects were observed. At the 10q21 locus, the risk of NSCL/P is estimated to increase with additional copies of the minor allele for females, but the opposite effect for males. This locus lies within an intergenic region downstream of *IPMK*, whose impact on craniofacial development is unknown. Our observation that the impact of genetic variants on orofacial clefting risk differs for males and females at this locus may further our understanding of the genetic architecture of orofacial clefting and and the sex differences underlying clefts and some other birth defects. R01DE016148, X01HG007485, K99DE024571, R01-DE011931.

2186T

Identification of 16q21 as a modifier locus for orofacial cleft phenotypes. *E.J. Leslie[1], J.C. Carlson[2], J. Standley[3], A. Petrin[4], J.R. Shaffer[5], A. Butali[6], C.J. Buxo[7], E. Castilla[8], K. Christensen[9], F.W.D. Deleyiannis[10], J.T. Hecht[11], L.L. Field[12], A. Garidkhuu[13], L.M. Moreno-Uribe[4], N. Nagato[14], I.M. Orioli[15], C. Padilla[16], F. Poletta[8], S. Suzuki[14], A.R. Vieira[1], G.L. Wehby[17], S.M. Weinberg[1], E. Feingold[2], J.C. Murray[3], M.L. Marazita[1].* 1) Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA; 4) Department of Orthodontics, College of Dentistry, University of Iowa, Iowa City, IA; 5) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 6) Department of Oral Pathology, Radiology and Medicine, Dows Institute for Dental Research, College of Dentistry, University of Iowa, Iowa City, IA; 7) Dental and Craniofacial Genomics Center, School of Dental Medicine, University of Puerto Rico, San Juan, Puerto Rico; 8) CEMIC: Center for Medical Education and Clinical Research, Buenos Aires, Argentina; 9) Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark; 10) Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, CO; 11) Department of Pediatrics, McGovern Medical School and School of Dentistry UT Health at Houston, Houston, TX; 12) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 13) Division of Oral and Maxillofacial Surgery, Graduate School of Dentistry, Tohoku University, Japan; 14) Division of Research and Treatment for Oral and Maxillofacial Congenital Anomalies, School of Dentistry, Aichi-Gakuin University, Japan; 15) ECLAMC (Latin American Collaborative Study of Congenital Malformations) at INAGEMP (National Institute of Population Medical Genetics), Rio de Janeiro, Brazil; 16) 17Department of Pediatrics, College of Medicine; and Institute of Human Genetics, National Institutes of Health; University of the Philippines Manila, Manila, Philippines; 17) Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA.

Orofacial clefts (OFCs) are common, complex birth defects with extremely heterogeneous phenotypic presentations. Two common subtypes – cleft lip alone (CL) and cleft lip and palate (CLP) - are often combined for genetic analysis (i.e. cleft lip with or without cleft palate, CL/P). In such studies of CL/P, the underlying assumption is that the effect on risk of cleft may be identical for both CL and CLP subgroups. However, such investigations cannot separate variants for which the risk of OFC differs by subtype. Despite substantial data supporting a shared etiology between CL and CLP, recent evidence suggests that subtype-specific risk factors and/or genetic modifiers may distinguish these two phenotypes. Therefore, identification of genetic factors that act as modifiers of cleft subtypes is critical for understanding the substantial variability of OFCs between individuals. We performed genome-wide association scans for genetic modifiers by directly comparing 450 CL cases with 1692 CLP cases recruited from 13 countries around the world. Gene-based scans of low-frequency variants (MAF<1%) revealed three genes (*C8orf34*, *TMEM246*, and *CDC42EP3*) significantly associated with cleft type. In the scan of common SNPs (MAF>1%), a 16q21 locus was strongly associated with cleft type (p=5.6×10^{-8}), which replicated in an independent sample of 360 CL and 725 CLP cases from Brazil, the Philippines, and Mongolia. Furthermore, the 16q21 alleles that increased CL risk were found at highest frequencies among individuals with a family history of CL compared to those with histories that included CLP (p=0.003). We further investigated the potential modifying behavior of the 16q21 locus on known CL/P risk loci with gene-gene interaction analyses and found significant interactions between the modifier locus and 8q21 (p=0.012) and 9q22 (p=0.023). In these interactions, the risk of CL increased only in the presence of both the CL/P risk allele and the 16q21 modifier; CLP risk was unchanged by the modifier. 16q21 contains several biologically plausible candidates including *LINC00922*, a long-non-coding RNA with putative gene regulatory function and *CDH11*, a craniofacially expressed protein that regulates extracellular matrix production. This study demonstrates the existence of common and low-frequency phenotypic modifiers for OFCs and further elucidates the complex genetic architecture of OFCs by identifying biologically plausible elements responsible for phenotypic heterogeneity.

2187F

Ballooning of redundant myelin sheaths in DS may resolve with Ca^{2+} as seen in the mouse. *A.N. van Hoek, A. Ramirez, M. Sauer, L. Dai, J.R. Korenberg.* University of Utah, Salt Lake City, UT.

The conduction speed of an action potential is greatly enhanced by myelination and is significantly related to cognition. In DS (trisomy 21) conduction is disturbed, reflected by abnormal fMRI, and in the Ts65Dn and DS mouse models for Chr 21 genes (Olig1, Olig2, Claudin 8) myelin and conduction are abnormal. The extent to which myelin is abnormal or related to brain or cognitive dysfunction, is unknown. Reasoning that myelin thickness is linearly related to conduction speed, we examined myelin in DS. We determined myelin structure and immunohistochemical (IHC) distribution of MBP (myelin basic protein), claudin 11 (tight junction protein tethering neighboring myelin layers), and GFAP (glial fibrillary acidic protein) by confocal and electron microscopy in Ts65Dn and control mice, in the non-human primate Macaca fascicularis, and in a brain kindly donated from a 76-year-old woman with DS. The IHC results indicated that the pattern of antibody labeling was similar to the macaque brain. EM revealed subsets of axons with compact myelin. However in the macaque, and exaggerated in DS, ballooned myelin with "onion" shapes of disorganized lamellae (>35, DS) were seen surrounding small axons. In both DS (fixed with 4%-paraformaldehyde (PA) & 0.2% glutaraldehyde in PBS) and macaque (fixed with 4%-PA in PBS), the pattern of ballooning revealed an elegant helical pattern, maintained by radial spokes of tight junctions that tether points in the layers, preventing total disruption of myelin. In mouse, ballooning but not helices were seen when the primary fixative (4% PA) included PBS, but not when 100 mM calcium acetate replaced PBS, revealing mostly compact myelin. These data show that the presence of calcium may be adequate to alter and possibly reverse ballooning. We propose that the ballooning in DS tissue may be in part present in vivo, caused by gene dosage effects of chr 21 genes affecting myelin and amplified by DS hypocalcemia. The low calcium ballooning may indicate that divalent cations may be needed to neutralize the negatively charged polysialic acids present in the oligodendrocyte glycocalix in order to form closely apposed myelin layers. In support of this is that mouse models overexpressing polysialic acids show myelin ballooning. The results indicate a mechanism of age-related myelin disturbance in DS that would decrease conduction velocity and cognitive processing, and it suggests a novel calcium mediated target for therapeutics of DS and DS-AD.

2188W

Variants in the degron motif of AFF3 cause a multi-system disorder with skeletal dysplasia and severe neurologic involvement. *N. Voisin[1], R.E. Schnur[2,3], S. Douzgou[4,5], A.J. Tanaka[6], C.F. Rustad[7], S.M. Hiatt[8], K. Kosaki[9], C. Mignot[10], C. Nava[10,11], E. Del Giudice[12], A. Mikhaleva[1], B. Yalcin[1], D. Donnai[4,5], N. Brunetti-Pierri[12,13], A. Reymond[1], W.K. Chung[6], The DDD Study.* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, VD, Switzerland; 2) GeneDx, Gaithersburg, MD, United States; 3) Cooper Medical School of Rowan University, Division of Genetics, Camden, NJ, United States; 4) Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Sciences Centre, St Mary's Hospital, Manchester, United Kingdom; 5) Institute of Human Development, University of Manchester, Manchester, United Kingdom; 6) Department of Pediatrics, Columbia University, New York, NY, United States; 7) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 8) HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States; 9) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 10) APHP, GH Pitié Salpêtrière, Department of genetics, Paris, France; 11) Sorbonne Universités, UPMC Univ Paris 06, UMR S 1127, Inserm U 1127, CNRS UMR 7225, ICM, Paris, France; 12) Department of Translational Medicine, Section of Pediatrics, Federico II University, Naples, Italy; 13) Telethon Institute of Genetics and Medicine (TIGEM), Naples, Italy.

The ALF domain defines the AFF1, AFF2, AFF3 and AFF4 transcription factor paralogs, which are components of the transcriptional super elongation complex and regulate expression of genes involved in neurogenesis and development. We identified six individuals with *de novo* missense variants in the ALF degron, a signal for protein degradation by the ubiquitin-degradation system. They present with a recognizable pattern of anomalies including microcephaly, global developmental delay, intellectual disability, brain atrophy, seizures, dysmorphic features, horseshoe kidney, a mesomelic form of skeletal dysplasia resembling Nievergelt/Savarirayan type and other skeletal features, only partially overlapping the *AFF4* variants-associated CHOPS syndrome. A seventh *AFF3* patient who has a *de novo* missense in the *AFF3* ALF but outside of the degron exhibits a milder phenotype with autistic behavior and no dysmorphic features. The three described CHOPS individuals and two more patients we identified carrying *de novo* variants in the degron of AFF4 allow to define that obesity and congenital heart defects are specific to individuals with *AFF4* variants, whereas *AFF3* degron variants are associated with seizures, failure to thrive and a distinctive skeletal dysplasia. Consistent with a causative role of *AFF3* variants in this syndrome, a previously reported individual with a microdeletion encompassing the ALF-domain of AFF3 exhibited overlapping clinical features. Since the corresponding mice deletion better recapitulates this patient's bones malformations than the ablation of the entire orthologous gene, we hypothesize a gain-of-function mechanism in the AFF3 patients. Murine models fully lacking the orthologous gene displayed skeletal anomalies, kidney defects, and brain and neurological anomalies. In conclusion, although ALF proteins were reported to be partially redundant, the phenotypes associated with *AFF3* and *AFF4* variants are clinically distinct and show only minimal phenotypic overlap.

2189T

Microdeletion of Xp22 encompassing *SHOX* and *ARSE*, showing incomplete penetrance and variable expressivity. *B. Simpson, H. Saal, R. Hopkin.* Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

The phenotype of deletions at Xp22.33-p22.31 are rare and poorly delineated. We evaluated a 15-month-old patient who presented with facial dysmorphism, developmental delay, hypotonia, noisy breathing, left inguinal hernia, strabismus and short stature. SNP microarray revealed a 6.9Mb deletion of Xp22.33-p22.31, which encompasses over 30 genes, five of which (*CSF2RA, XG, SHOX, ARSE*, and *NLGN4X*) have been associated with clinical phenotype in humans. The patient had dysmorphic features consistent with chondrodysplasia punctata 1, which can be caused by deletions of *ARSE*. He had depressed nasal bridge, short nasal septum and midface hypoplasia. He had short stature (Z=-2.33) and a familial history of short stature, which is consistent with a *SHOX* deletion. The *CSF2RA* gene has been associated with pulmonary alveolar proteinosis, which is not suspected in this patient at this time. Interestingly, a half-brother was also found to have the same deletion without the typical facies or clinical findings besides short stature (Z=-3.03) and mild developmental delay. The patient's parents are not available for evaluation. Our patient does have other siblings, all half-brothers, some with similar facial features, though no clinical testing has been performed to date. One half-brother with similar facies and strabismus also carries the diagnosis of autism spectrum disorder (ASD), which has been associated with the *NLGN4X* gene. Our patient, though young, does not have any early features of an ASD diagnosis. These observations suggest a broad constellation of findings can be seen with deletions in this region of the X chromosome. However, variable expressivity and non-penetrance may be seen as demonstrated by the asymptomatic sibling. We are in the process of evaluating our proband's siblings as well as characterizing 18 other individuals followed in our center with overlapping deletions to help more clearly define the anticipated phenotype(s).

2190F

CRISPR/Cas9 engineering to generate an isogenic model of the 3-Mb 22q11.2 syndromic deletion. *Y.T. Lin, N.P. Paranjape, A.P. Wiita.* Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA.

Copy number variation (CNV) in the genome, due to gains or losses of specific DNA regions, can be causative of human disease. However, unlike single gene disorders, CNV-associated syndromes can arise from complex underlying genetic bases that translate to diverse phenotypic outcomes that are not fully understood. Recently, CRISPR/Cas9 has been employed to engineer CNVs up to 2 Mb to study the implication of these structural changes in disease (Kraft et al. 2015, Cell Rep 10:833), but it remains unclear if engineering larger CNVs is feasible. As a proof-of-principle, we sought to develop an isogenic model of the 22q11.2 deletion syndrome (DS), the most common microdeletion disorder in humans. The 22q11.2 locus is highly susceptible to genomic rearrangements due to the presence of low copy repeats (LCR), with 85% of 22q11.2DS patients carrying a single-copy ~3-Mb deletion flanked by the two repeat regions LCR22A and LCR22D. Using a recently-described strategy (Tai et al. 2016, Nat Neurosci 19:517), we took advantage of these regions of nearly-identical sequence to design a single guide that targets both LCRs simultaneously to generate the microdeletion via non-homologous end joining on the resulting breakpoints. In initial gene dosage screening by qPCR, we achieved an encouraging success rate of nine out of 44 (20%) isogenic clones harboring the deletion. We further validated the ~3 Mb deletion by SNP microarray. We hypothesized that even clones with the same genetic change may exhibit patient-relevant variable phenotypic effects. We therefore characterized three deletion-positive clones using label-free quantitative proteomics. As expected, we found a significant decrease in the mean expression of proteins encoded in the deletion locus, with moderate clone-to-clone variation, compared to the mean expression of other genes. Notably, we identified 27 and 14 genes outside the 22q11.2 region that were up- and down-regulated, respectively, by more than two-fold in the deletion clones versus wild-type. While we are currently replicating this analysis, these initial findings suggest our proteomic approach may reveal genes outside the CNV region that could play a role in the variable outcomes of 22q11.2DS. Taken together, we show that the single guide strategy can efficiently generate CNVs up to ~3 Mb. These approaches can be used for downstream functional studies in cell line, stem cell, and murine models to reveal the spectrum of effects of syndromic CNVs.

2191W

Exome sequencing-based pipeline identifies functional variants within chromosome 1 associated with the risk of non-syndromic cleft palate. *S. Beiraghi[1], W. Looman[2], H. Stessman[3], U. Radhakrishna[4], B. Van Ness[3], A.K. Mitra[3].* 1) Pediatric Dept, 6-150 Moos, Univ Minnesota, Minneapolis, MN; 2) School of Nursing, University of Minnesota, Minneapolis, MN, USA; 3) Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, USA; 4) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, Michigan, United States of America.

Background Nonsyndromic cleft palate only (nsCPO) is a common congenital craniofacial developmental malformation in humans. Genetics variants governing the risk of nsCPO has yet to be unraveled completely. **Methods:** We performed exome sequencing on 4 subjects (2 affected sons, mom and dad) to identify novel genetic variants associated with the risk of non-syndromic nsCPO. Recent studies have identified variants in chromosome 1, including the gene GRHL3, with genome-wide significant association with nsCPO. Therefore, our focus was to discover additional risk associated single nucleotide polymorphisms (SNPs) within this chromosome. Peripheral blood was obtained following informed consent, genomic DNA was isolated and Exome sequencing was performed on Illumina's HiSeq2500 next-generation high-throughput sequencing system. Exome sequencing data was processed using Galaxy and variants were called using SAMtools and GATK Unified Genotyper. Common overlapping variants were filtered out using Western European (CEU) genomes in the 1000 Genomes Project. Variants were annotated using human reference database (GRCh37.75), snpEFF and ClinVar. **Results:** Exome sequencing identified 349,569 variants in Chromosome 1. When we looked for variants for which the Proband and affected sibling were homozygous and the parents are heterozygous, 8 variants followed our strict filtering criteria (QUAL>30; with dbSNPv137 rs IDs and present in 1000G database; likely effect predicted by snpEFF; clinical annotation available in ClinVar). Among the SNPs found significant are genes implicated in congenital diseases, including Charcot-Marie-Tooth. This disease is also characterized by hammer toe. Interestingly, both the affected children have Morton's toe (which we will be demonstrating in the pedigree figure). **Conclusions:** Our results thus confirmed the presence of functional variants within chromosome 1 as risk factors for nsCPO.

2192T

Genotype-phenotype correlation in WAGR syndrome: Large deletion of chromosome 11p in a patient with classical and non-classical symptoms of the syndrome. *P.N. Moretti[1,2], V.S. Souza[2,3], G.C.R. Cunha[2,3], S.F. Oliveira[1,3,4], J.F. Mazzeu[4,5,6], A. Pic-Taylor[2,3,4].* 1) Departamento de Morfologia e Genética, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brazil; 2) Laboratório de Embriologia e Biologia do Desenvolvimento, Departamento de Genética e Morfologia, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brazil; 3) Programa de Pós-graduação em Biologia Animal, Universidade de Brasília, Brasília, Brazil; 4) Programa de Pós-graduação em Ciências da Saúde, Universidade de Brasília, Brasília, Brazil; 5) Hospital Universitário, Universidade de Brasília, Brasília, Brazil; 6) Faculdade de Medicina, Universidade de Brasília, Brasília, Brazil.

   WAGR syndrome affects different body systems and is clinically characterized by the presence of Wilms' tumor, aniridia, genitourinary alterations and intellectual disability (ID). The molecular diagnosis for the syndrome is based in the deletion of the *WT1* and *PAX6* genes, which have essential function in the urogenital system and ocular development, respectively. Here we describe a patient with clinical findings associated with WAGR syndrome such as neuropsychomotor development delay and aniridia, and other malformations not previously described in patients with the syndrome. We performed blood collection, followed by karyotypic analysis and Chromosome Microarray Analysis using Cytoscan 750K (Affymetrix). The karyotype was normal (46, XX) and microarray analysis revealed a 10 Mb deletion in the short arm of chromosome 11, different than others reported in the literature. The deletion encompasses 41 coding genes, and include the critical region for the WAGR syndrome, with total deletion of *WT1* and *PAX6* genes, confirming the diagnosis of the syndrome. As a contiguous gene syndrome, the size and content of the deletion may account for phenotypic variations observed among patients with WAGR syndrome. Our patient present non-classical symptoms such as hypotonia/hypertonia; inguinal hernia; syndactyly/clynodactyly and patent foramen ovale, and other symptoms not described in other subjects with deletion, like pulmonary stenosis, retrognathia and short stature. We are currently analyzing gene functions and interactions to determine which genes could explain the phenotypic variation observed in our patient in comparison to other patients with WAGR syndrome.

2193F

Syndromic cleft genes implicated in non syndromic forms: Towards translational phenotypes? *B. Demeer[1,2,5], M. Basha[1], O. Boute[3], L. van Malderg[4], M. Mathieu6dramard[5], R. Helaers[1], C. Gbaguidi[6], N. Revencu[1,7], B. Bayet[8], B. Devauchelle[6], M. Vikkula[1,7].* 1) 1Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) EA 4666, LNPC, Université Picardie Jules Verne Amiens, France; 3) human genetics department, CLAD nord de france Amiens University hospital, Amiens, France; 4) Human genetics unit, CHU Besançon, Besançon, France; 5) Human genetics unit, CLAD nord de France, CHU Amiens Picardie, Amiens, France; 6) Maxillo-facial department, CHU Amiens Picardie, Amiens, France; 7) Center for Human Genetics, Cliniques universitaires Saint Luc, Brussels, Belgium; 8) Centre Labiopalatin, Division of Plastic Surgery, Cliniques universitaires Saint Luc, Brussels, Belgium.

   Cleft lip with or without cleft palate (CL/P) and cleft palate (CP) are the most common craniofacial birth defects with an approximate incidence of ~1/700. In 30% of cases, clefts are seen in syndromic forms likely caused by genetic factors. Non syndromic forms (NS) are believed to be caused by a combination of genetic and environmental factors. Several genome-wide association studies have proposed a few loci in NSCL/P. A few genes have been implicated in both syndromic and non-syndromic forms of CL/P. Moreover, there are families with NSCL/P following Mendelian transmission with low penetrance and variable expressivity. These two points suggest that a strategy of studying genes or pathways associated with syndromic forms as the cause of NS clefts could be productive. To this end we decided to perform whole exome sequencing on patients with NS cleft, after having ruled out *IRF6* mutations and cytogenetic anomalies. Patients from 12 families with CP, 10 families with CL/P, 2 families with velopharyngeal insufficiency, and 6 sporadic CP cases were selected. Several likely causative variants were identified in five families: in *GHRL3, TP63, LRP6* and *TBX1*. Clinical reassessment confirmed the isolated occurrence of cleft with variable expressivity in affected patients, as well as low penetrance, given the number of unaffected carriers. Our study illustrates the involvement of genes known to be mutated in syndromic clefts, in non syndromic CL/P. It raises the question of an important part of the "missing heritability" of NSCL/P possibly being explained by modest-to-medium penetrant variants in such genes.

2194W

Genetic variants in a patient with pancreatitis after propofol administration. *J.B. Cordero[1], R.J. Rooney[2], J.C. Han[2,3,4], C.L. Simpson[1].* 1) Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN; 2) Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN; 3) Department of Physiology, University of Tennessee Health Science Center, Memphis, TN; 4) Children's Foundation Research Institute, Le Bonheur Children's Hospital, Memphis, TN.

WAGR syndrome is a rare genetic condition caused by heterozygous chromosome 11p13 deletion. This region contains *PAX6*, which is vital to embryonic development of many systems, including the endocrine pancreas. Propofol is a commonly used sedative hypnotic agent administered as a fat emulsion for procedural and critical care sedation. Hepatic enzymes mainly responsible for propofol metabolism are UGT1A9, CYP2B6 and CYP2C9. Despite its wide use, case reports indicating acute pancreatitis induced by propofol poses concerns about its safety and serious adverse effects in patients with inherited metabolic disorders such as glycogen storage disease and children with preexisting hypertriglyceridemia. We studied the genetic variants in a 12-year-old patient with WAGR, obesity and hypertriglyceridemia who developed pancreatitis after propofol infusion. We sought to determine if variations elsewhere in the patient's genome or in the hemizygous alleles within the WAGR deletion could have contributed to the adverse reaction to propofol. Whole exome sequencing (WES) was performed on the patient. Variants were annotated using wANNOVAR and SNVs with biologically meaningful functional annotation were prioritized. SNVs were then analyzed individually for association with drug absorption, distribution, metabolism, excretion and links to pancreatitis, hypertriglyceridemia and propofol metabolism. Gene relations to phenotypes were derived from GWAS catalog, OMIM, GeneCards, PharmGKB, NCBI Gene and PubMed. From the WES variant data, we identified 5 associated genes for propofol metabolism, 24 for hypertriglyceridemia and 15 for pancreatitis. Heterozygous variants within *CYP4F2, APOE, CASR, ABO, FUT2,* and *TPMT* were predicted to have protein pathogenicity by a majority of prediction algorithms. We also found a rare variant of *CYP2B6* (rs58871670) in close proximity to a known *CYP2B6* variant (rs3745274) which is associated with decreased propofol metabolism. Several variants related to propofol metabolism, hypertriglyceridemia and pancreatitis appear to have placed the patient at higher threshold of risk to develop acute pancreatitis. However, we are not certain what combination of variants contributed to the patient's adverse reaction to propofol. The findings suggest that the patient's adverse drug reaction is not due to the WAGR region deletion, but a complex interaction of factors including genetic variations and underlying risk factors such as hypertriglyceridemia.

2195T

Whole-exome analysis of late-onset Alzheimer's disease reveals novel candidate genes involved in cognitive function. *C. Preuss[1], X. Wang[1], G. Ananda[1], S. Neuner[1], V. Phillip[1], A. Saykin[2], K. Nho[2], B. Lamb[2], A. Uyar[1], B. Lodgson[3], K. Karaturi[1], C. Kaczorowski[1], M. Sasner[1], G. Howell[1], G. Carter[1].* 1) The Jackson Laboratory, Bar Harbor, ME; 2) Stark Neurosciences Research Institute, Indiana University, IN; 3) Sage Bionetworks, Seattle, WA.

Background: Alzheimer's disease (AD) is the major cause of dementia in the aging population worldwide. Despite a strong heritable component for this complex polygenic disease, the underlying causes for cognitive decline in late-onset AD remain largely unclear. **Methods:** Here, we conducted whole-exome analysis using data from the Alzheimer's Disease Sequencing Project (4840 cases, 4294 controls) to identify the individual burden of common and rare pathogenic coding variants in AD. Single variant association, collapsed burden and gene set enrichment analysis were combined to identify novel candidate genes and pathways associated with cognitive function. We validated our association results in an independent longitudinal cohort of sporadic AD with 818 participants. Mouse models with cognitive decline were used to translate findings and prioritize candidate genes for functional validation. **Results:** We identified common coding variants in a novel locus (*SLC6A17*, $p<1\times10^{-6}$) with exome-wide significance and validated variants in known candidate genes (*APOE, MAPT, MS4A6A*) previously associated with cognitive function in AD. The lead SNP (rs41281364) in *SLC6A17* was replicated in a cohort from the Alzheimer's disease neuroimaging initiative (p=0.008) measuring CSF tau burden, an independent predictor of cognitive decline in AD patients. The glutamate transporter correlated with amyloid precursor expression in the hippocampus of a mouse model of AD (rho=0.618, p=7.3E-09) and is associated with cognitive decline (p=0.006, n=9 per group (impaired vs. unimpaired)) in aged AD mice. In addition to the common variant burden, gene-based analysis of rare deleterious variants identified novel candidates implicated in disease (*CEACAM1, CD93, PLA2G4D*, $p<1\times10^{-7}$). Rare deleterious variants are significantly enriched (p<0.05, Bonferroni adjusted) in pathways associated with cognitive function, including serotonergic signaling and amyloid-β clearance. **Conclusions:** Taken together, our results highlight that both common and rare coding variants contribute to the complex architecture of AD. Moreover, we demonstrate that variants with moderate-to-large effects in new biological candidates (*SLC6A17, PLA2G4D, HTR4*) along the functionally interacting glutamatergic and serotonergic signaling pathways affect cognitive decline in AD.

2196F

African haplotypic background mitigates the effect of *APOE* ε4 risk allele in Alzheimer disease. *F. Rajabli[1], B.E. Feliciano-Astacio[2], K. Celis[1], K.L. Hamilton-Nelson[1], L.D. Adams[1], A.R. Betancourt[2], H. Acosta[3], A. Chinea[4], G.S Bird[5], C. Reitz[6], R. Mayeux[6], J.M. Vance[1,7], M.L. Cuccaro[1], J.L. Haines[8], M.A. Pericak-Vance[1,7], G.W. Beecham[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL,USA; 2) Universidad Central del Caribe, Caguas, PR, USA; 3) Clinica de la Memoria, San Juan, PR, USA; 4) Universidad Central del Caribe, Bayamon, PR, USA; 5) Department of Biology, North Carolina A & T State University, Greensboro, NC, USA; 6) Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, NY,USA; 7) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA; 8) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA.

   Background: Late-onset Alzheimer's disease (LOAD) is a progressive neurodegenerative disorder. The *apolipoprotein E gene (APOE)* ε4 allele (*ApoE4*) is a major genetic risk factor for LOAD in non-Hispanic white (NHW) populations but has attenuated effects in African American (AA) and Hispanic (HI) populations. The underlying reasons for differential *ApoE4* effects across the populations are not clear. Our goal in this study is to assess the relevance of the local ancestral background to the differential effect of the *ApoE4* in AA and HI, due to the admixed nature of these populations. **Methods**: *ApoE4* and genome-wide genotyping (with IlluminaMEGA and GSA arrays) were performed in 2,937 AA (1,083 cases, 1,854 controls) and 286 Puerto Rican (PR) HI samples (162 cases, 124 controls). Local ancestry was calculated using SHAPEIT (with 1kGP reference) followed by RFMix (using HGDP reference panels). Association between the affection status and the *ApoE4* genotype in the presence of the local ancestry was analyzed using Fisher's Exact test. **Results:** *Overall: ApoE4* dose associated with LOAD risk in AA (heterozygotes-OR=2.23, CI:1.8-2.7, pv<2.2e-16; homozygotes-OR=7.65, CI:5.1-11.6, pv<2.2e-16) and PR (heterozygotes-OR=2.12, CI:1.0-4.6, pv=0.033; homozygotes-OR=7.97,CI:1.1-347.1, pv=0.02) populations, but with reduced effect size compared to NHW populations (heterozygotes-OR=3.2,CI:2.8-3.8; homozygotes- OR=14.9,CI:10.8-20.6*). Local Ancestry:* Results showed that PR and AA individuals with European local ancestry have a stronger *ApoE4* effect than those with African local ancestry (Table 1). The local ancestry was nominally associated with AD in logistic regression model.

Table 1. Association of *ApoE4* and AD in AA and PR						
Ethnic Groups	***APOE* Genotype**	**Local Ancestry**	**p-value**	**OR**	**2.50%**	**97.50%**
AA	ε3ε4 vs ε3ε3	African	2.06E-9	2.11	1.64	2.72
		European	5.81E-9	2.88	1.99	4.21
	ε4ε4 vs ε3ε3	African	<2.2e-16	5.85	3.74	9.35
		European	2.72E-13	20.7	7.11	82.47
PR	ε3ε4 vs ε3ε3	African	1	1.07	0.29	4.25
		European	0.008	3.7	1.28	12.52

Conclusion: These results suggest those PR who inherited an African *APOE* region have a lower risk of LOAD, similar to that described for AA. This supports the hypothesis that the differential effect of *ApoE4* in ethnic groups is due to ancestral variation around *APOE*. Thus, AA and PR individuals with the *APOE* region from African populations may harbor common protective factors that may help mitigate the effect of the detrimental *ApoE4*.

2197W

Dissecting the sex-specific basis of APOE ε4 allele effect on longevity. *P.R.H.J. Timmers[1], P.K. Joshi[1], J.F. Wilson[1,2].* 1) Centre for Population Health Sciences, University of Edinburgh, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom.

   The Apolipoprotein E e4 allele is linked to higher risk of developing Alzheimer's disease (AD) as well as a higher risk of overall mortality. Recent research suggests the *APOE* effect on lifespan and cognitive decline is stronger in women (Joshi et al. 2016; Lipnicki et al. 2017); however, it is currently unclear whether the difference in lifespans is due to sex-specific aetiology, or is a consequence of fewer men surviving until the onset of AD. Using UK Biobank data on 224,166 parental lifespans, we separate men and women by socioeconomic strata, and combine this with 13,927 cases of parental AD (4863 fathers, 9064 mothers) to investigate how the effects of *APOE* vary depending on expected lifetime. We found that the mortality hazard ratio of *APOE* increases from 2.03 to 2.08 in men and 2.08 to 2.15 in women when comparing the bottom and top strata. Curiously, we found no evidence of an increase in self-reported parent prevalence of AD in the higher mortality ratio groups. Further analysis suggests that groups with the same expected lifespan will have the similar *APOE* mortality hazard ratios, regardless of sex. In other words, the observed sex-specific effect of *APOE* on lifespan is primarily driven by the sex-specificity of lifespan itself and the age-related effects of APOE e4, rather than sex-specific aetiology of AD. **References:** Joshi, P.K. et al., 2016. Variants near CHRNA3/5 and APOE have age- and sex-related effects on human lifespan. *Nature Communications*, 7, p.11174. Lipnicki, D.M. et al., 2017. Age-related cognitive decline and associations with sex, education and apolipoprotein E genotype across ethnocultural groups and geographic regions: a collaborative cohort study B. L. Miller, ed. *PLOS Medicine*, 14(3), p.e1002261.

2198T

Age-related changes in white blood cell gene expression associated with skeletal fragility. *E. Quillen[1], J.P. Glenn[1], J. Foster[1], A. Sheldrake[1], L.A. Cox[1], D.P. Nicolella[2], T.L. Bredbenner[2].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Mechanical Engineering Division, Southwest Research Institute, San Antonio, TX.

Inflammaging – increased chronic inflammation in the elderly – is a frequently-cited contributor to diseases of aging, a growing segment of overall morbidity and mortality. The goal of this study is to determine if age-related changes in gene expression of white blood cells are associated with increased bone fragility using a baboon model. No work was done on living animals and no animals were euthanized specifically for this study. Total RNA was extracted from white blood cells collected at necropsy and mRNA libraries were generated using the KAPA Stranded mRNA-Seq kit for sequencing on the Illumina HiSeq2500. Reads were aligned to the *Papio anubis* genome using the STAR aligner, normalized to read counts per sample, and converted to a log-fold scale. Welch's t-tests were used to identify transcripts with significantly different expression between 38 middle-aged female baboons (14.7±2.7 years, human equivalent 44.1) and 44 elderly female baboons (25.8±1.6 years, human equivalent 77.4). 949 transcripts with nominally significant (α = 0.05) results from the t-tests were evaluated for association with age-adjusted measures of skeletal fragility in an independent sample of 54 female baboons. Skeletal fragility is a major risk factor for fracture but is incompletely indexed by the most common clinical indicator, bone mineral density. Alternative measures of bone quality of the femur were generated post-mortem including: structural stiffness in the habitual loading direction; fracture strength in a non- habitual loading direction equivalent to a side ways fall; collagen content and cross-linking; and, tissue density and percent mineralization in cortical bone. All analyses were performed while controlling for age and body size of the baboons. Eight transcripts were linked to measures of skeletal fragility and bone loss using a significance cutoff of $p < 1\times10^{-4}$. These include the baboon homologues of *RALGPS2*, *CMTR*, *SLC16A6*, and *E2F7* linked to gross measures of bone robustness. The majority of these genes have previously been linked to apoptosis or cell differentiation. Expression of *XIAP*, a regulator of apoptotic cell death was also associated to the tensile strength of the bone while expression of *RAB23*, a regulator of chondroprogenitor cells was associated with fracture resistance in a fall simulation. Finally, *PI4K2A*, a regulator of cell growth was associated with collagen content in the bone. This work was funded by NIH/NIAMS R01 AR064244 and R01 AR060341.

2199F

The genomic basis of human lifespan. *P. Joshi[1], N. Pirastu[1], K. Kentistou[1], K. Fischer[2], T. Esko[2], Z. Kutalik[3,4], J.F. Wilson[1,5] on behalf of the LifeGen consortium.* 1) Centre for Global Health Research, University of Edinburgh, Edinburgh, United Kingdom; 2) Estonian Genome Center, University of Tartu, University of Tartu, Tartu, 51010, Estonia; 3) Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, 1010, Switzerland; 4) Swiss Institute of Bioinformatics, Lausanne, 1015, Switzerland; 5) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.

Genomic analysis of longevity offers the potential to disentangle the underlying effects of lifestyle and disease on mortality and to illuminate the biology of human aging, but until now power for discovery has been limited. Here, using 29 population cohorts with genome-wide data and information on 606,000 **parent lifespans**, we discover two new, genome-wide significant variants affecting longevity (near HLA-DQA1/DRB1 and LPA). We also recapitulate previously identified associations with longevity at APOE and CHRNA3/5 and validate, at a lower level of significance, previous suggestions that common variants at CDKN2A/B, SH2B3 and FOXO3A influence longevity. Next we show that, amongst two hundred traits considered, giving up smoking, educational attainment, openness to new experience and HDL cholesterol levels are most positively genetically correlated with lifespan; whilst susceptibility to coronary artery disease (CAD), cigarettes smoked per day, lung cancer, insulin resistance and body fat are most negatively correlated. We suggest that the effect of education on lifespan is principally mediated through smoking, whilst the effect of obesity appears to act via CAD. Using instrumental variables, we quantify the effect of disease susceptibility and lifestyle factors, and suggest that an increase of one body mass index unit reduces lifespan by 4 months, whilst one year of education adds 11 months to expected lifespan.

2200W

Genomics of the aging hematopoietic system. *E. Bader[1,2], V. Bruat[1], E. Gbeha[1], P. Awadalla[1,2].* 1) Informatics, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Molecular Genetics, University of Toronto, Toronto, Ontario,Canada.

Hematopoiesis is a tightly controlled, dynamic process that continuously replenishes the populations of cells that circulate in blood. Dysregulation of circulating blood cells can negatively impact health, as they are essential for immune functions and oxygen transport. Increased age is associated with dysfunctional blood cell phenotypes and an overall decline in immune functioning, known as immunosenescence, which likely contributes to an increased incidence of age-related disease. Although age is generally associated with poor blood health and disease, a small population of elderly individuals remain healthy throughout their lifetime. Characterizing the genomic and transcriptomic landscape of blood cells in this population can provide insights into the mechanisms that preserve blood health. We exploit Canada's largest population cohort, the Canadian Partnership for Tomorrow Project (CPTP), a resource that provides extensive health and environment data as well as biological samples, to identify elderly individuals with superior blood health. We evaluate the health of an individual's blood using information from complete blood count (CBC) data, which includes blood phenotypes such as the proportions of different cellular populations and red cell distribution width. We observe an association between abnormal blood phenotypes and disease prevalence, demonstrating that blood phenotypes are informative of an individual's overall health. Additionally, individuals with poor blood health often develop abnormal blood phenotypes for a subset of CBC variables acquired sequentially, suggesting an initial event may contribute to downstream events. Differential gene expression analyses that compare young and old individuals with extremely healthy or poor blood health demonstrate that age and blood health alter expression of different sets of genes. Differentially expressed genes are regulated by non-coding variants, which may explain their differences in expression between individuals across age categories with varying blood phenotypes. To further characterize regulatory mechanisms that maintain or diminish blood health we are measuring chromatin accessibility in multiple cell populations. Our results suggest that some of the blood phenotypes reported in a CBC are better indicators of blood health irrespective of age, and therefore can potentially be used for more appropriate stratification and risk assessment of patients.

2201T

Identifying genetic variants associated with leukocyte telomere length in African Americans. *A. Little[1], J. Wilson[2], A. Aviv[3], A. Reiner[4,5], T. Thornton[1].* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216; 3) Center of Human Development and Aging, Rutgers, The State University of New Jersey, New Jersey Medical School, Newark, NJ 07103, USA; 4) Department of Epidemiology, University of Washington, Seattle, WA 98195, USA; 5) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

Leukocyte telomere length (LTL) is a complex trait that is influenced by genetic factors as well as environmental factors such as smoking and sedentary lifestyle. Genome-wide association studies (GWAS) of LTL, performed largely in European ancestry cohorts, have identified LTL-regulating genes that provide mechanistic insights into the potential roles of LTL dynamics (birth LTL and its age-dependent shortening). However, little is known about LTL-regulating genes in African Americans (AfAs). Recent studies have established that AfAs have longer LTL than white Americans on average. Racial differences in LTL and LTL-regulating variant genes that determine LTL dynamics may have a role in the predilection to common diseases such as CVD and cancer. To gain better insight into the genetic determinants of LTL in AfAs, we performed a meta-GWAS in a total of 3408 AfAs using whole genome sequence data from the NHLBI TOPMed project (N= 2410 AfAs from the Jackson Heart Study) along with 998 AfAs from Women's Health Initiative, where genotyped data was imputed to 1000 Genomes. Two loci had $P<5 \times 10^{-8}$ in AfAs: chromosome 10 (*OBFC1*) and a potentially novel inter-genic SNP on chromosome 9. An additional suggestively associated locus for LTL in AfAs ($P=6 \times 10^{-8}$) was identified on chromosome 4 intronic to *ZNF827*, which interestingly is involved in remodeling of telomeric chromatin. Of the 12 SNPs previously associated with LTL in European populations, only two replicated in our study of LTL in AfAs (*OBFC1* and *TERT*). Replication studies of the two putatively novel LTL-associated loci are ongoing in additional AfA samples. These results provide evidence that there are both shared and distinct genetic components contributing to LTL across multi-ethnic populations.

2202F

Meta-analysis of GWAS elucidates genetic architecture of dental caries.
D. Shungin[1,2], S. Haworth[3], A. Teumer[4], V. Alaraudanjoki[5], K. Divaris[6,7], K. Grinde[8], G. Hindy[2,9], B. Holtfreter[10], M.K. Lee[11], P. Pesonen[12], C. Agler[19], V. Anttonen[5], T. Kocher[10], M.L. Marazita[11,13,14], S. Offenbacher[15], M. Orho-Melander[9], J.R. Shaffer[11], T. Sofer[8], P.W. Franks[16,17,18], N.J. Timpson[3], I. Johansson[1]. 1) Umea University, Umea, Sweden; 2) Broad Institute of Harvard and MIT, Boston, MA USA; 3) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 4) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 5) Research Unit of Oral Health Sciences, University of Oulu, Oulu, Finland; 6) Department of Pediatric Dentistry, School of Dentistry, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA; 7) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA; 8) Department of Biostatistics, University of Washington, Seattle, WA, USA; 9) Department of Clinical Sciences, Diabetes and Cardiovascular Disease Genetic Epidemiology, Lund University Diabetes Center, Skåne University Hosptial, Malmö, Sweden; 10) Unit of Periodontology, Department of Restorative Dentistry, Periodontology, and Endodontology, University Medicine, Ernst-Moritz-Arndt University Greifswald, Greifswald, Germany; 11) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; 12) Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland; 13) Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 14) Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 15) Department of Periodontology, School of Dentistry, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA; 16) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden; 17) Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; 18) Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA; 19) Oral and Craniofacial Health Sciences, School of Dentistry, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA.

Dental caries is a heritable trait but evidence about implicated genomic regions has emerged only from small studies. We performed a fixed effects inverse-variance weighted meta-analysis (MA) of GWAS for dental caries in 26,792 individuals from 8 studies of European ancestry and 1 study of Hispanics/Latinos. Association analyses used age- and sex-standardized residuals of clinically assessed dental caries DMFS index, the sum of decayed, missing and filled tooth surfaces, excluding third molars. Genotypes were imputed in each individual study to 1000 Genomes v3 imputation panel. Three novel loci showed evidence for association with caries (*PITX1-c4orf66, CA12* and *AHCTF1, P<*5×10^{-8}). Using approximate conditional analysis we further identified additional independent signal in the *PITX1-c4orf66* region. Gene-based analysis (MAGMA) of MA results additionally highlighted the *TBL2* gene as being implicated in caries (*P<*2×10^{-6}). The *PITX1-c4orf66* and *CA12* associations were supported in 146,341 individuals with self-reported oral health data from UK biobank (question "Do you have any of the following? - Dentures"; *P=*2.0×10^{-19} and *P=*2.0×10^{-4} respectively), while the third SNP and its proxies were not available in UK biobank dataset. The *PITX1* locus has been previously shown in mice to exhibit preferential expression in the developing mandible, pituitary gland and teeth covering the tooth-forming region in the mandible. Loss of function mutations in *CA12* is reported to result in defects in fluids and carbonate secretions including such disease as pancreatitis and xerostomia (dry mouth syndrome). Using GREML, we estimated SNP-based heritability for the DMFS index in two participating studies (h^2_{SNP}=0.18, se=0.09 and h^2_{SNP}=0.39, se=0.14). We applied linkage-disequilibrium score regression to calculate genetic correlation (rg) between dental caries and 77 disease traits with available summary association statistics from GWAS. The correlations with waist-hip ratio adjusted for body mass index (rg=0.25, se=0.06, *P=*5.2×10^{-5}), waist and hip circumference (rg=0.22, se=0.05, *P=*1.6×10^{-5} and rg=0.15, se=0.05, *P=*0.005, respectively), years of schooling (rg=-0.13, se=0.05, *P=*0.004), body mass index (rg=0.14, se=0.06, *P=*0.01), and inflammatory bowel disease (rg=0.18, se=0.07, *P=*0.01) were significant at FDR<0.05 This meta-analysis and follow-up investigations highlighted genetic architecture of dental caries and well as its relationship with other disease traits.

2203W

Understanding relationships between longevity and physical senescence using Mendelian randomization approach. *S. Ukraintseva, K. Arbeev, Y. Wang, H. Duan, I. Zhbannikov, A. Kulminski, A. Yashin.* Duke University, Durham, NC.

Mendelian randomization (MR) is an important approach to studying human aging since it can help clarify biological mechanisms underlying statistical associations between the principal outcome of aging (lifespan/longevity) and various health and senescence related phenotypes. MR uses genetic variants as instrumental variables to establish whether such phenotypes are *causally* related to longevity. In this study, we employed MR approach to assess the causal relationships between several traits (mediators) characterizing physical senescence, and survival to the age 85 and beyond (outcome), using logistic regression, and genotyping and phenotypic data available from three longitudinal human studies (Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), and Health and Retirement Study (HRS)). SNPs in NRG1 and NRG3 genes that have previously been associated with the phenotypes of physical senescence were used as instrumental variables. We found that the rate of decline in BMI between ages 65 and 85; peak ages (ages at maximum) of the weight and BMI, after which they start to decline towards extreme ages; and Rosow-Breslau index of physical ability to walk half a mile, climb the stairs and do a housekeeping work, at ages 65+, were significant mediators of longevity in our study. These results suggest that manifestation of physical senescence even without an accompanying disease is major limiting factor for longevity. .

2204T

Allele-specific expression in healthy centenarians. *L.C. Tindale[1,2], N. Thiessen[1], S. Leach[1], A.R. Brooks-Wilson[1,2].* 1) Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada; 2) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, Canada.

The genetic basis of healthy aging and longevity remains largely unexplained. One hypothesis as to why long-lived individuals do not appear to have a decreased number of common-complex disease risk alleles, is that despite carrying these risk alleles, they express the disease-linked variants at a lower level than the wild type allele. We have sequenced the transcriptomes of four healthy centenarians and four mid-life controls using whole blood in a pilot study. We hypothesized that the centenarians will have an overall pattern of decreased expression of disease-associated alleles. This would effectively allow them to favour expression of the non-risk allele, despite also carrying the disease-causing variant. Heterozygous SNPs act as indicators to determine if one copy of the DNA is more highly expressed. This difference in expression between the two haplotypes of a diploid individual is referred to as allele-specific expression (ASE). The ASE analysis is currently in progress. Molecular signatures may distinguish between disease states, and many conditions have an inflammatory component that affects immune cells. CIBERSORT was used to estimate cell fractions with neutrophils being the most abundant source of RNA, followed by CD8 T cells, resting NK cells, and monocytes. In the present study, whole blood is an appropriate tissue in which to achieve our goal of conducting a hypothesis generating analysis about overall disease risk and patterns of ASE, which may be contributing to disease.

2205F

The PhenX Toolkit: Adding a resource for geriatric research. *M. Phillips[1], M. Sano[2], S. Cummings[3], C. Depp[4], J. Guralnik[5], T. Harris[6], D. Houston[7], C. Kawas[8], D. Kiel[9], R. Mayeux[10], L. Farrer[11], E. Ramos[12], W. Rossi[6], L. Kilpatrick[1], T. Hendershot[1], D. Maiese[1], A. Riley[1], H. Pan[1], W. Huggins[1], C. Hamilton[1].* 1) RTI International, Research Triangle Park, NC; 2) Icahn School of Medicine at Mount Sinai, New York, NY; 3) California Pacific Medical Center, San Francisco, CA; 4) University of California, San Diego, San Diego, CA; 5) University of Maryland School of Medicine, Baltimore, MD; 6) National Institute on Aging, Bethesda, MD; 7) Wake Forest School of Medicine, Winston-Salem, NC; 8) University of California, Irvine, Irvine, CA; 9) Institute for Aging Research, Hebrew SeniorLife and Harvard Medical School, Boston, MA; 10) Columbia University, New York, NY; 11) Boston University, Boston, MA; 12) National Human Genome Research Institute, Bethesda, MD.

The PhenX Toolkit (https://www.phenxtoolkit.org/) is a web-based catalog of standard measures to facilitate biomedical research. Use of PhenX (consensus measures for Phenotypes and eXposures) measures enables cross-study analyses for increasing statistical power and replicating results. In May 2017, PhenX convened nine experts to develop content for a new Geriatrics domain, which will include common measures for research in older adults. The Working Group will address the scope with 15 priority measures; identify open source protocols that pose low burden for study participants and investigators; identify measures that address phenotypes specific to geriatrics; and enhance existing PhenX measures with content for older adults. The measures recommended by the Working Group and brief descriptions will be presented. We will also present PhenX Toolkit bioinformatics tools and features, and collaborative efforts with NIH Common Data Element (CDE) Initiatives and the database of Genotypes and Phenotypes (dbGaP). The PhenX Toolkit currently includes over 500 measures from 24 research domains, with additional depth for Substance Abuse and Addiction, Mental Health, Tobacco Regulatory, and Sickle Cell Disease research. For each measure, the Toolkit provides detailed instructions (e.g., required training and equipment), features (e.g., Spanish language versions), and bioinformatics support (e.g., data dictionaries, REDCap compatible modules) to facilitate implementation and data analysis. The PhenX Toolkit is widely used, with more than 2,500 registered users worldwide and 755 visitors per day on average. This web-based resource has been recommended in more than 200 NIH funding opportunity announcements. We are celebrating the 10-year anniversary of the PhenX Toolkit and 5 additional years of funding that will allow us to continue to be responsive to the evolving needs of the research and clinical communities. *Funding provided by a Genomic Resource Grant (U41 HG007050-05) from NHGRI, with co-funding from the National Institute on Drug Abuse (NIDA) and the Office of the Director, National Institutes of Health (OD).* .

2206W

Does parent-to-offspring transmission of telomeres contribute to telomere length heritability in humans? *D.A. Delgado[1], C. Zhang[1], L.S. Chen[1], F. Jasmine[1], M.G. Kibriya[1], S. Roy[1], R. Paul-Brutus[1], H. Ahsan[1,2,3,4], B.L. Pierce[1,2,3].* 1) Department of Public Health Sciences, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL 60615; 3) Comprehensive Cancer Center, University of Chicago, Chicago, IL 60615; 4) Department of Medicine, University of Chicago, Chicago, IL 60615.

Leukocyte telomere length (TL) shortens with age and is a potential biomarker of risk for age-related diseases. TL is a heritable trait with two potential sources of heritability (h^2): inherited variation in non-telomeric regions (e.g., SNPs) and inherited variability in the lengths of telomeres themselves. Prior family-based and SNP-based h^2 studies have not attempted to disentangle these two sources. Here, we attempt to detect "direct" inheritance of telomeres by studying the association between identity-by-descent (IBD) sharing at chromosome ends and phenotypic similarity in TL. We measured genome-wide SNPs for a sample of Bangladeshi adults with substantial cryptic relatedness (n=5,069). TL was measured for each individual using qPCR or Luminex-based methods. We estimated IBD sharing (0, 1, or 2) at each chromosome end using 180 uncorrelated SNPs to approximate the total number of telomeres shared IBD for each participant pair (T_{shared}). We then estimated the association between T_{shared} and the difference in TL between relative pairs (TL_{diff}), adjusting for the pair's kinship coefficient (Φ) estimated from genome-wide SNPs. Among 583 sibling pairs, T_{shared} showed a non-significant inverse association with TL_{diff} (P=0.57), as did Φ (P=0.37). Among 945 cousin pairs, T_{shared} was inversely associated with TL_{diff} (P=0.04), while the inverse association for Φ was not significant (P=0.22), suggesting that sharing telomeres IBD contributes to pairwise similarity in TL. In analyses including all relative pairs (Φ >0.05), the association between T_{shared} and TL_{diff} was inverse for both T_{shared} (P=0.003) and Φ (P=0.40), but interpretation of this result is complicated by 1) the strong correlation between T_{shared} and Φ when analyzing all relative pairs and 2) the possibility that embryonic telomere reprogramming dilutes the contribution of shared telomeres to TL similarity for distant relatives (who are separated by many reprogramming events). SNP-based h^2-partioning analyses of all participants (using GCTA) had similar limitations, although we observed evidence of enrichment of h^2 at chromosome ends. In summary, we provide preliminary evidence that sequence similarity at chromosome ends (a proxy for sharing telomeres IBD) contributes to TL h^2, implying that variation in TL in parental germ cells influences TL in offsprings' embryonic cells despite telomere "reprogramming" during embryonic development.

2207T

Telomere length across many human tissues. *K. Demanelis[1], L. Chen[1], J.A. Doherty[2], J. Shinkle[1], M. Sabarinathan[1], H. Ahsan[1], F. Jasmine[1], M.G. Kibriya[1], B.L. Pierce[1], Genotype-Tissue Expression (GTEx) Consortium.* 1) Department of Public Health Sciences, University of Chicago, Chicago, IL; 2) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT.

Telomeres shorten over the life-course in most human cell types, and telomere length (TL) is a potential biomarker of aging and risk for age-related disease. However, the relationships among TL, aging, and disease have been primarily studied using leukocyte TL. In order to better understand variation in TL across tissue types and tissue-specific aging, we measured relative TL (RTL, telomere abundance relative to a diploid gene) in 33 tissue types from 435 individuals (age 20-70) in the Genotype-Tissue Expression Project (GTEx). RTL was measured for 3,123 tissue samples (up to 12 tissue types per donor) using a Luminex assay and validated using whole blood TL estimates from whole genome sequencing. We analyzed the relationships among RTL and covariates using linear mixed models. Mean RTL was shortest in whole blood and longest in testis. RTL did not differ by sex, but African Americans had longer RTL compared to Caucasians (p=0.07). The percent variation explained by individual and tissue was 16% (p< 10^{-16}) and 34% (p< 10^{-16}), respectively. Among the 20 tissues with >50 samples measured, RTL negatively correlated with age in 17 tissues, 13 of which were statistically significant (P<0.05), ranging from -0.42 (p=3x10^{-15}) for whole blood to -0.20 (p=0.02) for prostate. No tissue showed a significant positive association between RTL and age. Tissues with stronger associations between RTL and age had shorter mean RTL. The age-adjusted partial correlations between tissue-specific RTL and whole blood RTL were positive for all tissues (except ovary) and statistically significant (P<0.05) for 10 tissues, with the strongest correlations observed for suprapubic skin (r=0.48, p=3x10^{-6}), muscle (r=0.36, p=0.02) and pancreas (r=0.32, p=7x10^{-5}). Using GTEx RNA-Seq data, we observed no clear association between expression of components of telomerase and RTL within any tissue. However, in testis, expression of *TERT* (telomerase reverse transcriptase) was strikingly high compared to other tissues, consistent with its high mean RTL. Our results describe, for the first time, substantial heterogeneity in TL and age-related TL shortening across many human tissues. The cross-tissue correlations among RTL measures provide a foundation for the interpretation of epidemiological studies of leukocyte TL, aging, and disease. Ongoing and future analyses of these data include studies of gene expression correlates of RTL and the impact of genetic variation on RTL within and across tissues.

2208F

GWAS replicates known asthma variants validating self-reported childhood asthma diagnosis in the COPDGene Study. *L.P. Hayden[1,2], M.H. Cho[2,3], B.A. Raby[2,3], T.H. Beaty[4], E.K. Silverman[2,3], C.P. Hersh[2,3] on behalf of the COPDGene Investigators.* 1) Division of Respiratory Diseases, Boston Children's Hospital, Boston, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA; 4) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

    PURPOSE Childhood asthma is a phenotypically and genotypically distinct asthma subtype, with an estimated heritability of 80%. While childhood asthma is a known risk factor for development of reduced lung function and chronic obstructive pulmonary disease (COPD) in adulthood, few studies have evaluated self-reported childhood asthma in subjects with COPD. We hypothesized that a history of childhood asthma would be associated with decreased lung function and increased risk for COPD in adult smokers, and that a genetic association study would re-identify previously described variants associated with asthma.   **METHODS** We evaluated current and former adult smokers in the COPDGene Study. Subjects were of non-Hispanic white (NHW) or African American (AA) race and 45-80 years of age. Childhood asthma was defined by self-reported asthma, diagnosed by a medical professional, with onset at <16 years or during childhood. Subjects with and without history of childhood asthma were compared on measures of lung function. Genome-wide association was performed in NHW and AA populations separately, and subsequently combined in meta-analysis. Eight established asthma SNPs were examined for association with childhood asthma in this GWAS.   **RESULTS** Among 10,199 adult smokers, 730 (7%) reported a history of childhood asthma. Childhood asthmatics had reduced lung function and increased odds of developing COPD (OR 2.51, 95% CI 2.07-3.05). Among NHWs, 391 (6%) had childhood asthma, and GWAS identified one genome-wide significant association in *KIAA1958* (rs59289606, p=3.7×10^{-8}). However, among AAs, of which 339 (10%) had childhood asthma, and in a meta-analysis combining NHW and AA subjects, no results reached genome-wide significance. Four of eight previously established asthma SNPs examined were associated with childhood asthma in the current GWAS: *IL1RL1* (rs3771180) and *ORMDL3/GSDMB* (rs12936231) in NHWs, *KCNIP4* (rs4697177) in AAs and *LRRC32* (rs7130588) in meta-analysis (Bonferroni adjusted p<0.02 for all comparisons).   **CONCLUSIONS** Self-report of childhood asthma in adult smokers from COPDGene identifies a meaningful population who demonstrate the demographic, clinical, and genetic characteristics known to be associated with childhood asthma. This validates use of self-reported childhood asthma in this population. This GWAS identified a new variant in *KIAA1958* as associated with childhood asthma in NHWs, which has not previously been seen in respiratory disease.

2209W

Newborn metabolomics and risk of episodic wheezing in childhood: Findings from the INSPIRE study. *K.K. Ryckman[1], B.M. Donovan[1], P.J. Breheny[2], E.K. Larkin[3], T. Gebretsadik[4], K.N. Turi[3], T.V. Hartert[3].* 1) Department of Epidemiology, University of Iowa College of Public Health, Iowa City, IA; 2) Department of Biostatistics, University of Iowa College of Public Health, Iowa City, IA; 3) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN.

    Asthma is a complex, multifactorial disease that results in significant illness and death worldwide. Mitochondrial dysfunction has been implicated in the etiology of asthma, but there are critical gaps in our understanding of its contributing role. Newborn metabolic screening represents a unique data source that can be leveraged with perinatal and environmental exposures to further our understanding of mitochondrial dysfunction and asthma etiology. Our objective was to determine if newborn metabolite levels are associated with an increased risk of experiencing episodic wheezing in early childhood. Demographic and clinical data was collected from the Infant Susceptibility to Pulmonary Infections and Asthma Following RSV Exposure Study (INSPIRE) and linked to newborn metabolite data measured 24-72 hours after birth as part of the Tennessee Newborn Screening Program. Over 1,900 infants born during the summer months were enrolled over a two year period (June 2012-March 2014) from pediatric practices located in Middle Tennessee, with on-going follow-up. Episodic wheezing at 1, 2, and 3 years of age was categorized into 4 groups (no wheezing, 1-3 episodes, 4-12 episodes and >12 episodes) by number of wheezing episodes experienced in the past twelve months per ISAAC standards. The association between 32 amino acid and acylcarnitine metabolites with the categorical wheezing outcome was assessed at 1, 2 and 3-years of age using separate proportional odds regression models. Several unadjusted metabolite associations were significant after Bonferroni correction (p<0.001). Increasing levels of C10:1 were associated with increasing number of wheezing episodes at 2-years (OR 2.15, 95% CI 1.43-3.23) and 3-years (OR 2.57, 95% CI 1.60-4.13) of age, while increasing levels of C18:2 were associated with increasing number of wheezing episodes at 1-year (OR 1.37, 95% CI 1.11-1.70) and 2-years of age (OR 1.49, 95% CI 1.18-1.87). Linoleoylcarnitine (C18:2) is derived from linoleic acid, a polyunsaturated fatty acid, that has been implicated in airway hyperresponsiveness, epithelial injury and asthma severity. Our next step will be to link existing genome-wide association data for this cohort to the newborn metabolite measurements to further examine this pathway and its role in childhood wheezing and asthma. Identification of specific metabolites and their related genetic pathways may provide valuable information about the etiology of asthma and targets for primary prevention.

2210T

Influence of guideline adherence and ADRB2 SNPs inpredicting exacerbation frequency in asthma patients. *A. Santani[1,2], N. Sood[1], J. Connolly[1], F. Mentch[1], L. Vazquez[1], P. Sleiman[1,3], H. Hakonarson[1,3].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia,; 2) Department of Pathology and Laboratory Medicine, University of Pennsylvania; 3) Department of Pediatrics, University of Pennsylvania.

Asthma is the leading chronic disease in children. Several studies have identified genetic biomarkers associated with susceptibility and severity in both adult and pediatric cases. In this study, we evaluated outcome in 470 African American (AA) and European American (EA) pediatric cases all of whom were regular users of controlling medications, such as inhaled corticosteroids and leukotriene inhibitors, and frequent users of reliever drugs, such as short acting beta agonists. Patients were, stratified by genotype using two single nucleotide polymorphisms (SNPs) in the beta-2 adrenergic receptor (ADRB2) gene – rs1042713 and rs1042714, previously associated with asthma drug response outcome. We controlled for adherence to National Heart, Lung and Blood Institute (NHLBI) guidelines using deep mining of electronic health record (EHR) data to determine treatment course. Of the 470 cases, 446 had an updated Asthma Care Plan (ACP), 421 had PFT data, all 470 had a prescription for at least one ICS medication and for B2AR agonist drug. Data analyses were performed in two stages – 1) we separately compared the effects of genotype on outcome in AA and EA cases and 2) we combined AA and EA cohort in a meta-analysis of the same outcome. In all instances the main outcome (i.e. dependent variable) was mean frequency of exacerbations per year. Relevant independent variables in all instances were genotype (both rs1042713 and rs1042714), asthma severity and treatment. Additionally, for the combined meta-analysis, race was included as a variable in the combined meta-analysis. Using exacerbation frequency as the dependent variable, treatment according to the guidelines was relatively significant ($P = 0.060$), whilst mutations rs1042713 and rs1042714 were markedly statistically significant (with $P = 0.007$ and $P = 0.004$, respectively). Genotyping results demonstrated that homozygous carriers of the rare alleles for both SNPs were significantly more likely to experience asthma exacerbations ($P<0.01$), and this effect was largely driven by patients of African American ancestry ($P<0.01$). Taken together, our results demonstrate that patients homozygous for the minor alleles of rs1042713 and rs1042714 genotypes may derive less benefit from corticosteroid medications, particularly those of African American ancestry.

2211F

A simulated evaluation of data-driven algorithms for addressing clinical heterogeneity in complex traits. *A.O. Basile[1], A. Ulloa[2], A. Lucas[2], A.T. Frase[2], V. Abedi[2], H.L. Kirchner[2], C.B. Manney[2], J.B. Leader[2], M.D. Ritchie[1,2].* 1) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park , PA; 2) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA.

Complex diseases are often heterogeneous in nature; instead of representing a single condition, they may be comprised of multiple disorders, with varying symptoms, clinical presentations, and biological etiology. Trait heterogeneity is a confounding variable often overlooked in genetic approaches. It complicates patient diagnoses, and decreases statistical power to detect associations and risk attributed to susceptibility variants. The copious amounts of biomedical data in Electronic Health Records (EHRs) can be leveraged using data-driven approaches to overcome heterogeneity and detect patient subgroups. However, for an algorithm to be appropriate, it must be robust to noisy extraneous variables and missingness, appropriate for mixed data, and computationally practical. We performed *in silico* evaluations using data simulated with the Madelon algorithm to examine the performance of machine learning approaches in homogeneous data classification. 5,000 subjects with known outcome labels predicted by variables drawn from a Gaussian distribution with varying proportions of noise were simulated. Evaluated algorithms include icluster, similarity network fusion (SNF), generalized low rank models (GLRM), autoencoders, and manifold learning algorithms including local linear embedding (LLE), local tangent space alignment (LTSA), Hessian eigenmapping (HLLE), modified local linear embedding (MLLE), isomap, multi-dimensional scaling (MDS), spectral embedding, and t-distributed stochastic neighbor embedding (t-SNE). Algorithmic performance was evaluated using the average silhouette coefficient. The top methods, in order of performance, were LLE, spectral embedding, SNF, Isomap, t-SNE with PCA, and GLRM with PCA. These successful algorithms will be used to identify clinically meaningful subgroups in 49,000 patients with Chronic Obstructive Pulmonary Disease (COPD) from the Geisinger Health System MyCode® Community Health Initiative.

2212W

Disease associations of the zinc receptor GPR39 in the DiscovEHR Study cohort. *G.E. Breitwieser[1], S. Krishnamurthy[1], J. Overton[2], J. Reid[2], J. Robishaw[1], D. Carey[1], R. Metpally[1], DiscovEHR collaboration.* 1) Functional and Molecular Genomics, Weis Center for Research/Geisinger Clinic, Danville, PA; 2) Regeneron Genetics Center, Tarrytown, NY.

Zn[2+] is a transition metal essential for life. Deficiencies, seen in 2-3% of the United States population, cause hyosmia, hypogeusia, dermatitis, alopecia, susceptibility to infections and seizures, slow wound healing, and compromised gonadal function. Zn[2+] is both a first and second messenger, and physiological concentrations of extracellular Zn[2+] activate GPR39, a Family A G protein-coupled receptor most closely related to the ghrelin subfamily. *GPR39[-/-]* mice are viable, but have depression and anxiety-like behavior, increased seizures in response to kainate, immune system dysfunction, slowed wound healing, and increased fat accumulation in response to high fat diet. GPR39 plays an antagonistic role to ghrelin and motilin receptors in the stomach and intestine, slowing gastric emptying and secretion, and gastrointestinal motility. To define the importance of GPR39 in human disease, we identified rare missense, frameshift, premature termination and synonymous variants in whole exome sequences for 51,289 individuals in the DiscovEHR cohort. We found patients having rare heterozygous mutations in the zinc binding site, the pH sensor, cysteines participating in disulfide bonds, residues in highly conserved Family A motifs, and putative phosphorylation sites. Optimized Sequence Kernel Association Tests (SKAT-O) were run on binned LOF/missense (LOF/MIS) variants and separately on rare synonymous (SYN) variants having significant changes in codon usage bias. Only 2 patients were in both the LOF/missense and rare synonymous groups, and they were removed prior to SKAT-O analyses. All ICD9 codes in the electronic health record with at least three independent instances per individual were used for SKAT-O. Both groups yielded a common set of phenome-wide disease associations, including benign prostate hyperplasia with urinary obstruction (LOF/MIS p<1.06E-09; SYN p<1.13E-09), benign hypertension (LOF/MIS p<9.8E-06; SYN p<1.3E-05), psoriatic arthropathy (LOF/MIS p<1.4E-06; SYN p<1.5E-06) and diseases of hair or hair follicles (LOF/MIS p<1.3E-04; SYN p<9.7E-07). The convergence of results for both LOF/missense and rare synonymous variants suggests that haplo-insufficiency of *GPR39* may underlie the identified disease associations. Results argue that rare synonymous variants may be an untapped resource for both validating and extending the associations derived from rare LOF/missense variants. Supported by funds from Geisinger Health System & Regeneron.

2213T

Low frequency genetic variation in *TP53* is associated with final head circumference. *B. St Pourcain[1,2,3], V. Vitart[4], E. Zeggini[5], G. Dedoussis[6], V. Jaddoe[7,8,9], C.E. Pennell[10], K. Bønnelykke[11], D. Martinez[12], K. Walter[5], C.S. Shapland[1], A. Jackson[4], P. Timmers[4], J.F. Wilson[4,14], G. Davey Smith[2], J.L. Min[2], V. Iotchkova[5], S.E. Fisher[1,3], N. Timpson[2], D. Fernandez[13], J. Sunyer[13], T.S. Ahluwalia[11,12], C. Wang[10], J. Felix[7,8,9], I. Tachmazidou[5], C. Hayward[4], S. Haworth[2], ENIGMA Consortium, CHARGE Consortium, UK10K Consortium.* 1) Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 2) MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol BS8 2BN, UK; 3) Donders Institute, 6525 EN Nijmegen, The Netherlands; 4) MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH16 4UX, UK; 5) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK; 6) Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens 17671, Greece; 7) The Generation R Study Group, Erasmus Medical Center, University Medical Center, Rotterdam 3000 CA, the Netherlands; 8) Department of Epidemiology, Erasmus Medical Center, University Medical Center, Rotterdam 3000 CA, the Netherlands; 9) Department of Pediatrics, Erasmus Medical Center, University Medical Center, Rotterdam 3000 CA, the Netherlands; 10) Division of Obstetrics and Gynaecology, The University of Western Australia, WA-6009, Australia; 11) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, DK-2820 Copenhagen, Denmark; 12) Steno Diabetes Center Copenhagen, Gentofte, Denmark; 13) ISGlobal, Center for Research in Environmental Epidemiology , 08003, Barcelona, Catalonia, Spain; 14) Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK.

Background: Craniofacial development gives rise to the closely related traits of head circumference (HC) and intracranial volume (ICV). Final HC is largely determined by 6 years of age. Previous studies have identified common genetic variants for infant HC and adult ICV. This study aims to identify common and lower frequency variants for both final HC and final craniofacial dimension (as assessed through a joint analysis of HC with ICV). **Participants and methods:** Genome-wide association analysis was performed in 11 population based cohorts with either whole genome sequencing information (n=1,762; UK10K consortium), genotype imputation to a joint UK10K/1KG panel (n=8,285) or genotype imputation to the Haplotype Reference Consortium (HRC) panel (n=8,834). All participants were aged 5.5 years or older and of European ancestry. The entire HC meta-analysis included 18,881 participants, with a replication in a further independent sample of 973 adults, aged 18-85 years. Joint analysis of HC and ICV (craniofacial dimension) included 26,577 additional participants aged 9-98 years from the CHARGE / ENIGMA consortia with a total sample size of 45,458. **Results:** After conditioning on known variants, association with head circumference was seen at 3 loci, including a signal at a low frequent variant in the 5' UTR of the Δ133 isoform of *TP53* (rs35850753, EAF 0.022, beta 0.21 SD, p value 2.0e-08 discovery, 4.5e-05 replication, 9.4e-10 combined). Joint analysis of a final intracranial dimension identified also association at 6 loci at the genome-wide significant level in addition to those identified for either HC or ICV alone. **Discussion:** *TP53* is known to co-ordinate cranial neural crest growth and differentiation in animal models. p53-dependent neuroepithelial apoptosis, as a result of mutations in *TCOF1*, has been identified as primary mechanism underlying the pathogenesis of Treacher Collins syndrome, a congenital disorder of craniofacial development. rs35850753 tags both rs35850753 (r²=1) and rs78378222 (r²=0.43) that have been robustly linked to neuroblastoma. rs78378222 is also involved in a range of other human cancers and its hypomorphic allele is known to impair proper termination and polyadenylation of *TP53* transcripts . These results support a role for *TP53* transcripts in human craniofacial development.

2214F

Pathogenic and likely pathogenic mutations identified in apparently normal individuals of Arab descent. *A. Alkhateeb.* Biology Department, Jordan University of Science and Technology, Irbid 22110, Irbid, Jordan.

Arab genome is one of the least studied as compared to other ethnicities. Lots of variants that might be unique to the Arab population are still to be characterized. Here, and as a spin off another study, we exome-sequenced a set of unrelated Arabic individuals. Six exomes were analyzed through Ingenuity Variant Analysis tool. We limited our analysis to extremely rare mutations with a frequency of less than 0.001. We identified four damaging mutations; three are novel and one that is reported only once. None of the four mutations are present in the Exome Aggregation Consortium (ExAC) database. Two of the mutations are classified by Ingenuity as pathogenic; c.3955-1G>C in NALCN gene, and p.M296* in the PALB2 gene. The other two are classified as likely pathogenic; p.D959fs*5 in CACNA1S gene, and p.S200fs*7 in PMS1 gene. The presence of these debilitating mutations in apparently normal individuals shows the uniqueness of the Arab exome and the need to characterize it more to have a better understanding of its genetic background.

2215W

Widespread prevalence of a CREBRF variant amongst Māori and Pacific children is associated with weight and height in early childhood. *S.D. Berry[1], C.G Walker[1], K. Ly[1], R.G. Snell[2], P.E. Atatoa Carr[3], D. Bandara[1], J. Mohal[1], T.G Castro[1], E. Marks[1], S.M.B. Morton[1], C.G. Grant[1,4,5].* 1) Centre for Longitudinal Research He Ara ki Mua, University of Auckland, Auckland, Auckland, New Zealand; 2) School of Biological Sciences, University of Auckland, Auckland, Auckland, New Zealand; 3) National Institute of Demographic and Economic Analysis, University of Waikato, Hamilton, New Zealand; 4) Department of Paediatrics: Child and Youth Health, University of Auckland, Auckland, New Zealand; 5) Starship Children's Hospital, Auckland District Health Board, Auckland.

Investigating a large and ethnically diverse cohort from the Pacific region, we aimed to replicate and extend the recently reported findings that a CREBRF genetic variant is strongly associated with body mass index in Samoans. A birth cohort of more than six thousand children was utilised. In this study, genotyping of two markers (rs12513649 and rs37386328) was undertaken in Māori, Pacific, European and Asian individuals in the cohort. We report that these CREBRF genetic variants are not confined to Samoans but are prevalent in all other Pacific populations sampled, including Māori. We found that the rs37386328 variant was significantly associated with growth at four years of age. On average, we observed allele-specific increases in weight (p=0·004, +455 g), height (p=0·007, +0·70 cm) and waist circumference (p=0·004, +0·70 cm) at four years of age. The rs37386328 variant was not associated with birthweight (p=0·129). We replicated the finding that a CREBRF variant is associated with increased body mass. We then built on the original findings by demonstrating the prevalence of the rs12513649 and rs37386328 variants in multiple Pacific population groups and by demonstrating that the rs37386328 variant is associated with growth in early childhood. Pacific population groups experience a disproportionately high burden of obesity, starting in early childhood. This new knowledge offers potential for evidence-based interventions aimed at establishing healthy growth trajectories from the earliest possible age. .

2216T

The impact of MUC5B and KLK4 genes on dental caries. *T. Cavallari, H.M Salomão, S.J Moysés, S.T Moysés, R.I Werneck.* Pontifícia Universidade Católica do Paraná, Curitiba,Brasil.

The objective of this study was to identify the participation of *KLK4* and *MUC5B* genes polymorphisms on the susceptibility to dental decay. The studied population sample was composed of individuals who were affected (case) and unaffected (control) by dental decay, with 12 years old or more, paired by age and gender. The participants were recruited at the PUCPR (Pontifícia Universidade Católica do Paraná), using ICDAS criteria. Buccal cells were collected; the DNA was extracted and amplified using PCR. Uni-, bi- and multivariate analyses were performed. Two hundred patients were recruited, 122 (61%) women and 78 (39%) men. During the descriptive analysis, statistically significant results were found for ethnicity (p = 0.014, 0R = 0.036), biofilm (p < 0.001, OR = 10.887) and gingivitis (p < 0.001, 0R = 5.444). For genetic analyses, statistically significant results were found in the additive model for markers *rs*2735733 (p < 0.001), *rs*2249073 (p < 0.001) and *rs*2857476 (p < 0.001) for *MUC5B* gene and *rs*224267 (p = 0.009) for *KLK4* gene. The results from this study suggest that genetic polymorphisms of *MUC5B* and *KLK4* gene can influence in the development of dental caries.

2217F

Leveraging tissue specific omics data to estimate the disease/traits-related tissues. *R. Chen, Q. Wei, Q. Wang, H. Yang, Y. Ji, X. Zhong, B. Li.* Vanderbilt genetic institute, Vanderbilt University, Nashville, TN.

Genome-wide association studies (GWAS) have identified numerous causal loci associated with human complex traits and diseases. However, translating the discoveries into the understanding of mechanisms of complex traits and diseases remains a major challenge. Genes usually exert their functions in the tissue-specific manner, suggesting that the underlying molecular mechanisms are cell-type specific in general. Many efforts of linking causal genes to specific phenotype require tissue specific information to assure the accuracy and enhance the power in the detection of casual signals. Given the complexity of gene regulation and functionality in biology systems, the relevant tissues in which the genetic causality is manifested are obscure and have not been comprehensively explored. In this study, we systematically explored omics data across all tissues in GTEx to identify complex traits/diseases associating tissues. In particular, we focused on gene expression for this purpose, reasoning that expression of traits/disease associating genes have distinct expression patterns in relevant tissues and thus are able to infer the relevant tissues. We leveraged the extensive genetic findings in GWAS Catalog and assigned appropriate genes as the causal genes. Since the causal genes are generally unknown, we assign genes that are either proximal to GWAS loci or genes in which GWAS loci are located as the causal genes. We observed that the causal genes for a specific trait are highly expressed in the relevant tissues than others, exhibiting profound tissue-specificity. For example, using GTEx data, the top tissues with the greatest tissue-specificity for schizophrenia are all brain tissues, and the tissue with the most tissue-specific expression for blood traits is liver. To further test whether non-protein coding genes exhibit similar patterns, we explored psudo-genes and lncRNA genes but these genes were not able to identify relevant tissues, indicating that these non-coding genes are unlikely to play significant roles in genetic etiology of complex diseases. Finally, using coding genes, we comprehensively inferred tissues in GTEx for complex traits with >20 GWAS loci, and using blood traits as example we found that utilizing correct tissues improves the prioritization of known causal genes, implying its potential usefulness in identifying causal genes from GWAS discoveries.

2218W

Will big data close the missing heritability gap? *G.A. de los Campos[1,2], H. Kim[3], A. Grueneberg[1], A.I. Vazquez[1], S. Hsu[4,5].* 1) Epidemiology & Bio-statistics, Michigan State University, East Lansing, MI; 2) Statistics & Probability, Michigan State University, East Lansing, MI; 3) Duke Clinical Research Institute, Duke University, Durham, NC; 4) Physics and Astronomy, Michigan State University, East Lansing, MI; 5) Vice President for Research and Graduate Studies, Michigan State University, East Lansing, MI.

Despite the important discoveries reported by Genome-Wide Association Studies (GWAS), for most traits and diseases the prediction R-squared (R-sq.) achieved with genetic scores remains considerably lower than the trait heritability—a problem often described as the missing heritability (MH) gap. The gap between the trait heritability and prediction R-sq. achieved with genomic scores has two components. First, the amount of variance captured by a set of SNPs can be smaller than the trait heritability due to imperfect linkage disequilibrium between the alleles at the SNPs used for prediction and those at causal loci. A second component of the gap between the trait heritability and prediction R-sq. is given by the accuracy of the estimated effects. According to the framework just outlined, the use of sequence data and of very large datasets for model training should lead the end (or to a substantial reduction) of the MH gap. Until recently this hypothesis could not be tested empirically; however, this situation is rapidly changing as very large biomedical data sets from biobanks become available. In this study we propose a methodology that uses data and feature (i.e., SNPs) partitions to produce (an estimate of) a surface response relating prediction R-sq. with model complexity (e.g., # of SNPs) and the size of the data used to train the model. The proposed approach combines results from single-marker-phenotype associations with Bayesian whole-genome regression models and yields an estimate of the maximum prediction R-sq. that could be achieved with a SNP set. We applied the proposed methodology to data from the interim release of the UK-Biobank using various statistical methods for selecting SNPs and estimating their effects. Focusing on human height as a model trait and using 80,000 records for model training we achieved a prediction R-sq. in testing (n=22,221, unrelated to the training set) of 0.24 (95% CI: 0.23-0.25). Our estimates show that prediction R-sq. increases with sample size reaching a plateau at values that ranged from 0.1 to 0.37 for (GWAS-selected) SNP sets of 500 and 50,000, respectively. Soon much larger data sets will become available. Using the estimated surface response, we forecast sizable improvements in prediction R-sq. with larger sample size. We conclude that Big Data will lead to a substantial reduction of the MH gap, thus paving the ground towards a more intensive use of genomic information in precision medicine.

2219T

Performance of polygenic scores across ancestrally diverse populations. *L. Duncan[1], H. Shen[1], J. Pritchard[2], M. Feldman[3], K. Harris[4], B. Domingue[5].* 1) Psychiatry, Stanford University, Stanford, CA; 2) Genetics, Stanford University, Stanford, CA; 3) Biology, Stanford University, Stanford, CA; 4) Sociology & Public Policy, University of North Carolina at Chapel Hill, NC; 5) Education, Stanford University, Stanford, CA.

Polygenic scores (i.e. weighted metrics of phenotype-associated alleles) are now widely used in research and by companies like 23&Me. The majority of the genome-wide association studies (GWAS) on which these scores are based were conducted in European ancestry (EA) individuals. Questions remain regarding how well scores work in non-EA populations and whether, like Mendelian phenotypes, ancestry-related differences in phenotype-associated alleles may partially explain phenotypic differences across populations. To address these questions, we extracted data about predictive performance of polygenic scores (measured in phenotypic variance explained) from a review of 450 published papers and analyzed genotype and phenotype data (BMI, height) in the Adolescent Health dataset. We also tested the correlation between means of polygenic score distributions (for each of the 1000Genomes populations) with epidemiological data about mean height for relevant countries of origin, for the 1000Genomes populations. Across multiple phenotypes, polygenic prediction in African Americans (AA) was less strong than in EA individuals (AA on average 31% of variance explained in EA). The same was not true for Latino and Asian samples; predictive performance was usually better than in EA samples. Across the 1000Genomes populations, mean polygenic scores for each population were positively correlated with average height of individuals in regions from which the 1000Genomes data were obtained (r=.63, p=.009). Consistent with population genetic models (e.g. Scutari et al. 2016), prediction was worse in AA individuals when using EA GWAS data to weight polygenic scores. The same was not true for Latino and Asian samples. This could be due to sampling error, but also suggests that the relative importance of genetic and environmental effects may vary across populations. The positive correlation between mean polygenic scores for height, and mean phenotypic height, suggest that differences in polygenic scores across populations may contribute to population level differences in phenotypes. However, without evaluation of relevant environmental factors, it would be inaccurate to conclude that genetic differences explain observed phenotypic differences. Here we use the example of height, but this caveat is just as important for other complex genetic phenotypes.

2220F

Two novel loci detected and 10 known loci confirmed for estimated glomerular filtration rate in over 56,000 African Americans: The Million Veteran Program. *T.L. Edwards[1], J. Hellwege[1], A. Giri[1], O. Wilson[2], E.S. Torstenson[1], C.P. Kovesdy[3], K.A. Birdwell[4], C. Roumie[5], E. Siew[2], M. Matheny[6], C. Chung[5], D.R. Velez Edwards[8], C. O'Donnell[7], A. Hung[2] On behalf of The VA Million Veteran Program.* 1) Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/ Vanderbilt University, Nashville, TN; 2) VA TVHS Nashville, Division of Nephrology & Hypertension, Vanderbilt University, Nashville, TN; 3) Nephrology Section, Memphis VA Medical Center, Memphis TN; 4) Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) VA TVHS Nashville, Department of Medicine, Vanderbilt University, Nashville, TN; 6) VA TVHS Nashville, Department of Bioinformatics, Vanderbilt University, Nashville, TN; 7) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 8) Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center; Tennessee Valley Health Systems VA, Nashville, TN.

African Americans (AAs) have a higher risk of progressing to end stage renal disease (ESRD) than non-Hispanic European Americans. Genetic variants influencing kidney function have been identified through genome-wide association studies (GWAS) of chronic kidney disease and ESRD, as well as estimated glomerular filtration rate (eGFR) in AAs, however, small sample sizes have hindered discovery. We have performed a GWAS of eGFR in 56,253 AAs from the Million Veteran Program (MVP). eGFR was regressed on to common genetic variants (minor allele frequency > 1%) imputed to the 1000 Genomes reference panel adjusted for age, sex, body mass index and the top ten principal components. Analyses were performed by strata of diabetes and hypertension status, estimates from which were aggregated with fixed effects meta-analysis. A total of 799 SNPs representing 12 loci exceeded genome-wide significance. The most significant association was at a previously known locus, *SPATA5L1/GATM* on chromosome 15 (p-value = 2.02×10^{-39}). Two novel loci were detected between *TRIM69* and *C15orf43* (rs144803907; p-value = 4.50×10^{-9}) and near *OLFR690* (rs75113983; p-value = 2.25×10^{-8}). Other previously reported signals detected include: *HBB* (p-value = 1.54×10^{-18}), *FBXL20* (p-value = 7.51×10^{-11}), *UNCX* (p-value = 9.29×10^{-11}), *SLC47A1* (p-value = 3.92×10^{-10}), *SLC22A2* (p-value = 7.50×10^{-10}), and *UBE2Q2* (p-value = 5.12×10^{-9}). Among the known loci, this is the first significant observation in African ancestry populations, emphasizing the transethnic nature of genetic variation contributing to eGFR levels. Evaluation of genetically predicted gene expression (GPGE) with MetaXcan identified increasing expression of *DUOX2* in left ventricle of the heart (p-value = 1.16×10^{-13}) with decreasing eGFR, not *SPATA5L1* or *GATM*, even though these genes fall under the same GWAS peak. GPGE associations where increasing gene expression is associated with increasing eGFR include *SHF* in subcutaneous adipose (p-value = 1.60×10^{-9}), *SH2D2A* in visceral adipose (p-value = 4.19×10^{-13}), and *PNMT* in tibial nerve (p-value = 3.32×10^{-8}). Overall, this is the largest GWAS of eGFR in AAs to date, which replicates previously identified loci from whites, detects two novel loci, and is the first to evaluate genetically predicted gene expression associations with eGFR in AAs.

2221W

Slit2-Robo1 signaling may play a role in spontaneous preterm birth. *M. Karjalainen[1,2], H. Tiensuu[1,2], A.M. Haapalainen[1,2], M. Ojaniemi[1,2], M. Rämet[1,2], M. Hallman[1,2].* 1) PEDEGO Research Unit, and Medical Research Center Oulu, University of Oulu, Oulu, Finland; 2) Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland.

Background and aims: Spontaneous preterm birth (SPTB) is a major global healthcare problem associating with significant mortality and morbidity. SPTB has a genetic background with a potential role for both maternal and fetal genetic factors. Placental tissue has a regulatory role in the onset of labor and delivery. Here, our aim was to identify biological mechanisms leading to SPTB by combining the findings of a genome-wide association study (GWAS) with those of transcriptomics of placental tissue. **Methods:** We started by performing a GWAS to find susceptibility genes for SPTB. The study population comprised SPTB infants (gestational age, GA <36 weeks, *n* = 247) and infants born at term (GA 38-41 weeks, *n* = 419) from Oulu and Tampere University Hospitals, Finland. DNA samples were genotyped with the Illumina Human-CoreExome BeadChip. Association analysis was performed with SNPTEST, v. 2.5.2, under the additive model. The arising candidate genes were compared to genes showing significant expression level changes in SPTB placentas. Expression changes of selected genes were verified by qPCR in SPTB (*n* = 24) and term (*n* = 27) placentas. Immunohistochemistry was further used to visualize the expression patterns in the placental tissue. **Results:** In GWAS, we detected several suggestive association signals ($p < 10^{-5}$). The most significant pathway was Axon guidance ($p = 2 \times 10^{-7}$). When we screened the transcriptomic data for overlapping findings, we detected that one of the genes with suggestive signals, *SLIT2* encoding slit guidance ligand 2, was upregulated in SPTB placentas. Using qPCR, we further verified that *SLIT2* and the gene encoding its receptor, roundabout guidance receptor 1 (*ROBO1*), are upregulated in SPTB compared to term placentas. By immunohistochemistry, we detected that both Slit2 and Robo1 are localized to villous and extravillous trophoblasts. **Conclusions:** *SLIT2* was identified as a potential susceptibility gene for spontaneous preterm birth both in GWAS and in transcriptomics of placental tissue. We propose that Slit2-Robo1 signaling plays a role in SPTB. To replicate the finding, we will analyze the most promising *SLIT2* SNPs in another study population of SPTB and term infants. Further studies include silencing of *SLIT2* and *ROBO1* in trophoblasts to identify the downstream genes and their functions.

2222T

Adult height and risk of cardiometabolic disease. *E. Marouli[1], F. Del Greco M.[2], C. Astley[3], Z. Kutalik[4], R.J.F. Loos[5], J. Hirschhorn[6], P. Deloukas[1].* 1) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 2) Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy; 3) Boston Children's Hospital and Broad Institute, MA, USA; 4) Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland; 5) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 6) Broad Institute of Harvard and MIT, Cambridge, MA, USA.

Consistent with epidemiological observations that suggest phenotypic correlation between height and other traits or disease, many common variants associated with human adult height in GWAS have been reported to have pleiotropic effects. The latest large-scale genetic study of height identified 520 associated variants (of which 83 are low-frequency or rare), which were enriched for nominal associations with several cardiometabolic traits (in the opposite direction). To further investigate the observed pleiotropy between height and cardiometabolic traits including disease and test whether there is any causal relationship, we considered all 864 height-associated variants, which explain ~27% of height heritability, in ~150,000 individuals from the UK Biobank (UKBB) study. We calculated an unweighted height genetic score and tested its association with BMI, blood pressure (BP), smoking status, hypertension, coronary artery disease (CAD) and type 2 diabetes (T2D). We observed that individuals with more height-increasing alleles had a reduced risk for both CAD, odds ratio (OR) of 0.83 for top vs bottom quintile (95% CI 0.91-0.76, P=4.2E-05), and T2D, OR=0.87 (95% CI 0.81-0.94, P=5.77E-04). Initial analyses suggest that the effect of the height genetic score on CAD is not mediated through BMI, T2D, smoking status, or hypertension, as the adjusted model did not result in attenuation of the signal. We found significant associations of the height score with known CAD risk factors (P<0.0001) such as BMI, blood pressure and WHR, but not with smoking (ever-smoking vs. never smokers; P=0.208). We further performed Mendelian Randomization (MR) analyses in UKBB in order to investigate whether the effect of height on CAD and T2D might be causal. The estimate from the MR-Egger method for a causal effect between genetically determined height (inverse normal rank transformed) and CAD was: beta -0.14 (95% CI, -0.25 to -0.02, P=0.016), but no significant causal effect was observed for T2D: beta 0.021 (95% CI, -0.11 to 0.15, P=0.75). In conclusion, our results so far indicate that that higher height status (instrumented by the 864 variants), reduces the odds of CAD risk by 13%, but has no effect on T2D risk. The observed association may be mediated by shared biological processes for height and CAD development, including lipid levels (data to become available in Q4).

2223F

GWAS of early childhood caries in an Appalachian population. *E. Orlova[1], M.K. Lee[3], E. Feingold[1,2,3], D.W. McNeil[4], R.J. Crout[5], R.J. Weyant[6], M.L. Marazita[1,3,7], J.R. Shaffer[1,3].* 1) Dept. of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Dept. of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Center for Craniofacial and Dental Genetics, Dept. of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 4) Depts. of Psychology, and Dental Practice & Rural Health, West Virginia University, Morgantown, WV; 5) Dept. of Periodontics, School of Dentistry, West Virginia University, Morgantown, WV; 6) Depts. of Dental Public Health, and Information Management, School of Medicine, University of Pittsburgh, Pittsburgh, PA; 7) Clinical and Translational Sciences Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA.

Early childhood caries (ECC) is a complex, multifactorial disease with multiple environmental and genetic etiological factors, and serious consequences for the child's subsequent oral health and quality of life. It is the most common chronic disease of childhood, affecting almost one in four children nationwide, and disproportionately impacts vulnerable populations including ethnic/racial minorities and low SES groups. Although ECC is known to be heritable in part, few associated genetic loci have been identified. Here, we report a GWAS of ECC in 447 Caucasian, non-Hispanic children of up to six years of age, recruited through the Center for Oral Health Research in Appalachia study (COHRA-1). ECC was assessed by intraoral examination by a dentist or research dental hygienist. DNA from saliva samples was genotyped on the Illumina Human610 Quadv1 B array and imputed to the 1000 Genomes Project phase 1 (June 2011) reference. Logistic regression while adjusting for age, sex, and the first principal component of ancestry was used to test association under the additive genetic model for 3.37 million common (MAF > 5%) SNPs passing quality filters. No significant associations (p-value <5x10[-8]) were found, but 11 loci were associated at a suggestive level (p-value <10[-5]). The top result of the ECC GWAS was rs1582413 (p-value 1.48x10[-7]; OR 2.76; 95% CI 1.89-4.04), within an intron of the gene *MYT1L*, which encodes a neuronal transcription factor. Identifying genetic variants predisposing to caries in primary dentition may ultimately inform downstream caries treatment and prevention efforts, and help alleviate disparities in disease burden currently impacting disadvantaged groups. This work was made possible by the following grants: U01-DE018903, R01-DE014899, and R03-DE024264.

2224W

Novel genes identified by integrating genome-wide association analysis with transcriptomics in severe chronic obstructive pulmonary disease and quantitative emphysema. *P. Sakornsakolpat[1,2], J.D. Morrow[1], P.J. Castaldi[1], C.P. Hersh[1], E.K. Silverman[1], A. Manichaikul[3], M.H. Cho[1].* 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 3) Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA.

   Background: Genome-wide association studies (GWAS) have revealed multiple loci associated with chronic obstructive pulmonary disease (COPD [MIM 606963]) and related traits. Analyses combining GWAS with gene expression data, using a gene expression model in reference datasets, may provide insight into genetic regulation at these loci and provide genes for further functional analysis. We hypothesized that applying these methods to COPD and related traits could provide evidence of genetic regulation of previously identified and potentially novel loci. **Methods:** We obtained GWAS summary statistics from a meta-analysis of severe COPD (n = 3,497 cases and 5,704 controls), emphysema assessed as low attenuation area at -950 Hounsfield units (HU) (%LAA-950), and emphysema assessed by HU at 15th percentile of the CT lung density histogram (Perc15) (n = 12,031). We applied the PrediXcan / MetaXcan model to using transcriptome reference databases based on whole blood (n = 6,588 genes) and lung (n = 6,448 genes) from the Genotype-Tissue Expression (GTEx) project. We used a Bonferroni threshold correcting for the number of genes in each database. To provide further evidence of colocalized signals, we applied Bayesian colocalization to significant genes. We also performed pathway enrichment analysis on the top associated genes. **Results:** We identified genes regulated at previously identified loci from GWAS of COPD and related traits, and also identified novel associations (at sub-genome-wide significant loci), including *WNT3* for severe COPD, *SNTB2* and *DCBLD1* for %LAA-950, and *PSMA4*, *SNTB2*, and *ITGA1* for Perc15. However, most GWAS and expression quantitative loci (eQTL) showed poor evidence of colocalization, suggesting that the GWAS and eQTL may be driven by distinct variants. Pathway analysis of nominally significant genes for severe COPD was suggestive of over-representation of genes regulating Wnt signaling pathway and ferrous iron binding, while TCR signaling and signaling events of B cell receptor were over-represented in %LAA-950 analysis (false discovery rate < 0.05). **Conclusions:** Integrating a reference transcriptome may improve power and identify genes regulated by phenotype-associated variants, though some of these associations may be driven by LD. Further work to confirm these findings and identify the role of specific genetic variants in regulating gene expression may lead to new insights regarding how GWAS loci confer susceptibility to disease.

2225T

Ayurveda based deep phenotyping, a likely game changer for gene hunt in complex traits. *B.K. Thelma[1], G. Juyal[1,6], A. Pandey[1], N. Rana[1], S. Negi[2], U. Kumar[3], B. Bhat[4], R.C. Juyal[5].* 1) Department of Genetics, University Delhi South Campus, New Delhi-110021, India; 2) National Institute of Pathology, Safdarjung Hospital Campus, New Delhi-110029, India; 3) Department of Rheumatology, All India Institute of Medical Sciences, New Delhi-110029, India; 4) Department of Ayurveda, Holy Family Hospital, New Delhi-110025, India; 5) Regional Centre for Biotechnology, Gurugram, Haryana-122008, India; 6) Current Address: School of Biotechnology, Jawaharlal Nehru University, New Delhi-110067, India.

   Objective: To date, an appreciable amount of genome wide association studies (GWASs) have been performed for several complex traits. Despite their impeccable design and large cohorts, our knowledge on genetic architecture of diseases has been limited and thus hindering the progress in the area of predictive, preventive and personalised medicine (PPPM). One of the reasons for such restricted success is attributed to clinical heterogeneity demanding exploration of complementary approaches which will enable effective phenotypic characterization of individuals to obtain homogeneous study cohorts. Apart from OMICS approach, we believe Ayurveda, the traditional system of medicine practised in India adopts/uses effective deep phenotyping and this may propel a revolution in the outcome of complex trait genetics research. In this pilot study, we performed a GWAS of Rheumatoid arthritis (RA) (Amavata in Ayurveda) patients and healthy controls, who were deeply phenotyped into vata, pitta and kapha prakriti on the basis of Ayurveda. **Methods:** Genotyping data of 246 RA cases and 310 controls was retrieved from the total genotype data generated on Illumina Human660W Quad Bead-Chip genotyping platform, used in a recent RA GWAS study performed in the laboratory. Prakriti-wise test of association and Random forest analyses were employed. **Results:** 244 (vata=49; pitta=117; kapha=78) RA cases and 293 (vata=33; pitta=175; kapha=85) controls were included in the analysis. High quality genotype data for 480418 SNPs passed quality control. Prakriti-wise test of association and Random forest analyses revealed prakriti specific functionally relevant RA genes with significant effect sizes. **Conclusions:** Our findings suggest that Ayurveda based stratification of patients may enable investigations of homogeneous subsets of patients which in turn may aid in a better development of multi-omics signature based prognostic and diagnostic markers. It is also expected to generate insights into personalized nutritional and therapeutic intervention strategies.

2226F

Genomic features of loci associated with multiple complex traits in genome-wide association studies. *Y.-F. Wang[1], T.-Y. Wang[1], Y. Zhang[1], H. Zhang[1], J.J. Shen[1], M. Guo[1], J. Yang[1,2], P.C. Sham[3], Y.L. Lau[1,2], W. Yang[1,2,3].* 1) Paediatrics & Adolescent Medicine, The University of Hong Kong, Hong Kong, Hong Kong, Hong Kong; 2) Shenzhen PID Diagnostic and Therapeutic Laboratory, HKU-Shenzhen Hospital, Shenzhen, China; 3) Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong.

Genome-wide association studies (GWAS) have revealed many genomic regions associated with multiple traits, known as cross-phenotype (CP) associations. These regions may harbor genes and regulatory elements of special significance to human diseases and may help reveal shared disease mechanisms. Making use of the rich data from GWAS, we demonstrated that sequences in CP regions are more conserved, especially in the 3'-untranslated regions. As expected, functional elements, such as DNase-I hypersensitive sites, transcription factor binding sites and histone modification regions, were more enriched in CP regions. Based on these observations, we further found that the trait-associated variants in CP loci generally exert larger effects on disease susceptibility across a range of different diseases and traits. From the GWAS Catalog, we found that the variants associated with multiple diseases are more likely to be nonsynonymous, whereas independent variants in the same locus associated with different traits may function under different cellular context, suggesting complex regulatory mechanisms underlying multiple disease associations. In summary, these results may provide new insight into the genomic features of CP loci, and help better understand regulatory mechanisms underlying CP loci.

2227W

A novel relationship between GWAS-identified sleep traits loci in sleep duration variation in healthy adults. *X. Xu, A.T. Magis, J.C. Lovejoy.* Arivale Inc., 710 2nd Avenue, Suite 410, Seattle, WA, 98104.

Background: Sleep is increasingly recognized as an important lifestyle contributor to health. Chronic sleep disturbances (e.g., sleep deprivation, insomnia, large night-to-night sleep duration variation) are significantly associated with obesity, cardio-metabolic diseases, psychiatric disorders and all-cause mortality. Although sleep duration and disruption are affected by various environmental factors, heritability has been estimated at ~40% for sleep duration and ~25-45%for insomnia. Recently, GWAS analyses identified a few genetic loci associated with sleep duration and insomnia. The goal of this study is to further explore the effects of those SNPs in sleep duration variation. **Method and Results**: De-identified data from individuals in a commercially-available lifestyle coaching program (Arivale Inc., Seattle, WA) were collected starting from July 2015. The Arivale program involves health coaching on exercise, nutrition, stress management and other wellness goals, including sleep. Data from a total of 1832 Arivale program individuals were included in the present analysis. The study population included 50.6% females and 85.4% European-Americans. Their average age was 49.7 years, and average BMI was 26.6 kg/m^2. Upon joining the Arivale program, each individual was provided a Fitbit Charge HR$_{TM}$ (San Francisco, USA) to record daily activity and sleep information. Fitbit Charge HR$_{TM}$is a wristband that tracks all-day activity and sleep. Whole genome sequencing (WGS) was performed using whole blood samples and Illumina HiSeq X10 with sequencing mode PE150 (Illumina Inc., San Diego, USA). We collected nightly sleep data from each individual for 12-month time period, and calculated the individual's habitual sleep duration and sleep duration variation. A total of 19 previously identified sleep related SNPs were tested. General linear regression was performed to evaluate the relationships with sleep duration variation. It has been reported that rs113851554 in *MEIS1* gene is the strongest association for insomnia symptoms. In our study, rs113851554 also shows significant (FDR < 0.05) effect on sleep. In general, carriers exhibit 16-minutes higher sleep duration variations than homozygous. **Conclusion:** This study replicated previous identified sleep associated loci, and further explored the relationships with sleep duration variation. Our novel findings provide new insights into the biology of sleep and circadian rhythms in humans.

2228T

Structural variation influencing complex traits and metabolomic measurements. *A. Sabo[1], P. Mishra[1], W. Salerno[1], T. Chiang[1], B. Yu[2], A. Morrison[2], G. Metcalf[1], D. Muzny[1], R. Gibbs[1], E. Boerwinkle[1,2].* 1) Baylor College of Medicine, Houston, TX; 2) Human Genetics Center, University of Texas Health Science Center, Houston, TX.

Previous studies of the genetic architecture of complex traits have focused on the role of single nucleotide variation (SNV), but in fact structural variation (SV) covers more of the genome than single nucleotide variation. We performed Nimblegen whole exome capture followed by Illumina sequencing for 9,343 individuals from a large sample of European-Americans (EA) and African-Americans (AA) from the longitudinal ARIC cohort study. The data were processed using the integrated Mercury pipeline, followed by SV analysis using the XHMM software to identify rare copy number variants (CNVs). Although there has been intense effort on the detection and calling of SVs, there is very little work on modeling the genotype-phenotype relationship of SVs. We identified rare CNVs in 6,951 samples of European descent and 2,392 samples of African American descent, and investigated their impact on multiple core phenotypes and metabolomic measurements. We identified 36,390 CNVs, of which 11% were deletions, 26% duplication, and 63% were copy number variations - with both deletions and duplication detected in different samples. Most of the CNVs detected were relatively small and rare; 82% were smaller than 100kb, and 77% had a combined deletion and duplication MAF <1%. We have implemented both CNV level and gene level genotype-phenotype analysis of 245 metabolomic measurements and 20 core phenotypes from the ARIC study. We have transformed the phenotype to normality and included principal components in the model, and analyzed the two cohorts separately. We have confirmed previously identified association of proline dehydrogenase 1 copy numbers with plasma proline levels. We have identified several statistically significant novel findings in the AA cohort, which were replicated with nominal significance (pval < 0.05) in the EA cohort. These include association of CNVs overlapping *UGT2B11* gene with arginine levels, *ANKRD30B* with creatinine levels, and *FAM21B* with homocitrulline levels. The confirmation of these results is currently ongoing.

2229F

Genetic polymorphisms of LIN28B and MKRN3 in association with precocious puberty. *K. Lee[1], J. Yee[2], H. Kim[1], J. Kim[1], H. Gwak[2].* 1) College of Pharmacy, Chungbuk National University, Cheongju, Daejeon, South Korea; 2) College of Pharmacy & Division of Life and Pharmaceutical Sciences, Ewha Womans University, Seoul, Korea.

Precocious puberty is clinically defined as the development of secondary sexual characteristics before the age of 8 years in girls and 9 years in boys. Puberty timing is known to be regulated by complex interactions of environmental, genetic, nutritional and socioeconomic factors. Although the function of genes is not completely known, MKRN3 and LIN28B are regarded as important genes for precocious puberty. But, there are few studies conducted on these genes for precocious puberty in Korea and no study done in boys. Therefore, we aimed to investigate the effect of MKRN3 and LIN28B gene mutations on precocious puberty in both girls and boys. The follow-up of the Ewha Birth & Growth Cohort study was conducted annually since 2005 including children age 3, 5, 7, 8 and 9 years. The first clinical follow-up of puberty was assessed in 9-year-olds in 2011 and children aged 7 to 9 years at 2011 to 2012 were recruited for this study. Four SNPs of MKRN3 gene (rs2239669, rs12441827, rs657457 and rs12148769) and two SNPs of LIN28B gene (rs314280, rs314276) were selected based on the data from previous studies and the minor allele frequency (MAF) of 15% or above. A total of 110 girls and 90 boys were included in the analyses. There were 96 girls (87.3%) in Tanner stage 1, 14 girls in stage 2 (12.7%), and no one belonged to stage 3. Among boys, 44 were found to be in stage 1 (48.9%), 34 (37.8%) in stage 2, and 12 (13.3%) in stage 3. The genotyping assay results indicated linkage disequilibrium (LD) between rs2239669 and rs6576457 in MKRN3 and rs314280 and rs314276 in LIN28B. High probability of precocious puberty was shown in girls with G allele for rs314280 and C allele for rs314276 and in boys with G allele for rs6576457 and TT homozygous for rs12441827. We identified that SNPs of MKRN3 and LIN28B were partially associated with precocious puberty. Exact mechanisms involved in MKRN3 and LIN28B mutations need to be investigated and an ongoing effort to find other factors associated with the puberty should be maintained.

2230W

Finland, combining the population isolate structure with nationwide health care data for gene discovery. *A. Palotie[1,2,3,4], J. Koskela[1,2,3], I. Surakka[1], P. Häppölä[1], P. Heikkula[1], M. Alanne-Kinnunen[1], M. Perola[1,5], V. Salomaa[5], M.J. Daly[1,2,3], S. Ripatti[1].* 1) Institute for Molecular Medicine, Finland, University of Helsinki, Helsinki, Finland; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts General Hospital, Boston, MA, USA; 5) Institute of Health and Welfare, Helsinki, Finland.

Finland is the largest population isolate in Europe, which provides benefits in identification of rare and low frequency variants. Due to its small founder population, multiple genetic bottlenecks and rapid population growth a set of rare, deleterious variants have been enriched and are present in a frequency of 0.005-0.05. The 36 recessive diseases, called the Finnish Disease Heritage, best demonstrate this bottleneck effect. However, this feature is not limited to recessive, Mendelian variants; the power to identify deleterious variants that are potential drug targets has been demonstrated using population and birth cohorts as well as disease collections linked to national health records (Lim et al, 2014, Stoll et al 2013, Flannick et al 2014). The Nordic Countries have unique health care registers that cover health information of every resident for their entire life span. The 11-digit personal identification number links data from all different registries. The Finnish phenotype data resources combined with GWAS, WES and WGS data and efficient variant imputation (Surakka et al doi:https://doi.org/10.1101/080770) enable unique study designs. For example, in a PheWAS against 200 clinical outcomes we identified a Finnish enriched loss of function variant in the ANPTL8 gene protective for type 2 diabetes and high triglycerides. Another example is prediction of mortality using phenotypic cross disease categories like life time drug usage and life time cumulative number of hospitalization (YODA Score, Years of Drugs Applied), in population cohorts of 23 000 individuals (Finrisk) (hr=1.023 per YODA year, p<2e-16; hr=1.074 per month of hospitalization, p<2e-16 respectively). These analyses and abilities to re-contact study subjects enable to identify potential shared genetic backgrounds over traditional diagnostic boundaries. Potentially they provide tools for preventive public health measures and can identify low frequency variants that can pinpoint new cellular pathways that contribute to disease pathogenesis and as a consequence might provide potential drug discovery targets. To further strengthen the Finnish data resource, we have, together with international collaborators, initiated a study that aims to combine genome and health register data from up to 500 000 Finnish individuals. This study is likely to provide a widely usable data resource for genetic analyses, drug target discovery and reference data for implementation of genomics in health care.

2231T

Whole exome sequencing reveals a novel candidate gene, *HSPA1L*, for spontaneous preterm birth. *J.M. Huusko[1,2], M.K. Karjalainen[1], G. Zhang[2], B.E. Graham[3], E.G. Farrow[4], N.A. Miller[4], J.C. Murray[5], B. Bedell[5], F.L. Bødker[6], K. Christensen[6], S.M. Williams[3], M. Rämet[1], S.F. Kingsmore[7], K.K. Ryckman[8], M. Hallman[1], L.J. Muglia[2].* 1) PEDEGO Research Unit and Medical Research Center Oulu, University of Oulu and Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland; 2) Division of Human Genetics, Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, March of Dimes Prematurity Research Center Ohio Collaborative, Cincinnati, OH; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH; 4) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO; 5) Department of Pediatrics, University of Iowa, Iowa City, IA; 6) Institute of Public Health, University of Southern Denmark, Odense, Denmark; 7) Rady Children's Institute for Genomic Medicine, Rady Children's Hospital, San Diego, CA; 8) Department of Epidemiology, College of Public Health and Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA.

Background and aims. Preterm birth is the single leading cause of infant mortality and morbidity. The etiology of preterm birth is complex, and both environmental and genetic factors contribute to the overall risk. However, the genetic bases for preterm birth remain largely unknown. Our aim was to investigate presence of rare, likely damaging nucleotide variants in families with recurrent spontaneous preterm birth (SPTB). **Methods.** We performed a whole exome sequencing (WES) for 17 mothers from seven families with recurrent SPTB. Mothers were of Northern Finnish origin, and DNA samples were collected in Oulu University Hospital, Finland. WES was performed using Illumina HiSeq 2500, followed by generation and alignment of sequence data. Variant calls resulting from sequenced exomes were annotated and filtered using Ingenuity Variant Analysis. Variants passing the quality requirements were prioritized according to their frequency, pathogenicity and mode of inheritance. Only genes harboring rare (MAF <1% in European populations), likely damaging variants were subjected into pathway analysis (Ingenuity). Findings were replicated using additional set of WES data including 192 mothers (giving birth preterm) of European origin. **Results.** We identified multiple rare exonic missense variants in genes common for several affected mothers belonging to different Finnish families. One of these genes encodes the heat shock protein family A (Hsp70) member 1 like (*HSPA1L*). *HSPA1L* belongs to the Glucocorticoid receptor signaling pathway that was the most significant pathway within a group of mothers with recurrent preterm births. Rare, likely damaging, variants within *HSPA1L* were also seen in the replication set. Furthermore, *in silico* functional analysis of one of the rare variants, seen in two Finnish families, suggested that the variant generates an additional phosphorylation site that could affect the protein-binding kinetics. **Conclusions.** We found rare likely damaging missense variants in *HSPA1L* within four Finnish families with recurrent SPTB. This finding was supported by additional replication data as well as by *in silico* assessments. One of the rare damaging variants potentially alters the post-translational modification of the protein and thus could affect the protein-binding properties.

2232F

Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. *D. Qiao[1], C. Lange[4], A. Ameli[1], A. Sharma[1], N.M. Laird[4], H. Chen[5], D. Prokopenko[1], M. Parker[1], T.H. Beaty[6], J.D. Crapo[7], K.C. Barnes[8], D.A. Nickerson[9], M. Bamshad[10], C.P. Hersh[1,2], J. Morrow[1], B.D. Hobbs[1,2], R. Busch[1], D.A. Lomas[11], Y. Liu[12], M. Spitz[12], A. Agusti[13], B.J. Make[7], P.M.A. Calverley[14], C.F. Donner[15], E.F. Wouters[16], J. Vestbo[17], P.D. Paré[18], R.D. Levy[18], S.I. Rennard[19], R. Tal-Singer[20], V.M. Pinto-Plata[21], N. Marchetti[22], R. Bueno[3], B.R. Celli[2], G.J. Criner[23], E.K. Silverman[1,2], M.H. Cho[1,2], Lung GO, COPDGene Investigators, NHLBI Exome Sequencing Project, University of Washington Center for Mendelian Genomics.* 1) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 3) Division of Thoracic Surgery, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 4) Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; 5) University of Texas Health Science Center at Houston, Houston, TX, USA; 6) Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 7) National Jewish Health, Denver, CO, USA; 8) Division of Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA; 9) Department of Genome Sciences, University of Washington, Seattle, WA; 10) Division of Genetic Medicine, Department of Pediatrics, University of Washington and Seattle Children's Hospital, Seattle, WA, USA; 11) University College London, London, UK; 12) Dan L. Duncan Comprehensive Cancer Center, Department of Medicine, Baylor College of Medicine, Houston, TX, USA; 13) Thorax Institute, Hospital Clinic, IDIBAPS, University of Barcelona, CIBERES, Barcelona, ES; 14) University of Liverpool, Liverpool, UK; 15) Mondo Medico di I.F.I.M. srl, Multidisciplinary and Rehabilitation Outpatient Clinic, Borgomanero (NO), IT; 16) University Hospital Maastricht, Maastricht, NL; 17) University of Manchester, Manchester, UK; 18) Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, BC, CA; 19) University of Nebraska Medical Center, Omaha, NE, USA; 20) GSK Research and Development, King Of Prussia, PA, USA; 21) Department of Critical Care Medicine and Pulmonary Disease, Baystate Medical Center, Springfield, MA, USA; 22) Department of Thoracic Medicine and Surgery, Temple University School of Medicine, Philadelphia, PA, USA; 23) Division of Pulmonary and Critical Care Medicine, Temple University School of Medicine, Philadelphia, PA, USA.

   Introduction: Alpha-1 antitrypsin deficiency demonstrates that rare coding variants of large effect can influence chronic obstructive pulmonary disease (COPD) susceptibility. To identify such rare coding variants, we utilized an extreme-phenotype study design and conducted whole exome sequencing analysis in multiple cohorts of severe COPD. **Methods:** We sequenced 1019 unrelated subjects from the COPDGene study; cases had $FEV_1 < 50\%$ predicted, and controls had normal lung function and age > 55. We also sequenced 1579 related subjects from the Boston Early-Onset COPD Study (probands age < 53, $FEV_1 < 40\%$ predicted) and the International COPD Genetics Network study (ICGN) (probands age < 65, $FEV_1 < 60\%$ predicted). We conducted three complementary analyses, including gene- and pathway-based association tests for 1)very rare variants (MAF < 0.1%), 2) predicted deleterious variants (MAF < 1%), and 3) a family-based analysis using the gene-based segregation method (GESE). Unrelated subjects were sequenced in two batches and analyzed using SeqMeta and the resampling method in SKAT combined with Fisher's method. Family-based data were analyzed on lung function using MONSTER. We additionally sought evidence for overlap between case-control and family-based data using network methods. **Results:** In the analysis of very rare and rare predicted deleterious variants, we found that SeqMeta gave spurious results in genes with multiple low-count variants; no genes were significant using Fisher's method. Top pathways in the COPDGene data included *GATA3* and Th2 cytokine gene expression (p = 6.7e-06) for very rare variants, and the Jak-STAT pathway for rare deleterious variants (p=3.0e-05), neither of which replicated in family-based analyses. Using GESE in the family-based data, we identified two significant genes segregating in multiple pedigrees: *TBC1D10A* and *RFPL1*. We found no significant overlap of top results in case-control and family studies; however, network analysis found genes with p-value < 0.05 were significantly close to each other in a protein-protein interaction (PPI) network. We also observed several rare variant in genes previously identified that were nominally significant after conditioning on the known GWAS variant. **Conclusions:** These results likely reflect heterogeneity of genetic risk, insufficient power, and limitations of annotation. Ongoing efforts include expanding the sample size and further examination of GWAS loci related to COPD and lung function.

2233W

Enhanced methods to investigate the role of *Trans*-eQTL to complex traits. *C. Giambartolomei[1], N. Mancuso[1], H. Shi[1], B. Strober[2], A. Battle[2], B. Pasaniuc[1].* 1) UCLA, Department of Pathology & Laboratory Medicine, Geffen School of Medicine, 10833 Le Conte Ave, Los Angeles, U.S.A; 2) Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, U.S.A.

Most GWAS and eQTL overlap have largely focused on the role of local (*cis*) regulation of gene expression to complex traits. Unfortunately, only a limited fraction of GWAS signals can be explained by *cis*-eQTL variants. On the other hand, *trans*-eQTLs may have broader effects on the transcriptome and may have important phenotypic consequences as they are known to be tissue-specific. The contribution of distal *trans* gene expression regulations to GWAS is not well explored, mostly because of the difficulty in assessing trans regulation (lower effects coupled with the multiple-testing burden imposed by the number of genome-wide tests performed). Here we introduce new methods to co-localize *trans*-eQTL signals with GWAS overlap and quantify the contribution of *trans*-eQTLs to disease risk. Our approaches build on recently proposed methods for *cis*-eQTL co-localization that allow for multiple causal variants to yield estimates of local correlation in the *trans*-effects on gene expression and effects on complex traits. Our approaches only require summary association statistics (both eQTL and GWAS). We validate our methods using extensive simulations starting from real data and show that our methods outperform existing approaches while also providing an interpretation of *trans*-eQTL/GWAS co-localization. We analyze large-scale GWAS of more than 50 complex traits joined with gene expression across 44 tissues (GTEx) to identify GWAS loci explained by *trans* regulation of putative risk genes. We also quantify the *cis* versus *trans* contribution to disease using genetic correlation and examine the location of the disease-relevant eQTL by distance of the regulated gene. This help us better understand the location of disease variants influencing expression and the relevance of the *cis* versus *trans* distinction.

2234T

Connection of RAB8A and MED16 with implications on 5'-deoxy-5-fluorouridine response. *C.M. Murtagh, C. Lucas, A.L. Stark.* University of Notre Dame, Notre Dame, IN.

Treatment plans can be individualized to optimize drug response in cancer patients with a better understanding of how genetics influences protein expression and drug response. Our goal was to utilize genetic associations between protein levels and drug response to identify functional candidates for capecitabine therapy optimization. Capecitabine is an oral prodrug used in breast cancer treatment; 5'-deoxy-5-fluorouridine is the activated form of the drug used for *in vitro* studies. We identified three SNPS localized to the short arm of chromosome 19 in RAB8A that were associated with MED16 protein expression and 5'-deoxy-5-fluorouridine response in lymphoblastoid cells. RAB8A is a member of the RAS superfamily, highly conserved, and implicated in tumorigenesis. Past studies have revealed that MED16 is a coactivator involved in the transcriptional regulation of many RNA-polymerase II dependent genes and plays a role in Vitamin D reception, which may affect calcium homeostasis, cell proliferation, and cell differentiation. Thus, expression of MED16 may significantly alter transcription levels of RNA-polymerase II dependent genes and affect functions such as cell proliferation which are abnormal in cancerous cell cycles. No current literature links MED16 or RAB8A with chemotherapy or cancer survival. We knocked down RAB8A using siRNA in MCF7 breast cancer cells and assessed MED16 expression and capecitabine response. Decreased expression of MED16 is hypothesized to lead to poorer capecitabine outcomes, suggesting that MED16 is a contributor to capecitabine response. Results from this study hope to indicate that further research should be performed regarding the potential of MED16 to serve as a prognostic marker for individuals treated with capecitabine.

2235F

Cross-altitude analysis suggests a turning point at the elevation of 4,500m for polycythemia prevalence in Tibetans. *C. Cui[1], H. Zhang[1,2], Y. He[2,4], NA. Ouzhuluobu[1], NA. Baimakangzhuo[1], NA. Duojizhuoma[1], NA. Dejiquzong[1], NA. Bianba[1], NA. Gonggalanzi[1], Y.Y. Pan[1], NA. Qula[1], NA. Kangmin[1], NA. Cirenyangji[1], NA. Baimayangji[1], C.J. Bai[1], W. Guo[1], NA. Yangla[1], Y. Peng[2], X.M. Zhang[2], K. Xiang[2], Z.H. Yang[2], S.M. Liu[3], X. Tao[3], NA. Gengdeng[3], W.S. Zheng[1,5], Y.B. Guo[1,5], T.Y. Wu[3], X.B. Qi[2], B. Su[2].* 1) School of Medicine, Tibetan University, Lhasa, Tibet Autonomous Region, China; 2) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 3) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining 810012, China; 4) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100101, China; 5) College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China.

Tibetans are well adapted to hypoxic environment at high altitude. Compared with lowlanders moving to high altitude, Tibetans have relatively low hemoglobin concentrations that is considered a protection from high altitude polycythemia (overproduction of red cells). But how high can Tibetans live remains an open question. We analyzed hemoglobin profiles of nearly 9,000 Tibetan individuals from 20 geographic populations permanently residing at elevations from 2,227m to 5,018m. With the use of a nonlinear regression model, we identified an elevation turning point around 4,500m showing sharp increases of hemoglobin concentration and polycythemia prevalence. This elevation turning point likely marks the altitude limit for Tibetans. Our findings of an Hb turning point in Tibetans have several implications. First, although Tibetans are likely the most adapted population for high altitudein the world, an elevation of more than 4,500m still serves as a big challenge for them. Second, the identification of an altitude limit implicates an incomplete adaptation of Tibetans to high altitude although they have been living there since late Paleolithic time. This seems to explain the not-yet-fixed adaptive mutations (~80%)previously identified in the two hypoxic genes (*EPAS1* and *EGLN1*). In other words, natural selection has not finished its task and may still be in action. Finally, as hemoglobin concentration is only one of the adaptive physiological traits in Tibetans, other traits like resting ventilation, blood oxygen saturation and pulmonary vasoconstriction response should also be studied to further test the proposed turning point of 4,500m.

2236W

Polymorphisms associated with skin, hair and eyes color for forensic phenotyping purposes in Brazilian population. *C. Fridman, F.A. Lima, F.T. Gonçalves.* Dept of Legal Medicine, Ethics and Occupational Health; Medical School, University of São Paulo , São Paulo, Brazil.

Human pigmentation is a variable and complex trait determined by exposure to ultraviolet radiation, age diseases, hormones, and genetic factors. Some polymorphisms in pigmentation genes have been associated with the phenotypic diversity of skin, hair and eyes color in homogeneous populations. Forensic DNA Phenotyping (FDP) is benefiting forensic science in several countries, helping in criminal investigations due to its ability to suggest, with good accuracy, the possible phenotypes for externally visible characteristics (EVCs) in samples of unknown origin. Herein, we evaluated the associations between the SNPs present in the genes *SLC24A5* (rs1426654; rs16960620; rs2555364), *TYR* (rs1126809) and *ASIP* (rs6058017) with skin, hair and eyes color in individuals of the Brazilian population in order to point out the possible use of these markers in forensic practice in admixed populations. The volunteers answered a questionnaire in which they self reported these characteristics for comparison between genotypes and phenotypes. The results showed that for the SNPs rs2555364 and rs1426654 the ancestral allele was associated with characteristics of black skin color, brown or black hair and brown eyes. In addition, the ancestral allele of the SNP rs6058017 was significantly associated with black skin color and brown eyes. Inversely, the variant alleles of these SNPs are correlated with fair pigmentation characteristics for the evaluated EVCs, corroborating the previous studies performed in Different populations. These results show that molecular information may be useful for the inference of EVCs, and the FDP technique is are important tools for forensic studies in Brazil.

2237T

Whole genome sequencing in severe chronic obstructive pulmonary disease. *D. Prokopenko[1,2], P. Sakornsakolpat[1,2], H. Loehlein Fier[3,4], D. Qiao[1,2], C. Lange[1,4], T.H. Beaty[5], J.D. Crapo[6], E.K. Silverman[1,2], COPDGene Investigators, NHLBI TOPMed Investigators.* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, USA; 2) Harvard Medical School, Boston, USA; 3) Working Group of Genomic Mathematics, University of Bonn, Bonn Germany; 4) Department of Biostatistics, Harvard School of Public Health, Boston, USA; 5) Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA; 6) National Jewish Health, Denver, USA.

Whole-genome sequencing (WGS) offers comprehensive coverage of the genome, which gives several advantages over exome sequencing, including improvements in calling of coding regions and interrogation of impactful variants in non-coding regions. As part of the NHLBI Trans☐Omics for Precision Medicine (TOPMed) program, we submitted DNA samples from the COPDGene and Boston Early-Onset COPD studies for sequencing from 830 severe COPD cases with a mean forced expiratory volume in 1 second(FEV$_1$) of 33±11. Current and ex-smoker controls (n=1070) had FEV$_1 \geq$ 85% predicted. For association analysis we identified population substructure in the samples using the rare variant Jaccard method and grouped variants into non-overlapping windows using a spatial clustering approach. We performed single-variant, indel, and region-based analysis both genome-wide and in candidate genes and controlled for effects of population stratification. We further assessed the overlap of variants between sequencing and Haplotype Resource Consortium imputation using reference panels. Results: We performed analyses in 1794 total subjects, with separate analyses in non-Hispanic White (NHW) and African American (AA) individuals. We confirmed a known association region near *HHIP* (combined p=1.6x10^{-9}) and identified near-significant variants near *TNS1* in NHW and *SERPINA6* in AA subjects. The top association in *SERPINA6* in the AA is a common variant and not explained by the known *SERPINA1* COPD susceptibility Z or S alleles (conditional P = 9x10-6) nor other known *SERPINA1* variants. No association was seen in non-Hispanic whites. In regions around 22 variants recently described as genome-wide significant in COPD, we found nominally significant variants which were not present in imputation reference panels. In general for rare variants, even with a high imputation quality, we see a drop in concordance when compared with sequenced data. Conclusions: Whole-genome sequencing can identify large numbers of potentially functional and deleterious variants, and will serve as an important resource for identifying causal variants at known and novel loci for COPD. We anticipate including additional 7300 subjects into our analysis, which will increase our power. Future plans include testing COPD-related phenotypes and fine-mapping of known signals.

2238F

GWAS of fingerprint patterns. *E. Feingold[1], J. Chernus[1], J. Bowser[1], J. Roosenboom[1], S. Rajagopalan[1], K. Neiswanger[1], E.J. Leslie[1], J. Carlson[1], C. Sanchez[1], T. Soejima[1], A. Czeizel[2], J.T. Hecht[3], E.E. Castilla[4], J. Salamanca[5], I.M. Orioli[6], F.A. Poletta[7], S. Weinberg[1], M.L. Marazita[1].* 1) University of Pittsburgh, Pittsburgh, PA, USA; 2) FCCHD, Budapest, Hungary; 3) University of Texas McGovern Medical Center, Houston, TX, USA; 4) ECLAMC, Rio de Janeiro, Brazil; 5) Hospital Infantil Universitario Niño Jesús, Madrid, Spain; 6) Federal University of Rio de Janeiro, Rio de Janeiro, Brazil; 7) CEMIC: Center for Medical Education and Clinical Research, Buenos Aires, Argentina.

Human fingerprint patterns are a highly heritable phenotype known to be associated with important diseases such as schizophrenia and heart disease. Understanding fingerprint genetics may help dissect the underlying biology of such diseases, as well as providing insight into more general developmental processes. Fingerprints can be conceived of as a high-dimensional phenotype with various measures on each finger. Beyond classification into common pattern types (arches, loops, and whorls), more refined phenotypes such as ridge counts have been developed. Using the correlation structure of pattern types across study subjects' ten fingers, we constructed several multi-finger phenotypes and performed genome-wide association scans on them. The subjects are a set of approximately 1400 individuals recruited from the Pittsburgh Orofacial Cleft Study—a larger study of the genetics of cleft lip and palate. After imputation, genotypes were available for seven million SNPs. Since subjects include cleft cases, their cleft-free relatives, and control families from several populations, a mixed model was used to account for population structure and relatedness in performing the GWAS. Results showed differing levels of genetic control for different fingers and patterns. Signals included replication of genes previously reported for other fingerprint phenotypes as well as several newly-nominated loci. This work was supported by NIH DE016148.

2239W

Replication of HLA association with podoconiosis in diverse Ethiopian ethnic groups. *T.T. Gebresilase[1,2], C. Finan[7,8], D. Suvegas[6], T. Sisay[2], A. Aseffa[1], A. Adeyemo[4], C. Rotimi[4], G. Davey[3], E. Zeggini[6], M. Newport[3], F. Tekola-Ayele[5].* 1) Bioinformatics Directorate, Armauer Hansen Research Institute (AHRI), Addis Ababa, Ethiopia; 2) Unit of Health Biotechnology, Institute of Biotechnology, College of Natural and Computational Sciences, Addis Ababa University, Ethiopia; 3) Wellcome Trust Brighton and Sussex Centre for Global Health Research, Brighton and Sussex Medical School, Falmer, Brighton, UK; 4) Centre for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 5) Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA; 6) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 7) Institute of Cardiovascular Science, Faculty of Population Health, University College London, London, U.K; 8) Farr Institute of Health Informatics, University College London, London, U.K.

Podoconiosis is a non-infectious geochemical disease of the lower leg resulting from long term exposure of bare feet to clay soil of volcanic origin. Little is known about the soil component that triggers the disease but progress has been made towards understanding host susceptibility. Previous Genome-Wide Association Study (GWAS) conducted in Wolaita, Southern Ethiopia, indicated variants in the *HLA* class II region to be associated with podoconiosis. In the current study, we aimed to validate this finding in the original and two other independent ethnic groups in Ethiopia using a replication samples consisting of 943 cases and 949 controls. DNA obtained from saliva was genotyped using the Illumina HumanOmni25-8 v.1-2 chip at Wellcome Trust Sanger Institute (WTSI), UK. A total of 22 SNPs in the *HLA-II* region reached genome-wide significance (p < 5E-8) with rs9270911 being the most highly associated variant (P = 5.512E-10; OR1.53; 95 % CI 1.34 -1.74). Gene-set analysis using MAGMA showed enrichment of KEGG Type I diabetes mellitus (FDR-adjusted p-value = 3.17E-09) pathways. Overall, our findings confirm an important role of antigen-presentation and T cell mediated reponse in the pathogenesis of the disease, and future work should focus in elucidating functional consequences of these variants.

2240T

Fetal genome-wide meta-analysis of gestational age and preterm delivery. *X. Liu[1], D. Helenius[2], A. Buil[2], B. Feenstra[1] for the iPSYCH-BROAD Group and Early Growth Genetics Consortium.* 1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Institute of Biological Psychiatry, Sankt Hans Psychiatric Hospital, Roskilde, Denmark.

BACKGROUND: Preterm birth, defined as birth at less than 37 weeks of gestation, is a major cause of infant morbidity and mortality. In European countries, the preterm birth rates in 2010 ranged from 5.2-10.4% and 60-75% of infant deaths occurred among babies who were born prematurely. Despite evidence that genetic factors contribute to preterm delivery (PTD) and gestational age (GA) at delivery, robust genome-wide significant associations with genetic variants have not yet been identified. **OBJECTIVES:** As a prime example of a complex phenotypic outcome, preterm birth is influenced by a multitude of environmental and genetic factors each of which has a small effect. The present study aim to discover genetic loci in the fetal genome associated with birth timing through a large genome-wide meta-analysis of infants with GA information. **MATERALS AND METHODS:** Our study design is a meta-analysis of genome-wide association studies in a large number of cohorts within the Early Growth Genetics consortium, as well as other studies such as the Initiative for Integrative Psychiatric Research (iPSYCH). So far, the studies committing to the meta-analysis allow us to include genetic data from more than 60,000 infants. Within each discovery cohort, the participants' genotypes will be imputed from the Haplotype Reference Consortium (HRC) or 1000G Phase3 reference panels. Association testing will be performed in each cohort, and results will be combined using an inverse-variance weighted fixed-effects approach. **OUTCOME:** Collectively data from the contributing cohorts represent the largest study sample to date aimed at identifying genetic variants associated with PTD/GA. The meta-analysis is currently underway and results will be presented. These efforts will allow us to gain a better understanding of the human pregnancy and the timing of birth; knowledge that may ultimately lead to improved measures of prevention and therapy.

2241F

Genome-wide association meta-analysis of usual sleep duration in 31,703 Japanese population. *T. Nishiyama[1], C. Wang[1], M. Nakatochi[2], A. Hishida[3], M. Naito[3], K. Wakai[3], M. Kubo[4], A. Goto[5], M. Iwasaki[5], A. Hozawa[6], M. Yamamoto[6], T. Hachiya[7], A. Shimizu[7], K. Tanno[7], M. Sasaki[7], S. Suzuki[8].* 1) Public Health, Aichi Medical University School of Medicine, Nagakute, Japan; 2) Center for Advanced Medicine and Clinical Research, Nagoya Univ. Hospital, Nagoya, Japan; 3) Department of Preventive Medicine, Nagoya Univ. Graduate School of Medicine, Nagoya, Japan; 4) Laboratory of for Genotyping Development, RIKEN, Yokohama, Japan; 5) Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; 6) Tohoku Medical Megabank, Tohoku University, Sendai, Japan; 7) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank, Iwate, Japan; 8) Department of Public Health, Nagoya City Univ. Graduate School of Medicine, Nagoya, Japan.

Usual sleep duration is heritable with heritability estimates being around 40%. Candidate gene association studies for sleep duration identified a number of potential loci, though the results were inconsistent. A GWAS of sleep duration have identified and replicated two SNPs near PAX8 and CBWD2 [Gottlieb DJ 2015]. Another large-scale GWAS using the UK biobank replicated association with PAX8 and identified new locus, VPK2, and confirmed the association with VPK2 in other samples [Jones SE 2016]. However, any locus has neither reached genome-wide significance in the other GWAS nor been replicated in independent samples presumably due to limitations in sample size. Therefore, we conducted a genome-wide association meta-analysis of sleep duration, including GWAS from the biggest three cohort studies in Japan, the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study, the Japan Public Health Center-based prospective (JPHC) Study and the Tohoku Medical Megabank Project (TMM). Phenotype information was obtained from self-administered questionnaires at the entry into each cohort study. The ethics committees of all participating institutions approved the protocol for each cohort study, and all participants provided written informed consent with some exceptions (details omitted). Genotyping was performed with the Illumina Infinium OmniExpressExome-8 BeadChip. Imputation of autosomal SNPs was performed using IMPUTE2 and Minimac3 to a reference panel of the 1000 genome Project (phase 3) in the J-MICC and JPHC studies, and TMM, respectively. Population substructure was captured by principal component (PC) analysis on the sample of typed genotypes. After stringent quality control, GWAS of sleep duration as a quantitative trait was performed in each cohort adjusted for age, gender and the first 5 PCs, using SNPTEST and a self-made software based on mixed-effect model for the J-MICC and JPHC Studies, and TMM, respectively. In total, 6,730,804 SNPs in 31,703 participants passed our quality control criteria. In this meta-analysis, we found one locus significantly associated with sleep duration with genome-wide level (P-value < $5×10^{-8}$). This include novel genes, such as CUX2, ALDH2 and HECTD4 at 12q24.1, where the smallest P-value = $1.54×10^{-8}$ was found at rs76892715. We also found two suggestive loci (P-value < $5×10^{-7}$), KCNK2 at 1q41 and FABP6 at 5q33.3.These loci were not identified in the former GWAS of sleep duration and warrants further exploration.

2242W

Genome-wide association study of complement activity pathways: The Cooperative Health Research in South Tyrol (CHRIS) study. *D. Noce[1,2], L. Foco[2], C. Fuchsberger[2], M. Gögele[2], A. De Grandi[2], R. Würzner[1], C. Lass-Flörl[1], P.P. Pramstaller[2,3,4], D. Orth-Höller[1], C. Pattaro[2].* 1) Division of Hygiene and Medical Microbiology, Medical University of Innsbruck, Austria; 2) Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy; 3) Department of Neurology, General Central Hospital, Bolzano, Italy; 4) Department of Neurology, University of Lübeck, Lübeck, Germany.

Purpose: The complement system (CS) is a fundamental component of the innate immune response. It enhances antibodies' and phagocytic cells' ability to remove pathogens from organisms. The CS can be activated via three distinct pathways known as classical, lectin, and alternative. Deficiencies in each pathway are associated with chronic renal disease, age-related macular degeneration, immunological diseases and atypical hemolytic uremic syndrome, among others. To gain insight on the genetic basis of CS activity, we conducted genome-wide association studies (GWAS) of the three pathways in a general population sample. **Methods:** CS activity was assessed in 4990 18-to-93 year old participants (55% female) of the Cooperative Health Research in South Tyrol (CHRIS) study, using the WIESLAB commercial assay. For each pathway, activity was regressed over sex, age, plate, and date of experiment using linear mixed effect models (LMMs). GWAS of the LMM residuals were performed on the dosage levels imputed based on the Haplotype Reference Consortium v1.1 reference panel. GWAS were run with EPACTS on SNPs with minor allele frequency of >0.01 and imputation quality r^2 of >0.3, and accounting for population relatedness. **Results:** We observed and accounted for modest genomic inflation: λ from 1.00 to 1.06. Three loci were identified for the alternative pathway on chromosomes 1q31 (p = 8.8e-10), 6p22.1 (p = 4.7e-14), and 12p12.2 (p = 1.8e-9), respectively. The 6p22.1 locus was also associated with the classical pathway (p = 1.4e-35). Three loci were identified in association with the lectin pathway on chromosomes 1p36.22 (p = 7.3e-11), 9q34.2 (p = 7.6e-45), and 10q21.1 (p = 1.4e-260), respectively. All identified loci map to complement proteins responsible for the activation of the single pathways, and explain over 3.2%, 4.5%, and 34.9% of the variability of the classical, alternative, and lectin pathways, respectively. **Conclusions:** This is the largest genome-wide screening of the CS activity conducted to date. Associations were identified at loci carrying genes known to be involved in CS activity. The large variance explained by the identified loci opens up the opportunity for Mendelian randomization studies aimed at assessing causality between CS activation and complex diseases in the general population.

2243T

GWAS reveals loci associated with velopharyngeal insufficiency. *J. Roosenboom[1], J. Chernus[1], E.J. Leslie[1], M. Ford[2], E. Feingold[3], J.R. Shaffer[3], S.M. Weinberg[1], M.L. Marazita[1,3].* 1) University of Pittsburgh, Center for Craniofacial and Dental Genetics, Pittsburgh, PA, USA; 2) Childrens Hospital of Pittsburgh, Cleft Craniofacial Center, Pittsburgh, PA, USA; 3) University of Pittsburgh, Department of Human Genetics, Pittsburgh, PA, USA.

Velopharyngeal insufficiency (VPI) often results in hypernasal speech. VPI is caused by a failure to seal off the opening in the pharynx where the nasal cavity communicates with the oral cavity. VPI is often associated with orofacial clefting, especially when the secondary palate is affected. However, VPI can also be present in the absence of an overt cleft palate, suggesting an abnormality the structure or function of the soft palate musculature or an anatomical discrepancy in the size/shape of the pharynx. The genetic basis of VPI as an isolated trait is poorly understood. We performed a GWAS on 1689 individuals who had been evaluated for VPI. Our cohort was comprised of individuals with isolated lip clefts (no secondary palate involvement), unaffected relatives from families with a history of clefting and unaffected population controls. Assessment of VPI was made by a speech pathologist, based on either in person interviews or evaluation of standardized speech videos. A total of 92 participants showed evidence of VPI. The association analysis was carried out in EMMAX on ~350K genotyped and imputed SNPs (MAF > 5%), with age, sex and spoken language indicated as covariates in the model. Our preliminary results revealed two genome-wide significant loci (p < 5E-08): rs3781583 in *LRP5* on chr11q13.2 and rs12922822 on chr16p12.3. *LRP5* is a co-receptor for *WNT* signaling and required for cranial neural crest cell migration in zebrafish. Furthermore, it is associated with oropharyngeal skeletal disease and therefore a good candidate gene for VPI. The second SNP was located in a gene desert, and probably has some long-range effects on other genetic regions. These are the first reported genetic association results for isolated VPI.

2244F

A genome-wide association study of anorexia nervosa suggests a risk locus implicated in dysregulated leptin signaling. *X. Wang[1], H. Hakonarson[1,2,3], D. Li[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Anorexia nervosa (AN) is a complex and often chronic eating disorder characterized by the inability to maintain a normal healthy body weight and a persistent fear of weight gain, resulting in extreme emaciation and even death in some cases. Previous genetic and epidemiological studies have indicated a multifactorial etiology, where both genetic and environmental factors contribute to disease risk. As sample sizes have increased, genome-wide association studies (GWASs) of AN have begun to identify risk variants. To further elucidate the genetic architecture of AN, we performed a GWAS using data from our previously published study consisting of 1,033 AN cases by excluding 212 patients with AN who experienced diagnostic crossover during the course of their illness. We hypothesized that this reduction in phenotypic heterogeneity, despite the fact that AN and BN may share some genetic risk factors, would enhance gene discovery. Analysis of phenotypic variability led to the identification of a specific genetic risk factor that approached genome-wide significance (rs929626 in EBF1 (Early B-Cell Factor 1); P=2.04 × 10[-7]; OR = 0.7; 95% confidence interval (CI) = 0.61-0.8) with independent replication (P = 0.04), suggesting a variant-mediated dysregulation of leptin signaling may play a role in AN. Multiple SNPs in LD with the variant support the nominal association. This demonstrates that although the clinical and etiologic heterogeneity of AN is universally recognized, further careful sub-typing of cases may provide more precise genomic signals. In this study, through a refinement of the phenotype spectrum of AN, we present a replicable GWAS signal that is nominally associated with AN, highlighting a potentially important candidate locus for further investigation.

2245W

Genome-wide association studies of eye color in Han Chinese and Uyghur populations. *L. Wang[1,2], S. Wu[1,2], J. Tan[2], Y. Yang[2,3], K. Tang[1], S. Xu[1], L. Jin[1,2,3], S. Wang[1,2].* 1) Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, 200031 Shanghai, China; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, 200433 Shanghai, China; 3) Fudan-Taizhou Institute of Health Sciences, 1 Yaocheng Road, 225300 Taizhou, Jiangsu, China.

The iris controls the amount of light reaching the retina and functions as the diaphragm of the eye. Previous studies have identified genetic variants in several genes associated with human eye color in Europeans. However, there is no report of eye color related variants in East Asians. In this study we performed a genome-wide association study on eye color (assessed by Martin-Schultz eye color scales) in 2938 Han Chinese. We found two non-synonymous SNPs in *OCA2* independently associated with eye color (rs1800414: p=5.34×10e-28; rs74653330: p=3.39×10e-15). The findings were replicated in a second genome-wide scan on 690 Uyghur samples (rs1800414: p=8.58×10e-10; rs74653330: p=1.35×10e-8). Interestingly, we discovered a strong interaction between rs74653330 and rs12913832, a previously reported pigmentation related variant affecting the expression of *OCA2* (p=1.74×10e-8, beta=-1.93). The effect of rs74653330 could significantly increase when individuals also carry derived alleles of rs12913832. We also demonstrated that despite of the subtle differences of eye color in East Asians, a multinomial logistic regression model could also be used to predict eye color in Uyghur (AUC is 0.77 for Dark brown, 0.72 for light brown and 0.92 for green surround brown).

2246T

Pinpointing GWAS signals: Indels vs. SNPs. *S.A. Gagliano[1], S. Sengupta[1], C. Sidore[2], A. Maschio[2], F. Cucca[2,3], D. Schlessinger[4], G.R. Abecasis[1].* 1) Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA; 2) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy; 3) Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy; 4) Laboratory of Genetics, National Institute on Aging, US National Institutes of Health, Baltimore, Maryland, USA.

It is unclear whether insertion and deletion variants (indels) are more likely to influence complex traits than more abundant SNPs. By definition, each indel adds or omits nucleotide(s), potentially increasing its impact. On the other hand, some indels are in low complexity regions (LCRs) that are less likely to include genes, potentially decreasing impact. We sought to understand whether SNPs or indels are more likely to impact human health, and to quantify the gain by routinely including indels in GWAS. We comprehensively characterized indels, and assessed association with 120 quantitative traits (related to cardiovascular and blood traits, among others) from the SardiNIA study. Then, at trait associated loci we assessed the relative enrichment of indels to SNPs. In brief, SardiNIA consists of a genotyped and richly phenotyped set of over 60% of the adult population in Sardinia's Lanusei valley. Each sample was genotyped on four different Illumina Infinium arrays. We imputed individuals successfully genotyped on all arrays (N=6602) using Minimac3 based on a Sardinian reference panel (4x coverage; N indels=1.1M; N SNPs=24M) of 3839 individuals. 928,605 polymorphic autosomal indels remained for analysis after imputation quality thresholds were applied. Only 8% (72,218) of the indels mapped to LCRs, and 5% (33,016) were singletons. For each of the 120 traits, we ran association analyses in EPACTS, using inverse-normalized residuals in the Efficient Mixed Model Association eXpedited (EMMAX) test, adjusting for age and sex. 51 of the traits tested had SNPs and/or indels (MAF≥1%) reaching genome-wide significance (p≤5E-8), 33 of which had at least one indel reaching significance. There were 495 indels associated with at least one trait. We then investigated whether indels are more likely than SNPs to be "causal". To obtain an estimate of indel enrichment among potentially causal variants we assessed the proportion of indels to SNPs within 1Mbp of associated loci (N=89) compared to the rest of the genome for variants MAF≥1%. We weighed the proportions by the association signals. Indels were not significantly enriched (estimate=1.28; p=0.7). Assuming indels and SNPs have equal causality, we obtained 95% credible sets of potentially causal variants. One set contained solely a frameshift indel (rs200748895, posterior probability=0.99). Of variants with a posterior probability ≥0.1, 6% (10/170) were indels. Our results suggest potential gain in adding indels to GWAS.

2247F

Mutual allelic constraint in a human taste receptor gene cluster shows evidence of gene-gene interaction and may influence human reproductive partner compatibility. *R. Subaran, S. Munne.* CooperGenomics, Livingston, NJ.

Gene-gene interactions in humans have been notoriously difficult to study, often calling for a number of statistical comparisons so large that the effect and sample sizes required are prohibitive. Here, we sought to identify gene-gene interactions that might affect reproductive phenotypes, limiting our pairwise comparisons, a priori, to a set of alleles we hypothesized to have strong effects on reproductive phenotypes. We scanned a large a cohort of individuals from Northern Europe for pairs of alleles, within this filtered set, that bear the signatures of gene-gene interaction. This led us to identify multiple, separate, syntenic allele pairs where the underrepresentation of specific two-locus genotypes was statistically significant and could not be readily explained by canonical models of distance-based linkage disequilibrium (LD) and/or population admixture. One of these allele pairs spans a set of genes that are implicated in reproductive partner selection in mammals, and that are already known for long and unusual patterns of LD. Another pair is made up of alleles at distal ends of a human taste receptor gene cluster. We found that the presence of the minor allele in one gene in this cluster, strongly constrains the frequency of the major allele in the partner gene several hundred kilobases away, despite the absence of strong general LD between the two loci. To rule out the possibility of population stratification driving this observed effect, we tested for the presence of this pattern across multiple 1000 Genomes populations and found it to be recapitulated in all out-of-Africa populations. Therefore we conclude that alleles at these taste receptor loci have undergone continued mutual incompatibility for tens of thousands of years and may represent a gene-gene interaction involved in reproduction or reproductive partner choice.

2248W

High-risk genetic variants in genes involved in endothelial cell glycocalyx function in thrombotic storm. *K. Nuytemans[1], L. Gomez[1], N. Hoffman[1], P. Whitehead-Gay[1], N. Alves[1], A. Beecham[1], N. Dueker[1], C.S. Kitchens[2], D. Erkan[3], L.R. Brandao[4], A.H. James[5], R. Kulkarni[6], T.L. Ortel[5], J.M. Vance[1], M.A. Pericak-Vance[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami FL; 2) University of Florida, Gainesville; 3) Hospital for Special Surgery, Weill Medical College of Cornell University, New York, NY; 4) Hospital for Sick Children,Toronto, ON, Canada; 5) Duke University Medical Center, Durham, NC; 6) Michigan State University, East Lansing.

Thrombotic storm (TS; www.thromboticstorm.com) is a rare and severe clinical phenotype that occurs in a small subset of patients with venous thromboembolic disease. It is characterized by more than two acute arterial and/or venous thromboemboli, frequent events in unusual locations, progressive/recent unexplained recurrence and frequent refractory and/or atypical response to therapy. We hypothesize that rare high-risk genetic variants contribute to the underlying hypercoagulability state in TS. We performed whole exome sequencing (WES) in 1 multiplex family of 3 affecteds, 18 trios and 26 additional probands. Due to the rarity and severity of the phenotype, variants are filtered on functionality (nonsense, missense and splice), conservation and frequency (<5% in 1000genomes). We performed sharing analyses in the family and tested two different models in the trios; *de novo* and recessive model. We screened the remaining probands for variants in identified candidate genes. We identified 52 rare functional de novo variants and 85 rare functional compound heterozygous variant combinations. Interestingly, two of the *de novo* variants are in genes (*SLC26A2, CHPF2*) involved in the formation of chondroitin sulfate (CS), while a compound heterozygous variant combination affects *STAB2,* a CS metabolism gene. Sharing analyses in the family identified rare variants in known thrombotic factors as well as in *PAPSS2,* a gene working with *SLC26A2* to supply sulfate groups for molecule sulfation. Screening the remaining probands led to the identification of additional variants in these genes in another nine patients. In summary, we identified CS variants in approximately a quarter of the TS patients. CS plays an important role in the endothelial cell layer, the glycocalyx, protecting the blood vessel and harboring multiple molecules of the blood anticoagulation systems. The sulfated CS sidechains on proteoglycans create the negative charge of the barrier, preventing permeability and binding of platelets. They also bind absorbed (anticoagulant) molecules from the plasma, such as antithrombin. Additionally, CS on thrombomodulin increases its efficiency at inactivating thrombin. Our data suggests that variants involved in chondroitin sulfate formation are conferring risk for TS. Additional analyses evaluating burden of variants in other glycocalyx genes in all probands are ongoing. .

2249T

Genetic typing of DC-SIGN in recurrent vulvovaginal candidiasis. *N. Kalia[1], S. Kaur[2], J. Singh[1], M. Kaur[2].* 1) Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar, India; 2) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India.

   Statement of purpose: Vulvovaginal Candidiasis (VVC) is a disease condition characterised by the excessive growth of *Candida* species, a dimorphic yeast normally present in the vagina in small numbers. Its symptoms include excessive secretion of curdy white vaginal discharge and severe itching. It can be acute or recurrent, where recurrent VVC (RVVC) is defined as four or more episodes of symptomatic acute VVC per year. This disease condition occurs in 75% women of child bearing age with a global annual incidence of 1-2 % of all women. Several molecules of innate and adaptive immune system are implicated in etiology and pathogenesis of RVVC. Dendritic cell specific intracellular adhesion molecule-3 grabbing non-integrin (DC-SIGN), also known as CD209, is a pattern-recognition receptor which bind to carbohydrate residues in the cell wall of many pathogens including fungi. It is involved in processing and presentation of these antigens resulting in generation of immune responses. Objectives of the present study were to evaluate DC-SIGN gene polymorphism in susceptibility to RVVC. **Methods used**: The study protocol was approved by Institutional Ethics Committee. After obtaining written informed consent, blood samples were collected from 42 cases, diagnosed by a gynaecologist and 42 age-matched healthy controls without any recurrent vaginal complaints. For the diagnosis of RVVC, European (IUSTI/WHO) guidelines on the management of vaginal discharge were used. Genomic DNA was isolated from the blood samples using inorganic method. Genotypic analysis of rs4804801 SNP of DC-SIGN gene was carried by PCR-RFLP using restriction enzyme HincII. All statistical analyses were performed using SPSS version 16.0. **Summary of Results**: High frequency of variant allele 'T' and heterozygous genotype was observed in cases (19.05%; 38.10%) as compared to controls (14.28%; 28.57%). Frequency of wild allele 'A' and its homozygous genotype was observed to be low in cases (80.95%; 61.90%) comparative to controls (85.71%; 71.43%). Homozygous variant genotype was completely absent in both the studied groups. However, none of these differences were found to be statistically significant (p>0.05), which can be attributed to the small sample size employed in the present study. This variant of DC-SIGN may be associated with risk towards RVVC.

2250F

The Million Veteran cohort: A high quality and diverse resource of genetic data for large-scale analysis to enable personalized medicine. *Y. Shi[1], H. Hunter-Zinck[1], M. Li[1,9], N. Sun[3,8], J. Huang[1], C. Pan[2,6], J. Gelernter[3,8], P.S. Tsao[2,6], C.J. O'Donnell[1,5], T.L. Assimes[2,6], E.R. Hauser[4,7], H. Zhao[3,8], S. Pyarajan[1,5] on behalf of the VA Million Veteran Program.* 1) VA Boston Healthcare System, Boston, MA; 2) VA Palo Alto Health Care System, Palo Alto, CA; 3) VA Connecticut Healthcare System, West Haven, CT; 4) Durham VA Medical Center, Durham, NC; 5) Harvard Medical School, Boston, MA; 6) Stanford University School of Medicine, Stanford, CA; 7) Duke University, Durham, NC; 8) Yale School of Public Health, New Haven, CT; 9) University of Utah, Salt Lake City, UT.

   The Million Veteran Program (MVP) aims to integrate a patient's genetic data with the clinical record for personalized medicine in the care of U.S. veterans. The MVP is enabling large scale genome-wide association studies (GWAS) of complex human diseases. A successful GWAS requires rigorous quality assurance (QA) and quality control (QC) to minimize genotyping errors. Furthermore, large cohorts such as the MVP can provide evidence to improve best practices in genotyping technology and algorithms. Here, we describe the QC/QA on genotypic data for the first 396,197 samples in MVP cohort. Approximately 4,500 samples genotyped with MVP 1.0 array were grouped into a batch to call genotypes using the Affymetrix Power Tools software. After the advanced marker QC with batch specific information for accurately calling rare markers, 657,459 (95.74%) of the original 686,682 genotyped markers were retained. A rigorous sample QA/QC procedure including gender mismatch, genotype concordance, PCA, and relatedness analysis produced a cohort with above 95% call rate. The median sample concordance for 5,000 pairs of intentional duplicates is > 99.8%. Concordance is excellent for common markers (MAF > 10%) with 94% of these markers having discordance rates of less than 1%. Even for rare markers (MAF < 0.1%), 74% of markers in this MAF bin have discordance rates less than 1%, demonstrating that we can consistently call rare variants. Although the vast majority of the MVP samples are close to reference populations of European origin in their genetic makeup, many MVP sample exhibit admixture with respect to the African and the Asian population supergroups in 1000 Genomes Phase 3 references samples. The overlapping markers between the MVP population and ExAC database are stratified by their MAF values in the major continental populations to evaluate coverage of functional variants across the genomes of in diverse populations. All five-ancestry groups have similar in common variants counts. The African population has slightly more common variants due to the Genome wide imputation booster for African ancestry population and the Non-Finnish European population has the most rare variants. The completed QC analysis of the 396,197 samples demonstrates that the MVP is a high quality and diverse resource for researchers seeking to understand the complex interactions between genetics, the environment, and medicine. .

2251W

Gut microbiota composition in children and adults: *Bacteroides* vs *Blautia*. D. Radjabzadeh[1], S.A Beth[2], C.M Gomez[1,3,5], C.G Boer[1], P. van der Wal[1], J.C Kiefte-De Jong[2], M.A.E Jansen[2], S.R Konstantinov[4], M.P Peppelenbosch[3], J.P Hays[4], V.V.W Jaddoe[3,5], F. Rivadeneira[1,3,5], J.B.J van Meurs[1,5], H.A Moll[2], A.M Ikram[1,3,5], A.G. Uitterlinden[1,3,5], R. Kraaij[1]. 1) INTERNAL MEDICINE, ERASMUS MC, ROTTERDAM, Netherlands; 2) Department of Paediatrics, Erasmus MC, Rotterdam, Netherlands; 3) The generation R Study, Erasmus MC, Rotterdam, Netherlands; 4) Department of Gastroenterology and Hepatology, Erasmus MC, Netherlands; 5) Dept Epidemiology, Erasmus MC, Erasmus MC, Netherlands.

    Introduction: It is known that the microbiome is dynamic through the life course and is still in developing phase during childhood, although this has not been studied yet in large population-based cohort studies. To examine the shifts in gut microbiome in children and elderly, we profiled the gut microbiome of two extensively phenotyped cohorts: the Rotterdam Study (RS, elderly) and the Generation R birth cohort (GenR). **Materials & Methods**: Over 1,700 faecal stool samples in the RS cohort and over 3,000 in the multi ethnic GenR cohort were collected and stored at -20°C. The V3-V4 regions of 16S ribosomal RNA gene were sequenced using Illumina MiSeq technology. Reads were clustered into Operational Taxonomic Units (OTUs). Comparative analysis of microbiome diversities and composition of the two cohorts was done in R and R packages vegan and MaAsLin. In both cohorts only samples from subjects with North-European ancestry and in the RS cohort only samples within the age range of 51-62 years were included in the analyses. We adjusted for BMI, sex, technical covariates, and multiple testing by FDR (q<0.05). **Results**: In total, 1,081 RS and 1,463 GenR samples were included in the analysis (n= 2,544). After quality filtering 770 genus-level OTUs were obtained. We observed significant clustering of both cohorts based on Bray-Curtis dissimilarity (PERMANOVA, p= 0.001) and a higher alpha diversity (Shannon Index) in RS than in GenR (ANOVA, p < 2.2e-16). Our analysis revealed that 175 OTUs were more abundant in RS, while 137 OTUs were more abundant in GenR (q<0.05). Largest differences were observed for genera from family *Lachnospiraceae* including *Anaerostipes* and *Blautia* that were more abundant in RS, and genera from families *Ruminococcaceae* and *Bacteroidaceae* including *Faecalibacterium* and *Bacteroides* that were more abundant in GenR. **Conclusion:** Comparing the gut microbiome compositions in 2,544 children and adults showed significant clustering of the cohorts based on beta diversity and significantly different alpha-diversities between children and adults. A total of 312 OTUs were found to have differential abundances between the two cohorts with genera from the *Lachnospiraceae* family being more abundant in adults and genera from the *Ruminococcaceae* and *Bacteroidaceae* families being more abundant in children. Further analysis is needed in order to understand the mechanisms of these bacteria and their relationship with their host metabolism pathways.

2252T

Genome-wide association study of cranial vault shape reveals novel loci at 15q26.3 and 17q11.2. M. Lee[1], J. Roosenboom[1], J.C. Carlson[2], E.J. Leslie[1], E. Feingold[3], M.L. Marazita[1], J.R. Shaffer[1], S.M. Weinberg[1]. 1) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

    Genome-wide association studies (GWAS) have identified about two dozen loci associated with aspects of human facial shape. The cranial vault, a region comprised of interlocking flat bones surrounding the cerebral cortex, is strongly influenced by brain size and shape. Although cranial vault morphology has both clinical and evolutionary relevance, little is known about the genetic basis of normal cranial vault shape in humans. Following the paradigm observed for other craniofacial traits, we hypothesize that normal variation in the cranial vault may be influenced by genes that lead to cranial defects but different genetic variants within such genes. We performed GWAS on three vault measures [maximum cranial width (MCW), maximum cranial length (MCL) and cephalic index (CI)] in a sample of 4,389 healthy individuals of European ancestry from two cohorts: 2,260 unrelated individuals (TDFN) genotyped on the Illumina OmniExpress + Exome v1.2 array and 2,129 related and unrelated individuals (OFC) genotyped on the Illumina HumanCoreExome array. Both cohorts were imputed to the 1000 Genomes phase 3 reference panel. All measures were adjusted by sex, age and body size, then subjected to GWAS either in PLINK adjusting for ancestry principal components (TDFN cohort) or in EMMAX (OFC cohort). We then performed genome-wide meta-analyses using Stouffer's p-value method. Significant associations at two loci were observed: one was at 15q26.3 (lead SNP rs2924767, p = 2.107 × 10^{-8}) for MCW and the other was at 17q11.2 (lead SNP rs72841279, p = 5.29 × 10^{-9}) for MCL. No significant associations were observed for CI. The genes *MEF2A*, *ADAMTS17*, and *NLK* at the associated loci are known to play roles in craniofacial development, making them potential candidates. These results provide further insight into genetic pathways and mechanisms underlying normal variation in human craniofacial morphology. This study represents the first reported genome-wide significant associations for human cranial vault shape. This research was supported by an NIDCR grant U01-DE020078 (SMW, MLM), R01-DE016148 (MLM, SMW), and X01-HG007821 (MLM, SMW, EF).

2253F

Pharmacogenomic study of antithyroid drug-induced cutaneous reaction. *P. Chen[1,2,3,4], T. Chang[2].* 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan, Taiwan; 2) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University, Taipei, Taiwan, Taiwan; 4) Graduate Institute of Clinical Medicine, National Taiwan University, Taipei, Taiwan, Taiwan.

Graves' disease (GD) is an autoimmune thyroid disorder with clinical and academic importance. Treatments for patients with GD usually include antithyroid drugs (ATDs, including methimazole, carbimazole and propylthiouracil), radioactive iodine, and/or surgery. The adverse reactions of ATDs range from fever, rash, gastrointestinal distress to more severe ones, such as hepatotoxicity, vasculitis, as well as the potentially fatal complication, agranulocytosis. We took a pharmacogenomics approach to study the most common adverse effect, ATD-induced cutaneous reaction (TiCR, including skin rash, urticarial and itching). Patients with TiCRs were initially identified according to chart reviews (first-stage study) based on the discontinuation of the initial medication within two months after use as well as narratives recorded. Patients with TiCR have later been actively searched for from the outpatient department at NTUH for confirmation (second-stage study). We identified 24 patients with TiCRs between 2001 and March 2014 in our first-stage study. We used 427 GD patients without cutaneous reactions as the controls (GD controls). Then, we actively recruited additional 24 TiCR patients between March 2016 and June 2016 for the second-stage study. The results were compared with other 470 GD controls. We performed genome-wide SNP genotyping, HLA imputation using SNPs and direct HLA genotyping. The analyses were carried out by using PLINK. GWAS using Affymetrix Axiom Genome-Wide CHB 1 single nucleotide polymorphism (SNP) Array (a total of 549,828 autosomal SNPs and 10,728 X-chromosomal SNPs after quality control process) showed an excess of small P values at the tail of the quantile-quantile distribution plot, but only limited evidence of inflation caused by population stratification (genomic inflation factor λ_{GC} = 1.0154). We discovered that the association on chromosome 3p21.31 ($P_{first-stage}$ = 0.000251, $P_{second-stage}$ = 0.03455, $P_{combined}$ = 5.25 × 10^{-5}) might deserve reporting. The HLA region was thoroughly investigated; we detected no significant association signal within the HLA region on chromosome 6p21. In conclusion, we discovered that SNPs on chromosome 3p21.31 might be associated with TiCR. However, the association signals did not reach genome-wide significance level. Further increase of sample size is needed to achieve a definitive conclusion.

2254W

Characterization of *CYP2D6* by whole genome sequencing. *A.L. Halpern[1], N. Gonzaludo[2], O. Schulz-Trieglaff[3], E. Fritzilas[3], J. O'Connell[3], E.H. Margulies[2], M.A. Eberle[1], R.J. Taft[1].* 1) Illumina, San Diego, CA; 2) Illumina. Mission Bay, San Francisco, CA; 3) Illumina, Chesterford Research Park, UK.

CYP2D6 plays a critical role in drug metabolism and many known 'star allele' haplotypes are predictive of drug response. Clinical use of *CYP2D6* data requires not just determination of single-variant genotypes but resolution of diplotypes (i.e. pairs of full-gene haplotypes). Though sequencing can provide more information about phasing of variants than traditional genotyping methods such as qPCR or microarrays, approaches employing short read sequencing struggle to resolve haplotypes over multiple kilobases. A catalog of accurate, known haplotypes, however, could facilitate inference of sample diplotypes. We show by simulation that diplotypes of official star alleles can be distinguished from one another with accuracy >99.5%. However, such an approach critically depends on the completeness and accuracy of the haplotype catalog. To assess the accuracy and completeness of existing *CYP2D6* star alleles, we generated whole-genome sequencing (WGS) data at >30x depth for 96 samples characterized by the Genetic Testing Reference Materials Coordination Program (GeT-RM). Sequence data for the 70 samples that are consented for public release can be obtained from the European Nucleotide Archive (PRJEB19931). Using just the protein-altering variants we were able to identify diplotypes consistent with the previous GeT-RM consensus results in 94% of samples. More than half of the time, however, this criterion results in genotype conflicts at >=5 known-variant loci, and results in ambiguity between alternative diplotypes with different protein products 25% of the time, often with consequences for predicted drug response. A parsimony-based criterion using all *CYP2D6* variants still produces known-variant genotyping conflicts in >80% of samples, including conflicts in predicted proteins >50% of the time. These conflicts are not driven by unknown genetic variation - the median number of novel variants not represented in the star alleles is one. Variant calls made in >3000 additional samples with WGS data exhibit similar trends. These data indicate that *CYP2D6* calling from WGS data can be used to reproduce existing test results, but also underline the incompleteness of current star allele definitions. Calling diplotypes based on the existing nomenclature system for *CYP2D6* (with WGS data or other test methods) can produce ambiguous or potentially incorrect functional predictions, emphasizing the need for a comprehensive characterization of *CYP2D6* haplotypes.

2255T

Translating pharmacogenetics: An electronic phenotyping algorithm and survey study of diverse BioMe biobank patients treated with ACE inhibitors. *H. Naik[1], M.Y. O'Connor[2], S.C. Sanderson[3,4], N. Pinnell[1], M. Dong[1], A. Killie[1], A. Owusu Obeng[5,6], O. Gottesman[5], N. Abul-Husn[5], S.A. Scott[1,7].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 2) Department of Medical Education, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Clinical Genetics Department, Great Ormond Street Hospital, London, WC1N 3JH; 4) Department of Behavioural Science and Health, University College London, WC1E 6BT; 5) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 6) Department of Pharmacy, the Mount Sinai Medical Center, New York, NY 10029; 7) Sema4, a Mount Sinai venture, Stamford, CT 06902.

Previous studies have determined that the general public is interested in receiving pharmacogenetic information; however, little is known about knowledge and attitudes on pharmacogenetics among patients treated with medications that have a known adverse event risk. To assess these perspectives, we surveyed patients enrolled in the BioMe biobank at Mount Sinai who were treated with angiotensin converting enzyme inhibitors (ACEi), as ACEi are a very commonly prescribed antihypertensive, yet up to 1% of patients develop life-threatening angioedema (ACEi-AE). Electronic phenotyping algorithms were developed for the BioMe biobank patient data, which identified 199 ACEi-AE cases and 876 controls, defined as patients on an ACEi for ≥2 years without evidence of AE (plus other exclusions). Chart review determined that the case and control algorithms had sensitivities of 100% (95% confidence interval: 95-100%) and positive predictive values of 98.6% and 97.3%, respectively. Notably, cases differed from controls by age (61 vs. 58 years, p<0.01), gender (67% vs. 49% female, p<0.01) and self-reported race (62% vs. 27% African American, p<0.01), consistent with known risk factors for ACEi-AE. An anonymous survey was mailed to all 1075 patients and 91 were returned (9% response rate). Respondents were comparable by gender (52% female), ethnically diverse (47% European, 23% African, 24% Hispanic, and 6% other), had a mean age of 61.9 years, and 14 (15%) had confirmed ACEi-AE. Over a third of patients reported that prescribing physicians had not discussed with them the concepts of interindividual drug response variability (40%) or adverse event risk (including ACEi-AE and cough; 34%), and 73% of patients were previously unaware of pharmacogenetics. However, most patients agreed they would be interested in having pharmacogenetic testing if the results could improve dose selection (82%), prevent side effects (88%), provide a cheaper medication alternative (79%), and/or affect their family members (68%). Moreover, the majority of patients (67%) indicated that pharmacogenetic testing results would positively influence their compliance to medication use. In addition to identifying an innovative approach to define biobank cohorts for pharmacogenomic studies on rare adverse events, these results indicate that awareness of pharmacogenetics among patients is low, but their interest in pharmacogenetic testing is high, which could translate to improved medication compliance.

2256F

Defining the opportunity: The IGNITE CPIC Prescribing Study. *L. Wiley[1], A. Holmes[2], M. Rosenman[3], J. Murray[4], J. Hicks[5], F. Fitzhenry[6], M. Matheny[6], J. Cimino[7], N. Limdi[7], C. Aquilante[8], L. Cavallari[9], G. Lipori[9], A. Winterstein[9], A. Beitelshees[10], D. Lempkin[10], S. Tuteja[11], P. Empey[12], H. Ong[13], M. Levy[13], J. Denny[13], J. Peterson[13], IGNITE Network Pharmacogenomics Working Group.* 1) University of Colorado School of Medicine, Aurora, CO; 2) Richard M. Fairbanks School of Public Health, IUPUI, Indianapolis, IN; 3) Indiana University School of Medicine, Indianapolis, IN; 4) Meharry Medical College, Nashville, TN; 5) Moffitt Cancer Center, Tampa, FL; 6) Nashville Veterans Affairs Medical Center, Nashville, TN; 7) University of Alabama Birmingham, Birmingham, AL; 8) University of Colorado School of Pharmacy, Aurora, CO; 9) University of Florida, Gainesville, FL; 10) University of Maryland School of Medicine, Baltimore, MD; 11) University of Pennsylvania, Philadelphia, PA; 12) University of Pittsburgh, Pittsburgh, PA; 13) Vanderbilt University, Nashville, TN.

The Clinical Pharmacogenetics Implementation Consortium (CPIC), established in 2009, has developed clinical practice guidelines for use of 30 drug-gene pairs. However, the overall value of using pharmacogenomic variants to tailor therapies depends on the incidence of prescribed medications with an actionable pharmacogenomic association. Prescribing trends across the US can change rapidly due to economic and medical stimuli. This study seeks to assess the opportunity for tailoring prescriptions with CPIC Level A variants across multiple types of health systems and population demographics within the Implementing GeNomics In pracTicE (IGNITE) Network Pharmacogenomics Working Group. To accomplish this objective, we will use retrospective prescription data for adult patients at participating health systems to assess both total opportunity (e.g., number of patients receiving targeted drugs) and prescribing trends over time. For all adult patients prescribed at least one CPIC Level A drug, each site will extract patient level data for all CPIC Level A drugs and a selection of alternative therapies for these drugs for the past 5 years starting January 1st, 2011 (or earliest point after this date for which prescribing data are available). Monthly and total number of inpatient and outpatient visits will also be collected to adjust total prescription count by total healthcare system volume. Preliminary data from Vanderbilt University Medical Center show that between Jan 2011 and Sep 2015, 395,723 individuals had exposure to one or more CPIC Level A drugs with 162,819 (41%) patients exposed to two or more drugs. Of those patients exposed to multiple CPIC Level A drugs, there was a median of 2 drugs per patient (max = 14) and 28,106 (7%) having four or more CPIC level A drugs. The three most common drugs across the entire study period were: oxycodone (n=175,551), simvastatin (n=90,114) and citalopram (n=51,430). The number of patients exposed to 1+ CPIC drugs per year increased 29% from 2011 to 2014 (last full year of study) while the average number of unique patients seen monthly increased 15% over the same time-period (2011-2014). Data collection is ongoing at multiple IGNITE and Affiliate sites and should be available for presentation by fall. In summary, there is increasing volume of patients receiving CPIC Level A drugs over the past 5 years at the Vanderbilt University Medical Center.

2257W

Phenotypic consequences of the genetic regulation of expression of the Mendelian disease gene *CFTR*. *X. Zhong, E.R. Gamazon, P. Straub, A. Konkashbaev, P. Evans, N.J Cox.* Vanderbilt Genetics Institute, Nashville, TN.

A disease-causing gene can increase disease risk through two complementary modes: disrupting either the protein coding or the regulation of the gene. Recessive mutations in the gene *CFTR*, which encodes a membrane protein and chloride channel, are known to cause cystic fibrosis (CF), one of the most common Mendelian diseases. It is less clear, however, whether non-coding regulatory variants of *CFTR* exert roles in disease predisposition. Here, we report results of studies on *CFTR* using BioVU, a biobank linked to complete medical records. The effects of common regulatory non-coding variants were aggregated into their target gene's genetically-regulated gene expression (GReX) using PrediXcan. We calculated GReX of *CFTR* in tissues with quality predictors, and tested for associations with disease in ~10K Caucasians. Our results demonstrated three major phenotypic patterns: 1) CF and its canonical phenotypes ('bronchiectasis', 'pneumonia', 'diseases of pancreas', 'chronic sinusitis', and 'nasal polyps' etc.) emerged as the top associations, but further adjusting for ΔF508, a coding mutation that is the most common cause of CF, these previously top associations were no longer significant; 2) CF and its canonical phenotypes remained top associations even after conditioning on ΔF508 but the associations were attenuated; 3) a distinct set of phenotypes (other than CF-related phenotypes) emerged as top associations (e.g., increased expression of *CFTR* in 'Esophagus-GastroesophagealJunction' was associated with cardiac phenotypes such as cardiac conduction disorders, cardiac arrhythmia, paroxysmal tachycardia etc.), and stratification on ΔF508 did not attenuate the associations. These patterns suggested different scenarios regarding the SNP-based tissue-specific expression predictors: 1) the predictor captured mainly the effect of ΔF508; 2) the predictor captured ΔF508 mutation as well as additional genetic contributions to CF-related phenotypes; and 3) the predictor captured genetic contributions independent of ΔF508 that relate to non-CF phenotypes. In conclusion, we demonstrated that regulatory variants of *CFTR* contributed to disease risk in a tissue-specific manner, leading to different phenotypic consequences. Our analyses supported the complementary roles of coding and noncoding variants in disease susceptibility as well as the continuum from Mendelian to common diseases with complex inheritance.

2258T

Designing of an efficient genotyping chip for discovery and pan-disease screening in the VA's Million Veteran Program. *S. Pyarajan[1,5], Y. Shi[1], M. Li[1,9], J. Moser[10], S. Muralidhar[10], L. Fiore[1,4,7], T. O'Leary[10], J. Concato[3,8], J.M. Gaziano[1,5], P. Tsao[2,6] on behalf of the VA Million Veteran Program.* 1) VA Boston Healthcare System, Boston, MA; 2) VA Palo Alto Health Care System, Palo Alto, CA; 3) VA Connecticut Healthcare System, West Haven, CT; 4) Boston University School of Medicine, Boston MA; 5) Harvard Medical School, Boston, MA; 6) Stanford University School of Medicine, Stanford, CA; 7) Boston University School of Public Health, Boston MA; 8) Yale School of Public Health, New Haven, CT; 9) University of Utah, Salt Lake City, UT; 10) VA Office of Research and Development, Washington DC.

The Million Veteran Program (MVP) is a national program initiated by the Department of Veterans Affairs (VA) to collect consented biosamples for personalized medicine research from one million Veterans. The MVP has already collected over 570,000 samples and is expected to reach over one million in the next 4-5 years. To genotype all MVP participants at low cost, a custom high-density genotyping array, was designed on the Affymetrix Axiom® platform for genotyping the MVP samples. The MVP 1.0 array was designed enable investigation of markers of potential clinical significance for a wide variety of human conditions and diseases. We started with the Axiom® Biobank Genotyping Array as the backbone and designed a set of custom markers for inclusion on the array. The groups of curated marker that were included include primary SNPs and indels based on association with diseases or traits of interest including markers from the genes reported by ACMG for incidental findings, Psychiatry associated markers, candidates marker for RA association, genome-wide imputation booster for AA ancestry etc. The MVP 1.0 array includes 723,305 probesets investigating 686,682 markers. Markers from over 3,400 genes regions have been included. Additionally, over 800 clinical conditions, with an emphasis on conditions of high prevalence in US Veterans, are covered by potentially associated markers on the array. Markers for genome wide coverage for common (MAF > 5%) genetic variation of individuals of European and African ancestry were selected by an imputation aware algorithm. Our results indicate good performance for most of the markers including call rates across samples, concordance across duplicates for even very low frequency markers. Our genotyping data quality assessment indicates that overall, the MVP 1.0 Array is an efficient low cost genotyping array that can be used not only as a pan-disease screening array to design future more comprehensive and potentially expensive "omic" assays for MVP samples but also for discovery with efficient imputation.

2259F

Integration of a tool for patient self-assessment within primary care to enable precision prevention. *K. Rageth[1], C. McNamara[2], J. Zhao[3], S. Jurgens[1], M. Seiber[3], J. Angal[3], J. Shi[4], M. Moore[1], M. Mboob[1], M. Haendel[5], T. Groza[2], C. Boerkoel[1], A. Elliott[3], C. Hajek[1], I. De Vivo[4], M. Sincan[1], S. Puumala[3].* 1) Sanford Health, Sioux Falls, SD; 2) Genome.One, Sydney Australia; 3) Sanford Research, Sioux Falls, SD; 4) Harvard T.H. Chan School of Public Health, Boston, MA; 5) Oregon Health Sciences University, Portland, Oregon.

Health care disparities, which arise in part from inadequate access and knowledge, in underserved rural settings frequently lead to delayed diagnosis and inadequate management of complex chronic and rare diseases. These disparities are magnified by the consequential patient disengagement. We hypothesized that engaging patients to become agents of their own health through self-assessment and computational analysis partially addresses the knowledge and access gaps contributing to rural health care disparities. To address this, we converted validated questionnaires to an interactive electronic self-assessment tool embedded within the patient portal of the electronic medical record. To enable computational analysis, questionnaire and intake form responses were mapped to appropriate ontologies. Following refinement with the medical system customer relations and focus groups, we selectively launched this tool as part of a precision prevention initiative within a rural primary care setting. We present the process and logic underlying this engagement method as well as our initial results on patient interaction with the tool, educational initiatives encouraging patient uptake, and clinician response to the tool. .

2260W

High-density imputation identifies rare and low-frequency variants associated with human height in Japanese population. *M. Akiyama[1], K. Ishigaki[1], Y. Momozawa[2], Y. Okada[1,3,4], M. Kanai[1,3], M. Hirata[5], K. Matsuda[6], A. Takahashi[1,7], M. Sasaki[8], K. Tanno[8,9], A. Shimizu[8], M. Naito[10], K. Matsumoto[11], H. Mikami[12], T. Wakai[10], T. Yamaji[13], M. Iwasaki[13], N. Sawada[13], S. Tsugane[14], A. Hozawa[15,16], N. Minegishi[15,16], M. Yamamoto[15,16], M. Kubo[17], Y. Kamatani[1,18].* 1) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan; 4) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan; 5) Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 6) Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 7) Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Osaka, Japan; 8) Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Iwate, Japan; 9) Department of Hygiene and Preventive Medicine, School of Medicine, Iwate Medical University, Iwate, Japan; 10) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan; 11) Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan; 12) Division of Cancer Registry, Prevention and Epidemiology, Chiba Cancer Center Research Institute, Chiba, Japan; 13) Division of Epidemiology, Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; 14) Center for Public Health Sciences, National Cancer Center, Tokyo, Japan; 15) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 16) Graduate School of Medicine, Tohoku University, Sendai, Japan; 17) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 18) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Human height is a highly heritable trait. Although hundreds of genetic loci associated with adult human stature have been successfully identified in the recent large-scale genetic association studies, these findings were mostly derived from European population. Particularly, the contribution of rare and low-frequency variants has been largely unknown in East Asian population, and they are supposed to be population specific. To identify rare and low-frequency variants in Japanese population, we constructed an imputation reference panel (N = 3,541) by integrating whole genome sequence data obtained from 1,037 Japanese individuals (average depth: x30) with that of 1000 genome project (N = 2,504). After confirming the improved accuracy of the imputed variants, we performed genome-wide association study in 159,095 Japanese individuals who participated in the BioBank Japan project using > 27 million imputed sequence variants across autosomes. We detected 354 significantly and independently associated loci (P < 5.0 × 10^{-8}), and 64 of them were novel. Among the 354 lead variants including 26 insertion-deletions, 8 (2.3%) were rare (minor allele frequency [MAF] < 1%), and 15 (4.2%) were low-frequency (1% ≤ MAF < 5%). We evaluated these lead variants in 30,303 individuals in the independent sets of three Japanese population-based cohorts, and observed 344 variants (97%) showed the same directional effects, and nominal associations at 246 variants (70%) in the same direction. Conditional analysis of the associated loci revealed that 127 identified loci (36%) contained two or more independent association signals ($P_{condition}$ < 5.0 × 10^{-8}). Furthermore, we identified genes with multiple associated rare and low-frequency nonsynonymous variants at *SLC27A3* and *CYP26B1* genes (P for SKAT-O = 1.94 × 10^{-38}, and 1.33 × 10^{-45}, respectively), which were also replicated by evaluating three Japanese population-based cohorts. These results underscore the advantages of studying diverse population to reveal rare and low-frequency variants associated with human complex traits, and provide insights into the genetic components of adult human height.

2261T

Novel genes and mutations in patients affected by recurrent pregnancy loss. *P. Quintero-Ronderos[1], E. Mercier[2], M. Fukuda[3], R. González[4], CF. Suárez[4,5], MA. Patarroyo[4,6], D. Vaiman[7,8], JC. Gris[2], P. Laissue[1].* 1) Centro de Investigación en Genética y Genómica-CIGGUR. Grupo GENIUROS. Escuela de Medicina y Ciencias de la Salud. Universidad del Rosario. Bogotá, Colombia; 2) Department of Haematology, University Hospital, Nîmes. Faculty of Pharmacy and Biological Sciences and Research Team EA 2992, University of Montpellier, France; 3) Laboratory for Drug Discovery. National Institute of Advanced Industrial Science and Technology (AIST). Tsukuba city, Ibaraki, Japan; 4) Fundación Instituto de Inmunología de Colombia (FIDIC), Bogotá D.C., Colombia; 5) Universidad de Ciencias Aplicadas y Ambientales (UDCA), Bogotá D.C., Colombia; 6) Basic Sciences Department, School of Medicine and Health Sciences, Universidad del Rosario, Bogotá D.C., Colombia; 7) Institut Cochin, Université Paris Descartes, CNRS (UMR 8104), Paris, France; 8) Inserm, U1016, Paris, France.

Recurrent pregnancy loss (RPL) is a frequently occurring human infertility-related disease. It is considered a public health problem because of its high incidence and prevalence (1% - 5% of women worldwide). There are known causes of RPL such as antiphospholipid syndrome, chromosomal abnormalities, uterine anatomical abnormalities and thrombophilia. However, the 50% of all cases are idiopathic suggesting that genetic mechanisms may be involved in the development of the disease. RPL is considered a complex pathology because of the variety of molecular pathways and genes implicated in the reproductive success. Concerning the molecular aetiology of RPL several studies have had limitations in identifying the disease's genetic causes. In such scenario, next generation sequencing (NGS) provides a potentially interesting tool for research into RPL causative mutations. The present study involved whole-exome sequencing, for the first time, in 49 unrelated women affected by idiopathic RPL. We performed a bioinformatic analyses searching for non-synonymous sequence variants within a subset of 234 RPL genes made from a systematic review of literature. Interspecific alignments were performed to look for variants within conservative domains. All conserved gene variants were validated by Sanger sequencing. We identified 27 coding variants (22 genes) potentially related to the phenotype (41% of patients). The affected genes, which were enriched by potentially deleterious sequence variants, belonged to distinct molecular cascades playing key roles in implantation/pregnancy biology such as trophoblast-endometrium adhesion, coagulation pathways, angiogenesis, immunological modulation, metabolism, extracellular matrix remodeling, cell proliferation, differentiation, apoptosis and steroidal nuclear receptor activation. These results suggest that RPL has polygenic inheritance because mutations in many genes may be contributing to the phenotype. The NGS and bioinformatics approaches presented here represent an efficient way to find mutations, having potentially moderate/strong functional effects, associated with RPL aetiology. After the *in vitro* functional validation, we expect that these mutations and genes could represent molecular biomarkers in women with idiopathic RPL.

2262F

Predictive modeling of gene expression in ethnic minorities. *K. Keys[1], C. Gignoux[2], A. Mak[1], W. Eckalbar[1], M. White[1], C. Eng[1], S. Oh[1], J. Liberto[1], D. Hu[1], S. Huntsman[1], S. Salazar[1], J. Rodríguez-Santana[3], N. Zaitlen[1], M. Seibold[4], E. Burchard[1,5].* 1) Medicine, UCSF, San Francisco, CA; 2) Medicine, University of Colorado, Denver, CO; 3) Centro de Neumología Pediátrica, San Juan, Puerto Rico; 4) Pediatrics, National Jewish Health, Denver CO; 5) Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA.

The GTEx database is a large public resource of transcriptomic data with linked genotype data, currently including 8,555 biological samples spanning 53 tissue types from 544 individuals. This resource has proven invaluable for numerous investigations into human biology, including as a resource of gene expression using software such as PrediXcan for powerful gene-based association tests. While the data is of high value, it is notable that 84% of the subjects in GTEx are of European descent, in contrast to the genetic diversity present in the USA and beyond. Therefore, it is of great interest to quantify the extent to which PrediXcan predictive models trained on GTEx perform on ethnically diverse subjects. To do so, we compare a range of models from the PrediXcan PredictDB and other predictive models derived from recent developments in machine learning to measured expression on 39 African American and 20 Puerto Rican transcriptomes derived from peripheral blood. We discuss our findings in the context of prediction accuracy and consequences for using PrediXcan on non-European study subjects. Finally, we apply our best-performing validated training models across a set of blood-based and non-blood-based phenotypes.

2263W

Genome-wide association study of asthma in individuals of mixed African ancestry reveals a novel association with markers on chromosome 2q14. *S. Chavan[1], M. Daya[1], N.M. Rafaels[1], A.M. Levin[2], M. Boorgula[1], T.M. Brunetti[1], A. Shetty[1], C.R. Gignoux[1], G. Wojcik[3], R. Johnston[4], EG. Burchard[5], L. Caraballo[6], G.M. Dunston[7], M.U. Faruque[8], E.E. Kenny[9], J.M. Knight-Madden[10], D.A. Meyers[11], E.R. Bleecker[11], C. Ober[12], C. Rotimi[13], J.G. Wilson[14], L.K. Williams[15], H. Watson[16], Z.S. Qin[17], M.A. Taub[18], T.H. Beaty[18], I. Ruczinski[18], R.A. Mathias[18], K. Barnes[1], CAAPA consortium.* 1) Colorado Center for Personalized Medicine, University of Colorado, Aurora, CO, USA; 2) Department of Public Health Sciences, Henry Ford Health System, Detroit, MI, USA; 3) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 4) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA; 5) Department of Bioengineering and Therapeutic Sciences , University of California, San Francisco, CA, USA; 6) Institute for Immunological Research, Universidad de Cartagena, Cartagena, Colombia; 7) Department of Microbiology, Howard University College of Medicine, Washington, DC, USA; 8) National Human Genome Center, Howard University College of Medicine, Washington, DC, USA; 9) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 10) Caribbean Institute for Health Research, The University of the West Indies, Kingston, Jamaica; 11) Department of Medicine, University of Arizona College of Medicine, Tucson, AZ, USA; 12) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 13) Center for Research on Genomics & Global Health, National Institutes of Health, Bethesda, MD, USA; 14) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 15) Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, MI, USA; 16) Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, Barbados; 17) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 18) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

    RATIONALE: Asthma is a complex disease with striking disparities across racial and ethnic groups, which may be partly attributable to genetic factors. One of the main goals of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to discover genes conferring risk to asthma in populations of African descent. **METHODS:** We performed a genome-wide meta-analysis of asthma association across CAAPA studies (using a total of 4,827 cases and 5,397 controls), genotyped on the African Diaspora Power Chip (ADPC), and imputed using the whole genome sequence reference panel from CAAPA itself. Statistical models appropriate to each study design were used to perform tests for association, and results were combined using the weighted Z-score method. We tested for replication of genome-wide significant results in two European cohorts (218+324 and 28,399+128,843 cases and controls respectively) and African ancestry subjects from the BioMe biobank (910+3,325 cases and controls). **RESULTS:** SNPs rs787160 and rs17834780 (chromosome 2 hg19 base pair position 144650140 and 144638809, respectively) were significantly associated with asthma case-control status (p=5.07e-09 and =2.44e-08 respectively). These SNPs were not associated with asthma in Europeans, but rs787160 was marginally associated with asthma in African ancestry subjects from the BioMe biobank (p=3.15e-02, with the same direction of effect as seen in the discovery data set). The combined p-value for rs787160 across both the CAAPA discover and BioMe replication is 3.7e-9 (Fisher's method). SNPs rs787160 and rs17834780 are intergenic to the *Rho GTPase Activating Protein 15 (ARHGAP15)* and *Glycosyltransferase like Domain Containing 1 (GTDC1)* genes. *ARHGAP15* is expressed in high levels in lymph node cells, peripheral blood mononuclear cells, CD8 T-cells, and B-lymphocytes, while *GTDC1* is expressed in high levels in lung and peripheral blood leukocytes. Glycosyltransferase plays a role in adhesion of environmental factors to epithelial cells, which could in turn play a role in risk to asthma. **CONCLUSIONS:** We identified a novel association signal for asthma and markers on chromosome 2q14, which may be specific to African ancestry. Currently further replication is underway in an additional 6,607 subjects of African ancestry.

2264T

Association study of placebo-treated patients from 35 clinical trials suggests genetic contribution to the placebo response. *A. Wuster[1,2], A. Haug-Baltzell[1,3,4], T. Bhangale[1,2], T.W. Behrens[1], R.G. Graham[1].* 1) Human Genetics Dept, Genentech Inc, South San Francisco, CA; 2) Bioinformatics and Computational Biology Dept, Genentech Inc, South San Francisco, CA; 3) Arizona Biological/Biomedical Sciences Program, University of Arizona, Tucson, AZ; 4) Genetics GIDP, University of Arizona, Tucson, AZ.

    The most common cause of Phase III clinical trial failure is an indistinguishable response rate between placebo and treatment arms. Since placebos are generally regarded as adequate controls, this has historically been accepted as indicative of a lack of drug efficacy. However, recent findings implicating specific genetic variants with a higher propensity for response under placebo treatment (*e.g.* rs4680, which causes a 3-4x reduction in the activity of the dopamine metabolism gene *COMT*) suggest a possible genetic basis for the placebo response. This indicates a potential for trial failure due to inadequately matching likely placebo responders between trial arms and highlights the need for further research into the genetic basis of the placebo response and its interaction with drug molecular pathways. Unfortunately, most work to date has been limited to analgesia and does little to uncover a generalized understanding of placebo genetics. We performed whole genome sequencing to 30x coverage of 11,817 patients from 35 clinical trials spanning 10 diseases (Rheumatoid Arthritis (RA), Asthma, Age-related Macular Degeneration, Colorectal Cancer, Idiopathic Pulmonary Fibrosis, Diabetic Retinopathy, Chronic Obstructive Pulmonary Disease, Inflammatory Bowel Disease, Systemic Lupus Erythematosus and Interstitial Lung Disease). To identify variants associated with placebo response, we performed genetic association analysis on patients who had been assigned to the control arm of each trial, contrasting individuals who responded to treatment, as defined by each trial's primary clinical endpoint, to non-responders. We then compared and combined trial-specific results using a meta-analysis approach. Preliminary results suggest that a variety of genetic variants affect a patient's propensity for response in a placebo arm, but the variants are largely disease-specific. For example, response in the control arm of RA trials is potentially associated with variation near *RHBDL3* (Rhomboid Like 3), *EBF2* (Early B-Cell Factor 2) and, intriguingly, *OPCML*, which encodes a likely opioid-binding protein. However, none of the previously reported analgesia-related variants, including rs4680, were significantly associated with RA placebo response. Our findings ultimately suggest that the biological origin of the placebo response is variable between disease areas.

2265F

The Precision Medicine Initiative All of Us Research Program: Innovative access to unprecedented data. *A. Ramirez[1], A. Philippakis[2], A. Kho[3], H. Xu[4], G. Abecasis[5], G. Hripcsak[6], P. Harris[1], S. Katherisan[2], D. Glazer[7], J. Denny[1].* 1) Medicine, Vanderbilt Univ, Nashville, TN; 2) Broad Institute, Cambridge, MA; 3) Northwestern University, Chicago, IL; 4) UT Health Science Center, Houston, TX; 5) University of Michigan, Ann Arbor, MI; 6) Columbia University, New York City, NY; 7) Verily Life Sciences, South San Francisco, CA.

The Precision Medicine Initiative All of Us Research Program is designed to recruit and engage at least one million participants reflecting the diversity of the United States. These participants will complete health surveys and undergo baseline physical measures as well as have the option to share their electronic health record data and biospecimens for future genetic and other biomarker testing. The Data and Research Center (DRC) was charged with storing, curating, and securing data and making it available to a broad community of traditional and non-traditional researchers. The DRC team includes Vanderbilt University Medical Center, Broad Institute, Verily Life Sciences, Columbia University, University of Michigan, and other subawardees. The DRC working with All of Us partners including many enrolling Health Care Provider Organizations, the Participant Center led by Scripps Research Institute, the Participant Technologies Systems Center led by Vibrent Health, the world's largest research-cohort biobank at the Mayo Clinic, and a strong leadership team at the National Institutes of Health has developed a robust internal operational system and research data stack. Beta testing of the infrastructure began with real participants on May 31, 2017. All data in the DRC is hosted in a cloud based platform to facilitate research in an open and collaborative manner. Data initially arrive in a raw data repository (RDR) and then is processed into a curated data repository (CDR) using structured data ontologies. The Observational Medical Outcomes Partnership (OMOP) common data model is being used for EHR data. To enable this structure, broad data access policies are being developed that leverage the cloud infrastructure. Researchers will engage in a dynamic user platform including cohort builder tools and notebook based analysis capabilities that facilitate collaboration and scale to large analyses. Furthermore, the DRC will be the primary source of support for researchers accessing data and is developing training and outreach efforts to engage a diverse spectrum of users. The DRC aims to support world class research optimizing this unparalleled data source for the benefit of human health. This presentation is a key part of efforts to engage world-class researchers including the ASHG community in development of this intentionally dynamic framework.

2266W

Data-driven approach to dietary phenotypes for nutrigenomics in UK Biobank. *J.B. Cole[1,2,3], J.N. Hirschhorn[3,4].* 1) The Broad Institute of MIT and Harvard, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Boston Children's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA.

Diet is a complex multifactorial trait of specific and highly related measures which together have demonstrated strong genetic underpinnings ($h^2 \sim 30\%$). However, previous studies on the genetics of diet have been limited in power in terms of both sample size and the proportion of genetic variance dietary variables can explain individually. Therefore, we utilized dimensionality reduction techniques on comprehensive UK Biobank diet data to optimize phenotypes for genetic discovery. A final genotyping QC dataset of 139,185 individuals was obtained by removing UK Biobank and UK BiLEVE recommended genomic exclusions (N=408 and N=406, respectively), sex mismatches (N=191), and non-European-derived Caucasians as defined by a thorough analysis of genetic ancestry (N=13,543). We have begun extensive efforts on curating over 300 relevant diet fields from both the lifestyle and environment and 24-hour diet recall questionnaires. Each field for any given participant can have both multiple answers (up to 24) and multiple instances (up to 5), contributing to the complexity in parsing appropriate and informative diet-related variables, all the while emphasizing its vast potential. Preliminary GWAS analysis on 10 UK Biobank univariate dietary traits (e.g. overall frequency of fish intake, cheese intake, alcohol intake, etc.) demonstrated 0 genome-wide significant signals. Thus, moving forward we plan to conduct two distinct, yet complementary, unbiased data-driven approaches to define and optimize traits for genome-wide association study (GWAS) analysis. The final phenotypic dataset, with over 1000 individual variables, will then be subjected to a principal components analysis (PCA) to extract orthogonal summary variables explaining greater phenotypic variance than any individual variable alone. We will additionally conduct a PCA on the genetic-covariance matrix of significantly heritable variables using a linear mixed model approach to maximize the genetic variance explained by top PCs. The top PCs derived from these two complementary PCA approaches will then be used as GWAS phenotypes. We will return these new composite dietary phenotypes to the UK Biobank resource. This analysis will not only help us understand the phenotypic and genetic relationships between the different foods we eat, but is also robustly suited to identify novel genetic determinants of diet and provide valuable information for future research efforts in metabolism and medicine.

2267T

A broad survey of the relationship between autozygosity and fitness-related and sociodemographic traits in the UK Biobank. *E.C. Johnson[1,2], M.C. Keller[1,2].* 1) Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO; 2) Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO.

The body of literature on the association between very distant inbreeding (measured using F_{ROH}, the proportion of the genome contained in runs of homozygosity (ROHs)) and complex traits in humans is growing, but remains smaller than the body of literature examining the relationship between inbreeding depression and Mendelian disorders. Characterizing the relationship between F_{ROH} and complex traits can lead to a better understanding of their evolutionary history and genetic architectures. Using genome-wide single nucleotide polymorphism (SNP) data from over 100,000 individuals in the UK Biobank project, we explored the relationship between autozygosity (as measured by F_{ROH}) and more than 30 traits, including several potentially related to fitness. Preliminary results indicate that increased autozygosity is significantly associated with several traits likely related to fitness, such as birth weight, height, and fluid intelligence, consistent with a prediction of purifying selection biasing causal variants toward being (partially) recessive; further, the relationships between inbreeding and birth weight and height remained significant after controlling for average household income and years of educational attainment. F_{ROH} was also related to several sociodemographic traits such as income and socioeconomic status. The direction of causation for these relationships is less clear, because these traits may directly influence level of outbreeding in parents and be passed down culturally or genetically to offspring. In general, interpretation of relationships between F_{ROH} and traits must be made with a good deal of caution, because of the potential for third variable or reverse causation explanations of the relationship.

2268F

Fine-mapping of tobacco and alcohol associated loci in around 900K participants. *M. Liu, GSCAN Consortium.* Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO.

Alcohol and tobacco use are common causes of preventable deaths. Genome-wide association studies (GWAS) have found a handful of significant functional variants (such as *CHRNA5* and *GCKR* for nicotine and alcohol respectively) that are now under being tested in biological models. However, to discover these variants, we would require a well-powered GWAS with large enough sample sizes to produce meaningful results and reduce false positives. The GWAS & Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) have conducted a GWAS meta-analysis with around 1 million participants and found a total of greater than 200 novel loci amongst 6 phenotypes: Cigarettes per Day (CPD, N=259,485), Age of Initiation (AI, N=257,272), Smoking Initiation (SI, N=966,878), Smoking Cessation (SC, N=467,073), Drinker non-Drinker (DND, N=787,983) and Drinks per Week (DPW, N=695,140). Most of the results, except for a few studies that used either 1000G phase 3 or their own internal panel, used the Haplotype Reference Consortium panel. We characterized these loci by applying a conditional analysis to find independent variants in these loci. We have also begun the process of fine-mapping within the loci. The conditional analysis was done with a sequential forward selection method to identify independent variants within the significant loci (defined as 1MB around the significant variant). In the conditional analysis, we found 1 independent variant out of 1 loci for AI, 49 independent variants out of 18 loci for CPD, 81 independent variants out of 22 loci for SC, 431 independent variants out of 134 loci for SI, 106 independent variants out of 67 loci for DND and 152 independent variants out of 47 loci for DPW. The average imputation quality for common variants (minor allele frequency (MAF) > 1%) was 0.89 (SD=0.11) and the imputation quality of variants with a MAF of < 1% but greater than .1% was 0.74 (SD=0.19). From here, we are confident that despite using imputed results, we would still find plausible and rare (with MAF as low as .1%) variants within the 18 loci. Furthermore, we also have plans to explore these significant loci and expect to present pathway analysis results at the conference. We anticipate that our results will show new promising loci/variants that can be used in understanding the mechanisms of addiction.

2269W

Genetic predictors of biomarker levels derived from prospective epidemiologic cohorts applied to electronic health records to identify new biomarker-disease associations. *J.D. Mosley[1], Q. Feng[1], Q.S. Wells[1], S.L. Van Driest[1,2], C.M. Shaffer[1], T.L. Edwards[3], L. Bastarache[4], W.Q. Wei[4], L.K. Davis[4], C.A. McCarty[5], W. Thompson[6], C.G. Chute[7], A.S. Gordon[8], M.R. Palmer[8], D.R. Crosslin[9], E.B. Larson[8,10], D.S. Carrell[10], I.J. Kullo[11], J.A. Pacheco[12], P.L. Peissig[13], J.G. Linneman[13], B. Namjou[14], M.D. Ritchie[15], K.M. Borthwick[15], S.S. Verma[15], J.H. Karnes[16], T.J. Wang[4], C.M. Stein[1], J.C. Denny[1,4], D.M. Roden[1,4,17].* 1) Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN; 3) Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN; 4) Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 5) University of Minnesota Medical School, Duluth campus,, Duluth, MN; 6) Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 7) Schools of Medicine, Public Health, and Nursing, Johns Hopkins University, Baltimore, MD; 8) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 9) Departments of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 10) Kaiser Permanente Washington Health Research Institute, Seattle, WA; 11) Department of Cardiovascular Diseases, Mayo Clinic, Rochester MN; 12) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 13) Biomedical Informatics Research Center, Marshfield Clinic Research Institute, Marshfield, WI; 14) Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 15) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 16) Department of Pharmacy Practice and Science, University of Arizona College of Pharmacy, Tucson, AZ; 17) Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN.

Background: Biomarkers enable early disease detection, risk stratification and provide insights into disease mechanisms. Typically, biomarkers are identified using prospective epidemiologic studies which are long, expensive and have low power. We hypothesized that we could apply genetic predictors of biomarker variability derived from an epidemiological cohort to an electronic health record (EHR) population and use phenome-wide scanning to predict novel biomarker-disease associations. **Methods:** We selected 53 biomarkers measured in the Atherosclerosis Risk in Communities study (n=7,740 subjects) and used Bayesian sparse linear mixed modeling to calculate weightings across 739,681 SNPs. These weightings were used to compute a genetically predicted value for each biomarker in a population of 37,153 genotyped individuals with EHRs. We tested the association between each genetic predictor and 1,139 clinical EHR diagnoses. **Results:** The underlying genetic architectures of the biomarkers varied considerably and were reflected in the distributions of the predicted biomarkers in the EHR cohort. These distributions ranged from trimodal for biomarker levels that were heavily driven by SNPs with relatively large effect sizes (e.g. vWF protein levels) to approximately normally distributed for highly polygenic phenotypes (e.g. waist circumference). We selected 42 biomarker-disease associations *a priori* as positive controls, and 28/42 were significant at FDR q<0.1. There were 349 additional associations (known and novel), including pack-years of smoking (associated with alcohol use and obesity), waist circumference (psoriasis and venous thrombosis), white blood cell count (anxiety disorder and bronchitis), magnesium levels (kidney stones and chondrocalcinosis), platelet count (antibiotic poisoning), monocyte count (tonsillitis), FEV1/FVC (emphysema), vWF protein levels (venous and arterial thrombosis), triglycerides (arterial diseases) and LDL cholesterol (septicemia). We validated a novel association between low LDL and septicemia using a separate EHR cohort (n=22,281) and confirmed an increased risk of septicemia in low LDL-cholesterol (LDL-C<60 mg/dl) versus normal LDL-C (>90 and <130 mg/dl) subjects (OR=3.54 [90% CI: 2.81-4.46], p<2x10^{-16}). **Conclusions:** The design we describe leverages the strengths of epidemiological and EHR data sets. By testing for cross-data set associations, the design can efficiently identify novel biomarkers of disease risk.

2270T

New insights into the genetic architecture of complex human traits from Bayesian mixture model analyses in a large dataset. *J. Sidorenko[1,2], L.R. Lloyd-Jones[1], A. Metspalu[2], T. Esko[2,3], P.M. Visscher[1,4], M.R. Robinson[1,5].* 1) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia; 2) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 3) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge 02142, MA, USA; 4) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia; 5) Department of Computational Biology, University of Lausanne, 1011 Lausanne, Switzerland.

Introduction of linear mixed models (LMM) to human genomics has demonstrated that a large proportion of the complex trait heritability can be captured by jointly accounting for all common genetic variants. LMM can be used to obtain SNP markers effects with best linear unbiased prediction properties (BLUP), which have been shown to improve genomic prediction over genetic predictors obtained from large-scale meta-analyses. The classical LMM, however, assumes that there is a single distribution of genetic effects, implying that all markers contribute to SNP-based heritability. This approach can be improved through the use of a Bayesian mixture model framework, where a mixture of normal distributions of SNP effects, including a zero effect, is adopted instead. Here, we gain new insights into the underlying genetic architecture of complex human traits by applying a Bayesian mixture model (BayesR) with five normal distributions of genetic effects (each variant potentially explaining 1, 0.1, 0.01, 0.001% or none of the trait genetic variance) to 17 quantitative traits in the UK Biobank (140,720 European participants and 954,662 common SNPs with MAF > 1%). We demonstrate a high degree of polygenicity for all traits, with most of the SNP-heritability (38-85%, mean 65%) attributable to genetic markers that each contribute just 0.001% to the total genetic variance. We find evidence of large effects, cumulatively explaining 3 to 6% of the SNP-heritability, for traits such as age at menopause, menstrual cycle length, and heel bone mineral density. We demonstrate how this Bayesian framework can be used to test for sex-specific effects, finding limited evidence across all traits for regions of the genome that contribute differing amounts of variance among the sexes. We replicate our findings in a sample taken from the Estonian Biobank (EGCUT, N= 15,299) and show that a genetic predictor created from a BayesR analysis in a large cohort with individual-level data outperforms genetic predictors created from a large-scale meta-analysis for height, with increased prediction accuracy into independent cohort over 30% (R^2 0.24 vs. 0.16). Together, these results highlight the potential of Bayesian mixture models to provide better inference for SNP detection, flexibly test for interactive effects and improve genomic prediction.

2271F

Rare coding variant association analysis for anthropometric traits using more than 25,000 exome-sequenced samples from ExAC. *S. Vedantam[1,2], A. Nomura[3], A. Mahajan[4], J.J. Lee[5], C.M. Lindgren[2,4,6], A.P. Morris[4,7,8], D.G. MacArthur[2,9], M.I. McCarthy[4,10,11], S. Kathiresan[2,12,13,14], J.N. Hirschhorn[1,2,12], D. Saleheen[5,15,16] on behalf of the ExAC consortium.* 1) Center for Basic and Translational Obesity Research, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Department of Cardiovascular and Internal Medicine, Kanazawa University Graduate School of Medicine, Japan; 4) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK; 5) Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA; 6) The Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 7) Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) Department of Biostatistics, University of Liverpool, Liverpool, UK; 9) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 10) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 11) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK; 12) Harvard Medical School, Boston, MA, USA; 13) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 14) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 15) Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA; 16) Center for Non-Communicable Diseases, Karachi, Pakistan.

Genome-wide association studies (GWAS) have identified thousands of common variants associated with polygenic traits and diseases. These common variants are largely non-coding and thus present challenges for definitively implicating genes within an associated locus. Identifying causal rare coding variants can directly point to causal genes, but generally requires sequencing data, which, until now, has only been done in sample sizes smaller than that of GWAS. Thus, the role of rare variants in contributing to heritability and implicating causal genes remains largely unknown. Initial exome-sequencing studies suggest that the required sample sizes for detecting robust associations will be at least comparable to those needed for GWAS, but a systematic study to assess the actual sample size requirements has not been conducted. To better understand the role of rare coding variation in polygenic traits and the implications for future sequencing studies, we analyzed exome-sequencing data for three anthropometric traits where GWAS has been successful: height, body mass index (BMI) and waist-hip ratio adjusted for BMI (WHR). We performed single variant and gene-based tests for these traits in 25,774 samples across five different ancestries that contributed to the Exome Aggregation Consortium (ExAC), from the GoT2D, T2DGENES and MIGen consortia. We performed both single variant association tests and aggregate gene-based tests of rare variants. We were able to replicate multiple single-variant associations with known common variants, but no individual rare variant has reached exome-wide significance. In gene-based tests for height, we identified exome-wide significant associations with rare variants in a known skeletal dysplasia gene (*ACAN*), and another novel gene (*EIF1AD*). Our results for BMI and WHR adjusted for BMI did not show any exome-wide significant variants or genes, although we observed an enrichment of lower p-values in genes known to cause syndromes with obesity as a feature. Together, these results indicate that the current sample of >25,000 individuals is substantially underpowered to discover rare coding variants associated with the same traits where GWAS was successful at comparable sample sizes. Much larger sample sizes will be needed in order to identify the rare coding variants associated with polygenic traits that could directly implicate biologically relevant genes.

2272W

Copy number variation associated with white blood cell phenotypes in the eMERGE Network. *M.R. Palmer[1], A. Gordon[1], I.B. Stanaway[2], D.S. Carrell[3], J.T. Glessner[4], H. Hakonarson[5], P. Piessig[6], E. Larson[3], G.P. Jarvik[1], D.R. Crosslin[2], The electronic Medical Records and Genomics (eMERGE) Network.* 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 3) Kaiser Permanente Washington Health Research Institute, Seattle, WA; 4) Molecular Neurogenetics Unit and Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, and Department of Neurology, Massachusetts General Hospital, Boston, MA; 5) The Children's Hospital of Pennsylvania, Philadelphia, PA; 6) Marshfield Clinic Research Institute, Marshfield, WI.

Background White blood cell counts (WBC) and cell type differentials are collected in the course of standard medical care, as lower or higher levels from baseline could be an indicator of a variety of diseases. Genome wide association studies have identified variants associated with WBC phenotypes in cohorts of European and African ancestry (Nalls et al. 2011; Keller et al. 2014; Crosslin et al. 2012; 2013). Structural variation is another possible source of genetic influence on WBC phenotypes, but few association studies have been carried out with copy number variants (CNVs). **Methods** We extracted total leukocyte counts and differential proportions of neutrophils, lymphocytes, monocytes, eosinophils, and basophils of outpatients from electronic health records for participants in the Electronic Medical Records and Genomics (eMERGE). We used PennCNV to call CNVs from Illumina Human660 W-Quadv1 genotyping data. We performed linear regressions of median leukocyte counts, lymphocytes, monocytes, neutrophils and logistic regressions of basophils (controls 0-0.6%, cases 0.61-6.0%) and eosinophils (controls 0-1.8%, cases 3.0-34.6%) with ParseCNV-called CNV Regions (CNVR) in PLINK (Glessener et al. 2013). We adjusted for the covariates age, sex, median BMI, and eigenvectors 1-10 derived through principal component analysis. **Results** A total of 8447 participants met all inclusion criteria (Crosslin et al. 2012), passed quality control, and were included in at least one WBC phenotype. With a P-value threshold of 5×10^{-4} to correct for multiple testing (Glessner et al. 2013), twelve CNVRs from leukocyte counts association met this threshold, eleven from neutrophils, five from lymphocytes, seven from monocytes, three from eosinophils, and six from basophils. A deletion at 10q23.33 (776.3kbp; MAF=0.3%) was associated with neutrophil differential (ß=8.40; p=3.0×10^{-4}). This region contains the genes *CYP2C18*, *CYP2C19*, *CYP2C8*, *CYP2C9*, *HELLS*, and part of *TBC1D12*. Mutations in *HELLS* have been linked to cases of immunodeficiency-centromeric instability-facial anomalies syndrome (Thijssen et al. 2015). A 3.0kb duplication at 20q13.33 (MAF=5.7%; ß=0.32, p=2.3×10^{-4}) was associated with monocyte differential. This region overlaps genes *LIME1*, *SLC2A4RG*, *ZGPAT*, which are all expressed in hematopoietic tissues. **Conclusions** We found associations of CNVRs for all the phenotypes, including novel and replication findings. Some include genes associated with inflammatory and immune phenotypes.

2273T

Utilizing protein quantitative trait loci to identify functional candidates from genome-wide association studies. *S. Sivertson, A. Stark.* University of Notre Dame, IN.

While genomewide association studies (GWAS) have focused on identification of single nucleotide polymorphisms, there remains a paucity of information on how these associations, including repeated and validated associations, are connected to function. Quantitative trait loci (QTL) mapping has been utilized for mRNA, protein, methylation, microRNA in order to help annotate these findings. We interrogated 198 protein QTL (pQTLs) associations published in 2014 with the NHGRI-EBI GWAS database in the fall of 2016 in order to find any recent associations that could be attributed to a protein. Over 9% of the pQTLs identified had further trait associations identified, including HIV-1 control, hematological parameters and red blood cell traits, colorectal cancer, the age at onset of menarche, height, Parkinson's Disease, and fasting glucose-related traits. We then compared the protein with the GWAS results and prioritized interesting targets for functional follow-up based on literature and known pathways. Interestingly, rs10505477 has shown repeated association with colorectal, gastric, and prostate cancer and is a pQTL for *MPND*. Previously, these GWAS associations had hypothesized that it was acting through *MYC*, the nearby major coding gene. However, the SNP itself is located within the *CASC8* gene and we used siRNA to knockdown *CASC8* and assessed *MPND* expression as well as *MYC* expression in order to demonstrate the importance of *MPND* in this association. Future research could be done on the remaining traits and demonstrate the importance of utilizing QTL data in association interpretation.

2274F

Integrative fine-mapping of 34 complex phenotypes. *R. Johnson[1], G. Kichaev[2], K. Burch[2], B. Pasaniuc[3,4].* 1) Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90024, USA; 2) Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90024, USA; 3) Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90024, USA; 4) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA 90024, USA.

Genome-wide association studies (GWAS) have identified thousands of variants in the genome that reproducibly associate with complex disease. However, most variants uncovered in GWAS are not directly biologically causal, but rather, are correlated with the true causal variant(s) through linkage disequilibrium. Statistical fine-mapping aims to prioritize Single Nucleotide Polymorphisms (SNPs) in regions surrounding a significant association peak to discern the true causal variants. Here, we apply PAINTOR, an integrative probabilistic fine-mapping framework that combines summary association statistics (i.e. Zscores) with functional genomic annotation data, to prioritize possible causal variants for a set of traits and diseases. As not all functional annotations are relevant to a trait under study, we first identify informative epigenetic marks using forward selection for each trait. We then construct a trait-specific integrative functional model to refine the set of SNPs in every GWAS region. Our study analyzes GWAS from 34 complex phenotypes which consist of over 4,347 regions in total. We demonstrate in this data that by applying leveraging functional data, we can reduce the number of SNPs required for follow-up in order to capture 95% of the causal variants. We provide an atlas of "credible sets" that we make available to the research community.

2275W

Probabilistic assignment of causal genes at transcriptome-wide significant risk loci. *N. Mancuso[1], G. Kichaev[2], H. Shi[2], C. Giambartolomei[1], A. Gusev[3,4], B. Pasaniuc[1,2,5].* 1) Dept of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles, CA; 2) Dept of Bioinformatics, UCLA, Los Angeles, CA; 3) Dept of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 5) Dept of Human Genetics, UCLA, Los Angeles, CA.

　　Transcriptome-wide association studies (TWASs) have identified thousands of genes whose locally-regulated expression is associated to complex traits and diseases. Since complex correlations between predicted expression levels at a given locus can yield multiple associated genes in TWAS, for the vast majority of risk loci it remains unclear which gene is causally impacting disease risk. Here we introduce a new probabilistic framework to estimate posterior probabilities for genes to be causal at TWAS significant loci. Our approach mirrors the recently introduced methods for statistical fine-mapping of causal genetic variants to yield sets of genes that contain the causal gene at a given credibility (e.g., 90%) to increase accuracy in functional follow-up. Our proposed framework accounts for both the correlation structure among genetic variants (i.e. linkage disequilibrium, LD) as well as correlations in the expression of genes (i.e. co-expression of genes at a given region in the genome). In detail, we model the distribution of TWAS statistics using a multi-variate normal distribution parametrized by LD, eQTL weights and causal gene effect sizes. We integrate effect sizes within a Bayesian approach and compute the marginal posterior probability of a gene being causal. We validated our approach through extensive simulations starting from real genotype data in 1000 Genomes and realistic heritability levels for gene expression data. We find that our method outperforms existing co-localization methods with the largest performance gain when multiple causal genes at a locus impact trait. In the analysis of TWAS results for lipids, we find that several genes with well-studied functional impact on HDL are highly ranked using our method. For example, APOA1 and LIPC, genes believed to play a role in HDL levels, were ranked with marginal posterior probabilities of > 0.99. Overall, we find a single gene in the 90% credible set for 30 / 94 risk loci for HDL, LDL, and triglycerides with 2.6 genes on average (s.d. = 1.8). Finally, we analyze 65 large-scale TWAS across multiple traits and diseases to localize the credible set of causal genes at known GWAS risk loci.

2276T

The Macaque Genotype and Phenotype (mGAP) database: a novel resource to support genetic disease model development and translational research in nonhuman primates. *B. Ferguson, B.N. Bimber, L. Shorey, H. Rosoff, M. Yan, L. Carbone, E. Spindel, A. Vinson.* Primate Genetics Section, ONPRC, Oregon Health Sciences Univ., Portland, OR.

　　Rhesus macaques, the most broadly studied nonhuman primate (NHP), offer outstanding opportunities for the translational study of complex human genetic diseases. Leveraging the similar natural disease spectrum as humans, a diverse range of rhesus macaque disease models has been established, including models of cardiovascular disease, addiction, age-related macular degeneration, developmental disorders, cognitive decline, colitis, and infectious diseases such as zika and simian immunodeficiency virus (SIV). The discovery of similar genetic disease associations in macaques and humans underscores the value of macaques for the genetic analysis of complex traits, and for the potential testing of pharmacogenomics treatments. To support expanded use of the rhesus macaque for genetic analysis, model characterization, and therapeutic development, we are genomically characterizing an expansive, pedigreed rhesus macaque colony housed at the Oregon National Primate Research Center (ONPRC). We selected a set of 2,245 Indian-origin rhesus macaques, members of a single, 9 generation pedigree, for genomic analysis. The study subjects include those with phenotypic data collected on cardiovascular disease endophenotypes, age-related macular degeneration, cognitive measures, colitis, endometriosis, S/HIV response, and behavioral measures, among others. We are using a combination of whole genome sequencing (WGS) and genotyping-by-sequencing (GBS) to identify genome-wide sequence variants in these subjects. We have established a custom, publically accessible database to share these data: the <u>M</u>acaque <u>G</u>enotype <u>a</u>nd <u>P</u>henotype database (mGAP; https://mgap.ohsu.edu). In the future, the genotype database will be expanded to include clinical measures extracted from electronic health records. Thus the mGAP database will enable investigators to 1) identify potential new disease models, based on predicted functional variants, 2) identify animals with particular sequence variants of interest, 3) identify genotypes in animals with selected phenotype measures, and 4) query phenotype measures among macaques based upon selected genotypes. <u>Conclusion</u>: mGAP provides a unique opportunity for investigators to identify potential NHP genetic disease models to advance the study and treatment of human disease.

2277F

Examining sex differences in genetic effects on subcortical brain structures. *K. Grasby on behalf of The Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium.* Psychiatric Genetics, Queensland Institute of Medical Research, Herston, QLD, Australia.

Individual differences in subcortical brain morphometry are highly heritable. Several genetic variants have been identified that influence the volume of subcortical structures, and these variants are structure specific rather than general in effect. Sex differences have also been observed in the volume of subcortical structures. Here we assess if genetic variants influencing subcortical brain structures differ between females and males. We present preliminary results from sex-stratified GWAS meta-analyses on seven subcortical structures derived from T1-weighted magnetic resonance imaging scans from ~6000 females and ~5400 males from 22 cohorts from the ENIGMA consortium. This sample is a subset of the larger ENIGMA consortium analyses on variation in subcortical structures among ~30,000 individuals. The subcortical structures analysed were the nucleus accumbens, amygdala, caudate nucleus, hippocampus, globus pallidus, putamen, and thalamus. The volume of these structures were calculated with validated and freely available software, and mean volume calculated from the left and right hemispheres. Genotypic data were imputed to the 1000 genome reference panel. Preliminary results in this discovery meta-analysis identified four genome-wide significant loci, three were novel. One intergenic locus (14q22.3), which we have previously reported as significant in a meta-analysis that combined females and males, was associated with putamen volume in females and in males in these stratified analyses. Tests of heterogeneity showed that the top hit in males was significantly stronger in effect in males than in females (p <1.658e-10). Of the novel loci, one was an intronic locus (11q15.2) within SPON1 associated with nucleus accumbens volume in females. The second was an intronic locus (2q34) within ERBB4 associated with amygdala volume in males. The third was an intergenic locus (13q21.31) associated with putamen volume in males. Several loci passed a suggestive level of significance (p < 1e-6), each of these suggestive associations were trait and sex specific. We plan to extend these findings with data from the UK Biobank. We will present SNP heritability and genetic correlation with psychiatric and cognitive traits using the results from the second round of meta-analysis.

2278W

***In silico* evaluation of a more comprehensive pharmacogenetic profile for predicting opiate metabolizer phenotype.** *F.R. Wendt[1,2], A. Sajantila[3], R. Chakraborty[1], G. Pathak[1], A.E. Woerner[2], R.S. Moura-Neto[4], B. Budowle[1,2,5].* 1) Institute for Molecular Medicine, University of North Texas Health Science Center, Fort Worth, TX; 2) Center for Human Identification, University of North Texas Health Science Center, Fort Worth, TX; 3) Laboratory of Forensic Biology, Department of Forensic Medicine, University of Helsinki, Helsinki, Finland; 4) Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; 5) Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia.

Purpose: The pharmacogene encoding CYP2D6 is highly polymorphic and confers poor, intermediate, extensive, and ultrarapid phase I metabolism of many endogenous toxins and foreign compounds, including opiates. The pharmacogenetics of opiate metabolism is particularly important due to the relatively high incidence of addiction and overdose of opioid analgesics. Recently, trans-acting opiate metabolism/analgesic response enzymes (UGT2B7, ABCB1, OPRM1, and COMT) have been included in pharmacogenetic studies for better prediction of metabolic response of ante- and post-mortem patients. Many studies are limited in that subject selection favors presence/absence of phenotype and places less emphasis on individual demography, resulting in substructured sample cohorts. Additionally, studies that attempt to predict phenotypic state often rely on single nucleotide polymorphism (SNP)-targeted genotyping approaches which inherently cannot detect novel/rare polymorphisms. In contrast, gene-targeted massively parallel sequencing (MPS) can identify additional polymorphisms that refine previous pharmacogenetic predictions. **Methods:** 1000 Genomes Project data for ~15,000 SNPs were analyzed in two phases: (1) Describe population genetic variation and summary statistics for five opiate metabolism genes in self-reported healthy individuals from five super- and 26 sub-populations; and (2) Assess how comprehensive full-gene information refines metabolizer phenotype (MP) estimates. Both phases were performed using RStudio, Excel-based workbooks, Genetic Data Analysis, and TreeView. **Results:** A summary is provided of population statistics, variant effect predictions, and clustering of populations based on polymorphisms associated with opiate metabolism. Comparisons of standard inferences using SNP-targeted versus full-gene predictions indicate that a full-gene approach provides significantly decreased CYP2D6 activity scores (qualitative measure of phenotype based on *CYP2D6* diplotype) (p < 0.001). These results indicate that some extensive metabolizers may be incorrectly classified due to damaging polymorphisms elsewhere in the gene. Finally, logistic regression indicates little association between CYP2D6 MP and genotype-predicted activity of downstream enzymes. **Conclusions:** These results provide a population genetic baseline for pharmacogenetically relevant polymorphisms and demonstrate that full-gene information should be considered for phenotype predictions.

2279T

Drug target genes associated with clinical phenotypes in the genetically isolated population of Finland are more likely to succeed in pharmaceutical development. *R. March[1], M. Cunha[1], C. Elks[1], A. Platt[1], M. Alanen-Kinnunen[3], S. Lemmelä[3], V. Salomaa[5], P. Jousilahti[5], M. Daly[4], D. Goldstein[1,2], A. Palotie[3,4].* 1) Centre for Genomics Research, IMED Biotech Unit, AstraZeneca, Cambridge, United Kingdom; 2) Institute for Genomic Medicine, Columbia University, USA; 3) Institute for Molecular Medicine, Helsinki, Finland; 4) Broad Institute of MIT and Harvard, USA; 5) Institute of Health and Welfare, Helsinki, Finland.

Drug target genes (DTG) associated with disease are believed to be more likely to succeed in pharmaceutical development. Recent research, however, suggests that most meaningful genetic associations are driven by missense or loss of function variants. Such variants are often rare and may be more easily detected in populations that have undergone population bottlenecks, such as Finland. In contrast, genes associated with drug response may escape negative selection and remain at relatively high prevalence. We tested these hypotheses through a partnership between AstraZeneca's Centre for Genomics Research and the University of Helsinki. First, we compared DTG from 17 AstraZeneca programs for which the pharmaceutical outcome (successful/unsuccessful FDA approval) was known. We identified one or more genetic variants associated with drug response and/or relevant clinical phenotypes for most (10, 59%) of the DTG (PubMed, AstraZeneca data on file). Consistent with earlier reports, genetic associations were present in 7 (78%) of successful drug programs versus 3 (33%) of unsuccessful drug programs. The associated gene variants included missense and predicted loss of function variants in ExAC, a database of whole exome sequence from 60,706 unrelated individuals. The mean MAF of DTG variants associated with pharmaceutical success in all populations was 0.14 (95% CI -0.02 to 0.2). These variants were modestly enriched in Finns (mean 1.8, 95% CI 0.17 to 1.29, top: *PARP* rs11571833 with an enrichment odds ratio of 1.8). Second, we identified clinical phenotypes associated with variants in DTG for candidate drugs and compared the frequency of these variants in Finns with those for non-genetically isolated populations. We scanned for associations between genetic variants in selected genes and clinical phenotypes in over 20,000 Finns, comparing the allelic frequencies from the top-hits with those available from ExAC. We identified one or more genetic variants associated with a clinical phenotype for all genes, except for *IL33*. Most of these variants were enriched in the Finnish population (top: *CDK12* rs200488621 with an enrichment odds ratio of 26.6), suggesting that due to founder events, variants with phenotypic effects may have increased frequency in Finland. Our results suggest that DTG associations, especially in genetically isolated populations such as Finland, can increase success in pharmaceutical development. Validation studies are required to confirm these results.

2280F

Transferability of polygenic risk prediction across diverse and admixed populations. *A.R. Martin[1,2,3], D. Palmer[1,2,3], R.K. Walters[1,2,3], H. Huang[1,2,3], C. Chen[1,2,3,4], S. Ripke[1,2,3,5], M. Lam[6], C.R. Gignoux[7,8], G. Wojcik[7], J.S. Hleap[9], S. Gravel[9], C.D. Bustamante[7], E.E. Kenny[10], B.M. Neale[1,2,3], M.J. Daly[1,2,3].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 3) Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 4) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; 5) Department of Psychiatry, Charite Universitatsmedizin Berlin Campus Benjamin Franklin, Berlin, Germany; 6) Research Division, Institute of Mental Health, Singapore; 7) Department of Genetics, Stanford University, Stanford, CA 94305, USA; 8) Colorado Center for Personalized Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA; 9) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 10) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

Over the past decade, genome-wide association studies (GWAS) have successfully identified thousands of disease loci, but the vast majority of this research has focused on populations of European descent. As clinics have begun adopting prediction using genetic risk scores, a key open question surrounds the transferability of findings across populations. Population genetic theory indicates that effects of linkage disequilibrium (LD), allele frequencies, and genetic architecture will impact transferability of risk scores, even if all causal alleles have the same biological effect in all populations; however, there has been a dearth of empirical data to demonstrate this. We have recently identified directional inconsistencies in polygenic risk scores across populations for several well-studied traits and diseases, such as height and type II diabetes. Here, we use several approaches to understand and improve GWAS transferability: 1) we model risk prediction accuracy with coalescent simulations, 2) we empirically assess prediction accuracy in multi-ethnic schizophrenia GWAS cohorts, and 3) we develop novel statistical methods to improve cross-population prediction accuracy. Our complex trait simulation framework models genotypes and phenotypes for multiple continental populations, and shows that correlations between true and inferred risk are typically highest in the population from which summary statistics were derived. Additionally, scores inferred from European GWAS are unpredictably biased in other populations. Using results from the Psychiatric Genetics Consortium, we find that East Asian schizophrenia risk is better predicted by summary statistics from smaller East Asian cohorts (13k cases and 16k controls) than from ~3-fold larger European cohorts (37k cases and 113k controls, Nagelkerke's R^2 = 0.104 vs 0.066). To address these issues, we implemented a Kalman filter that uses LD to correct effect size estimates to improve inference of trans-ethnic polygenic scores from summary statistics. Extensions of this work to include local ancestry will be particularly relevant in recently admixed populations. Simulations and empirical data both indicate that risk scores from large-scale European GWAS may have limited value in non-European samples, owing to both mean and variance shifts, and highlights the need for the development and adoption of improved polygenic risk methods, as well as the inclusion of more diverse individuals in medical genomics.

2281W

Genomic characterization of the immunoglobulin heavy chain variable gene locus in individuals of African, Asian, and European descent reveals signatures of elevated haplotype diversity. *O. Rodriguez[1], W. Gibson[2], M. Emery[1], G. Deikus[1], M. Pendleton[1], W.A. Marasco[4,5], R. Sebra[1], E.E. Eichler[3], A.J. Sharp[1], M. Laird Smith[1], A. Bashir[1], C.T. Watson[2].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine of Mount Sinai, New York, NY; 2) Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Cancer Immunology & AIDS, Dana-Farber Cancer Institute, Boston, MA; 5) Department of Medicine, Harvard Medical School, Boston MA.

Genes encoding the heavy chains of antibodies, critical components of adaptive immunity, are located in the human immunoglobulin heavy chain locus (IGH). 67% of the ~1MB locus is composed of segmental duplications and 41% is composed of additional repeat elements. The locus remains under-characterized, due to the inability of standard high-throughput technologies to resolve the complexity, and despite increasing evidence supporting a role for IG polymorphism in the immune response. Therefore, the development of improved genomic resources and tools for more effectively studying IG genetic diversity is needed. As a critical first step, we have leveraged PacBio sequencing to generate high-quality haploid assemblies from fosmids spanning the IGH V gene cluster of an African (NA19240), East Asian (NA18555), and European (NA12878) individual, facilitating continued descriptions of novel germline variation, and the first direct comparison of locus-wide IGHV haplotypes from different ethnicities. A total of 4,307 single nucleotide variants (SNVs) were identified among these individuals, with varying numbers of unique SNVs per haplotype (range=105-948). 48% of the SNVs were not present in 1000 Genomes Project calls, including SNVs within coding and known IG regulatory regions, translating into the discovery of 16 novel IGHV alleles. Comparisons to hg19 and hg38 also revealed the presence of five large structural variants (SVs), ranging from 9.5-75 Kb. Considering both SNV and SVs, each of the characterized haplotypes were unique. The longest block of shared identical sequence between the individual haplotypes was 6.2 Kb, and per haplotype IGHV gene counts ranged from 42 to 55. Shared alleles among the haplotypes were observed for only 8 IGHV genes. Further demonstrating signatures of elevated diversity, intra-individual metrics revealed that locus-wide IGHV SNV densities within each haplotype were 2.2- to 4.2-fold higher than those observed outside of IGHV. Within IGHV, SNV densities were highest within segmental duplications, IGHV coding segments, and adjacent up/downstream non-coding regions. Together, our results provide a rich set of reference assemblies for the genomics community and indicate that germline haplotype diversity within the IGHV locus is among the highest in the human genome. Further studies are warranted to fully characterize the extent of IG germline variation at the population level and elucidate its role in Ab repertoire development and disease.

2282T

Isolated population from Runta, Boyacá: Findings related to founder effect of the nonsense variant of the HGSNAT gene. *L.J. Torres[1], HM. Velasco[2], Y. Sanchez[3], S. Santos[4], T. Vinasco[5].* 1) Torres Leandra. Master Human Genetics, Universidad Nacional De Colombia, Bogota, Bogota, Colombia; 2) Velasco Harvy M, MD, Master Human Genetics, Universidad Nacional De Colombia, Bogota, Bogota, Colombia; 3) Sánchez Yasmin, MD, Universidad Pedagogica y Tecnologica de Colombia, UPTC, TUNJA, Colombia; 4) Santos Sidney. PhD, Head, Laboratory of Human and Medical Genetics, Universidade Federal do Pará. Belém, Brazil; 5) Vinasco Tatiana, MS, Laboratory of Human and Medical Genetics, Universidade Federal do Pará. Belém, Brazil.

OBJECTIVE: To describe the phenomena of genetic isolate through the findings of the founder effect in a group of MPS III in Boyacá, Colombia. **METHODS:** Study of a large pedigree in Runta, Boyacá. Analysis of the pathogenic variant responsible for MPS III. Biochemical analysis for MPS IIIA/B. HGSNAT sequencing. Typification of 6 STRs adjacent to the variant found in 69 inhabitants of Runta and 78 healthy controls from Tunja. Haplotype analysis. Calculation of WHE. F-statistics. Carriers study on 198 settlers included in the group. Calculation of the time of the mutation onset through ESTIAGE and DMLE v 3.1. **RESULTS:** Seven related, alive and affected individuals were found on a large pedigree of 14 patients (7 deceased), detecting the 1360 C>T nonsense variant in HGSNAT, and thus confirming MPSIIIC. The ancestral haplotype of the mutation was present in 88.9% of Runta inhabitants with differences in Fst (0.16) between Runta and Tunja people. A very high frequency of heterozygotes (34%) was observed during the study of carriers. The average age of the ESTIAGE mutation was 41 (CI 95% 35-48) and 39 generations (CI 95% 35-46), which was estimated by using DMLE v 3.1. **CONCLUSIONS:** An ancestral haplotype was identified, although it was not found in the control population. This evidence supports the existence of a founder effect for the variant c.1360 C>T in the Runta village, and an approximate time of origin of the mutation of 1000 years. In consequence, this may be a pre Hispanic mutation, possibly of Muisca origin, that has managed to remain until the present day. This population could be classified as genetic isolate.

2283F

Local adaptation shaped the genetics of psychiatric disorders and behavioral traits in European populations. *R. Polimanti[1], M. Kayser[2], J. Gelernter[1].* 1) Yale University School of Medicine, West Haven, CT; 2) Erasmus University Medical Center, Rotterdam, Rotterdam, the Netherlands.

Recent studies have used genome-wide data to investigate evolutionary mechanisms related to behavioral phenotypes, identifying widespread signals of positive selection (i.e., variants with beneficial effects on individual fitness increase in population frequency). Here, we conducted a genome-wide investigation to study whether the genetics of these traits was affected by local adaptation (i.e., adaptation in response to selective pressure related to the local environment), performing a high-resolution polygenic risk score (PRS) analysis in 23 European populations with respect to variables related to geo-climate diversity (geographical coordinates, temperature, daylight, precipitation rate, and humidity), pathogen diversity (bacteria, protozoa, and virus), and language phonological complexity (consonants, segments, and vowels). The analysis was adjusted for the genetic variation of European populations to ensure that the differences detected would reflect differences in environmental exposures. The top finding was related to the correlation between winter minimum temperature and schizophrenia PRS (z = 3.84, p = $1.28*10^{-4}$). Additional significant geo-climate results were also observed with respect to bipolar disorder PRS (Maximum Sunny Daylight: z = -2. 93, p = $3.42*10^{-3}$), depressive symptoms PRS (Latitude: z = 3.47, p = $5.38*10^{-4}$), major depressive disorder PRS (Maximum Precipitation Rate: z = -3.21, p = $1.33*10^{-3}$), subjective well-being PRS (Minimum Relative Humidity: z = -2.95, p = $3.22*10^{-3}$). Beyond geo-climate variables, we observed three findings related to pathogen diversity and language phonological complexity: openness-to-experience PRS correlated with protozoan diversity (z = 3.56, p = $3.82*10^{-4}$) and language consonants correlated with conscientiousness PRS (z = -2.97, p = $2.98*10^{-3}$) and extraversion PRS (z = 2.87, p=$4.13*10^{-3}$). In summary, we report the first evidence regarding the role of local adaptation in shaping the genetic architecture of psychiatric disorders and behavioral traits in European populations. Specifically, our study identified several putative signals of local adaptation and many of them appear to be supported by known epidemiological and molecular data.

2284W

Detecting variation maintained by balance between recurrent mutation and selection in human populations. *N. Koelling[1,2], A.O.M. Wilkie[1], G. McVean[2], A. Goriely[1].* 1) MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom; 2) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, United Kingdom.

Highly penetrant *de novo* mutations (DNM) associated with certain congenital disorders, such as Apert syndrome, have been observed at levels up to 1000-fold higher than the background mutation rate and occur more commonly with increasing paternal age. A possible explanation for this is that these mutations confer a selective advantage to the spermatogonial stem cell in which they arise, leading - over time - to an enrichment of sperm carrying the mutation. However, the selfish spermatogonial selection (SSS) of these mutations is balanced by their deleterious phenotype in the offspring, resulting in their absence from healthy populations. We hypothesized that a similar balance of positive selection in the germline and negative selection on the organism may be acting on other sites, where the impact of a mutation may be less deleterious, and that we might detect these sites as hotspots of rare but recurrent mutations in population-level variation data. We developed a statistical method to identify sites with signatures of such recurrent independent mutation events, and validated it using simulations. Applying our method to data from the exome aggregation consortium (ExAC), we identified 343,962 candidate sites showing a signature of recurrent mutation. As expected, these recurrently mutated sites showed features such as an enrichment in C>T transitions at CpG dinucleotides, which are known to have elevated mutation rates. Other sites were associated with cell proliferation: recurrent mutation sites were found close to hotspots of cancer mutations from COSMIC (including hundreds of mutations at the exact same codon) and in genes involved in clonal hematopoiesis, such as *DNMT3A* and *JAK2*. In addition, we also recovered milder known selfish mutations, such as an *FGFR2* variant encoding p.S252L, which has been associated with Crouzon syndrome. Finally, these sites were also enriched in independent sets of DNM, confirming that we could identify locations recurrently mutated in the germline. We identified a subset of variant sites that may be exhibiting a balance between positive germline and negative organismal selection by accounting for nucleotide context, somatic mutations and technical artefacts. These included variants in genes associated with SSS as well as novel genes and thus support this strategy for the discovery of new candidate disease genes associated with increased paternal age.

2285T

New favored haplotype alleles in human adaptation to high altitude in Andes.

T. Stobdan[1], A. Akbari[2], P. Azad[1], D. Zhou[1], O. Poulsen[1], O. Appenzeller[3], G.F. Gonzales[4], A. Telenti[5,6], E.H.M. Wong[5], E.F. Kirkness[5], J.C. Venter[5,6], V. Bafna[7], G.G. Haddad[1,8,9]. 1) Division of Respiratory Medicine, Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA; 2) Department of Electrical & Computer Engineering, University of California, San Diego, La Jolla, CA 92093, USA; 3) Department of Neurology, New Mexico Health Enhancement and Marathon Clinics Research Foundation, Albuquerque, NM 87122, USA; 4) High Altitude Research Institute and Department of Biological and Physiological Sciences, Faculty of Sciences and Philosophy, Universidad Peruana Cayetano Heredia, Lima 31, Peru; 5) Human Longevity Inc., San Diego, CA 92121, USA; 6) J. Craig Venter Institute, La Jolla, CA 92037, USA; 7) Department of Computer Science & Engineering, University of California, San Diego, La Jolla, CA 92093, USA; 8) Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA; 9) Rady Children's Hospital, San Diego, CA 92123, USA.

Background: Human high altitude adaptation or mal-adaptation, such as chronic mountain sickness (CMS), is explored to understand the physiological and molecular mechanisms that underlie long-term (generational) exposure to hypoxia. This direct high altitude adaptation experiment-in-nature can be utilized to understand evolutionary processes in humans and to understand diseases where hypoxia plays part in their pathophysiology. **Results:** Here we report the results of an analysis of the largest whole-genome-sequencing of CMS and non-CMS individuals, identified candidate genes and functionally validated these candidates in a genetic model system (*Drosophila*). We used PreCIOSS algorithm that uses Haplotype Allele Frequency score to separate haplotypes carrying the favored allele from the non-carriers and accordingly prioritize genes associated with CMS or non-CMS. Haplotypes in eleven candidate regions were significantly different between CMS and non-CMS. A closer examination of individual genes in these regions revealed the involvement of previously identified candidates (e.g., *SENP1*) and also unreported ones *MC2R*, *SGK3*, *COPS5*, *PRDM1* and *IFT122* in CMS. The majority of SNPs were in non-exonic regions of these candidate genes. Remarkably, in addition to genes like *SENP1*, *SGK3* and *COPS5* which are hypoxia-inducible factor-dependent (HIF-dependent), our study revealed for the first time HIF-independent gene *PRDM1* and hormonal receptor gene *MC2R*, indicating an involvement of various systems and pathways in high altitude adaptation. Furthermore, we down-regulated the orthologs of each candidate gene in *Drosophila* and tested for hypoxia tolerance. We observed that down-regulating orthologs of *SGK3*, *COPS5*, *TCF54*, *PRDM1* and *IFT122* significantly enhanced their hypoxia tolerance. **Conclusions:** Our study reveals the involvement of both HIF-dependent and independent mechanisms in human adaptation to high altitude. The PreCIOSS algorithm applied on a large number of genomes identifies involvement of both new and previously reported genes in selection sweep, highlighting the involvement of multiple hypoxia response systems. Since the overwhelming majority of SNPs are in non-exonic (and possibly regulatory) regions, we speculate that this molecular adaptation allows for more flexibility in adjusting the physiological responses to environmental challenges such as hypoxia.

2286F

Genetic structure of pre-Columbian remains of the Malambo culture in the Lower Magdalena, Colombia.

M. Noguera[1], A. Gómez[1], I. Briceño[1,2], J. Bernal[3], W. Rivera[4], J. Martín[4]. 1) Bogotà, Colombia, Institute of Human Genetics, Faculty of Medicine, Pontificia Universidad Javeriana, Bogotá D.C, Colombia Kra 7 ·40-62; 2) Grupo Genética Humana.Facultad de Medicina,Universidad de la Sabana. Chía, Colombia; 3) Universidad Tecnológica de Bolívar. Cartagena, Bolívar, Colombia; 4) Grupo Memorias del Caribe, Facultad de Antropología.Universidad del Norte. Barranquilla, Atlántico. Colombia.

Introduction: Ancient DNA studies, allow determining population dynamics of the pre-Columbian cultures and clarifying the genetic continuity and discontinuity. In the Malambo culture in the department of Atlántico, preliminary findings of pottery and lithic objects, can be complemented with the genetic study of human remains and their relationship with the archeology of Northern South America and the Caribbean. **Objective:** To study the population structure of the comunities of Lower Magdalena and the Caribbean, based on bioarchaeological analysis and ancient DNA extracted from skeletal remains supplied by the MAPUKA Museum, at the Universidad del Norte in Barranquilla, Colombia. **Materials and Methods:** A bioarchaeological (genetic-archaeological) study of the Malambo tradition was carried out. The hypervariable-I region of mitochondrial DNA was analyzed after being successfully amplified and sequenced, Sequences were analized using CLC main Workbench software: Multiple sequence alignment with the revised Cambridge reference Sequence. HVR-I segments were aligned to the concensus sequence. Arlequín 3.5.2 was used to analize the intrapopulations and interpopulations levels. Phylogenetic Networks were constructed through the Median Joining (NJ) algorithm, comparing the study samples with other paleo-Colombians. **Results:** Only one haplogroup was observed in four individuals, in which the G16361A mutation was found. This mutation was recently described to define sub-haplogroup B2j) in Venezuela, which suggests a genetic continuity in Northern South America. **Conclusión:** Identification of subhaplogroup B2j in the individuals belonging to the pre-Columbian population of Malambo and recently reported in a contemporary inhabitant of the Venezuelan coast, could support a human migration route to or from the Orinoco river valley.

2287W

Comparative whole genome variation and architecture in a pediatric African American and Caucasian dataset. *P. Sleiman, L. Tian, H. Hakonarson.* Center Applied Genomics, CHOP, Philadelphia, PA.

Admixed populations, including African Americans, remain underrepresented in genetic research. We have generated whole genome sequencing data on 900 pediatric samples comprising 532 children of African American ancestry and 368 of European ancestry representing one of the largest pediatric and African American whole genome datasets generated to date. All samples were sequenced to a depth of ~30X using a PCR free library prep on an Illumina X10. Alignment to GRCh37 and variant calling was carried out using the DRAGEN cloud (Edico Genome). Variant annotations were generated using SnpEff, and Variant Effect Predictor (VEP) for RefSeq and Ensembl GENCODE gene models in addition to ANNOVAR outputs for the UCSC knowGene models. Structural variants (SVs) were called using MANTA (Illumina) and copy number variants (CNVs) using CANVAS (Illumina). Increased single nucleotide variation has been well established in African genomes, however, to our knowledge no large scale studies have systematically estimated the burden of variation in admixed populations taking local ancestry into account. Having generated local ancestry estimates across the admixed genomes we present a local ancestry-based overview of the distribution and burden of rare variants, SVs and CNVs in the African and European segments in the admixed AA individuals as well as a global comparison with the European ancestry samples in our cohort.

2288T

***GCH1* plays a role in high altitude adaptation of Tibetans.** *Y. He[1,3], Y. Guo[1,9], C. Cui[2], N. Ouzhuluobu[2], N. Baimakangzhuo[2], N. Duojizhuoma[2], N. Dejiquzong[2], N. Bianba[2], Y. Peng[1], C. Bai[2], N. Gonggalanzi[2], Y. Pan[2], N. Qula[2], N. Kangmin[2], N. Cirenyangji[2], N. Baimayangji[2], W. Guo[2], N. Yangla[2], H. Zhang[1], X. Zhang[1], W. Zheng[1,9], S. Xu[4,7,8], H. Chen[5], S. Zhao[9], Y. Cai[9], S. Liu[6], T. Wu[6], X. Qi[1], B. Su[1].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; 2) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa 850000, China; 3) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100101, China; 4) Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, CAS, Shanghai 200031, China; 5) Center for Computational Genomics, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China; 6) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining 810012, China; 7) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; 8) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China; 9) College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China.

Tibetans are well adapted to high-altitude hypoxia. Previous genome-wide scans have reported many candidate genes. But only a few of these candidates were studied. In this study, we reported a hypoxia gene (*GCH1*, GTP-cyclohydrolase I), involved in maintaining NOS function and normal blood pressure, harbors many potentially adaptive variants in Tibetans. We re-sequenced an 80.8kb fragment covering the entire gene region of *GCH1* in 50 unrelated Tibetans. Combined with the published data, we demonstrated that there were many *GCH1* variants showing deep divergence between highlander Tibetans and lowlander Han Chinese. Neutrality tests confirmed the signal of Darwinian positive selection on *GCH1* in Tibetans. Moreover, the association analysis indicated that the Tibetan version of *GCH1* is significantly associated with multiple physiological traits in Tibetans, including blood nitric oxide concentration, blood oxygen saturation and hemoglobin concentration. Taken together, we propose that *GCH1* plays a role in the genetic adaptation of Tibetans to high altitude hypoxia.

2289F

Improving genotype imputation in population isolates using identity by descent. *M. Abney, A. El Sherbiny.* Dept Human Gen, Univ Chicago, Chicago, IL.

Genotype imputation, using standard imputation software, for a sample from an isolated population can result in unexpectedly high error rates in certain classes of markers. In particular, using an external reference panel can result in a high rate of false positive calls while also failing to impute the potentially large number of private variants. Even though sequencing a strategically chosen internal reference panel can significantly improve accuracy, we find that there are still often a subset of markers for which imputation accuracy is low. Here we present a method that leverages the identity by descent information captured in the framework set of markers to further improve imputation quality, often dramatically for some sets of markers. We also provide a method for optimally selecting the individuals for the internal reference panel. Unlike other approaches the methods presented here work equally well when individuals are of known or unknown relatedness (i.e. cryptically related) and whether there is a known pedigree or not. Our imputation method can be used on its own or can integrate the output from standard imputation methods to provide the highest imputation accuracy. We apply our method to real genotype data from an isolated population and find, for instance, that heterozygous concordance rates improve significantly, and that, when used together with a method such as IMPUTE2 using 1000 Genomes reference data, our method outputs imputed genotypes that are as accurate as if an internal reference panel were also used. That is, without sequencing an internal reference panel our method obtains imputation quality approximately as accurate as if an internal reference panel had been sequenced.

2290W

Consanguinity, IBD, and ROH: Detangling their complex relationship. *A. Severson[1], N. Rosenberg[2].* 1) Genetics, Stanford University, Stanford, CA; 2) Biology, Stanford University, Stanford, CA.

Identity by descent (IBD) and runs of homozygosity (ROH) are two measures commonly used to describe genomic sharing. IBD evaluates sharing across individuals while ROH evaluates sharing within an individual. Empirical estimates of IBD and ROH have revealed that these measures are correlated. Moreover, it has been shown that increased quantities and lengths of IBD and ROH tracks correlate with increased consanguinity. It is not surprising that higher levels of inbreeding produce more ROH, but the increased IBD within the population is less intuitive. To investigate the effect of consanguinity on IBD and ROH, we first consider the effect on time to the most recent common ancestor (TMRCA). We employ a dioecious, diploid Wright Fischer model with different consanguinity regimes. We model sibling mating, cousin mating, double cousin mating, and uncle-niece mating. For each consanguinity regime, we are able to model the effect of the fraction of consanguineous mating pairs on TMRCA in the entire population. We find that with differing magnitudes that increasing consanguinity decreases TMRCA. This is not only true for the products of consanguineous unions, but also for more generally "unrelated" individuals in the population. Moreover, this result recapitulates empirical findings of the correlation between consanguinity, ROH, and IBD. Next, we will run simulations to explicitly extend the findings from TMRCA to ROH and IBD.

2291T

Evolutionary rate and the human transcriptome: Function, mechanism, and disease genes. *P. Evans, N. Cox, E. Gamazon.* Gen Med, Vanderbilt University, Nashville, IL.

The development of mechanistic explanatory models of protein sequence evolution has broad implications for our understanding of cellular biology, population history, and disease process. Several molecular features have been proposed as potential causal factors for the rate of protein evolution. Here I analyze a variety of gene properties, including expression level, expression breadth, gene age, network interaction, and heritability, and quantify their contribution to protein sequence divergence. We find that many factors contribute independently and we present strong evidence for the importance of gene function. Most notably, co-expression network connectivity and gene expression heritability provide the best univariate models for protein evolutionary rate. Heritability, a measure of the influence of the regulatory genome, is a novel and, important determinant of evolutionary rate. Furthermore, we use the GTEx human transcriptome resource to examine the forces that shape protein evolution in a multi-tissue framework; in particular, analysis of gene expression using this resource. We implement a self-organizing map (SOM) neural network to jointly analyze all examined potential factors and to account for nonlinear effects, utilizing it to develop a dispensability score for each gene. Methodologically, we demonstrate that genome-wide studies of disease may be enhanced by incorporating the determinants of evolutionary rate into the search for disease genes. Our study presents a comprehensive analysis of a range of factors that constrain molecular evolution and proposes a novel framework for the study of protein function and disease mechanism.

2292F

EP300 contributes to high altitude adaptation of Tibetans by regulating nitric oxide production. *X.B. Qi[1], W.S. Zheng[2], Y.X. He[1,4], C.Y. Cui[3], NA. Ouzhuluobu[3], NA. Dejiquzong[3], Y. Peng[1], C.J. Bai[3], NA. Duojizhuoma[3], NA. Gonggalanzi[3], NA. Bianba[3], NA. Baimakangzhuo[3], Y.Y. Pan[3], NA. Qula[3], NA. Kangmin[3], NA. Cirenyangji[3], NA. Baimayangji[3], W. Guo[3], NA. Yangla[3], H. Zhang[1], X.M. Zhang[1], Y.B. Guo[2], S.H. Xu[5,8,9], H. Chen[6], S.G. Zhao[2], Y. Cai[2], S.M. Liu[7], T.Y. Wu[7], B. Su[1].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; 2) College of Animal Science and Technology, Gansu Agricultural University, Lanzhou 730070, China; 3) High Altitude Medical Research Center, School of Medicine, Tibetan University, Lhasa 850000, China; 4) Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China; 5) Chinese Academy of Sciences Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, CAS, Shanghai 200031, China; 6) Center for Computational Genomics, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China; 7) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining 810012, China; 8) School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China; 9) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China.

The genetic adaptation of Tibetans to high altitude hypoxia likely involves a group of genes working in the hypoxic pathway, as suggested by previous studies. To test the adaptive role of the previously reported candidate gene *EP300* (histone acetyltransferase p300), we conducted resequencing of a 29.83kb gene region of *EP300* in 80 unrelated Tibetans. The allele-frequency and haplotype-based neutrality tests detected signals of Darwinian positive selection on *EP300* in Tibetans, with a group of variants showing allelic divergence between Tibetans and lowland reference populations including Han Chinese, Europeans and Africans. Functional prediction suggested the involvement of multiple *EP300* variants in gene expression regulation. More importantly, genetic association tests in 226 Tibetans indicated significant correlation of the adaptive *EP300* variants with blood nitric oxide (NO) concentration. Collectively, we propose that *EP300* harbors adaptive variants in Tibetans, which may contribute to high altitude adaptation through regulating NO production.

2293W

De novo genomic assembly and assessment of inbreeding in a severe bottleneck population. *A.F. Scott, D.W. Mohr.* Inst Gen Med, Blalock 1034, Johns Hopkins Med Sch, Baltimore, MD.

Population bottlenecks lead to small populations with increased inbreeding and random genetic drift, the loss of genetic variability, reduced reproductive success, and higher mortality as a result of a poorer ability to respond to environmental insults (e.g., infectious diseases). Once species enter "extinction spirals" their populations continue to decrease until they can no longer be sustained. As biodiversity in natural populations decreases world-wide due to climate change, habitat loss, illegal animal trade and other reasons the reduction of genetic plasticity will have wide-spread consequences.   We studied the aftermath of a population bottleneck in the Hawaiian monk seal (Neomonachus schauinslandi), a species with approximately 1400 animals, whose population declined to, perhaps, fewer than 100 individuals in the 1890s as a result of over hunting. In 1983, the National Oceanic and Atmospheric Administration (NOAA) launched a vigorous species recovery effort. We created a high-quality *de novo* genome assembly using a combination of optical mapping and linked-read sequencing (Mohr et al, 2017) in which we identified 98% of the expected protein coding genes in 170 scaffolds whose median length (N50) is over 29 Mb. The linked-read sequencing produced phased haplotype blocks that could be compared to the frequently sequenced human NA12878 "gold standard" genome. As expected, the Hawaiian seal has significantly reduced heterogeneity that is approximately ten percent of that found in humans. Although a higher than expected number of congenital abnormalities has not been observed in the wild we did note marked reduction of heterogeneity within the MHC genes in the seal, suggesting that these animals may be at a heightened risk of decline or extinction due to pandemic disease.   Reference: Mohr et al http://biorxiv.org/cgi/content/abstract/128348v2.

2294T

Dietary adaptation of *FADS* genes varied geographically in Indian populations. *D. Wang[1], K. Ye[1], O. Bar-Yosef[2], A. Keinan[1,3,4].* 1) Computational Biology and Biological Statistics, Cornell University, Ithaca, NY; 2) Department of Anthropology, Harvard University, Cambridge, MA; 3) Cornell Center for Comparative and Population Genomics, Cornell University, Ithaca, NY; 4) Center for Vertebrate Genomics, Cornell University, Ithaca, NY.

The fatty acid desaturase (*FADS*) genes encode rate-limiting enzymes for the biosynthesis of omega-6 and omega-3 long-chain polyunsaturated fatty acids (LCPUFAs) from plant-derived precursors. While LCPUFAs can be absorbed directly from meat/seafood-based diets, their biosynthesis is essential to compensate for their absence in plant-based diets. We have previously reported geographically and temporally varying patterns of *FADS* genetic adaptation to diet in Europe, with one haplotype (D) adaptive to plant-based diets in farmers by enhancing LCPUFAs biosynthesis and another haplotype (M2) adaptive to meat/seafood-based diets in pre-Neolithic hunter-gatherers by diminishing LCPUFAs biosynthesis. In South Asia, while we have reported adaptation of *FADS* genes to plant-based diets, it is still unknown whether the diverse demographic history and dietary habits among subgroups of Indians resulted in geographically varying adaptive patterns. Here, we compiled a large data set representing 52 populations across India and applied imputation to create a uniform set of variants. First, we observed drastic frequency variation across geography: opposite frequency patterns between Andaman Islanders (D: 3.13%; M2: 50.0%) and mainland Indians (D: 84.3%; M2: 5.75%); even among mainland Indians, the frequency of D varies, ranging from 66.6% in the Kharia population to 94.4% in the Kurumba population with a general trend of higher frequency in western Indians than in eastern ones. Moreover, we found that the frequencies of haplotype D in present-day Indian populations are significantly associated with the percentage of ancestral north Indian (ANI) ancestry ($p = 0.031$), and with their language families ($p = 0.041$). Formal haplotype-based selection tests reveal the presence of geographically varying selection signals while considering different demographic history. Overall, the geographically varying patterns of allele/haplotype frequency in South Asia are consistent with the geographically varying dietary patterns: the meat/seafood-based diets of Andaman Islanders, and the plant-based diets in mainland Indians with western Indians relying more heavily on plant and eastern Indians on seafood.   .

2295F

Detect co-evolution of genes in admixed populations with genome-wide data. *M. Zhou[1], H. Wang[2], X. Li[1], S. Redline[2], H. Tang[3], X. Zhu[1].* 1) Case Western Reserve University, Cleveland, OH; 2) Harvard Medical School, Boston, MA; 3) Stanford University, Stanford, CA.

Co-evolution of genes is an important topic in population genetics. However, detecting co-evolution in human population is often challenged owing to the experimental constraints. Our previous study suggested such co-evolution can be detected by testing the correlation of local ancestry in recently admixed populations such as African Americans or Hispanic populations. However, our previous study only examined co-evolution occurred on different chromosomes. Within a chromosome, a challenge of detecting co-evolution is to control the confounding due to admixture linkage disequilibrium (LD). In this study, we developed a method to eliminate the bias caused by admixture LD when searching for co-evolution on the same chromosome. We applied this method to search for co-evolution in three African American cohorts: (1) the Candidate Gene Association Resources (CARe) study; (2) the Family Blood Pressure Program (FBPP); (3) the Women's Health Initiative (WHI), with a total of 16,252 unrelated African Americans. We tested pairwise correlations of local ancestry on each chromosome for each of the three cohorts while excluding those pairwise loci with genetic distance less than 70cM in order to eliminate the high correlation because of admixture LD. The genomic control values were 0.90, 0.82 and 0.81 for CARe, FBPP, and WHI, respectively, suggesting our method leads to a conservative type I error rate. We next meta-analyzed the results from the three cohorts. The genomic control value for the meta-analysis was 1.03. Our results revealed 11 significant pair regions on multiple chromosomes after correcting for multiple tests (P-value<3e-6). Those pair regions located multiple genes that have been previously reported under natural selection pressure. We observed *LRRC7* was negatively correlated with Duffy blood locus *ACKR1,* and *DIP2C* was negatively correlated with *PCDH15*, a gene with a key role in development and function of sensory hair cells. Our analysis suggests our method is powerful to identify co-evolution in recently admixed populations.

2296W

Where is Brazil? Placing admixed Brazilian populations in a global genetic map. *A. Arcanjo Silva[1,2,3], S. Oliveira[2], N. Klautau-Guimaraes[2], J. Walker[3], M. Batzer[3].* 1) Colégio Marista de Brasília Ensino Médio, Brasília, Distrito Federal, Brazil; 2) Departamento de Genética e Morfologia, Universidade de Brasília, Brasília, Distrito Federal, Brazil; 3) Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, United States.

The state of the art of the knowledge about the relationship between genetic variability and the dynamics of human populations is a result of years of describing allelic and genotypic frequencies. The patterns of variation for genetic markers in coding and non-coding regions are not the same, since mutations in coding regions might have significant impact in one's ability to survive. These different patterns can help clarify the relationship between populations. Many markers were described that could efficiently identify samples with European ancestry from those with African ancestry and Asian ancestry. In a globalized world, where migrations occur frequently, genetic constitution and ancestry is not so well established. In some populations, especially the ones originated by recent colonization such as Brazil, the genetic constitution is a mix of the genetic constitution of its founders. The main goal of this work is to estimate the genetic variability that is accessed by neutral genetic markers, which describe the variability relationships among population groups in an unbiased way, in two knowingly admixed Brazilian populations, placing Brazil in a global genetic map. To achieve that, a panel of 100 Alu insertions was used in a set of 1125 worldwide samples, including 160 samples from two admixed Brazilian populations, Brasília and Kalunga. The insertion panel showed a tenfold increase in the number of linked loci to each locus in Brasilia, Kalunga and Utah when compared to African, European, Asian and Indian populations. Which is mainly due to the recent admixture of different populations in the making of those populations (250 years). Kalunga clusters with the African populations, but still shows differentiation due to being admixed, Brasilia clusters with both the European and the Middle-Eastern groups, being the last one more genetically similar to Brasilia than the first due to an African genetic component. The population groups cannot be considered homogeneous, despite at least one population of each group not showing population differentiation to the other populations of its group. In conclusion, Brazilian admixed populations tend to cluster with the populations in which the genetic constitution is more similar to their own: instead of clustering with African or European populations, Brasilia clustered with other admixed populations with similar genetic founders, while Kalunga clustered with African populations due to its great African ancestry.

2297T

HLA-G and _HLA-A_ extended haplotypes in a Brazilian population sample: The close relationship between _HLA-G_ promoters and _HLA-A_ coding alleles. *E.C. Castelli[1,2], M.A. da Paz[1], I.O.P. Porto[1], A.S. Souza[1], T.H.A. Lima[1], L.C. Veiga-Castelli[3], E.A. Donadi[4], C.T. Mendes-Júnior[5].* 1) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, School of Medicine, Botucatu, State of São Paulo, Brazil; 2) São Paulo State University (UNESP), School of Medicine, Pathology Department, Botucatu, State of São Paulo, Brazil; 3) University of São Paulo (USP), School of Medicine of Ribeirão Preto, Department of Genetics, State of São Paulo, Brazil; 4) University of São Paulo (USP), School of Medicine of Ribeirão Preto, Department of Medicine, State of São Paulo, Brazil; 5) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, SP, Brazil.

The human leukocyte antigen G (_HLA-G_) is a non-classical HLA class I gene which presents limited protein variability and restricted expression to certain tissues. HLA-G has immunomodulatory properties, inhibiting immune responses when interacting with specific Natural Killer and T cell receptors, such as KIR2DL4, ILT2 and ILT4. On the other hand, the _HLA-A_ locus is one of the most variable genes of the human genome and its encoded molecules are expressed in nearly all nucleated cells. HLA-A associates with endogenous peptides and presents then on the cell surface to T CD8+ cells. Balancing selection has been well documented for the _HLA-A_ coding segment, mainly at the region encoding the peptide-binding site, which is associated with the presence of many frequent and divergent coding alleles in worldwide populations. Likewise, many studies have detected signatures of balancing selection at the _HLA-G_ promoter. However, it is not clear whether balancing selection is indeed operating at the _HLA-G_ locus, or if it is a hitchhiking effect caused by the selective pressures acting on its neighbor gene, _HLA-A,_ since they are separated by approximately 100Kb. Here we evaluate the _HLA-A_ and _HLA-G_ variability in 405 individuals from the Southeast Brazil. Each _locus_ was entirely evaluated by using Next Generation Sequencing (NGS), and haplotypes were determined by a bioinformatics workflow that combines the physical associations detected directly at the NGS data and the computational inference from PHASE algorithm. _HLA-A_ and _HLA-G_ haplotypes were then combined as extended haplotypes. The results indicate that _HLA-G_ and _HLA-A_ are in strong linkage disequilibrium. For instance, 94% of the A*01 alleles are associated with the _HLA-G_ promoter 010102a, and 95% of the A*23 alleles are associated with the _HLA-G_ promoter 0104a. Although some _HLA-A_ alleles may occur with different _HLA-G_ promoters, each main _HLA-G_ lineage (G010101, G010102, G0103 and G0104) is primarily associated with a different set of _HLA-A_ coding alleles. Both the _HLA-G_ promoter and the _HLA-A_ coding segment present signatures of balancing selection, which is stronger for the _HLA-A_ coding region. Thus, it is possible that the balancing selection operating at the _HLA-A locus_ is also shaping the _HLA-G_ promoter frequencies, downplaying its direct role on _HLA-G_. This work was supported by FAPESP/Brazil (#Grant 2013/17084-2) and the Brazilian Council for Scientific and Technological Development (CNPq/Brazil).

2298F

Rewriting the genetic history of Austro-Asiatic communities in South and Southeast Asia. *M.E. Phipps[1], D. Tagore[2], F. Aghkhanian[1], R. Naidu[1], P. Majumder[2], A. Basu[2].* 1) Jeffrey Cheah School of Medicine & Health Sciences, Monash University Malaysia, Sunway City, Selangor, Malaysia; 2) National Institute of Biomedical Genomics, Kalyani, West Bengal, India.

There are large gaps in our knowledge about the peopling of India and South East Asia by AustroAsiatic (AA) language speakers. Today, the majority of AA speakers in Indochina are agriculturalists, and it has been assumed that the spread of rice cultivation is inextricably linked to the spread of the AA language family. However, it is intriguing that tribal communities such as the Gond, Ho, Santal, Bihor, Korwa and Khasi in India, and indigenous Negrito and some Senoi tribes in Malaysia who are all Austrosiatic speakers, represent the last hunter gatherer communities with negligable evidence of knowledge or practice of rice cultivation. We believe that ours is a larger and more comprehensive investigation to elucidate the genetic history and dispersal of AA speakers in this region. We genotyped 496 (Illumina Omni 2.5 and Affymetrix 6) unrelated individuals from eleven AA communities and neighbouring Austronesian, Tibeto Burman and Dravidian groups in Malaysia and India. Our genotype data was then compared with the HGDP dataset for further analysis. Contrary to our initial expectations, Principal Component Analysis and ADMIXTURE revealed that all AA speakers of India were more closely linked to the Senoi, rather than Negritos in Malaysia. In addition there was an appreciable association between these Senoi and Tibeto Burmans. Identical by descent (IBD) and Treemix analyses suggested that in earlier times, the AA of Malaysia and Tibeto Burmans may have initially split from a common ancestral population. It is probable that more recently, some AA groups in Southeast Asia migrated westwards giving rise to the present day Indian AA speakers. D-statistics that was performed provide evidence of extensive gene flow from East Asians (Han Chinese, Dai, Cambodians) towards AA Malaysia and Tibeto Burmans, representing south western migration events. We have not found evidence to support a previous report that linked rice cultivation and the spread of Austroasiatic languages. Therefore, we propose that later migrations and multiple gene flow events within the past two millennia between Austroasiatic, Sino-Tibetan and Dravidian speakers in South and Southeast Asia, have shaped the present day language and genetic structures of populations in this part of the world.

2299W

Genetic differentiation of Hispanics using ancestry informative markers.
C.H. Setser, R. Chakraborty, J.V. Planz, R. Barber, A. Eisenberg, D.S. Cross.
Molecular and Medical Genetics Department, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107.

   There are at least 10,500 unidentified human remains in the US as of August 2015, with 2,041 of presumed Hispanic origin (NamUs 2015). Conventional DNA analysis identifies an individual through comparison with reference profiles. For those with no reference, panels of ancestry informative single nucleotide polymorphisms (SNPs) exist (Kidd 2014, Seldin 2009), but they focus on global differentiation and are not suitable for ancestry determination of admixed populations (e.g. Hispanics). We hypothesize that a small panel of SNPs ascertained from appropriate populations with great genetic differentiation can distinguish ancestry within Hispanic populations. This bioinformatics study uses the Genomic Origins and Ancestry in Latinos (GOAL) dataset of 250 individuals with ancestry from Columbia, Cuba, Dominican Republic, Haiti, Honduras, or Puerto Rico genotyped using the Affymetrix 6.0 chip to develop an informative Hispanic SNP panel. Starting with 897,336 SNPs, we trimmed to 531,878 SNPs using linkage disequilibrium of 0.7. We then calculated pairwise F_{ST} for each SNP with each population pair using PLINK (Haiti excluded). SNPs that met 0.15 threshold for the four comparisons were included in a 1217 SNP panel. We used STRUCTURE to visualize population separation. To determine if a smaller SNP set could be utilized while retaining information we only used the top ten mean F_{ST} values from each population and added an extra five to try to distinguish Cuba vs. Dominican Republic for a condensed panel of 56 SNPs. Additionally, we combined 1000 Genomes data (108 Yoruba, 107 Iberian, 103 Chinese) and the GOAL data to verify whether the countries are separating by ancestral lines or geographic region. STRUCTURE analysis showed Honduras was easily distinguished from other countries in the 1217 and 56 SNP panels. Other countries were also separated based on contribution from ancestral populations, however the separation was less than ideal. Notably, Honduras contributed 71% of the SNPs in the 1217 panel. When analyzed with 1000 Genomes data, Honduras separated with the Chinese population for K=1-3, but was the first GOAL population to separate from the ancestral line. Utilizing an efficient SNP panel consistently separated Honduras from other populations demonstrating proof of concept. Greater separation of country of origin may be seen with a larger dataset and additional substructure information unveiled by principal components analysis.

2300T

One step for study of transposable element-associated structural variations (TASVs) using de novo assembled Korean genome. *S. Mun[1], S. Kim[1], W. Lee[1], K. Kang[2], H.S. Kim[3], K. Han[1].* 1) Department of Nanobiomedical Science, Dankook University, Cheonan-si, South Korea; 2) Department of Microbiology, Dankook University, Cheonan-si, Republic of Korea; 3) Department of Biological Sciences, Pusan National University, Busan 609-735, Republic of Korea.

   Due to advanced sequencing technology, personal genome sequencing is currently possible and indeed many human genomes have been sequenced by next-generation sequencing (NGS) technology. The comparison of the individual genomes revealed a huge genomic difference between human populations and even individuals from closely related ethnic groups. There are many causative agents for the genomic changes. Transposable elements (TEs) are known as one of major factor to cause genomic variations through various mechanisms including *de novo* insertion, insertion-mediated deletion, and TE-TE recombination-mediated deletion. In this study, we carried out *de novo* whole-genome sequencing with multiple insert-size libraries and phasing analysis using statistically aided, long-read haplotyping (SLRH) from one Korean individual (KPGP9). The *de novo* genome assembly results in 31,305 scaffolds with a scaffold N50 size of 13.23Mb. With sequencing results, we performed variant annotations of KPGP9 genome and comparison of non-synonymous SNPs to the seven individual human genomes (KPGP9, AK1, SJK, YH, Yoruba, Venter, and Watson). We identified 97 SNPs that only detected from the Korean individual genomes. Further, through computational data analysis and experimental verification, we revealed that 182 TE-associated structural variation (TASV) insertions and 89 TASV deletions contributed to 64,232 bp (NCAI: -3369 bp) sequence gain and 82,772 bp sequence loss in the KPGP9 genome, respectively. Biological function and clinical implication were annotated for SNPs and TASVs identified. As demonstrated in this study, we release another Korean *de novo* genome and propose Korean common genetic differences by comparison to several ethnic groups. Our finding highlights again the role of TEs as one of major drivers to create structural variations in human individuals.

2301F

Haplotype map of Russian population. *I.V. Evsyukov[1], D.V. Zhernakova[1,2], S.V. Malov[1], S. Kliver[1], N. Cherkasov[1], G. Tamazian[1], P. Dobrynin[1], A. Gorbunova[1], S. Kolchanova[1], A. Shevchenko[1], M. Rotkevich[1], A. Yurchenko[1], S. Sidorov[1], A. Komissarov[1], K. Krasheninnikova[1], A. Logachev[1], S. Simonov[1], D.E. Polev[3], A. Glotov[3], A. Novozhilov[4], V. Brukhin[1], S.J. O'Brien[1].* 1) Theodosius Dobzhansky Center for Genome Bioinformatics, Saint-Petersburg State University, St.Petersburg, Russian Federation; 2) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands; 3) Research Resource Center for Molecular and Cell Technologies, Research Park, Saint-Petersburg State University, St. Petersburg, Russian Federation; 4) Department of Ethnography and Anthropology, Saint-Petersburg State University, St. Petersburg, Russian Federation.

For the last few years a number of national genomic projects emphasized the diversity and similarity of world populations, shed light onto several aspects of population history and provided substantial assistance to the development of medical genomics. In spite of occupying almost 12% of the world's landmass, and having the ninth largest world population until recently the Russian Federation has not participated in large sequencing projects. To overpass this, Genome Russia project was launched striving to accomplish whole genome sequencing of more than 2,000 samples coming from diverse population groups and ethnic minorities within the Russia. One of the main aims of this project is to construct a haplotype map (HapMap) of the Ethnic Russian for further usage in medical studies.   Here we introduce the preliminary Russian haplotype structure based on 41 Western Russian samples (consisting of 12 trios and 5 unrelated individuals) from Novgorod region and Pskov region. We inferred haplotypes from SNP genotypes by using SHAPEIT2 tool [1] (using trios in the analysis allowed us to increase the haplotype phasing quality). Haplotype structure analysis was performed in Haploview [2]. We were able to analyze linkage disequilibrium and haplotype block structure, performed haplotype population frequency estimation and identified tag SNPs relevant to Russian population.   These results will allow us to better understand the Russian genome and to create population-specific genotyping microarrays for medical and other studies.   1. J. O'Connell, D. Gurdasani, O. Delaneau, et al. A general approach for haplotype phasing across the full spectrum of relatedness. 2014 2. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005.

2302W

Ancestry Hub: For whole-genome local ancestry analysis. *S. Jiang, L. Ma, Q. Song.* 4DGenome, INC, Atlanta, GA.

Population stratification is a growing concern in genetic-association studies. Admixture has created mosaic chromosomes in human populations; even within the same individual, different segments and different homologous chromosomes may have different ancestral origins. Averaged ancestry at the genome level (global ancestry) is insufficient for detecting population substructures and correcting population stratifications in association studies. Recently, we have reported a new approach for local ancestry analysis, which has been implemented into a software tool called aMAP (ancestry of Modern Admixed Populations). Here, we report that we build up a web-based platform of ancestry hub to do the ancestry analysis. Data analysis requires access to powerful computational infrastructure, high quality bioinformatics software, and personal skills to operate the tools. To solve this challenge, we present a practical solution to the local-ancestry data analysis. We will provide the web tool for using aMAP in ancestry analysis. Our reference panel will contain haplotypes from the 1000 Genomes Project. In the future, we envisage an increase of the reference panels in diversity of reference populations. We setup Hub Online at Amazon storage and S3 based storage to provide easy access to the genotype data and to seamlessly transfer the input data into our local server for analysis. Now it can analyze with Th26 reference populations. The ancestry hub allows users to upload their own genotype data or haplotype data. Currently, the hub supports VCF or 23andme format. Our method also exhibits a tolerance to missing ancestral reference panels, an applicability to genetically close populations, and a capacity for analyzing multi-way admixed individuals. The web-based hub is open source and free for non-commercial users through the link https://www.ancestry.4dgenome.com/.

2303T

Relationships between the regulation of gene expression, mutational burden, and recombination in a large population cohort. *H. Edgington[1,2], I. Alves[1,2], P. Awadalla[1,2].* 1) Informatics and Biocomputing, Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

Hill-Robertson interference refers to a process where, in the absence of recombination, beneficial alleles arising on different haplotypes will compete for selective advantage, each preventing the other from reaching fixation. Thus, a proposed advantage of recombination is its utility in bringing advantageous mutations together on one haplotype. In this same manner, deleterious mutations in regions of low recombination may not be purged as efficiently when they exist on different haplotypes. This model describes a purely genetic framework, and does not take into account the activity of genes in which mutations may accumulate. In this study we explore what the relationship is between expression (and consequential exposure to natural selection) and recombination in order to integrate gene action into the Hill-Robertson framework. In order to compare expression differences while accounting for among-individual biases we made use of a phenomenon known as allele-specific expression (ASE), wherein the two copies of the genome at a particular locus are expressed differently. This provides for a within-individual comparison of allelic expression, which controls for the influence of external variation on expression. Additionally, previous work has shown an enrichment of ASE at deleterious variants across the genome, causing those alleles to be downregulated. This suggests the potential for ASE to act as a mechanism whereby individuals are protected from the effects of deleterious mutations. We hypothesized that, similar to the impact of Hill-Robertson interference on population allele frequencies, we might see a positive correlation between linkage disequilibrium among deleterious mutations and the degree of allele-specific expression within a transcribed region. We used a simulation-based approach to estimate allele-specific expression and linkage disequilibrium in data from CARTaGENE, a population cohort of French Canadians. We compared these parameters in order to determine whether haplotypes carrying more deleterious mutations were expressed with lower frequency than more neutral haplotypes, as well as whether linkage among deleterious mutations correlated with the magnitude and consistency of allele-specific expression.

2304F

Tracing maternal lineage of Austronesian-speaking Melanesians and Micronesians in the Solomon Islands. *M. Isshiki[1], I. Naka[1], R. Kimura[2], T. Furusawa[3], K. Natsuhara[4], T. Yamauchi[5], M. Nakazawa[6], T. Ishida[1], R. Ohtsuka[7], J. Ohashi[1].* 1) Department of Biological Sciences, Graduate School of Science, The University of Tokyo; 2) Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus; 3) Graduate School of Asian and African Area Studies, Kyoto University; 4) Faculty of Nursing, The Japanese Red Cross Akita College of Nursing; 5) Faculty of Health Science, Hokkaido University; 6) Graduate School of Health Sciences, Kobe University; 7) Japan Wildlife Research Center,.

Accumulated archaeological, linguistic and genetic evidences suggest that modern Austronesian (AN)-speaking Melanesians are derived from the admixture of indigenous Non-Austronesian- (NAN-) speaking people in Near Oceania and Austronesian- (AN-) speaking people from Southeast Asia. In this study, we analyzed mitochondrial DNA (mtDNA) polymorphisms in D-loop region for two AN-speaking Melanesian populations (Munda and Kusaghe) and an AN-speaking Micronesian population (Rawaki) from New Georgia Island in the western Solomon Islands to trace the maternal lineage of AN-speaking Melanesians. The major mtDNA haplogroups in these three populations originated in Asia and the 'Polynesian motif', which is well-characterized mtDNA marker for Polynesians, was frequently observed in the two AN-speaking Melanesian populations but not in the AN-speaking Micronesian population in New Georgia Island. Principle component analyses also revealed genetic proximity between AN-speaking Melanesians (Munda and Kusaghe) and AN-speaking Polynesians in Tonga. These results suggest that Polynesian ancestors have considerably contributed to maternal gene pool of AN-speaking Melanesians in the Solomon Islands before their expansion to Remote Oceania.

2305W

Exploration of the ancestral genetic landscape of the Arabian Peninsula.
D. Platt[1], P. Zalloua[2,3]. 1) IBM T J Watson Research Ctr, Yorktown Hgts, NY; 2) 2Lebanese American University, School of Medicine, Beirut 1102 2801, Lebanon; 3) Harvard School of Public Health, Boston, MA, USA.

The extreme weather events of the Younger Dryas that left Arabia and most of the southern Levant dry drove Natufian expansion from the southern Levant to the Negev, North East Levant, and Mesopotamia while other cultures emerged in the Zagros and in Anatolia. The first Sumerian city states were established around 6 BCE, and the Dilmun culture in Arabia emerged as a major trading power connecting Sumer to the Harrupan civilization. Besides material remains, many of these cultures, including Natufians, left human remains whose genetic imformation provides geographical and chronological records of human migration and admixture starting nearly from the time of post glacial expansions to modern times. In this study, we sought to explore the population genetics of the Arabian Peninsula, identify likely ancient source populations, and investigate evidence of ancient migrations and admixture using PCA, ADMIXTURE, f_3, ALDER, and LDA. ADMIXTURE identified two major contributors to the modern Arabian peninsula populations: an early expansion marked by Natufian aDNA samples, and a component Persian Gulf associated component. PCA distinguishes between the Natufian genetics and the Arabian Pensinsula group that ADMIXTURE combines. Gumuz and Somali admixture impacts the Peninsula, Egypt, and the Levant. Admixture between the Natufian genetic component and East Africa show significant negative f_3 statistics admixing in Egypt, and similarly Egyptian and the Natufian genetic component admixing in East Africa in both directions. Ancient Anatolian genetics have penetrated the study region except for Yemen according to f_3. While we found little gene flow as a candidate for Indo-European admixture in any direction, we did see patterns of expansions carrying Natufian genetics that marked expansion and admixture aligns with that of the proto-Semitic language.

2306T

Evolutionary patterns of long non-coding RNAs with coding capacity of oligopeptides.
W. Lin[1,2], Z. Wang[1], Y. Li[1,3,4]. 1) Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, P. R. China; 2) University of Chinese Academy of Sciences, Beijing, P. R. China; 3) Shanghai Center for Bioinformation Technology, Shanghai Industrial Technology Institute, Shanghai, P. R. China; 4) Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, P. R. China.

Abstract Background Long non-coding RNAs (lncRNAs) are characterized with insufficient protein-coding capacity and weak evolutionary constraints. Recently, a fraction of lncRNA transcripts have been detected to bind with ribosomes by deep sequencing, which implied the translational capacity of some lncRNAs. Our study aimed to identify the lncRNAs with function of coding oligopeptides in human, and revealed evolutionary characteristics and selective signatures of these coding lncRNAs. **Results** Putative ORFs of human lncRNAs from NONCODE, LNCipedia and MiTranscriptome databases were aligned with human oligopeptides from EROP-Moscow database. After strict filtering of ambiguous matching, five human lncRNAs were identified to encode known oligopeptides, including *NONHSAT052368.2*, *NONHSAT052369.2*, *NONHSAT052370.2*, *NONHSAT134945.2* and *NONHSAT125858.2*. The lengths of the coding lncRNA ORFs were consistent with that of their encoded oligopeptides. The tissue-specific expression patterns of the coding lncRNAs were strongly associated with the functions of corresponding oligopeptides, such as fertilization and immunity. Human coding lncRNAs were Simians-specific or Hominidae-specific and their ORFs were originated only in Hominidae species, which suggested that they have evolved recently. In Simians-specific coding lncRNAs, substitution rates were decreased upon the emergence of ORFs, indicating signatures of purifying selection on the ORFs. Furthermore, population genetics study based on the 1000 Genomes Project revealed that in most of the coding lncRNAs, nucleotide diversity was decreased along with significant signatures of purifying selection, compared with their neighboring genomic regions. However, *NONHSAT125858.2*, which encoded an oligopeptide involving in antigen recognition, showed a different pattern of balanced selection among the populations. **Conclusion** These results provided the evidence of coding capability for oligopeptides of human lncRNAs. Our study thus proposed a mechanism for creation of functional short peptides from lncRNAs during the Hominidae evolution, which may contribute to Hominidae-specific genetic novelties by tinkering with pre-existing non-coding genes.

2307F

Ohana: Detecting selection in multiple populations by modelling ancestral admixture components. *J.Y. Cheng[1,2], F. Racimo[2,3], R. Nielsen[1,2].* 1) Department of Integrative Biology and Statistics, University of California, Berkeley, Berkeley, CA; 2) Natural History Museum of Denmark, University of Copenhagen, Denmark; 3) New York Genome Center, 101 Avenue of the Americas New York, NY 10013.

One of the most powerful and commonly used methods for detecting local adaptation in the genome is the identification of extreme allele frequency differences between populations. In this paper we present a new maximum likelihood method for finding regions under positive selection. The method is based on a Gaussian approximation to allele frequency changes, which allows it to account for interbreeding between populations and retains high power when the test populations are admixed. It can also simultaneously and efficiently compare multiple populations. We evaluate the method using simulated data and compare it to methods using summary statistics. We also apply it to a human genomic data set and identify loci with extreme genetic differentiation between major geographic groups. Most of the genes identified are previously known selected loci relating to hair pigmentation and morphology, skin and eye pigmentation, including the top two genes: *EDAR* and *SLC25A5*. In contrast to previous genomic scans, we include data from Aboriginal Australians, which provide us with additional power to detect selection specific to East Asians. Using this method, we can identify new candidate loci - like *CASC15*, involved in melanoma suppression - and narrow down on the likely causal SNPs in previously reported candidate regions - like *KCNB2*, involved in various neurological functions.

2308W

Peruvian Genome Project: A new reference of Andean haplotypes to study genome populations. *H. Guio[1], D.N. Harris[2,3], A. Shetty[2,3], W. Song[2,3], V. Borda[1,4], O. Caceres[1], C. Padilla[1], K. Levano[1], O. Trujillo[6], D. Tarazona[1], C. Sanchez[1], M. Galarza[1], S. Capristano[1], H. Montejo[1], P. Flores-Villanueva[1], E. Tarazona-Santos[4], T. O'Connor[2,3,5].* 1) Laboratorio de Biologia Molecular, Instituto Nacional de Salud-Peru, Chorrillos, Lima, Peru; 2) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; 3) Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 4) Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 5) Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 6) Centro Nacional de Salud Intercultural, Instituto Nacional de Salud, Lima, Perú.

INTRODUCTION: There are a limited number of studies from Latin America that have included native populations mainly due to geographic limitations and ethical considerations when interacting with native communities to enroll them in research studies. This project, initiated by the Peruvian National Institute of Health in 2010, represents the first major effort at studying specific native and mestizo communities across Peru, including demographic history, and population migration patterns forming the basis of precision genomic medicine for the Peruvian people. A total of 953 participants from 30 communities (13 mestizo and 17 native) were enrolled in this study, of which we have generated genotype array data for 171 and high coverage whole genome sequence data on 150. RESULTS: Our cohort of individuals, including the mestizos, has a much greater contribution from Native American ancestry than previous large-scale sequencing studies. We were able to construct a migration and diversity topography of Peru, which revealed that major cities harbor a high degree of genetic diversity independent of European contributions. Further, using identify-by-descent networks, we illustrate that during pre-Inca, Inca, and Spanish administration the Andean region was central to the population structure of Peru, while post-Spanish independence, the population dynamics seem to shift towards the coast, consistent with history. This is also consistent with high altitude adaptation leading to reduced gene flow into the Andes. We also demonstrate fine-scale population structure within the mestizo communities, identifying admixture between Native American communities in addition to an increase European contribution, both of which need to be considered when performing GWAS and personalized medicine programs. CONCLUSION: The genetic background of Native Americans has spread worldwide including in United States. It is expected that almost one million immigrants will come from Andean countries with high levels of Native American background. This population will be classified as "Hispano/Latino" in many studies in United States, but this group is a mixture of individuals with different levels of Native American ancestry and from different Native American sources. For this, we need to better understand the genetic variation and architecture of these native populations, migration patterns, and ethno geographical studies.

2309T

MixFit: Methodology for computing ancestry-related genetic scores at the individual level and its application to the Estonian and Finnish population studies. *T. Haller[1], L. Leitsalu[1], M. Nuotio[2], T. Esko[1], D. Boomsma[3], K. Kyvik[4], T. Spector[5], M. Perola[1,2,6], A. Metspalu[1].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) University of Helsinki, Helsinki, Finland; 3) Vrije University, Amsterdam, Netherlands; 4) University of Southern Denmark, Odense, Denmark; 5) Kings College London, London, United Kingdom; 6) National Institute for Health and Welfare, Helsinki, Finland.

Ancestry information at the individual level is a resource for personalized medicine, demographical and history research, and for tracing back personal history. We report a new method for quantitatively determining personal genetic ancestry based on genome-wide data. Numerical ancestry component scores are assigned to individuals based on comparisons with reference populations. These comparisons are conducted with an existing analytical pipeline making use of genotype phasing, similarity matrix computation and our addition - multidimensional best fitting by MixFit. The method is demonstrated by studying Estonian and Finnish populations in geographical context. We show differences in the genetic composition of these close European populations and how they have influenced each other. We determine ancestry component distribution by geographical region for Estonia and Finland to highlight how these populations have interacted with their neighbors. Sorting the individuals by the birth date allows investigation of time-dependent trends in ancestry component distribution. We perform association analyses between ancestry components and anthropometric traits and report several associations. Our analytical methods apply to studying specific individuals but can be extended to population studies. We map the ancestral composition of Estonia and Finland in the geographical and historical context. The analytical pipeline has been published, MixFit is available at www.geenivaramu.ee/en/tools/mixfit.

2310F

Discovering rare variants and deciphering a population structure of 386 Mongolian individuals by whole-genome sequencing. *C. Kim, S.K. Yoo, S. Kim, J.Y. Shin, J.S. Seo.* Genomic Medicine Institute, Seoul National University College of Medicine, Seoul, Gyeong-ki, South Korea.

Here, we present the whole-genome sequencing data for 386 Mongolian individuals which mainly composed of the Buryats and the Khalkha Mongols by average sequencing depth of 17X. We discovered 3.8 million novel single nucleotide polymorphisms (SNPs) which were not previously reported. Moreover, 965,663 SNPs which were rare (minor allele frequency < 0.5%) in the 1000 genome project phase 3 were low frequency or common in our dataset. Series of analysis demonstrated distinctive population structure of the Mongolians from the East Asians. We constructed robust imputation panel for Northern Asian populations which produces great imputation accuracy for rare (mean r2 of 0.82) and low frequency (mean r2 of 0.87) SNPs. We identified significant gene flow from the Buryats to the Finnish which was predicted to be occurred in 1,228 (±87) year. Moreover, 13.38% of Buryat admixture was predicted in the Finnish genome. In summary, this study illustrated advantage of whole-genome sequencing to build reference panels and to study population history.

2311W

Large-scale whole genome sequencing of the Estonian population reveals new insights into population history and recent natural selection.

M. Metspalu[1], L. Saag[1], G. Hudjashov[1,3], L. Pagani[1,4], L. Kushniarevich[1,6], V. Pankratov[6], M. Mitt[2], R. Andreson[1], M. Kals[2], D. Lawson[5], T. Esko[2], R. Mägi[2], A. Metspalu[2]. 1) Evolutionary Biology, Estonian Biocentre, Tartu, Estonia; 2) Estonian Genome Center, University of Tartu, Tartu, Estonia; 3) Statistics and Bioinformatics Group, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand; 4) Department of Biological, University of Padova, Italy; 5) Integrative Epidemiology Unit, School of Social and Community Medicine, 7 University of Bristol, Bristol BS8 2BN, UK; 6) Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus.

Altogether 2244 whole genomes of geographically diverse individuals from Estonia were sequenced to a median depth of 30x using Illumina HiSeq with TruSeq PCR-free library preparation method. We found 19M SNVs and 6.6M indel variants with allele count larger than two of which 8.4M were novel. To study the fine-scale genetic structure of the Estonian population, we concentrate on a subset of the genomes (N=436), which comprehensively cover rural Estonia to minimize the mixing effect of historical urbanization. We will make an effort to mine the genealogical trees to fine-tune the analysis and exclude outliers. We further combine these genomes with a pan Eurasian panel of high coverage genomes from hundreds of populations. Using haplotype and allele frequency based methods we show that the genetic structure within Estonia is largely in line with the division of *inland* vs. *maritime* Estonia what has been proposed based on archaeological findings. Furthermore, we identify and quantify the relative contributions of the three major genetic domains of the European gene pool in Estonians and estimate split times from linguistically and geographically adjacent populations. We use Finestructure and inter-population doubleton distribution to reveal patterns of genetic sharing between Estonians and other European populations and infer population history in a series of population splits and admixture events in pre-historic and historic times. We used IBDNe to estimate recent changes in effective population size across local communities. And finally, we use the novel Singleton Density Score approach to reveal very recent changes in allele frequencies/natural selection in the Estonian population and compare the results to those found in UK10K. .

2312T

Association of *VASP* polymorphisms and infectious disease burden in global populations and identification of candidate resistance haplotypes in sub-Saharan Africa.

M. Phun, B. Rosenn, D. Glenn, D. Rupasinghe, J. Church, D. Zuzga. Biology, La Salle University, Philadelphia, PA.

Pathogens employ numerous strategies to breach epithelial barriers. One critical target of pathogens is the host cell cytoskeleton. Pathogens induce cytoskeletal remodeling to disrupt tight junctions and increase the permeability of the epithelium to gain access to underlying tissue. Alternatively, the cytoskeletal machinery may be directly co-opted to facilitate the attachment, invasion, or intracellular motility of pathogens. The vasodilator-stimulated phosphoprotein (VASP) is a processive actin polymerase that induces membrane remodeling through the polymerization of actin cytoskeletal filaments. Interestingly, VASP function is regulated by numerous pathogens to induce remodeling of host cells. If VASP is indeed a conversed target of pathogens, it is plausible that evolutionary pressure on host cells would select for resistant variants of *VASP* that impair the ability of pathogens to co-opt VASP function. To investigate this possibility, we determined the minor allele frequency (MAF) of *VASP* single nucleotide polymorphisms (SNPs) across 26 global populations. A heatmap was then constructed to represent the minor allele frequency of >400 SNPs, which revealed increased minor allele frequency in Sub-Saharan African populations that also exhibit disproportionally high burdens of infectious disease compared to global populations. Next, SNPs exhibiting increased MAF in Sub-Saharan African populations and SNPs predicted to impact VASP function were employed in linkage disequilibrium (LD) analysis. SNPs found in LD were then used to reconstruct VASP haplotypes. These haplotypes may represent unique alleles that provide resistance to infectious disease. Future studies will examine the effect of *VASP* haplotype expression on host-pathogen interactions *in vitro.*.

2313F

Mapping the genetic diversity in indigenous Malays populations and cosmopolitan Malay. *W. Saw[1,2], K. Yusoff[3], T. Rahman[4], S. Xu[5,6,7], C. Khor[6,9,10], Y. Teo[1,2,8,11], B. Hoh[3,11].* 1) Life Science Institute, National University of Singapore, Centre for Life Sciences, Singapore, Singapore, Singapore; 2) Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive, Singapore 117549; 3) Faculty of Medicine and Health Sciences, UCSI University, Jalan Menara Gading, Taman Connaught, Cheras, 56000 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia; 4) Clinical Pathology Diagnostic Centre Research Laboratory, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh Campus, Selangor, MalaysiaEndFragment; 5) Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 6) School of Life Sciences and Technology, ShanghaiTech University, Shanghai, China; 7) Collaborative Innovation Center of Genetics and Development, Shanghai, China; 8) Genome Institute of Singapore, Agency for Science, Technology and Research Singapore, 60 Biopolis St, Singapore 138672; 9) Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117597; 10) Singapore Eye Research Institute, 20 College Road, Singapore 169856; 11) These authors contributed equally to this work.

Southeast Asian Malays has historically been considered as a genetically distinct population, although there is emerging evidence to suggest that the genetic architecture is an admixture of four major ancestral populations, comprising the Austronesian, Proto-Malay, East Asian and South Asian. Here we attempt to formally establish the genetic membership of different groups of Southeast Asian Malays, from city-dwelling cosmopolitan Malays to indigenous Malay groups that still rely on the practice of hunter-gatherer subsistence and reside in primary forests. By analyzing at least 2 million autosomal SNPs across the genomes of 40 individuals from 3 indigenous groups and 1 cosmopolitan Malay group, we quantified the extent of genetic differentiation between the city dwellers and forest inhabitants. Our results revealed that the indigenous Malays were genetically more distinct to cosmopolitan Malays, than Northern Gujarati Indians from Southern Tamil Indians, despite the population divergence time between two indigenous Negrito populations to be less than 2KYA, and between the Senoi and Negritos to be less than 4KYA. Thus, even though the divergence between indigenous Malay groups was very recent, differential environmental exposures have introduced substantial variations in the genomes of these indigenous groups.

2314W

Mitogenomes sequencing: Reveals huge diversity in human maternal lineages of Jammu and Kashmir, India. *I. Sharma[1], V. Sharma[1], P. Kumar[2], E. Rai[1], M. Vilar[3], S. Sharma[1].* 1) Department of Biotechnology, Shri Mata Vaishno Devi Ubiversity, Katra, Jammu and Kashmir,182320, India; 2) Institute of Human Genetics, University of Jammu, Jammu and Kashmir, 180006, India; 3) Human Genographic Project, National Geographic Society, USA.

Jammu and Kashmir (J&K) is a Northern state of the Indian subcontinent. The Societal structure of population is governed by a large number of religious and endogenous ethnic groups. It is believed that J&K had served as a corridor, due to its strategic location in the Himalayan region, for deep in time migrations as well as had been route of various invaders. It is expected that in each spate, invaders brought their own distinct type of gene pool leading to admixture and present day population structure of the region. It's been proposed many a times that most of the migrations, especially invasions, had been male specific thus maternal gene pool of India is relatively conserved while paternal gene pool is relatively diverse. We carried out this study to find the diversity in maternal gene pool of J&K-India. DNA from blood of 83 individuals was subjected to complete mitogenome sequencing, outsourced from Medgenome Labs Pvt. Ltd. DNA is used to perform targeted amplification of the mitochondrial genome by long-range PCR, library was prepared using Illumina TrueSeq kits, and then sequenced to more than 1000X coverage. The sequences obtained were aligned to revised Cambridge mitochondrial Reference Sequence (rCRS) to identify variants in genome. The variants were anylysed to assign maternal haplogroups to each individual. We found high diversity with some of new lineages, in the maternal gene pool of this region. Different ethnic groups are the amalgamation of M, U, H, W, R, K, F, D, T, A, C and I haplogroups. This provides genetic evidence that this region has acted as corridor deep in time as well to many recent events; however it has not been restricted to male mediated gene flow. IS*, ER and SS acknowledge Human Genographic Project and National Geographic Society, USA for providing grant for the study. .

2315T

Recent changes in contemporary effective population size from identical by descent segments. *A. Urnikyte[1], A. Molyte[1], V. Kučinskas[1], Z.A. Kučinskienė[2].* 1) Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariskiu St. 2, LT-08661 Vilnius, Lithuania; 2) Department of Physiology, Biochemistry, Microbiology and Laboratory Medicine, Faculty of Medicine, Vilnius University, M. K. Čiurlionio St. 21, LT-03101 Vilnius, Lithuania.

  Introduction: Analysing the effective population size (*Ne*) in different groups of a small specific population we can obtain an insight into the demographic history and dynamics of modern human populations as well as better understanding of known mechanisms of microevolutionary processes. The contemporary Lithuanian population is composed of a complex mixture of former Baltic tribes and states. In this study we aimed to estimate the recent effective population size, by using inferred long segments of identity by descent (IBD), and estimate the effective/census size ratio in the Lithuanian population and to reconstruct the recent past events between the two main ethnolinguistic groups (Aukštaičiai and ⬚emaičiai) of the Lithuania, since historically these groups probably developed over a long period of time as two independent Baltic tribes. **Methods:** Single-nucleotide polymorphism (SNP) genotyping of 295 samples was performed with an Illumina 770K HumanOmniExpress-12v1.1 array. To infer the history of the recent effective population size we used a non-parametric method, based on the Wright–Fisher discrete-generation model, implemented in the open-source IBDNe v. 04Sep15.e78 software package. **Results:** The recent *Ne* values were obtained for 50 generations ago with 95% confidence intervals (CI), assuming a generation time of 25 years. The evaluated *Ne*/N ratio was 0.125 (95% CI [0.077; 0.271] for g = 2 (corresponds to 1941), 0.114 (95% CI [0.065; 0.280] for g = 1 (corresponds to 1966), and 0.124 (95% CI [0.065; 0.341] for g = 0 (corresponds to 1991). The estimates of *Ne* were approximately one-tenth of the Lithuanian census population size. The reconstructed past events between the two main ethnolinguistic groups (Aukštaičiai and ⬚emaičiai) of Lithuania showed differences between the groups. **Conclusions:** The obtained values of Ne/N ratios are small (0.1) compared with other genetics-based estimates between 0.21-0.65. Natural levels of fluctuations such as variance in size, reproduction, sex ratio, and the degree to which generations overlap probably caused the small values of Ne/N. The true effective size might be contained within the bootstrap confidence interval. Our results will provide comprehensive implications for further fundamental genomic and evolutionary studies. This study is part of the LITGEN project, which was approved by the Vilnius Regional Research Ethics Committee No. 158200-05-329-79, date: 2011-05-03.

2316F

Selection analysis in Chileans identify adaptation signals in Native Americans, highlighting regulatory processes. *L. Vicuña[1], S. Eyheramendy[1], F. Martinez[2].* 1) Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Región Metropolitana, Chile; 2) Departamento de Antropologia, Pontificia Universidad Católica de Chile, Santiago, Región Metropolitana, Chile.

  Detection of positive selection signatures in populations around the world is helping to uncover human demographic history as well as the genetic basis of diseases. In addition, genetic admixture has been shown to play an important role in shaping human diversity. We look for local adaptation signatures left by ancient hard-sweeps in the admixed Chilean population by performing five statistical tests genome-wide. Further, we search for post-admixture signatures left by soft-sweeps by identifying regions with significant deviations in the mean local Native-American ancestry. These analyses identify novel and previously-reported candidate genes for selection. We also perform enrichment analysis of variant categories in region under selection, and find that signatures inherited from Native-Americans are especially enriched in categories affecting gene expression regulation. Moreover, we find that, for most tests and categories, selected regions have a higher mean local Native-American ancestry, with regulatory variants being among the highest. Finally, through structural spatial analysis, we find non-synonymous variants that are highly differentiated between Native-Americans, Europeans and Chileans, which are associated with high-prevalence diseases among Chileans, like Chagas disease and severe influenza-A (H1N1) infection.

2317W

Analysis and findings in high-depth target sequencing of over 20000 individuals in China. *H. Xu[1], L. Sun[2], X. Jin[1], J. Akey[3].* 1) BGI-Shenzhen, Shenzhen, China; 2) Department of Dermatology, Anhui Medical University, China; 3) Princeton university.

To comprehensively delineate population genetics in the Chinese population, we reanalyzed ~1300 genes of ~5Mb in the coding regions and Major Histocompatibility Complex (MHC) of 5-Mb in ~20000 individuals of Han Chinese ancestry with ~45x coverage. We identified ~450000 single-nucleotide variants (SNVs), the majority of which were rare, previously unknown and population-specific. We identified loss-of-function mutations and explored effects of rare protein-truncating variants, which might improve medical and functional interpretation in these regions. We showed fine-scale population structures within China by analyzing common and rare variants. Demography histories were inferred with these 20000 samples of target sequencing, comparing to the results of 3 population models combining large-scale Chinese, African-American and European-American samples. We are also investigating introgression of archaic segments in MHC regions, suggesting impacts of selected introgression of functionally advantageous genes in the immune system.

2318T

Standardized visualization of demographic history. *Y. Zhou, S. Browning.* University of Washington, Seattle, WA.

Standardized visualization of population history is an urgent need for both theoretical studies and data interpretation in the field of population genetics. Benefiting from the growing amount of available genetic data from populations all over the world, the demographic history of the target population(s) can be fully understood in relation to other reference populations. Usually, these studies need to give the historical relationship among related populations to interpret their conclusions, or as a part of their conclusions. The common solution for visualizing population history is to create the figures line by line with drawing software, and no existing software can provide a standardized visualization for population demographic history. In this study, an R package named *plotms* is developed to visualize the population demographic history in a standardized way. *plotms* is built on *ms*'s input language for demographic history, which can be applied to arbitrary numbers of populations with changing population sizes and any number of waves of migrations. *plotms* is also very easy to use and it currently can read the input of *ms*, MaCs, Cosi, and *msa*. The input from more simulation languages will be supported in the future..

2319F

Principal components analysis with sensible weighting of sequencing variants: Improved inference of fine scale population structure with whole genome sequencing data. *T.A. Thornton, J.L. Kirk.* Biostatistics, University of Washington, Seattle, WA.

There is great potential for rare variants to be a powerful resource to delineate fine-scale population structure patterns. Whole genome sequencing (WGS) data has an abundance of rare variants, and recent work has demonstrated that rare variants can have differential population structure and greater geographic clustering than common variants. Existing principal components analysis (PCA) methods, such as EIGENSTRAT (Price et al., 2006), have been shown to work well for population structure inference with common variants. However, these PCA-based approaches can fail when applied to WGS data due to the abundance of extremely low frequencies resulting in the genetic relatedness matrix (GRM) used for the PCA becoming unstable. Existing PCA-based approaches for population structure inference give weights to variants according to minor allele frequency (MAF), where variants with the lowest MAF receive the largest weights in the analysis. In reality, for most variant MAF is a poor proxy for ancestry. Here, we propose a new method for population structure inference using whole genome sequencing data where informative weighting of variants is used, where the weight of a variant is based on the amount of information provided for delineating population structure. The method incorporates both entropy and mutual information from classical information theory in statistics. In applications to the 1000 Genomes Project, we demonstrate that our new approach outperforms existing PCA-based approaches and is able to delineate both continental and fine-scale population structure.

2320W

Assessing human diversity patterns using in-silico discovered Alu and LINE-1 mobile insertion elements in the Simons Genome Diversity Project. *S. Watkins[1], J. Feusier[1], C. Goubert[1], S. Mallick[2], D. Reich[2], L.B. Jorde[1], The SGDP consortium.* 1) Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA; 2) Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

To assess the population distribution and characteristics of young Alu and LINE-1 mobile element insertions (MEIs), 296 human whole-genome sequences from the Simons Genome Diversity Project (SGDP) were analyzed with the mobile element locator tool (MELT). Jointly-called raw non-reference MEI insertions were screened to remove markers showing strong deviation from Hardy-Weinberg equilibrium, markers with multiple mis-inheritances in a three-generation pedigree, and markers lacking target-site duplications. This WGS sequence-based in-silico discovery process yielded 12,800 Alu and 1,982 LINE-1 elements that are strictly identity-by-descent markers and free from population ascertainment bias. These in-silico based Alu allele frequency estimates were concordant with estimates from traditional PCR-based MEI genotyping. Principal components analyses based on allele-sharing genetic distance estimates for Alu and LINE-1 data sets were congruent with CpG-free and transversion-only SNV data sets. The proportion of Alu and LINE-1 sites that were heterozygous in each individual was highest in Africa, lower in Eurasia, and lowest the New World. Outside of Africa, heterozygosity decreased in a linear fashion with increasing distance from the African centroid. Admixture analysis using polymorphic Alu insertions successfully quantified world admixture patterns, distinguished San and Mbuti from other Africans, and identified a cline of native American ancestry across northern Asia that was maximal in populations of northeastern Russia. These results demonstrate that MEIs can be reliably ascertained from whole-genome sequences and that they yield results highly concordant with those of other genotyping methods. .

2321T

Archaic-genome-agnostic detection of introgressed segments. *S.R. Browning[1], Y. Zhou[1], B.L. Browning[2].* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA.

We present new methodology and software for detecting archaic introgressed segments, such as those from Neanderthals or Denisovans, in modern human genomes. Our method is similar to S* [1] in that it uses a scoring algorithm based on linkage disequilibrium and does not require knowledge of the archaic genome sequence. This makes it possible to detect introgressions from as yet unsequenced archaic populations, and to detect introgression when the archaic reference genome has poor sequence coverage in a region or is diverged from the introgressed archaic genome. Unlike S*, our method can be applied to arbitrarily large samples, enabling detection of low frequency introgressed segments. Our scoring procedure utilizes a genetic map and adjusts for variant density, which permits a single score threshold to be used throughout the genome. In addition, our approach can analyze an entire chromosome without use of a sliding window and thus can gain power when introgressed segments are tiled across a longer region. To test our method, we simulated data on 4000 European individuals and 100 outgroup African individuals, following published models of human and Neanderthal demographic history, with 3% introgression. Our method found 1/3 of the introgressed material, at a false discovery rate of 7.2%. Although some introgressed segments cannot be detected because they are too short, our agnostic method captures a substantial portion of the introgressed sequence. As proof of principal, we apply our method to the 4000 whole genome sequenced individuals of European ancestry from the UK10K study, as well as to populations from the 1000 Genomes Project and to Papuans from the Simons Genome Diversity Project. In the UK10K data, the introgressed segments that we detect comprise on average 0.4% of each individual's genome, and 74% of the detected introgressed variants match the Neanderthal reference genome. [1] Plagnol, V. and J.D. Wall, *Possible ancestral structure in human populations*. PLoS Genetics, 2006. 2:e105.

2322F

Using ancient DNA from Sardinia to assess population stability from the Neolithic to present. *J. Marcus[1], C. Posth[7,8], L. Lai[11], C. Sidore[4,5,6], J. Beckett[14], H. Al-asadi[2,3], A. Furtwängler[9], M. Zoledziewska[4], C. Chiang[12,13], K. Dey[3], M. Gradoli[15], G. Abecasis[5], D. Schlessinger[9], R. Skeates[10], J. Krause[7,8], J. Novembre[1,2], F. Cucca[4,6].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Department of Ecology and Evolution, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL; 4) Istituto di Ricerca Genetica e Biomedica, CNR, Cagliari, Italy; 5) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI; 6) Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy; 7) Max Planck Institute for the Science of Human History, Jena, Germany; 8) Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany; 9) Laboratory of Genetics, NIA, NIH, Baltimore, MD; 10) Department of Archaeology, Durham University, Durham, England; 11) Department of Anthropology, University of South Florida, Tampa, FL; 12) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Behavior, University of California, Los Angeles, Los Angeles, California; 13) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California; 14) Independent Contractor, Cagliari, Italy; 15) School of Archaeology and Ancient History, University of Leicester, Leicester, United Kingdom.

Ancient DNA (aDNA) has provided a powerful tool for assessing the temporal stability of populations within a geographic locale. In much of mainland Europe, aDNA has revealed a relatively dynamic population history from the Neolithic period through the Bronze Age. Due to their relative isolation, island populations may in general not experience the same population dynamics as mainland populations. The island population of Sardinia in particular has been hypothesized to have a relatively stable and continuous population from the early Neolithic, largely on the basis of modern Sardinian DNA. Here we directly assess continuity using genome-wide capture data (~1.2 millions SNPs) of 26 ancient humans from the island of Sardinia spanning the Neolithic, Copper Age, and Bronze Age, including individuals from the Nuragic culture. Through analyzing read-level DNA damage patterns and estimating modern contamination levels, we authenticate these ancient DNA sequences and removed outlier loci and individuals for downstream analyses. Projecting these ancient individuals on to modern axes of genetic variation, as defined by principal component analysis on a large-scale reference dataset of modern human populations from Sardinia, Europe and the Middle East, reveals no obvious temporal structure within Sardinia within this long time frame. Consistent with previous hypotheses of early migrations in Europe, we observe clustering of these ancient individuals with previously published sequences of ancient humans associated with an "early farming" Neolithic culture. Through the application of multiple population genetic methods and exploratory data analysis tools we find that relative to mainland Europe there has been population stability within the island of Sardinia. Beyond shedding light on Sardinian population history, the relative stability we infer is important for understanding the local frequencies of disease susceptibility alleles.

2323W

Impacts of European colonization on an indigenous community in British Columbia. *A.C. Owings[1], B. Petzelt[5], J.S. Cybulski[4], R.S. Malhi[1,2,3].* 1) Program in Ecology, Evolution and Conservation Biology, University of Illinois at Urbana Champaign, Champaign Urbana, IL; 2) Department of Anthropology, College of Liberal Arts and Sciences, University of Illinois, Urbana-Champaign, IL; 3) Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana-Champaign, IL; 4) Research Division, Canadian Museum of History, Gatineau, Quebec, Canada; 5) Metlakatla Treaty Office, PO Box 224, Prince Rupert, British Columbia V8J 3P6, Canada.

Excavations in some of the archaeological sites in the Prince Rupert Harbour region, British Columbia, Canada, have yielded the skeletal remains of over 250 pre-European contact individuals. Their connection to modern living descendants in the area has been demonstrated through DNA analysis and shows genetic continuity back to the late Pleistocene. The sites also enable study of the genetic effects of European colonization. A recent autosomal DNA study found a 57% reduction in effective population size in the living descendants compared to the ancients. Here we present mitochondrial DNA (mtDNA) data for 65 of the ancients and 53 presumed living descendants using both Sanger sequencing and an Illumina MiSeq high-throughput methodology. We find shared haplotypes between the ancient and living groups as might be expected, but some unique haplotypes in each group as well, indicating that some ancient haplotypes may have been lost over time and some may have been gained through gene flow. In addition, we find that mtDNA diversity has decreased by ~50% in one of two living descendant communities, but increased in the other compared to the ancient population. We also present Y-chromosome STR data for 50 ancient and 44 living people obtained through the use of multiplexed Y-STR kits. All ancient individuals had Y-chromosome haplogroups that are typically found in the Americas. Genetic diversity measures such as thetaS and haplotype diversity for the ancient and living individuals show a decrease in Y-chromosome diversity in the living population by roughly 20%. However, the sample size of the living individuals is small due to the ~55% non-indigenous Y-chromosome haplogroups. The Bayesian Skyline Plot method in the program BEAST was used to simulate the demographic changes over time. This is among the first studies to quantify the genetic effects of colonization in an indigenous population in the Americas by the analysis of ancient and living individuals sampled through time. A similar study design can be applied to other indigenous communities to understand the impacts of European colonization - one of the most significant human demographic events in history.

2324T

Tracing the origin of ancient polynesian human genomes across the Pacific. *P. Salazar-Fernandez[1], M. Avila-Marcos[2], E. Hagelberg[3], K. Sandoval[1], C.D. Quinto-Cortes[1], M. Moraga[4], T. Parks[5], A. Mentzer[5], A. Moreno-Estrada[1].* 1) National Laboratory of Genomics for Biodiversity, Irapuato, Guanajauto, Mexico; 2) International Laboratory for Human Genome Research, Queretaro, Mexico; 3) University of Oslo, Norway; 4) University of Chile, Chile; 5) University of Oxford, United Kingdom.

In 2010, mitochondrial DNA extracted from two ancient skulls found in southern Brazil belonged to the haplogroup B4a1a1a, exclusively found in Polynesia. Radiocarbon analyses indicated that these individuals most probably died before the 19th century, prior to any registered transport of Polynesian people to South America by European vessels. Further genome-wide analyses showed a complete Polynesian ancestry for both samples, with Cook Islands as the closest source population. However, scarcity of genotyping data from modern Polynesian populations posed a major limitation for inferring a more specific place of origin for said skulls. Here, we re-analyze these ancient DNA samples using an extended reference panel that comprises over 475 genotyped samples from 18 different locations across the Pacific Ocean. With this data we explore the genetic affinities of the Botocudo skulls at a finer scale to potentially pinpoint their genetic origin, and we demonstrate the importance of assembling diverse genetic reference panels to shed light on the evolutionary past of human remains devoid of archeological context.

2325F

MHC-dependent mate selection within the Health and Retirement Study (HRS). *Z. Qiao[1], J. Powell[2], D. Evans[1,3].* 1) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia; 2) Institute for Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia; 3) Medical Research Council (MRC) Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom.

Disassortative mating refers to the phenomenon in which individuals with dissimilar genotypes and/or phenotypes mate with one another more frequently than would be expected by chance. Although the existence of disassortative mating is well established in plant and animal species, the only documented example of negative assortment in humans involves dissimilarity at the Major Histocompatability (MHC) locus. Unfortunately previous studies investigating mating patterns at the MHC have been hampered by limited sample size and contradictory findings. Inspired by the sparse and conflicting evidence, we investigated the role that the MHC region played in human mate selection using genome-wide association data from 597 European American spouses from the Health and Retirement Study (HRS). First we treated the MHC region as a whole, and investigated genomic similarity between spouses using three levels of genomic variation: SNPs, classical HLA alleles (both 4-digit and 2-digit classifications), and amino acid polymorphisms. The extent of MHC dissimilarity between spouses was assessed using a permutation approach. Second, we investigated fine scale mating patterns by testing for deviations from random mating at individual SNPs, HLA genes, and amino acids in HLA molecules. Third, we assessed how extreme the spousal relatedness at the MHC region was compared to the rest of the genome, to distinguish the MHC-specific effects from genome-wide effects. The multi-locus relatedness analyses showed no significant difference in MHC relatedness between spouses and non-spouse pairs at either the level of SNPs (relatedness coefficient, R=0.013, one-sided p=0.638) or classical MHC alleles (R=-0.025, p=0.258); however, analyses based on HLA amino acid polymorphisms revealed stronger evidence of dissimilarity (R=-0.046, p=0.095). In terms of fine mapping, we did not find evidence of excess dissimilarity between spousal pairs at individual SNPs, classical HLA genes or amino acids after adjustment for multiple testing. However, the MHC dissimilarity among spouses was extreme relative to the rest of genome (i.e. only 1.6% of the randomly sampled genomic windows exhibited more extreme dissimilarity than the MHC region), but was as extreme in opposite-sex non-spousal pairs, and therefore the dissimilarity could not be attributed to mate selection. Despite the long-standing controversy, our analyses did not support a significant role of MHC dissimilarity in human mate choice.

2326W

Pseudogenes in the mouse lineage: Transcriptional activity and strain-specific history. *P.M. Muir[1,2,3], C. Sisu[5,7], A. Frankish[8], I. Fiddes[9], M. Diekhans[9], T. Keane[8], M. Gerstein[4,5,6].* 1) Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT; 2) Systems Biology Institute, Yale University, West Haven, CT; 3) Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, CT; 4) Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 5) Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT; 6) Department of Computer Science, Yale University, New Haven, CT; 7) College of Health and Life Sciences, Brunel University London, Uxbridge, London, UK; 8) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 9) University of California, Santa Cruz, CA.

Pseudogenes are ideal markers of genome remodelling. In turn, the mouse is an ideal platform for studying them, particularly with the availability of transcriptional time course data during development (just completed in phase 3 of ENCODE) and the sequencing of 17 strains (completed by the Mouse Genome Project). Here we present a comprehensive genome-wide annotation of the pseudogenes in the mouse reference genome and associated strains. We compiled this by combining manual curation of over 10,000 pseudogenes with results from automatic annotation pipelines. Also, by comparing human and mouse, we annotated 327 new unitary pseudogenes in human with respect to mouse and 210 unitary pseudogenes in mouse with respect to human. (We make our annotation available through a resource website mouse.pseudogene.org.) The overall mouse pseudogene repertoire (in the reference and strains) is similar to human in terms of overall size, biotype distribution (80% processed, 20% duplicated) and top family composition (with many *GAPDH* and ribosomal pseudogenes). However, notable differences arise in the age distribution of pseudogenes with multiple retro-transpositional bursts in mouse evolutionary history and only a single one in human. Furthermore, in each strain ~20% of the pseudogenes are unique, reflecting strain-specific functions and evolution – e.g. the pseudogenization of taste receptors is linked to a change in the diet of the NZO strain. Finally, we find ~15% of the pseudogenes are transcribed, a fraction similar to human. Furthermore, we show that processed pseudogenes are commonly associated with highly transcribed genes. While this can be observed through all of mouse development, the relationship is strongest not at the early embryo stages but later on, after depletion of maternal RNA.

2327T

Inference of allele-frequency trajectory histories from present genomes. *Y. Field[1,2], J.K. Pritchard[1,2,3].* 1) Genetics, Stanford, Stanford, CA; 2) Howard Hughes Medical Institute, Stanford, Stanford, CA; 3) Biology, Stanford, Stanford, CA.

The fundamental objective of population genetics is to understand the dynamics, determinants and consequences of the change in frequency of genetic variants in populations over time. This has broad implications to the studies of human evolution, human history, and medical genetics. Methods have been developed to infer recent changes in allele frequencies from present-day genome data. However, whereas determinants of allele frequency trajectories—from demography to natural selection—are highly dynamic in nature, these methods are limited to infer a single frequency change integrated over an estimated time window and lack temporal resolution. Here we present TRAJECTORY, a coalescent-based statistical method that infers recent changes in allele frequencies from present-day genomes with user-defined temporal resolution. TRAJECTORY models the frequency trajectory as two population size history problems, corresponding to the haplotype groups of the target SNP. To this end we borrow modeling ideas from *fastNeutrino*, an efficient inference of population size histories. To account for recombinations and haplotype phasing error—both of which shuffle the haplotype histories of the derived and ancestral alleles—we extend on modeling ideas from the *Singleton Density Score*, a method that we previously developed to infer a single allele frequency change over a sample-dependent timescale. In addition to providing temporal resolution, TRAJECTORY improves on existing methods in terms of power and further allows for joint modeling of multiple sites to directly infer characteristic trajectories of polygenic adaptation. We will present the method, simulations, validations and example applications demonstrating the new temporal dimension of analysis provided by our method.

2328F

SG10K: Insights into the genetic architecture of Singaporeans. *C. Bellis[1], I.D. Irwan[1], C.A. Lin[1], T.H. Koh[1], C. Wang[2], W.W. Soon[3], A. Wilm[4], C.C. Shih[4], J. Liu[1], SG10K Consortium.* 1) Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research of Singapore (A*STAR), Singapore; 2) Computational and Systems Biology, Genome Institute of Singapore, A*STAR, Singapore; 3) Next Generation Sequencing Platform, Genome Institute of Singapore, A*STAR, Singapore; 4) Scientific and Research Computing, Genome Institute of Singapore, A*STAR, Singapore.

The unique ethnic diversity inherent within the Singaporean population opens it up as an opportune cohort for population genetics studies. The Singapore population consists of three major ethnic groups; Chinese, Malay, and Indian, which together represent ~80% of the genetic variation across Asian populations. In 2015, the "SG10K project" was initiated with an overarching aim of sequencing the genomes of 10,000 Singaporeans. To date our collaborative partners include SingHealth Duke-NUS Institute of Precision Medicine, Singapore Eye Research Institute, Centre for Personalised and Precision Health, Tan Tock Seng Hospital, National University Health System and several Translational and Clinical Research Flagship Programmes (Heart failure, Parkinson disease). Our main objectives are to (1) comprehensively characterize genetic variation in Singapore population; (2) create a WGS reference panel for accurate genotype imputation in Asian population; and (3) generate a large control dataset for WGS-based genetic association study of diseases. We have adopted a shallow-pass sequencing approach, which on average will cover each base at a depth of approximately 15×. Our analytical pipeline hosted by the National Supercomputing Centre (NSCC), Singapore incorporates GATK (v3.6) and follows GATK best practices. Our initial analytical pipeline test was undertaken on n=1,059 genomes and required approximately 2-3 weeks compute time running on 20 reserved nodes at NSCC. We anticipate providing insight into genetic variation in the local Singaporean population following the first genotype calling pipeline test. Upon completion, this study will provide valuable genetic information to facilitate precision medicine initiatives in Singapore and will empower genetic studies of Singapore and Asian-centric diseases.

2329W

SeleDiff: A scalable tool for testing and estimating selection differences between populations. *X. Huang[1,2], M. Wang[1,2], L. Jin[1,3], Y. He[1].* 1) Chinese Academy of Sciences Key Laboratory of Computational Biology, Chinese Academy of Sciences-Max Planck Society Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Shanghai, 200031, China; 2) Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China; 3) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200433, China.

Genome-wide scan for natural selection is a classical challenge in human genetics. Currently, there are two popular kinds of approaches for detecting signals of natural selection from human genomes. One is based on extended haplotype homozygosity; the other is based on genetic diversity θ. The time complexities of these approaches, however, are quadratic with respect to the number of variants or the number of samples. As more and more variants and samples become available in human populations, the quadratic time complexity would limit research community to fast detect signals of natural selection. Moreover, neither of these approaches could quantify the differences of the strength of natural selection between populations. Here, we implemented a scalable tool called SeleDiff for testing and estimating selection differences between populations. Using SeleDiff, we can analyze a simulated dataset containing 300,000 variants and 100,000 samples in 15.7 minutes (wall time) with less than four gigabytes of random access memory (OS: Red Hat Enterprise Linux Server release 6.3; CPU: AMD Opteron™ 6174). Running time analysis showed SeleDiff has linear time complexity with respect to both the number of variants and the number of samples. This linear time complexity would make SeleDiff a fast tool for quantifying signals of natural selection in genome-wide scan.

2330T

Distribution of common and rare variants in an underrepresented population in public genomic databases and the possible impact in precision medicine. *C.S. Rocha[1,3], B.S. Carvalho[2,3], I. Lopes-Cendes[1,3].* 1) Department of Medical Genetics, School of Medical Sciences - Unicamp, Campinas, SP, Brazil; 2) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing; Unicamp, Campinas, SP, Brazil; 3) BRAINN - Brazilian Institute of Neuroscience and Neurotechnology.

Medical treatment is set on the average patient, which leads to inefficiency of some treatments in a proportion of the population. Precision Medicine (PM) has emerged as a concept in which scientific knowledge and technology will come together to provide the basis for more effective and personalized medicine. To implement PM, it is important to determine the genetic profile of a population as the distribution of rare and common variants may have different downstream implications in different population backgrounds. Therefore, we aim to determine the genomic profile of Brazilian individuals and compared it with data available from other population. We performed whole exome sequencing (WES) in 106 subjects and genotyped 264 individuals using the Affymetrix SNP 6.0 array. These are normal/reference individuals from the region of Campinas, SP, BRAZIL. Genomic data produced was deposited in two public databases (www.bipmed.org), which are part of the Brazilian Initiative on Precision Medicine (BIPMed). Alternative allele frequencies (AAF) were calculated for each variant on both datasets. On the SNP-array study we compared the frequency from 4 populations provided by the manufacturer. Data from WES was compared with populations from the Genome Aggregation Database. In the WES study, we found 624,137 variants (SNVs and INDELs), among them 68,149 (10.9%) were not detected in any other population. Furthermore, 45.51% of the variants detected have AAF of less than 1%, 22.44% of the variants have AAF between 1% and 5%. Comparing variants from the WES study with Clinvar we found 193 variants classified as pathogenic and 33 variants classified as likely pathogenic. In the SNP-array study, the correlation analysis and the PCA showed that our dataset is closer to the CEU population, distant from CHB, JPT and more distant from YRI. We also found 9 SNPs in which the more frequent allele identified in other populations was the less frequent allele in the Brazilians or vice versa. Our results clearly indicate that there are differences in the genomic distribution of both rare and common variants in Brazilian individuals as compared to other populations, and demonstrate the importance of constructing a larger and geographically diverse panel of reference genomic datasets for both rare and common variants, especially for populations which are underrepresented in the currently available public genomic databases.

2331F

Deep learning for reference-free inference of archaic local ancestry. *A. Durvasula[1], S. Sankararaman[1,2].* 1) Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA; 2) Department of Computer Science, UCLA, Los Angeles, CA.

   Statistical analyses of genomic data from diverse human populations have demonstrated that modern human populations trace a small proportion of their genetic ancestry to archaic hominins such as Neanderthals and Denisovans. These analyses were enabled by the availability of archaic genome sequences. Several studies have suggested that archaic admixture has been common in the history of populations even though the ancestral archaic populations have not been identified. This observation motivates the problem of reference-free archaic local ancestry inference, i.e., inferring segments of the genome that trace their ancestry to an archaic population even in the absence of reference archaic sequences. Previous attempts at reference-free archaic local ancestry inference have relied on a small number of features or summary statistics (such as S*), which have limited power and a high false positive rate. Recent advances in deep learning permit the learning of complex, non-linear features that can be useful in a number of inferential tasks. Here, we present a deep neural network for archaic local ancestry inference. The architecture consists of 8 hidden layers (200, 100, 75, 50, 50, 50, 2, 2 nodes) that outputs to 2 neurons, a batch normalization layer, and a softmax activation layer. Each hidden layer is followed by a rectified linear unit activation and a dropout layer. We use coalescent simulations to train the network on a number of features that summarize patterns of genetic variation, amplifying the signal of deep divergence expected from archaic introgression. This allows fast and accurate inference of archaic introgression. We applied this method to 5 genomes from a sub-Saharan African population in a preliminary analysis and find that an average of 2.03% (SD: 0.38) of their genomes is labelled as archaic. This result is in line with previous estimates (Hammer et al. 2011) using less data and lower powered methods, but increases the resolution with which we can understand archaic hominin admixture and demonstrates the potential of directly using present-day human genomes to learn about archaic admixture.

2332W

Variation and genetic control of mutation rates in house mice. *B.L. Dumont.* The Jackson Laboratory, Bar Harbor, ME.

   Mutation provides the ultimate source of all new alleles in populations, including variants that drive evolutionary adaptation and cause disease. At the same time, the *de novo* mutation rate is itself a quantitative genetic trait that displays striking differences between species and among individuals. Despite the central significance of this variation for genetics and evolution, little is known about the genetic causes of mutation rate heterogeneity or inter-individual variation in mutation rate. Toward these goals, I am conducting two parallel bioinformatic analyses of whole genome sequences from house mice. First, I am utilizing the Collaborative Cross (CC) 8-way recombinant inbred mouse panel as a forward evolution resource to study the accumulation of mutations over ~30 generations of organized outcrossing and inbreeding. I show that the number of accumulated mutations in the genomes of different CC lines varies 3.5-fold. This variation is driven, in large part, by the unique, dynamic genome captured in each CC breeding funnel. Building on this recognition, I perform a genome-wide scan for mutation rate modifiers and identify multiple putative mutator alleles that function in DNA repair and the cellular metabolism of genotoxic compounds. Second, I am mining high-quality genome sequences from 36 inbred laboratory strains to identify variants that are private to individual strains. These private alleles reflect recent germline mutation events that collectively mirror the action of mutational processes at work on a specific genetic background. I use this insight to show that the allelic spectrum of inherited mutations is variable among strains, with closely related strains sharing similar profiles. These findings reveal a genetic component to the nucleotide distribution of de novo mutations in the mouse genome. Together, these on-going investigations have unveiled marked complexity in germline mutation rate variation, with multiple genetic factors shaping both mutation frequency and spectrum.

2333T

Patterns of shared signatures of recent positive selection across human populations. *K.E. Johnson[1], B.F. Voight[2,3,4].* 1) Genetics and Gene Regulation Program, Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia PA; 2) Department of Systems Pharmacologty and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA; 4) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Scans for positive selection in human populations have identified hundreds of sites across the genome with evidence of recent adaptation. These signatures often overlap across populations, but the question of how often these overlaps represent a single ancestral event remains unresolved. Signals can be shared across populations, such as the lactase persistence haplotype across Europe; or unique to an individual population, like high altitude-adaptive mutations in Tibetans. If a single selective sweep spread across the globe, the same sweeping haplotype should appear in multiple populations, implying a common selective pressure across diverse populations and environments. Identifying such shared selective events is of fundamental interest, pointing to genomic loci and human traits important in recent history across the globe. Additionally, newly available genomic annotations could help attach these signatures to a potential gene and molecular phenotype that may have been selected across multiple populations. We performed a scan for positive selection using the integrated haplotype score (iHS) on 20 human populations from the 1000 Genomes Project (phase 3), and propose an additional correction for local recombination rate to iHS scans on genomic sequencing data. To create a catalog of shared (i.e., a common ancestral event) and unshared overlapping selective sweeps in these populations, we employed fastPHASE to cluster these haplotypes across populations. We observed more sharing than expected for every within-continent population pair, and with population pairs from every inter-continental grouping. The majority of shared sweeps occurred within a single continent (71%), though we did find instances of sweeps shared across multiple continents. Using additional genomic annotations, we connected multi-population overlapping sweep intervals with possible biological mechanisms at several loci. We identified potential new sites of adaptive introgression, including a shared sweep across Europeans and South Asians at the lincRNA *CT64*. At the glycophorin locus containing *GYPA/GYPB/GYPE*, associated with malarial resistance, we identified unshared sweeps in all continental groups studied. Finally, at the alcohol dehydrogenase cluster on chromosome 4, an established target of recent positive selection in East Asians, we identified an independent sweep in an African population whose haplotype is associated with alcohol dependence in African Americans.

2334F

The landscape of genetic variation in Estonians. *M. Kals[1,2], T. Nikopensius[1], R. Mägi[1], N. Tõnisson[1], A. Reigo[1], T. Esko[1,3], A. Metspalu[1,4].* 1) Estonian Genome Center, University of Tartu, Tartu, Estonia; 2) Institute of Mathematical Statistics, University of Tartu, Tartu, Estonia; 3) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia.

To describe genetic variation in the Estonian population, we combined 2,420 whole-genome sequencing (WGS) and non-overlapping 2,356 whole-exome sequencing (WES) samples from the Estonian Biobank of the Estonian Genome Center, University of Tartu (EGCUT). WGS individuals were selected randomly by county of birth, followed a PCR-free sample preparation. Libraries sequenced on the Illumina HiSeq X Ten with the use of 150bp paired-end reads to 30x mean coverage. WES samples were enriched for target sequences (Agilent Technologies; Human All Exon V5+UTRs) according to manufacturer's recommendations to a median depth of 70x Combination of WGS and WES data provides better insight to coding regions. EGCUT is a population-based biobank, containing almost 52,000 samples of the adult population (aged ≥18 years), which closely reflects the age, sex and geographical distribution of the Estonian population. From WGS data only, we identified more than 32M single-nucleotide variants (SNVs) and 2.5M short indels. Among 13M (1.8) novel SNVs (indels), 10M (1.1M) were singletons/doubletons, resulting in average 6,600 novel variants per individual. Furthermore, we compared the overlap of novel variants between 1000 Genome Project Phase 3 EUR panel, UK10K and the Genome of the Netherland Consortium (GoNL) and identified that >98% of the low-frequency ($0.5 \leq MAF < 5\%$) and common ($MAF \geq 5\%$) variants exist in the reference databases, but only <40% of the rare variants ($MAF < 0.5\%$, MAC=1 and MAC=2 excluded) were observed in the other European populations. Adding WES samples and focusing on only coding regions (total sample size N=4,776) increases the proportion of singletons/doubletons, but number of novel variants remained close to 40%. In summary, we have investigated the genetic variability in the Estonian population. We further estimate the number of expected actionable incidental findings per individual, profile structural variants and focus on functional variants.

2335W

Gene expression predictive performance varies across diverse populations. *L.S. Mogil, A. Badalamenti, H.E. Wheeler.* Biology, Loyola University Chicago, Chicago, IL.

For many complex traits, gene regulation is likely to play a crucial mechanistic role. How the genetic architectures of complex traits vary between populations and subsequent effects on genetic prediction are not well understood, in part due to the historical paucity of GWAS in populations of non-European ancestry. We used data from the MESA (Multiethnic Study of Atherosclerosis) cohort to study the genetic architecture of gene expression. The goal of this work is to optimize genetic prediction of gene expression within and across populations for use in gene based mapping methods like PrediXcan. Genotype and monocyte gene expression were available in African Americans (AFA, n=233), Hispanics (HIS, n=352), and Caucasians (CAU, n=578). We used elastic net modeling (alpha=0.5) with 10-fold cross validation to optimize genotypic predictors of gene expression in each population. The numbers of genes with 10-fold cross-validated R^2 >0.1 for each population were AFA=1782, HIS=1899, and CAU=1660. We found the best predicted gene, *CHURC1*, was the same across populations with an R^2 > 0.87 in each population. Out of 327 genes with an absolute value difference of R^2 >0.5 between AFA and CAU, 184 (56%) were well-predicted genes in AFA that were poorly predicted in CAU. The *TOP1MT* gene had an R^2 =0.64 in AFA and R^2=0.008 in CAU. The driving SNP with the highest weight in the AFA population (rs2467941) had a high MAF in AFA (0.375) but was not present in CAU. Across genes, population pairwise predictive performance was correlated (R^2 > 0.5) with the highest predictive performance correlation between HIS and CAU (R^2 =0.604), reflective of the European admixture proportions in the HIS cohort. Using genotype weights trained in MESA to predict gene expression in 1000 Genomes populations as a test set showed that a training set with ancestry similar to the test set is better at predicting gene expression in test populations. The AFA training set prediction for Yoruba (YRI) showed a mean predicted versus observed R=0.104, while the CAU training set prediction for YRI showed a mean R=0.068. This work enhances our knowledge of genetic mechanisms underlying gene expression and complex traits in all populations.

2336T

Adaptive eQTLs in human populations. *M. Quiver, J. Lachance.* School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA.

Adaptive evolution is often due to regulatory sequence change, and expression quantitative trait loci (eQTLs) have allele frequencies that can vary across populations. To identify adaptive loci in human populations, we combined continental estimates of allele frequencies with eQTL data from the Genotype-Tissue Expression (GTEx) project and RegulomeDB. After controlling for the effects of linkage disequilibrium, we calculated population branch statistics (PBS) at each polymorphic site. This allowed us to test whether the strength of selection differs for eQTLs that affect many of tissues, determine which tissues are more likely to be adaptive, and identify outlier eQTLs. We find that eQTLs are more likely to have large allele frequency differences between populations than random SNPs from the 1000 Genomes Project. However, this pattern is largely due to SNP ascertainment bias. It is hypothesized that pleiotropic loci are expected to evolve slowly. Along these lines, our results indicate that pleiotropic eQTLs are less likely to be locally adaptive in human populations. We also find that signatures of positive selection are stronger for tissues that are involved in immune and metabolic function. eQTLs that regulate expression in blood, liver, or skin are enriched for PBS outliers. By contrast, eQTLs that regulate expression in the cerebrum or female-specific tissues have a relative lack of PBS outliers. Scans of positive selection also reveal that the strongest PBS signal in many adaptive regions of the genome is an eQTL. For example, rs2814778 is an adaptive eQTL at the Duffy locus that has experienced strong positive selection for increased malaria resistance in Africa. Taken together, our results suggest that allele frequency changes at eQTLs have played an important role in human adaptation.

2337F

A comparative study of endoderm differentiation in humans and chimpanzees. *L.E. Blake[1], S.M. Thomas[1], J.D. Blischak[1], C.J. Hsiao[1], C. Chavarria[1], M. Myrthil[1], Y. Gilad[1,2], B.J. Pavlovic[1].* 1) Human Genetics, University of Chicago, Chicago, IL., USA; 2) Genetic Medicine, University of Chicago, Chicago, IL., USA.

There is substantial interest in the evolutionary forces that shaped the regulatory framework that is established in early human development. Progress in this area has been slow because it is difficult to obtain relevant biological samples. Inducible pluripotent stem cells (iPSCs) provide the ability to establish in vitro models of early human and non-human primate developmental stages. Using matched iPSC panels from humans and chimpanzees, we comparatively characterized gene regulatory changes through a four-day timecourse differentiation of iPSCs (day 0) into primary streak (day 1), endoderm progenitors (day 2), and definitive endoderm (day 3). As might be expected, we found that differentiation stage is the major driver of variation in gene expression levels, followed by species. We identified thousands of differentially expressed genes between humans and chimpanzees in each differentiation stage. Yet, when we considered gene-specific dynamic regulatory trajectories throughout the timecourse, we found that 75% of genes, including nearly all known endoderm developmental markers, have similar trajectories in the two species. Interestingly, we observed a marked reduction of both intra- and inter-species variation in gene expression levels in primitive streak samples compared to the iPSCs, with a recovery of regulatory variation in endoderm progenitors. The reduction of variation in gene expression levels at a specific developmental stage, paired with overall high degree of conservation of temporal gene regulation, is consistent with the dynamics of developmental canalization. Overall, we conclude that endoderm development in iPSC-based models are highly conserved and canalized between humans and our closest evolutionary relative.

2338W

Determining the distribution of deleterious variation in population isolates using local ancestry and pedigree data. *J. Mooney[1], C. Huber[2], S. Service[3], J. Sul[4], C. Marsden[2], N. Freimer[3], K. Lohmueller[1,2], Costa Rica/Columbia Consortium for Genetic Investigation of Bipolar Endophenotypes.* 1) Department of Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Department of Ecology & Evolutionary Biology, University of California Los Angeles, Los Angeles, CA; 3) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA; 4) Department of Psychiatry and Biobehavioral Sciences, Semel Center for Informatics and Personalized Genomics, University of California Los Angeles, Los Angeles, CA.

Isolated populations generally exhibit less genetic diversity and greater genetic drift, which can impact patterns of deleterious genetic variation. However, most isolates examined to date were founded from a single ancestral population. Here we investigate genomic diversity of recently admixed population isolates from Latin America. Specifically, we use whole genome sequence data paired with extended pedigree data for Costa Rican and Colombian population isolates, which were derived from a limited number of founders during 16th-17th centuries, and recently expanded to about 3 million individuals. We used long runs of homozygosity, 2.0Mb or greater, to study how recent demographic history shaped the genomic background of Costa Rican and Colombian individuals. We found that the proportion of the genome that falls within a run of homozygosity (ROH) in Costa Rican and Colombian individuals was similar to 1000 Genomes CLM individuals. However, the individuals with the largest proportions of their genome within a ROH are from isolates. Furthermore, these individuals have some of the largest inbreeding coefficients. Additionally, we observed an increase in the number of homozygous derived deleterious variants within runs of homozygosity. Simulations suggest the significant enrichment of deleterious variation was due to recent inbreeding. Thus, we leveraged extended pedigree data from these isolates to further understand the impact of recent inbreeding on deleterious variation. We identified a significant correlation with an individual's pedigree inbreeding coefficient (F) and proportion of their genome within a ROH. Moreover, pedigree F was significantly correlated with the number of derived deleterious alleles within an individual's genome, suggesting that recent inbreeding has hindered removal of deleterious variants by purifying selection. Finally, we explored how admixture affected genetic diversity in the isolates. We found a positive correlation between pedigree F and the proportion of European (CEU) ancestry. Conversely, pedigree based F was negatively correlated with Native American and African ancestry. Overall, our results suggest that deleterious variants in population isolates may be affected by recent inbreeding and admixture.

2339T

The composition and intensity of *de novo* mutations in the Lithuanian exome. *L. Pranckėnienė, A. Jakaitienė, L. Ambrozaitytė, I. Domarkienė, I. Kavaliauskienė, R. Meškienė, T. Rančelis, Z.A. Kučinskienė, V. Kučinskas.* Department of Human and Medical Genetics, Vilnius University, Vilnius, Lithuania.

Family-based whole-genome and whole exome sequencing data allows the identification of *de novo* mutations that provide more direct measures of mutation rates. According to the Dutch, Danish, and Icelandic genome projects rate of *de novo* single nucleotide variant is about 1.2×10^{-8}–1.5×10^{-8} per nucleotide per generation and increase with father's age. Point mutations can have different effects depending on which site of the genome they occur. A mutation that damages protein structure does not necessarily lead to a detectable human-disease phenotype, and a mutation that predisposes an individual towards a disease is not necessarily evolutionarily deleterious. The main interest of this study was to evaluate the composition and intensity of *de novo* single nucleotide mutations based on the Lithuanian whole exome data. The data set consisted of the Lithuanian population samples. Sequencing data of 48 parent-offspring trios exomes generated by sequencer SOLiD 5500™ and primary analysis performed by Lifescope™. *De novo* single nucleotide variants (DNV) called by two alternative programs VarScan and VarSeq™. Called DNV were filtered, manually reviewed by the IGV and validated by Sanger sequencing. Functional annotation of all identified *de novo* single nucleotide variants were performed using ANNOVAR. Preliminary results of DNV rate before validation step is 7.7×10^{-8} with 95% CI [6.4×10^{-8};9.2×10^{-8}] for VarSeq, and 1.2×10^{-7} with 95% CI [1.1×10^{-7};1.4×10^{-7}] for VarScan. DNV rate varies between families and is higher for the Lithuanian exome data as compared to 1.5×10^{-8} estimated in other studies. The results raise the question of whether the reported higher DNV rate is at least partially due to a local sequence context and/or possible natural selection on the exome. Therefore, ongoing work is focused on the analysis of probable DNV clusters, influence of sequence context on DNV and the influence of mutation rates on human-chimpanzee divergence based on comparative genomics models. This is the first attempt to evaluate a composition and intensity of *de novo* variants for trios from the Lithuanian population. This study supported by the LITGEN project (VP1-3.1-□MM-07-K-01-013) funded by the European Social Fund under the Global Grant Measure.

2340F

Testing for local adaptation in populations of *Drosophila melanogaster* for olfactory receptor genes. *V. Ramesh, J. Pool.* Genetics, University of Wisconsin-Madison, Madison, WI.

The development of the field of population genetics and genomic sequencing has helped scientists better understand an organism's evolutionary history and genetic variation among populations, yet the correlation between genetic variation and the extent to which it can lead to evolutionary change remains unknown. Of particular interest is understanding the genetic architecture underlying variation between populations in regards to the sense of olfaction, as it is not fully understood. To bridge this grey area, this research project tested for the presence of local adaptation which may have acted upon 11 populations of *Drosophila melanogaster* for genes coding for olfactory receptors. This evidence was obtained by analyzing fully sequenced genomes of subjects representing 11 populations of *D. melanogaster* spread throughout the sub-Saharan African range (including a novel wilderness population) and Europe, through window and SNP F_{ST} statistics. From this analysis, the data collected was converted to quantiles distinguishing unusually high F_{ST} values in order to identify loci with elevated genetic differentiation that could result from local adaptation. Additionally, a new method was developed using Circos software to aid in graphically depicting genetic differentiation. Preliminary results indicate the lines of variation between the tested populations of *D. melanogaster* as reflective of their geographical deviation from the ancestral population, and located several candidate genes (the tandem paralogs *Or22a* and *Or22b*, plus *Or19b*, *Or33a*, *Or65b*, and *Or67a*) that are likely experiencing natural selection in select populations.

2341W

Genome-wide population analysis of 2,543 microsatellites and STR-SNP haplotypes reveals a novel class of highly diverse polymorphisms. *G. Shin[1], H. Lee[1], S.M. Grimes[2], M. Kubit[1], I.P. Vogelaar[1], H.P. Ji[1,2].* 1) Division of Oncology, Stanford University School of Medicine, Stanford, CA 94305; 2) Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304.

Microsatellites, also referred to as short tandem repeats (STRs), are multiallelic in terms of germline variation as a result of higher mutation rate than SNPs. As a result, STR genotyping is highly informative when characterizing genetic populations. However, significant gaps remain in our knowledge about the population characteristics of microsatellites, the linkage disequilibrium structure of microsatellites and STR-SNP haplotype structure in the human population. Addressing all of these questions, we developed and applied novel targeted sequencing technology (STR-Seq) [1], which can simultaneously genotype thousands of microsatellite loci and phased proximal SNPs (i.e. STR-SNPS haplotypes) with significantly higher accuracy than all other methods including whole genome sequencing. STR-Seq uses single-molecule sequencing in combination with targeted in vitro CRISPR-Cas9 fragmentation. For a given individual, one accurately genotypes thousands of microsatellite in a single assay. For this study, we evaluated 2,543 microsatellite loci from a population of 1,004 individuals [2]. Our analysis included 436 Marshfield loci [3], which has been used to characterize the population sample and an additional 1,915 loci where a proximal SNP was positioned within 100bp of the STR. Thus, over 2.5 million loci were evaluated across the human population. Overall, we identified approximately 2,000 STR genotypes and 1,000 STR-SNP haplotypes per individual. STR-SNP linkage was extremely low (mean r^2 < 0.1). We discovered an entirely new class of microsatellite/STRs which are highly polymorphic as noted by having 20 or more alleles. Finally, we identified a significant number of STR-SNP haplotypes and describe the geographical and population differences of these previously undescribed, novel genetic markers. **Reference** 1. Shin, G., *et al.*, **Nat Commun**, 2017, 8:14291. 2. Rosenberg, N.A., *et al.*, **Science**, 2002, 298(5602):2381-2385. 3. Pemberton, T.J., *et al.*, **BMC Genomics**, 2009, 10:612. .

2342T

Inverse correlation between mutational and selective forces in human coding regions with distance from gene ends leads to opposite patterns of synonymous and non-synonymous variant prevalence. *Y. Waldman[1], E. Halperin[2,3], A. Keinan[1,4,5].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA; 3) Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, Los Angeles, CA; 4) Cornell Center for Comparative and Population Genomics, Cornell University, Ithaca, NY; 5) Center for Vertebrate Genomics, Cornell University, Ithaca, NY.

Evolutionary rate varies not only between proteins but also within proteins. While studies, which mainly focused on non-synonymous variation, have associated intra-protein variation with functional and structural constraints, a large portion of the variation is not explained by current models. Here we examine both synonymous and non-synonymous variation in coding regions in humans and the mechanisms behind them. We observed clear but opposite patterns for synonymous and non-synonymous variation. For example, comparing the sequences of more than 10,000 human-mouse orthologous genes, we find that non-synonymous divergence rate decreases as a function of the distance from both '3 and '5 ends of genes while synonymous divergence increases as a function of the distance from both ends of genes. We obtained similar results for other organisms as well (Drosophila, Saccharomyces, Candida and more). Next, we examined what are the underlying mechanisms behind these opposite patterns by analyzing millions of synonymous and non-synonymous variants observed in more than 140,000 fully sequenced human exomes. Results indicate that the depletion in non-synonymous variation as a function of the distance from gene ends is a result of non-uniform selection: non-synonymous variants far from gene ends are much more likely to be damaging (ClinVar) and are therefore under selection. Functional and structural annotation data only partially explains the distribution of damaging non-synonymous variation across coding regions, suggesting that current annotation data is incomplete and/or that other constraints contribute to the observed pattern. Further support comes from non-synonymous somatic mutations in cancer (COSMIC) that show increase in frequency as a function of the distance from gene ends, similar to synonymous variants, presumably because the vast majority of these somatic mutations are under no or weak selection. Similarly, non-synonymous variants present in very low frequencies in human populations (gnomAD), which are presumed to be relatively recent and therefore with less time for selection to operate, also exhibit synonymous variation pattern. These and other results suggest that mutational forces increase the genetic variation as a function of the distance from gene ends while selection acts in opposite direction to remove variation located far from gene ends, hence providing additional and important angle in our understanding of gene architecture.

2343F

Patterns of genetic variation within the Genome Russia Project. *S. OBrien, G. Tamazian, D. Zhenakova, S. Kliver, N. Cherkasov, I. Evsukov, P. Dobrynin, A. Gorbunova, S. Kolchanova, A. Shevchenko, M. Rotkevich, A. Yurchenko, S. Sidorov, A. Komissarov, K. Krasheninnikova, S. Malov, A. Logachev, S. Simonov, D. Polev, T. Davydova, S. Sivtseva, A. Glotov, V. Osakovsky, A. Novozhilov, V. Brukhin. Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St Petersburg Ru,.*

Recently many national genomic projects highlighting the diversity and similarity of world populations have begun, providing insight into population history and aiding in medical genomics. Although Russia is the largest country in the world, we have not formally participated in large sequencing projects. To overcome this, Genome Russia Project was initiated to generate whole genome sequences of more than 2,500 samples coming from diverse populations throughout Russia. The principal aims of the project are: 1.)Characterize of genetic variation of different Russian ethnic groups; 2.)Develop reference database of DNA variation within Russia that can be assessed for medical import;3.)Identify variants that cause known diseases.4.) Develop a Russian HapMap; 5.) Explore population natural history in Russia and Eurasia. Sixty trio genome samples representing population groups from 3 regions(Novgorod, Pskov and Yakutia regions) were sequenced and analyzed. We have identified > 10.5 million SNPs and >2.8 million short indels in three populations . Approximately 95% of these are listed in dbSNP or 1000G Project database s. We combined the Russian SNP results with genomic variants from 206 recently released Eurasian genomes to assess the PCA plot of population structure of our samples within Eurasian populations. We annotated the identified variants, noting differences in MAF as compared to 1000G data for known variants . To investigate medically relevant variants we note significant MAF differences of trait- and disease- associated SNPs in Russian versus 1000G populations. Consider lactose intolerance which is prevalent in Asian peoples (~90%), while in Europeans is rather low (~5% in Northern Europe). SNP rs4988235 located in LCT locus is a functional SNP that mediates LCT activity, and its G allele tags the lactose intolerant haplotype. This allele is more frequent in European Russian populations as compared to Europeans from 1000G CEU (MAF in Pskov + Novgorod is 0.59; MAF in CEU is 0.26; p-value = 9.40 x 10-6) and pronounced in Yakutia (MAF ~0.96; p-value of the difference of Yakut samples vs 1000G CEU is 1.48 x 10-12). Thus , Novogrod and Pskov are more lactose intolerant then Europeans, and Yakut have an even higher prevalence of this condition. In sum, we identified variants shared and specific for each population, annotated their their presence in known databases; potential pathogenicity; known association with diseases and complex traits.   .

2344W

1000 high coverage whole-genome sequences representative of the Taiwanese population from Taiwan Biobank. *M. Su[1,2], C. Lin[1,2], C. Hsiung[1,2], H. Chu[1,2], P. Wu[1,2], C. Shen[1,2], F. Liu[1,2]. 1) Taiwan Biobank, Academia Sinica, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.*

The complex migration history (gene flow) among the Taiwanese populations and lack of Taiwanese samples in the 1000 genome reference panel result in lower imputation accuracy, and further hamper genetic research in Taiwan. The Taiwan Biobank (TWB) reports the whole-genome sequences of 1,000 healthy Taiwanese individuals and constructs a population-specific reference panel (1KTWB) for human genetic variation. We sequenced 1,000 samples by PCR-free sequencing with more than 30X coverage genome-wide, aiming to characterize genetic variation down to 0.1% minor allele frequency in the Taiwanese population. Samples were sequenced using Illumina HiSeq 2500 and 4000 sequencers according to the manufacturer's protocol. Sequence reads were aligned to the human reference genome with decoy sequences (version GRCh37d5/hg19d5) using Burrows Wheeler Aligner (BWA-MEM) 0.7.15-r1140. Single nucleotide polymorphisms (SNP) and insertion and deletion (INDEL) calling were performed on all sample's gVCF files jointly using GATK HaplotypeCaller v3.6. Variant Quality Score Recalibration (VQSR) was used to filter variant sites with know SNPs and INDEL variations datasets. 28 million, including 16 million novel, SNPs were characterized. We also detect structural variants, including 3.4 million insertions and deletions. Here we demonstrate the value of high coverage sequencing to boosts imputation accuracy at rare and low-frequency variants in Taiwanese population.

2345T

NHGRI Sample Repository for Human Genetic Research: Cell lines and DNA from the 1000 Genomes and HapMap collections. *E.M. Kelly, T. Dixon, B. Martinez, A.M. Resch.* Coriell Institute for Medical Research, NJ.

Understanding genetic variation across diverse human populations is critical for reconstructing human evolutionary history and identifying the genetic basis of disease. The goal of the NHGRI Sample Repository for Human Genetic Research is to support studies of the human genome, including genetic variation in human populations. The NHGRI Sample Repository maintains and distributes a collection of well-characterized cell line and DNA samples collected for the 1000 Genomes and International HapMap Projects. Whole blood samples collected from anonymized and de-identified donors were contributed with broad consent to data release for future studies including: genotyping and sequencing, gene expression, proteomic studies, and all other types of genetic variation research. Several distinct populations are represented in the 1000 Genomes and HapMap collections, making it one of the most comprehensive and unique datasets of its kind. The NHGRI Sample Repository distributes specimens from these collections to researchers around the globe. Samples are available to qualified researchers from academic, non-profit, government, and commercial organizations. The full catalog can be accessed at *https://catalog.coriell.org/NHGRI.*.

2346F

Increasing signal and refining population annotations using a common haplotypes co-association network. *L. Doroud[1], S. Hateley[1], O. Schaedel[1], H. Guturu[1], K. Noto[1], C. Ball[1], K. Chahine[2].* 1) AncestryDNA LLC, San Francisco, CA; 2) Ancestry.com DNA, LLC, Lehi, UT.

Common haplotypes are genomic regions shared by multiple individuals likely due to population structure rather than recent identity by descent. While segregation of haplotypes is correlated with ethnicity and genetic communities, analyzing these correlations individually for each haplotype is noisy and cumbersome. Taking advantage of the fact that many haplotypes frequently co-occur within populations, we can cluster co-occurring haplotypes to reduce noise and find refined sub-populations for more meaningful annotation. We propose a novel graph-based approach to group co-associated common haplotypes, ultimately allowing for better annotation of genetic communities. From our database of over four million customers, we used a dataset of over one million customers who have consented to research and constructed a co-occurrence network of observed common haplotypes, revealing 'haplotype families' within sub-populations. Identifying co-occurrence patterns of common haplotypes in the human population at large scale can reveal hidden insights into our species' recent history.

2347W

Partitioning heritability of low-frequency variants reveals relative strength of negative selection across functional annotations. *S. Gazal[1,2], A. Ganna[2,3], A. Schoech[1,2], P.R. Loh[1,2], A. Gusev[1,2], T. Esko[2,4], A. Palotie[2,3,5], B.M. Neale[2,3], S. Sunyaev[2,6], H.K. Finucane[1,2,7], A.L. Price[1,2].* 1) Department of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA,USA; 2) Broad Insitute of MIT and Harvard, Cambridge, MA, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 5) Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki FI- 00014, Finland; 6) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 7) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA.

Although it is widely known that common variant heritability is enriched in functional annotations and concentrated in non-coding variants, little is known about the distribution of the heritability of low frequency variants ($0.5\% \leq MAF < 5\%$). Here, we partitioned the heritability of both common and low-frequency variants in 41 common diseases and complex traits from the UK Biobank (average N = 104K) across a broad set of functional annotations. We accomplished this by extending stratified LD score regression (Finucane et al. 2015 *Nat Genet*) to low-frequency imputed variants, incorporating estimates of imputation accuracy and a UK10K reference panel, and showing via simulations that this approach produces robust results. Meta-analyzing enrichment results across 26 independent traits, we observed that the low-frequency variant enrichments (LFVE) of each functional annotation are highly correlated to the common variant enrichments (CVE) (r = 0.70; s.e. = 0.13), and tend to be even larger for annotations with large CVE (regression slope = 1.78; s.e. = 0.26). We observed very large LFVE for non-synonymous variants (0.45/0.27 % of low-frequency/common variants; LFVE = 42.0x vs. CVE = 7.1x; P = 3.9 x 10^{-8} for difference), and large LFVE for coding+UTRs variants (2.49/2.10 % of variants; LFVE = 12.2x vs. CVE = 4.7x; P = 2.7 x 10^{-4}) and regions conserved in primates (2.82/2.12 % of variants; LFVE = 16.5x vs. CVE = 9.4x; P = 4.8 x 10^{-3}). We observed an even stronger difference for some traits, such as non-synonymous variants (LFVE = 53.1x vs. CVE = 7.8x; P = 1.2 x 10^{-4}) and coding+UTRs variants (LFVE = 24.7x vs. CVE = 7.7x; P = 2.3 x 10^{-3}) for height. We performed forward simulations using SLiM2 (Haller & Messer 2017 *MBE*) to link these results to the action of negative selection. We determined that the CVE of an annotation primarily depends on its proportion of sites under selection, whereas its LFVE/CVE ratio primarily depends of the strength of selection. This implies that non-synonymous variants have a lower proportion of deleterious variants than regions conserved in primates, but a higher mean selection coefficient for deleterious variants. Finally, our analyses demonstrate that identifying non-coding annotations with high LFVE will help prioritize rare non-coding variants in whole-genome sequencing studies of common diseases and complex traits.

2348T

***HLA-A* extended promoter and coding variability in a Brazilian population sample by using massively parallel sequencing.** *T.H.A. Lima[1], I.O.P. Porto[1], L.C. Veiga-Castelli[2], C.T. Mendes-Junior[3], E.A. Donadi[4], E.C. Castelli[5].* 1) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unity, School of Medicine, Botucatu, State of São Paulo, Brazil, MS; 2) Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil, PhD; 3) Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Estado de São Paulo, Brasil, PhD; 4) Department of Medicine, School of Medicine of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, State of São Paulo, Brazil, PhD; 5) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unity, School of Medicine, Departmet of Pathology, Botucatu, State of São Paulo, Brazil, PhD.

HLA-A is one of the most polymorphic genes of the human genome. The HLA-A molecule is responsible for displaying small peptides at the cell surface, which in turn may be recognized by T cell receptors as foreign or damaged proteins. Thus, these are critical molecules for the activation of adaptive immune responses. Here we evaluate the *HLA-A* extended promoter and coding sequence variability and haplotypes in a Brazilian population sample (408 individuals from the Southeast Brazil) by using massively parallel sequencing and freely available softwares. Allele, promoter, coding and extended haplotype frequencies, plus the Linkage Disequilibrium (LD) along the *HLA-A* loci, were also calculated. The *HLA-A* variability between nucleotides -1605 and -1 were considered as encompassing the *HLA-A* extended promoter (only the last 300 bases of this segment have been considered at the IPD-IMGT/HLA database so far). At the extended promoter, we detected 100 variable sites, 90% of them presenting polymorphic frequencies, and all arranged into 25 different haplotypes. Each coding *HLA-A* allele group was associated with a unique or a small group of closely related extended promoter haplotypes. The most frequent haplotype detected was associated with coding alleles from the HLA-A*02 group. At the coding region (from -300 to +3203, which is the region tracked by the IPD-IMGT/HLA database), we detected 90 different sequences. Among them, 51 are identical to previous described sequences, and the remaining 39 may be considered new *HLA-A* coding alleles. However, these new sequences represent only 1.34% of the coding sequences here detected. The low number of haplotypes detected for the extended promoter and the close relationship among these promoters and the coding alleles are probably consequences of the high Linkage Disequilibrium (LD) along the entire *HLA-A* loci, with many variable sites in complete LD with each other. In addition, the presence of few but highly divergent haplotypes is compatible with the balancing selection signatures already described for this gene. Functional evaluations are still necessary to completely understand if these divergent extended *HLA-A* promoters are related to differential expression patterns. This study was supported by FAPESP/Brazil, Grant 2013/17084-2.

2349F

MC1R regulatory and coding polymorphisms and pigmentation in an admixed population from Brazil. *C.T. Mendes-Junior[1], A.L.E. Pereira[1,2], L. Marcorin[2], M.L.G. Oliveira[2], N.C.A. Fracasso[2], G. Debortoli[2], G.V. Silva[1], J.D. Massaro[3], A.L. Simões[2], E.A. Donadi[3], E.C. Castelli[4].* 1) Departamento de Química, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, SP, Brazil; 2) Departamento de Genética, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 3) Departamento de Clínica Médica, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 4) Departamento de Patologia, Universidade Estadual Paulista (FMB-UNESP), Botucatu, SP, Brazil.

Melanogenesis, the biochemical pathway for human pigmentation, is extremely complex and involves over 120 genes. Different skin, eye and hair phenotypes reflect the accumulation of melanin, which displays two main conformations: eumelanin, a darker pigment that ranges from black to brown, and pheomelanin, a lighter pigment that ranges from yellow to red. The activation of the melanocyte membrane-bound MC1R protein activates a series of reactions that stimulates eumelanin synthesis, while its inhibition results in the production of pheomelanin. Polymorphisms spread across the *MC1R locus* may influence the balance of pheomelanogenesis and eumelanogenesis. The aim of the present study is to identify *MC1R* polymorphisms and haplotypes that are associated with phenotypic traits. A sample of 340 unrelated individuals was obtained from the region of Ribeirão Preto city, northeast region of the São Paulo state, Brazil. The samples were stratified according to eye, skin and hair phenotypes. DNA libraries (HaloPlex Target Enrichment System, Agilent) including the regulatory and coding regions of *MC1R* were submitted to next-generation sequencing (MiSeq, Illumina). Considering the whole extension of the gene (chr16:89978527-89987385; hg19), 120 polymorphic sites were identified. After a rigorous filtering process for removing low quality genotypes and *singletons*, haplotypes encompassing 58 polymorphic sites were obtained. Eighteen SNPs presented phenotype associations, including six (rs11547464, rs1805005, rs1805006, rs1805007, rs1805008, rs1805009) that are currently used in the DNA-based eye and hair color prediction model called HIrisPlex. In general, associations hereby identified proved statistically very strong, reaching *p*-values as low as 2.68×10^{-16} and Odds Ratio of 382.500 for rs1805007*T and red hair. Many polymorphisms coincide with transcription factor or microRNA binding sites. In addition, associations of whole-gene haplotypes with blue eyes, red hair and freckles were also identified. Thus, the characterization of both coding and regulatory regions of *MC1R* corroborates its role in human pigmentation pathway, and highlights the regulatory mechanism encompassing this gene, as well as the impact that polymorphisms have in the pigmentation of admixed individuals. FINANCIAL SUPPORT: FAPESP (Grant 2013/154470) and CNPq/Brazil (Grant 448242/2014-1 and Fellowship 309572/2014 -2).

2350W

Allele frequencies of pathogenic single nucleotide variants in a Japanese population based on a whole-genome reference panel of 2,049 individuals. *Y. Yamaguchi-Kabata, J. Yasuda, Y. Suzuki, N. Fuse, F. Katsuoka, Y. Kawai, H. Kawame, K. Kojima, M. Nagasaki, T. Mimori, O. Tanabe, K. Kinoshita, S. Kure, M. Yamamoto, ToMMo Study Group.* Tohoku Medical Megabank Organization, Tohoku University, Aoba-ku, Sendai, Japan.

Clarifying population frequencies of pathological variants is essential to construct infrastructure for genomic medicine for different populations; however, rare variants that may have biological or medical effects have not been clarified for the Japanese population. Here, by analyzing a whole-genome reference panel of a Japanese population based on 2,049 individuals (2KJPN, available at iJGVD; http://ijgvd.megabank.tohoku.ac.jp/), we characterized the genomes of Japanese individuals in terms of allele frequency of functional and pathological variants. Among the 28 million autosomal single nucleotide variants (SNVs) in 2KJPN, 6,739 SNVs were identified as pathologically annotated variants that overlap with known pathogenic variants in the Human Gene Mutation Database. In addition, we found that missense and nonsense SNVs in 2KJPN included candidates of expected pathogenic variants. By focusing on 32 genes for 17 congenital metabolic diseases for newborn screening (NBS) in Japan, we identified reported pathogenic variants and estimated their carrier frequencies by variant filtering based on variant annotations and allele frequencies. Most of the NBS genes showed lower risk allele frequencies in 2KJPN than in those with European ancestries, which may help explain the genetic basis for lower incidence rates of congenital metabolic disorders in Japan. However, there are a few exceptions: *PCCA* for propionic acidemia and *SLC25A13* for citrullinemia showed higher risk allele frequencies in 2KJPN than in those with European ancestries. In addition, we identified genome-wide pathogenic SNVs showing higher allele frequencies in 2KJPN than in other ethnic groups, and these SNVs included clinically important variants (e.g. CDH1 for hereditary diffuse gastric cancer, APRT deficiency and *CETP* for cholesterol ester transfer protein deficiency) for personalized medicine and prevention for the Japanese. We are also working on evaluation of pathogenic variants of actionable genes recommended by American College of Medical Genetics and Genomics for incidental or secondary finding, by a more careful approach. Our results and ongoing activities on variant curation would lay the foundation for i) evaluating the relationships of genomic variants and disease prevalence, and ii) improving diagnostic strategies and genetic testing in Japan and East Asia.

2351T

Substantial fraction of genes under recessive selection illuminates a missing component of human variation in population genetics and model organism studies of human disease. *D.J. Balick[1], D. Jordan[2], S. Sunyaev[1], R. Do[2].* 1) Genetics Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

Much of our knowledge of the biology and prevalence of recessivity in humans stems from model organism research, under the assumption that the underlying mechanisms are highly conserved. However, relevant experimental techniques cannot be performed in humans directly. Complementary information about phenotypic recessivity in humans comes from pedigree studies, which are costly and suffer from ascertainment bias. Beyond phenotypes, patterns of recessive natural selection in humans must be inferred from highly non-equilibrium natural population data, and are thus fundamentally harder to obtain. Notably, existing population genetic and conservation statistics are largely insensitive to recessive selection, due to a reliance on heterozygosity and fixation rates that are similar in neutral and recessive variation. This leads to an unknown number of functionally important, yet uncharacterized genes under strong selection. To identify this "missing" component of biologically and medically relevant genes under strong recessive selection, we developed a statistic specifically tuned to distinguish between recessive and non-recessive variation under strong selection, and to further distinguish recessivity from weaker selection and neutrality. Using data from 34500 European individuals from the ExAC study, we classified "recessive genes" for which protein damaging variation is severely deleterious only in homozygote form. We find that a substantial fraction of human genes are under recessive strong selection, suggesting a significant impact on both the global distribution of selective effects and a wide range of human genetic applications. To assess the relevance of a large class of recessive genes to analogs of human genetics in other species, we estimate the fraction of recessive genes in orthologs in several model organisms. Relative to all genes under strong selection in humans, recessive genes are preferentially lost in orthologs, with highly significant depletions in distant species beyond vertebrata. This impacts studies of human disease in model organisms that rely on the identification of an ortholog. Genes used in fly models of human disease exhibit a stronger depletion than in all fly orthologs. This suggests a generic bias against human recessivity in studies of both Mendelian and complex human disease in distant model organisms, mirroring the large fraction of genes under strong recessive selection missing from human population genetic inferences.

2352F

Rapid detection of identity-by-descent from whole genome sequence data. *J.E. Hicks[1,2], M.A. Province[1].* 1) Division of Biostatistics, Washington University School of Medicine, Saint Louis, MO; 2) Division of Statistical Genomics, Washington University School of Medicine, Saint Louis, MO.

Detection of shared genome regions identical-by-descent (IBD) from a common ancestor has broad applications in population genetics and genetic epidemiology. Most existing methods for detecting IBD from genotypes were developed for the common variants found on SNP arrays. Because they only examine common variants, these methods require many markers to declare IBD and cannot reliably detect the more distant coancestry that results in IBD segments shorter than 2 megabases. Explosive population growth in recent human history has resulted in an excess of rare variants across the genome. Since they tend to have arisen recently, the presence of a rare variant in two individuals suggests recent coancestry at that locus. By focusing on rare variants, small IBD segments can be more reliably detected. Whole-genome sequencing (WGS) provides access to both rare and common SNPs throughout the genome. However, analyzing WGS data comes with two major obstacles: the large number of variants identified and the presence of error in them. We present a new algorithm for detecting IBD, **A**daptive **D**etection of **I**BD **O**ver **S**equences (ADIOS). In two individuals, only a small subset of sites are informative of the underlying IBD state: rare variants found in both individuals (suggests IBD), and sites where both individuals are homozygous but for different alleles (excludes IBD). These sites differ for each pair of individuals. ADIOS determines pairwise IBD states by including only these variants in a Hidden Markov Model that adjusts for genotyping error and does not require phase information. By restricting to a subset of markers, determining IBD with ADIOS is both CPU and memory efficient. We simulated IBD in pairs of individuals (n=5000) from the chromosome 10 data in the 1000 Genomes Project's European samples. For each segment size tested, a region of specified size from a randomly selected individual was copied onto two individuals and tested. ADIOS had high power to detect IBD (46.5% of IBD bases detected for 500 kb segments, 76.1% for 1 Mb segments, 91.4% for 2 Mb, 94.7% for 3 Mb). To test false positive rate, a dataset (n=100) of null data was generated by joining 10kb segments from the 1000 Genomes data. ADIOS had a low false positive rate, with 5.5% of examined bases incorrectly shown as IBD. ADIOS shows great potential to facilitate rapid and accurate identification of IBD for rare causative alleles which has become the new genomic focus for complex traits in the WGS era.

2353W

Mutation rate estimation from population data. *X. Tian[1], B.L. Browning[1,2], S.R. Browning[1].* 1) Department of Biostatistics, University of Washington, SEATTLE, WA; 2) Department of Medicine, University of Washington, SEATTLE, WA.

We present a new method for mutation rate estimation based on segments of identity by descent (IBD) shared among multiple individuals. The sharing of IBD segments among multiple individuals makes it possible to distinguish true mutations from genotype errors. Our method is applicable to sequence data with or without close relatives. We first assess the performance of our method on simulated data with known phasing information. We simulated genome-wide data having a mutation rate of $1.25*10^{-8}$ per base pair per generation, a constant population size of 10,000 and a sample size of 400 diploid individuals. The analysis of simulated data indicated that our method is unbiased and has high precision. With true IBD segments, actual effective population size, and no genotype error, our method's mutation rate estimate has a standard error of $0.01*10^{-8}$. With inferred IBD segments, inferred effective population size, and 0.01% genotype error added to the simulated data, our method's mutation rate estimate has a standard error of $0.02*10^{-8}$. We next applied our method to Framingham Heart Study data from the NHLBI TOPMed project. In this dataset, approximately 4,100 individuals of European ancestry across 3 generations are whole genome sequenced. We identified 701 complete parent-offspring trios. Parental information determines the haplotype phase in the 701 trio offspring at most positions. From the analysis on the 701 phased trio offspring, we estimated the genome average mutation rate to be $1.10*10^{-8}$ per base pair per generation with a 95% confidence interval of $[0.91, 1.24]*10^{-8}$ per base pair per generation.

2354T

Extremely rare variants reveal patterns of germline mutation rate heterogeneity in humans. *S. Zöllner[1,4], J. Carlson[2], J. Li[2,3], the BRIDGES Consortium.* 1) Dept Biostatistics, Univ Michigan, Ann Arbor, MI; 2) Dept Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Dept Human Genetics, University of Michigan, Ann Arbor, MI; 4) Dept Psychiatry, University of Michigan, Ann Arbor, MI.

Precise estimates of the single-nucleotide mutation rate and its variability are essential to the study of human genome evolution and genetic diseases. However, estimates using common variants are biased by selection and biased gene conversion while analyzing de novo variants provides insufficient observations to consider sequence context. Here we use ~36 million singleton variants observed in 3,716 whole-genome sequences to characterize the heterogeneity of germline mutation rates across the genome. These singletons arose very recently in the population, and are thus largely unaffected by confounding evolutionary factors. We show that nucleotide context is the strongest predictor of mutability, with mutation rates varying by >650-fold depending on the identity of three bases upstream or downstream of the mutated site. We estimate up to 60-fold difference between non-CpG motifs within a single mutation category. Histone modifications, replication timing, recombination rate, and other local genomic features further modify mutability. We see clear evidence of interaction between nucleotide context and genomic features: Presence of a genomic feature can increase or decrease the mutation rate, dependent on the sequence context. Based on our model, we generate a genome-wide distribution of mutation rates that spans over 4 orders of magnitude. We evaluate the estimated models in an independent dataset of ~46,000 *de novo* mutations and show that singleton-based estimates provide a more accurate prediction of the mutation patterns than estimates based on common variants used in previous approaches. Incorporating the effects of genomic features further improves the prediction; combined, sequence motives and genomic features account for ~88% of the explainable variance. Finally, we examine the distribution of mutation rates among know pathogenic mutations and demonstrate how precise mutation rate estimates can further our understanding of heritable diseases. The effects of sequence contexts, genomic features, and their interactions we present capture the most refined portrait to date of the germline mutation patterns in humans.

2355F

Detecting long-term balancing selection using allele frequency correlation. *K. Siewert[1], B. Voight[2,3].* 1) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Balancing selection occurs when it is beneficial for multiple alleles to be present in a population, which can result in the preservation of variants over long evolutionary time periods. A characteristic signature of this long-term balancing selection is an excess number of intermediate frequency polymorphisms near the balanced variant. However, the expected distribution of allele frequencies at these loci has not been extensively detailed, and therefore existing summary statistic methods do not explicitly take it into account. Using simulations, we show that new mutations which arise in close proximity to a site targeted by balancing selection accumulate at frequencies nearly identical to that of the balanced allele. In order to detect balancing selection, we propose a new summary statistic, Beta, which detects these clusters of alleles at similar frequencies. Our statistic has improved power to detect balancing selection over existing summary statistics under a wide range of simulated models and parameters. We compute Beta on 1000 Genomes Project data to identify loci potentially subjected to long-term balancing selection in humans. We report two balanced haplotypes. The first, located in the *CADM2* gene, is tightly linked to a GWAS SNP for cognitive functions and has been shown to have experienced positive selection in canine. The second, located in the *WFS1* gene, is in high linkage disequilibrium with a GWAS variant for Type II Diabetes. Both haplotypes are in tight linkage with eQTLs for these genes and fall in regions of high regulatory potential, indicating these haplotypes could be affecting regulation of their respective genes. These results suggest that balancing selection maintains variants which influence a wide variety of complex traits.

2356W

Recurrent adaptation of different haplotypes in *FADS* genes to plant-based and animal-based diets in a diverse worldwide set of extant and extinct human populations. *K. Ye[1], D. Wang[1], F. Gao[1], O. Bar-Yosef[2], A. Keinan[1,3,4].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Department of Anthropology, Harvard University, Cambridge, MA; 3) Cornell Center for Comparative and Population Genomics, Cornell University, Ithaca, NY; 4) Center for Vertebrate Genomics, Cornell University, Ithaca, NY.

The fatty acid desaturase (*FADS*) genes encode key enzymes for biosynthesizing omega -6 and -3 long-chain polyunsaturated fatty acids (LCPUFAs). While LCPUFAs can be absorbed from animal-based diets, biosynthesis is essential to compensate for their absence in plant-based diets. Adaptive signals have been reported for some *FADS* variants in Africans, South Asians, some East Asians, Greenlandic Inuit and Native Americans. Recently, we reported two haplotypes (called M2 and D), comprising of opposite alleles of the same *FADS* variants, were respectively adaptive to the animal-based diets of pre-Neolithic European hunter-gatherers by diminishing LCPUFAs biosynthesis and to the plant-based diets of European farmers after the Neolithic Revolution by enhancing biosynthesis. It seems that opposite alleles of the same causal variants were respectively adaptive to the two diets. However, another recent study evaluated linkage disequilibrium among genetic variants and suggested that selection has targeted different variants in Europeans, South Asians and Inuit. Here, we systematically investigated the genetic variants and haplotypes under selection in global present-day populations (*e.g.* 1000GP, HGDP, SGDP, UK10K) and in historical populations with ancient DNA (aDNA, compiled from over 25 studies). We recalled genotype for critical structural variants with high-depth sequencing data and with imputation we obtained a uniform set of variants across all data sets. We applied three types of selection tests: 1) solely based on present-day DNA (*e.g.* Tajima's D, Fay and Wu's H, iHS, nSL, and SDS), 2) compared frequency change between ancient and present-day DNA, and 3) inferred selection coefficient from allele frequency time series. We found the same *FADS* haplotypes under selection in multiple populations. Haplotype D carries adaptive signals in most traditional farming populations, ranging from South Asians (current frequency 82%), Europeans (63%), to Africans (40%) and some East Asians (43%). Two closely related haplotypes (M1 and M2) were adaptive in Eskimos (M1:73%), Native Americans (M1:39%; M2:54%), southeastern Asians (M2:53%) and in historical European hunter-gatherers (M2 increased from 29% to 50% from 30 to 7.5kya). We further applied simulation of selection with observed ancient and present-day DNA to prioritize multiple candidate variants. These complex patterns of recurrent dietary adaptation of different *FADS* haplotypes are unique in human evolution.

2357T

Genetic origins of Easter Island and remote Oceania. *A. Ioannidis[1], J. Blanco[2], C. Quinto[2], K. Sandoval[2], E. Hagelberg[3], M. Moraga[4], J. Miquel-Poblete[4], T. Parks[5], M. Ávila-Arcos[6], E. Burchard[7], C. Eng[7], A. Adams[8], A. Mentzer[5], C. Bustamante[8], A. Moreno-Estrada[2].* 1) Computational and Mathematical Engineering, Stanford University, Palo Alto, CA; 2) Human Population Genomics Lab, Laboratorio Nacional de Genómica para la Biodiversidad, Irapuato, Mexico; 3) Department of Biology, University of Oxford, Blindern, Norway; 4) Facultad de Medicina, Universidad de Chile, Santiago de Chile, Chile; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 6) International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México, Querétaro, Mexico; 7) University of California San Francisco, San Francisco, United States; 8) Department of Genetics, Stanford University, Palo Alto.

Reaching its easternmost extent in Easter Island less than one thousand years ago, the settlement of Polynesia represents a final chapter in the dispersal of humans across the globe. Though it occurred relatively recently, much remains unknown about this unique oceanic process in historical population genetics, including the sequences of islands settled, the timing of settlements, and the origins of the settlers. Using dense genome-wide SNP array data of 445 modern samples from seventeen key islands (or island clusters) spanning remote Oceania, we infer settlement patterns stretching across the Pacific to Easter Island, and we address the long lingering suggestion that Native Americans could have played a role in the ancient population history of Easter Island. We confirm this theory, but we show that the ancient contact with Native Americans took place in Polynesian "up stream" of Easter Island, that is, before its settlement, and that this component was then carried to Easter Island by its admixed founders. We use three different timing analyses, as well as evidence of differential origins within the Americas of distinct Native American genetic signatures that we find present in Polynesia, to support our conclusions. The isolated Polynesian islands, separated by vast Pacific distances, have provided us with a uniquely structured canvas on which to implement novel variants of ancestry deconvolution and migration analysis techniques; we will describe these techniques, which could be useful for analyzing similarly isolated populations from other understudied regions of the world. Our results demonstrate the important role that both recent and ancient admixture events have played in creating the rich diversity patterns that define modern Polynesian populations.

2358F

A web-based initiative to accelerate research on African ancestry in the Americas. *M.E. Moreno, K. Bryc, P.K. Penton, R. Rubenstein, L. Vermillion, J.L. Mountain, the 23andMe Research Team.* 23andMe, Inc., Mountain View, CA.

The lack of genetic research on people of African ancestry has greatly limited our understanding of their genetic diversity. Since little is known about the ancestors of African descendants living in the Americas, it is difficult to determine from which ethnic groups or geographical areas slaves were taken. A lack of written records has resulted in ambiguity for many African Americans regarding their ancestors' origins. In 2016, 23andMe launched the African Genetics Project, a web-based genetic research study of individuals with recent African ancestry. The goals were to improve 23andMe's ability to provide more detailed ancestry results for customers with some African origins and to contribute to the global understanding of how people historically migrated throughout Africa and from Africa to the rest of the world. The study was open to any adult in the United States whose four grandparents were born in a single high priority country (Angola, Benin, Burkina Faso, Cameroon, Gabon, Ghana, Guinea Bissau, Guinea, Côte d'Ivoire, Liberia, Republic of Congo, Senegal, Sudan, Togo) or were affiliated with a single ethnic or tribal group in those countries. Consenting participants were asked to provide a saliva sample for genetic analysis and complete a survey about their family's birthplaces. Over a period of seven months, over 1,100 eligible individuals enrolled in the study. To date, 70% of these participants have provided a saliva sample and 60% have completed all study steps. The countries with the largest enrollment rates were Ethiopia, Somalia, and Sudan. The countries with the smallest enrollment rates were Benin, Burkina Faso, Congo (Brazzaville), Gabon, Guinea-Bissau, and Togo. Recruitment efforts included posts on 23andMe's social media channels, including targeted Facebook advertisements, a 23andMe blog post, presentations at university-based African Studies centers, and an article in the online African-American culture magazine, The Root. Participants also drove recruitment by posting about the study on blogs, forums, and YouTube. The success of this study puts 23andMe in a position to study African populations that are underrepresented in genetic research and serves as a pilot for recruitment of other underrepresented populations. An important next step will be to compare the genetic data collected from the African Genetics Project with the data of consenting 23andMe research participants from across the Americas.

2359W

Origin and affinities of Lakshadweep Islanders. *M.S. Mustak[1], N. Rai[2], M. Naveen[1], S. Prakash[2], N. Pasupuleti[1], G. Chaubey[3], K. Thangaraj[2].* 1) Applied Zoology, Mangalore University, Mangalore, Karnataka, India; 2) CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana-500007, India; 3) Department of Evolutionary Biology, Estonian Biocentre, Riia23b, Tartu, Estonia.

    Archipelago Lakshadweep resides in a south-west part of India in the Arabian Sea. In addition to its geographical isolation, the gene pool of these islands encompasses the signatures of ancient human dispersal across the South Asian (SA) corridor. In order to reconstruct the population history of Lakshadweep population, we have analysed uniparental (mtDNA and Y chromosome) and biparental (750K autosomal loci) markers among 1359 individuals belonging to several ethnic groups of Lakshadweep Islands. We observed the overwhelming presence of mitochondrial haplogroup R30, whereas the Y chromosome major haplogroups were P267-R2a (16%), Y495-R1a2b (12%). Both mtDNA and Y chromosome signals showed a close genetic link between Lakshadweep populations with the mainland Indian populations. The allele frequency and haplotype-based autosomal analyses suggested their closest affinity with the Southern Indian state Kerala.

2360T

The genetic substructure of the Japanese population: Results from the Japan Multi-Institutional Collaborative Cohort Study. *M. Nakatochi[1], T. Nishiyama[2], A. Hishida[3], M. Naito[3], M. Kubo[4], K. Wakai[3], the Japan Multi-Institutional Collaborative Cohort Study.* 1) Statistical Analysis Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 2) Department of Public Health, Aichi Medical University School of Medicine, Nagakute, Japan; 3) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan; 4) Laboratory for Genotyping Development, Center of Genomic Medicine, RIKEN, Yokohama, Japan.

    Aims the Japanese population has been known to be grouped into super-clusters, which are Mainland and Ryukyu clusters, in the population genetics. Investigation of the super-clusters in the detail was very important to perform genetic disease association studies and reveal demographic history in the Japanese population. We examined the Japanese population substructure using genome wide genotyping data from the Japan Multi-Institutional Collaborative Cohort Study (J-MICC), including south west islands of Japan. *Methods* A total of 14,539 study subjects from the 12 areas of the J-MICC study including the Ishigaki and Amami Islands were genotyped at RIKEN Center for Integrative Medicine using a HumanOmniExpressExome-8 v1.2 BeadChip array. Subjects with discordant sex information and close relationship pairs were removed. First, Principal component analysis (PCA) with J-MICC subjects, 1000 genomes East Asian (EAS), and Pan-Asian SNP (PASNP) consortium genotype data were performed to examine the population substructure of east Asian including J-MICC subjects. We also performed ADMIXTURE analysis for J-MICC subjects. *Results* PCA with J-MICC subjects, 1000 genomes EAS, and Pan-Asian EAS showed three clusters (Mainland, Ryukyu, and Korean clusters) from J-MICC subjects. Of the J-MICC subjects, 0.3% were assigned to the Korean cluster. PCA with only J-MICC subjects further indicated the Amami cluster. The Amami cluster existed closest to Ryukyu cluster, but didn't exist closest to Korean and Taiwan populations. The result of ADMIXTURE analysis revealed that the minimum value of the cross-validation (CV) error was observed when the assumed number of ancestral populations (K) was 8. *Conclusion* We examined the population substructure of J-MICC study, and identified four sub-clusters (Mainland, Ryukyu, Amami, and Korean clusters). ADMIXTURE analysis indicated that the K=8 was minimum value of the CV error. Thus, the result suggests the possibility that more genetic clusters exist in Japanese population. These results are expected to be important information to do genetic disease association studies with the use of the Japanese population data.

2361F

IBD detection at biobank-scale. *A. Naseri[1], X. Liu[2], S. Zhang[1], D. Zhi[3].* 1) Department of Computer Science, University of Central Florida, Orlando, Florida; 2) Department of Epidemiology, Human Genetics & Environmental Science, University of Texas Health Science Center at Houston, Houston, Texas; 3) School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas.

Recently, very large genotyped cohorts (VLGCs) with rich phenotypes, such as UK Biobank, are becoming available. Unlike traditional cohorts, these cohorts contain samples about 0.1%-1% of the entire population. At this scale, genetic relatedness among these samples are ubiquitous and cannot be simply ignored as unrelated. Identical-by-Descent (IBD) segments, as direct evidence of common ancestry among individuals, are a reliable source for genetic relatedness. Therefore, having direct IBD information will enable a wide range of applications from association studies to population genetics. However, current IBD detection tools are too slow to be applied to biobank-scale data sets. We have developed a new method, RaPID, that is more than 100 times faster than GERMLINE, the fastest IBD detection tool for the past 8 years. The efficiency of RaPID is achieved by an innovative randomization strategy around Richard Durbin's Positional Burrow-Wheeler Transformation (PBWT), which enables PBWT to work with genomic sequences with errors and mutations. Thanks to PBWT's linear time complexity with the sample size, RaPID is the first computationally feasible method that can infer IBD segments among individuals in a biobank-scale cohort. We compared RaPID and GERMLINE by simulations using macs with a demographic history model of the Europeans. We not only verified the speed of RaPID, but also showed that for IBD segments with length >5cM, RaPID can detect more IBD segments (99%) than GERMLINE (85%) with a comparably higher accuracy (99% for both methods). Moreover, at >1.5cM, RaPID reported both more IBD segments (97% vs 65%), and a higher accuracy (92% vs 63%). Interestingly, RaPID also reported sharper segment boundaries (RMSE 0.4cM) than GERMLINE (RMSE 1.5cM) at >5cM. A similar trend is seen at 1.5cM: RaPID (0.5cM) vs GERMLINE (0.7cM). We applied RaPID to the data from 1000 Genomes project, and identified both sub-continental and inter-continental IBDs. Most notably, we identified a genetic relationship at 1.5cM among Finns and Asians, suggesting possible historical migration event. Overall, our results showed that RaPID is not only fast, but also accurate in terms of IBD detection and it may be an essential tool to uncover the rich genetic relatedness information in large genotyped biobank data.

2362W

Exploring the demographic and admixture history of Central Mexico. *A.W. Reynolds[1,2], J. Mata-Miguez[2], D.A. Bolnick[2,3].* 1) Department of Integrative Biology, University of Texas at Austin, Austin, TX; 2) Department of Anthropology, University of Texas at Austin, Austin, TX; 3) Population Research Center, University of Texas at Austin, Austin, TX.

The Spanish colonization of Mexico led to the creation of new communities containing individuals who had been born in three different continents (North America, Europe, and Africa) and who came from diverse cultural and linguistic backgrounds. Major population collapses also occurred in indigenous communities following colonization of this region. While previous research has explored these events in Mexico City and other urban areas, their impact on rural populations remains unclear. In Xaltocan, a small town in central Mexico, historical documents suggest that the Spanish rarely visited the town. To better understand the demographic history of this region and the genetic effects of Spanish colonial history, we collected samples from 47 present-day residents from Xaltocan. The samples were genotyped for >600,000 genome-wide single nucleotide polymorphisms (SNPs) using the Affymetrix Axiom® Human Origins Array. Individual ancestry estimates were calculated using ADMIXTURE and RFMix. Estimates of the number and timing of admixture events at Xaltocan were calculated using TreeMix and Tracts. Past changes in population size were modeled using dadi. We find a much lower average proportion of European ancestry at Xaltocan compared to previously sampled populations in the Americas. We also find evidence that a single admixture event between European and Native American source populations affected our study population at Xaltocan.

2363T

Signatures of multiple-mergers coalescence in genomic diversity data.
D.P. Rice[1], J. Novembre[1], M.M. Desai[2]. 1) Human Genetics, University of Chicago, Chicago, IL; 2) Organismic and Evolutionary Biology, Harvard University, Cambridge, MA.

The genetic diversity of a population reflects its demographic and evolutionary history. Methods for inferring this history typically assume that the ancestry of a sample can be modeled by the Kingman coalescent process. A defining feature of the Kingman coalescent is that it generates genealogies that are binary trees: no more than two ancestral lineages may coalesce at the same time. However, this assumption breaks down under several scenarios. For example, pervasive natural selection, rapid spatial range expansion, and extreme variation in offspring number can all generate genealogies with "multiple-merger" events in which more than two lineages coalesce instantaneously. Therefore, detecting multiple mergers is important both for understanding which forces have shaped the diversity of a population and for avoiding fitting misspecified models to data. Current methods to detect multiple mergers rely on the average site frequency spectrum (SFS). However, many of the signatures of multiple mergers in the average SFS are also consistent with a Kingman coalescent process with a time-varying population size. Here, we present a new method for detecting multiple mergers based on correlations in the SFS across the genome. Unlike the average SFS, these correlations depend mostly on the topologies of genealogies rather than their branch lengths and are therefore robust to most demographic effects. We apply this method to genomic diversity data from a variety of species, including humans and D. melanogaster.

2364F

Studying global variation of gene flow using geo-referenced genetic data. *S. Song[1], N.D. Berkowitz[1], Y. Wang[1], D.A. Turissini[1], H. Guturu[1], J. Byrnes[1], E. Hong[1], C. Ball[1], K. Chahine[2].* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

Geographic distances often cause genetic similarities to decay. Heterogeneity in the rate of decay is caused by various factors affecting gene flow. Understanding gene flow barriers or corridors is crucial to identify fine scale population structure and can yield insights on evolutionary processes that shape present day genetic diversity. A newly published method EEMS (Estimated Effective Migration Surfaces) models the relationship between genetics and geography and produces a visual representation of population structure highlighting potential regions of historic gene flow (Petkova Nature 2016). Using a research-consented subset for our database of over 4 million samples, we inferred the 'putative geographic origin' from which the individual's recent ancestry is derived based on pedigree information. We then applied EEMS using genome-wide SNP data and estimated effective migration rates among human populations in Europe and North America. We identified several known geographic barriers and regions showing high connectivity. We further investigated observed barriers/corridors and compared these with migration patterns found using pedigree data.

2365W

Genetic structure in Brittany highlights physical and cultural limits. *J. Giemza[1], M. Karakachoff[1], E. Charpentier[1], K. Rouault[2], A. Saint-Pierre[2], F. Simonet[1], S. Lecointe[1], P. Lindenbaum[1], C. Férec[2], H. Le Marec[1], S. Chatel[1], J.J. Schott[1], E. Génin[2], R. Redon[1], C. Dina[1].* 1) L'institut du thorax, INSERM, CNRS, UNIV Nantes, CHU Nantes, Nantes, France, Nantes, Pays de la Loire, France; 2) INSERM UMR 1078, CHRU Brest, UNIV Brest, Brest, France.

Background The genetic structure of human populations varies throughout the world, being influenced by migration, admixture, natural selection and genetic drift. Characterising such genetic variation can provide insight into demographical history and informs research on disease association studies, especially on rare recent variants. In this study, we examine the fine-scale genetic structure of Brittany and surrounding regions of France. Brittany is a region in the north-west of France, historically and culturally distinctive. Genetic proximity between Bretons and Irish has been shown in [1]. Currently, administrative Brittany covers only 80% of historical Brittany. Southern limits of historical Brittany extend further than the Loire River, the biggest physical barrier in the region. Eastern limits do not coincide with any significant geographical feature, potential genetic barrier could be thus a result of cultural and historical differences. Methods and Results We genotyped 1005 individuals from North-Western France, with at least three of their grandparents born within a 15 kilometres distance using Axiom™ Precision Medicine Research Array. Principal Components analysis revealed a high correlation between geographical position and components (p-value < 2e-16). Visualisation of PC1 (0.16 % of variance) on the map points to three subpopulations: one in the south of Loire River and two in the north, one of which overlaps with historical Brittany. Partial Mantel tests confirm that genetic differentiation is not uniform. We also approximate eastern border of "genetic Brittany" based on ADMIXTURE results and test the strength of the barriers with Fst statistic. Southern border, corresponding to Loire River, is more pronounced. Conclusion We here report both evidence for isolation by distance within at a very fine level and existence of two genetic barriers, the Loire River and the historical boundary of Brittany Duchy. Subsequently, we will verify and extend our findings with fineSTRUCTURE software and with analysis based on Identity By Descent. This fine-scale population structure may have consequence in association analyses, especially for rare variants which tend to be geographically clustered. These results support the need for a genetically matched panel of controls in gene mapping analyses in French population. 1. Karakachoff, M. et al. Fine-scale human genetic structure in Western France. Eur J Hum Genet 23, 831–836 (2015).

2366T

The Irish DNA Atlas: Revealing fine scale population structure and history within Ireland. *E.H. Gilbert[1], S. O'Reilly[2], M. Merrigan[2], D. McGettigan[2], A.M. Molloy[3], L.C. Brody[4], W. Bodmer[5], K. Hutnik[5], S. Ennis[6], D.J. Lawson[7], J.F. Wilson[8,9], G.L. Cavalleri[1].* 1) Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Leinster, Ireland; 2) Genealogical Society of Ireland, Dún Laoghaire, County Dublin, Ireland; 3) School of Medicine, Trinity College, Dublin 2, Ireland; 4) Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; 5) Weatherall Institute of Molecular Medicine and Department of Oncology, University of Oxford, Oxford OX3 7DQ, UK; 6) School of Medicine and Medical Science, University College Dublin, Dublin, Ireland; 7) University of Bristol, Department of Mathematics, University Walk, Bristol BS8 1TW, UK; 8) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, Scotland; 9) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, Scotland.

Demographically, Ireland has remained relatively isolated from mainland Europe, notwithstanding notable historical migrations including the Norse Vikings, Anglo-Normans, and the British Plantations. A considerable number of individuals outside Ireland share Irish ancestry including an estimated 33 million individuals in the United States. Although previous studies have shown the Irish population within Ireland to be situated at the end of a number of genetic clines within Europe, the extent of fine-scale population structure within Ireland is largely unknown and it is unclear to what degree historical migrations have impacted the modern Irish genome. Here we illustrate fine scale genetic structure across Ireland that follows sociological boundaries within Ireland and we present evidence of admixture events into Ireland from both the neighbouring British Isles and continental Europe. Utilising the 'Irish DNA Atlas', a DNA cohort (n = 194) of genealogically described Irish individuals with four generations of ancestry linked to specific regions in Ireland, we analysed the Atlas in combination with 2,039 individuals of regional British ancestry (the PoBI dataset) and show that the Irish population can be divided in 10 distinct geographically stratified genetic clusters; three are of shared British and Irish ancestry, and seven of predominantly 'Gaelic' Irish ancestry. This structure is remarkably homogenous, and is associated with few gene flow barriers within Ireland. Using a reference of 6,760 European individuals and two ancient Irish genomes, we quantified the ancestry of these Irish clusters in the context of neighbouring Europe as well as ancient Ireland. We show high levels of North West French-like and Norwegian-like ancestry within Ireland. We also show that our 'Gaelic' Irish clusters present homogenous levels of ancient Irish ancestries. Finally we detect admixture events into Ireland that coincide with the Plantations of Ulster, as well as provide evidence of Norse Viking gene flow into Ireland. Our work informs both on Irish history, as well as the study of Mendelian and complex disease genetics involving populations of Irish ancestry.

2367F

Characterizing heterogeneity within fine-scale population structure. *A. Hippen Anderson[1], A.R. Girshick[2], A. Boroomand[2], R.E. Curtis[1], B. Wilson[2], K. Noto[2], E.L. Hong[2], C.A. Ball[2], K.G. Chahine[2].* 1) AncestryDNA, Lehi, UT; 2) AncestryDNA, San Francisco, CA.

We have recently demonstrated the ability to leverage network analysis methods to discover genetic communities within a large identity-by-descent (IBD) network. Members of each community share more IBD with one another than with those outside of their community. We interpreted these genetic communities using pedigree annotations. Using this approach, we discovered, named, and localized over 300 genetic communities. Here we provide a richer understanding of these genetic communities as a group of people within a temporal and geographic context. Using machine learning techniques, we identified common features within 63 different genetic communities including genetic ethnicity patterns and other related meta-data. We found that some communities are more homogenous in this feature space than others. For instance, some genetic communities tend to have a dominant ethnicity while others are highly admixed. The more heterogeneous the genetic community, the more likely we were to find an identifiable sub-group within it. We were also interested in understanding the temporal structure of genetic communities. Communities differed in the distribution of birth years among members, possibly indicating compositional differences on extremely short time scales. We also analyzed these birth year distributions with respect to geographic location. These results provide a novel way to understand IBD network-based genetic communities in their temporal, spatial, and ethnicity contexts.

2368W

The genomic health of ancient hominins. *J. Lachance, T. Cooper, A. Berens.* School of Biology, Georgia Institute of Technology, Atlanta, GA.

The genomes of ancient humans, Neandertals, and Denisovans contain many alleles that influence disease risks. Using genotypes at 3180 disease-associated loci, we estimated the disease burden of 147 ancient genomes. After correcting for missing data, genetic risk scores were generated for nine disease categories and the set of all combined diseases. These genetic risk scores were used to examine the effects of different types of subsistence, geography, and sample age on the number of risk alleles in each ancient genome. On a broad scale, hereditary disease risks are similar for ancient hominins and modern-day humans, and the GRS percentiles of ancient individuals span the full range of what is observed in present day individuals. In addition, there is evidence that ancient pastoralists may have had healthier genomes than hunter-gatherers and agriculturalists. We also observed a temporal trend whereby genomes from the recent past are more likely to be healthier than genomes from the deep past. This calls into question the idea that modern lifestyles have caused genetic load to increase over time. Focusing on individual genomes, we find that the overall genomic health of the Altai Neandertal is worse than 97% of present day humans and that Ötzi the Tyrolean Iceman had a genetic predisposition to gastrointestinal and cardiovascular diseases. As demonstrated by this work, ancient genomes afford us new opportunities to diagnose past human health, which has previously been limited by the quality and completeness of remains.

2369T

FastNGSadmix: Admixture proportions and principal component analysis of a single low-depth sequencing sample. *E. Jørsboe[1], K. Hanghøj[2,3], A. Albrechtsen[1].* 1) The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen N, Denmark; 2) Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen K, Denmark; 3) Université de Toulouse, University Paul Sabatier (UPS), Laboratoire AMIS, CNRS UMR 5288, Toulouse, France.

We present fastNGSadmix, a method for fast and easy estimation of admixture proportions and principal component analysis (PCA), for a single low-depth next generations sequencing (NGS) sample using a panel of reference populations, with population specific allele frequencies. We show that fastNGSadmix has increased accuracy compared to established methods for estimating admixture from reference panels, such as iAdmix and ADMIXTURE. fastNGSadmix corrects for the bias, introduced by having a limited size reference panel, which is a substantial problem with other methods. Reference panels where populations are represented by fewer than 20 individuals has a large impact on the accuracy, if no correction for small reference panels is done. fastNGSadmix works for samples with very low sequencing depth, we show through downsampling that fastNGSadmix is able to estimate the correct admixture proportions for a sample with depth of coverage of less than 0.005X, equivalent of there on average being less than one read per 200 sites. This is because the method works with genotype likelihoods, this enables us to take the uncertainty of the genotypes into account in the proposed admixture model. A bootstrapping approach has been implemented for the admixture estimation, meaning the uncertainty on the admixture estimates can easily be obtained, making it easier to infer if the admixture proportions with regards to different populations differ significantly. We use the estimated admixture proportions to perform PCA incorporating both population structure and genotype uncertainty. Existing PCA methods based on NGS data do not model population structure and simply assume that genotypes between individual are independent. This method does not make this assumption and instead works by modeling the impact of population structure via the estimated admixture proportions from fastNGSadmix. It also works with genotype likelihoods modelling the genotype uncertainty. It is therefore well suited for doing PCA for admixed NGS samples. This method has for example been applied to ancient DNA samples, where it is easy to apply this method to the individual samples. Ancient DNA is usually characterized by high uncertainty on the data, why it is crucial to take genotype uncertainty into account. In addition the samples are modelled independently, which allows for analysing related samples.

2370F

Assessment of BAP1 germline and somatic alterations in uveal melanoma. *M.H. Abdel-Rahman[1,2,3], G. Boru[1], T. Grosel[1], R. Pilarski[2,3], J.B. Massengill[1], F.H. Davidorf[1], C.M. Cebulla[2,3].* 1) Department of Ophthalmology, The Ohio State University, Columbus, OH; 2) Division of Human Genetics, Department of Internal Medicine, The Ohio State University, Columbus, OH; 3) The Ohio State University Comprehensive Cancer Center, Columbus, OH.

Objective: BAP1 (BRCA1-associated protein 1) is a deubiquitinating hydrolase with tumor suppressor function through involvement of various biologic processes including chromatin remodeling, DNA damage response, cell cycle regulation, and cell growth. Germline mutation in BAP1 is associated with Tumor Predisposition Syndrome (BAP1-TPDS, OMIM #614327) with at least four main cancers: uveal melanoma (UM), cutaneous melanoma (CM), mesothelioma (MMe) and renal cell carcinoma (RCC). Somatic mutations in BAP1 are observed in many different cancers with UM the most common. The aim of this project is to assess the frequency and type of germline and somatic alterations in BAP1 in UM patients. Methods: Germline mutations in BAP1 were assessed in 160 UM patients with moderate to high risk for hereditary cancer predisposition. Somatic mutations were assessed in an additional 47 UM tumors. Mutations were assessed by direct (Sanger) sequencing. In a subset of patients somatic and germline large deletions and epigenetic inactivation were assessed by MLPA and pyrosequencing respectively. The expression of BAP1 was assessed in tumors and non-tumor choroids by qRT-PCR, immunohistochemistry and/or Western blot. Results: We identified a total of 22 patients with one or more germline variants. Six (3.8%) of these patients had truncating mutations. The other 14 patients had benign/likely benign variants or variants of uncertain significance. In patients with familial UM the frequency of truncating mutation was 6/32 (18.8%). Somatic variants were detected in 29/47 (61.7%) of the tumors including 5 (10.7%) with large deletions (>20bp). No germline deletion/duplication in BAP1 was detected in 98 patients. No significant evidence of somatic or constitutional promotor hypermethylation was detected in the samples tested. Two out of the 17 non-tumor choroid tissues showed significant decrease in BAP1 RNA expression compared to normal controls. No detectable germline genomic alteration or aberrant promoter methylation was identified suggesting other mechanisms for epigenetic BAP1 inactivation. Conclusions: BAP1 is an important candidate for hereditary predisposition to UM but additional candidate genes exist. Large deletion is not a major mechanism for germline inactivation of BAP1. Epigenetic mechanisms, other than promotor methylation, are possible for both germline and somatic inactivation of BAP1.

2371W

Paperwork matters! The importance of clinical phenotype information in variant interpretation. *M. Anderson, C. Pardo, H.C. Kang, J. Thusberg, R. Quintana, K. Erhard, S-H. Wang, J. Holle, D. Iacoboni, I. Wilson, K. Ouyang.* Invitae, San Francisco, CA.

The validity and utility of genetic testing require that variant classifications be evidence based, objective, and systematic. As genetic testing becomes available to a larger percentage of the population, detailed clinical information about patients and their family members becomes increasingly relevant for variant classification. Although sufficient clinical information is often provided in well-described case reports in the published literature, most classified variants are observed only in the clinical testing laboratory setting. Therefore, the ordering clinician becomes the sole source of phenotypic data, as provided on the test requisition form. To objectively incorporate this clinical data into our laboratory's pre-existing evidence-based variant classification framework called Sherloc, we defined point-based criteria and usage rules allowing us to evaluate the following: 1. clinical phenotype 2. variant segregation in families 3. variant de novo status As part of this process, we developed a set of predefined clinical criteria for approximately 130 oncology genes. For genes such as NF2, FH, and VHL, our interpretation criteria are nearly identical to the consensus clinical diagnostic criteria. For genes that lack a formal consensus, such as SDHB and SDHD, we took a rigorous, conservative approach in establishing internal criteria that consider age of onset, phenotypic specificity, penetrance, prevalence, and the existence of phenocopies. The application of our method is illustrated in cases of STK11 and PTCH1 variants, in which detailed phenotypic information provided by the clinician impacted classification. We also present MEN1 and CASR cases in which clinical information did not meet our threshold criteria for use in the variant classification process. These results highlight the need for ordering providers to share detailed clinical patient information, as it may influence variant classification and ultimately clinical care.

2372T

Mutational spectrum in BRCA1 and BRCA2 genes in Colombia. *I. Briceno[1], A. Gomez[2], N. Diaz[1], M. Noguera[2], D. Diaz[3], M. Casas[3].* 1) Bioscience, Universidad de La Sabana, Chia, colombia, Colombia; 2) Pontificia universidad Javeriana; 3) Instituto de referencia andino.

Introduction: The risk of developing breast and ovarian cancer is higher in families that carry one or more mutations in BRCA1 or BRCA2 genes. Detection of mutations in families and patients allows timely decisions in the field of preventive medicine. Objective: To identify the presence of mutations in the Colombian population and evaluate two testing strategies. Methodology: Tests referred for BRCA1 and BRCA2 analyzed by complete direct sequencing of both genes and partial sequencing based on founder mutations in a test designed by Instituto de Referencia Andino in Bogotá ("Profile Colombia"). Results: We found 110 patients carrying deleterious mutations in this group of 900 Colombian patients 65% where found in BRCA1, and 35% in BRCA2. Conclusions: The spectrum of 50 different mutations in Colombia as detected in our study is broader than the one previously reported for this country. "Profile Colombia" is a useful screening test to establish both founder and new mutations with a detection rate of 12% in cases with family history of breast cancer. Complete sequencing shows a detection rate of 14%, and should complement the study of a genetic basis for this disease. Key words: Breast cancer, BRCA1, BRCA2, sequencing, Profile Colombia.

2373F

Retrospective analysis of 36 fusion genes in 6170 patients of *de novo* acute leukemia and myeloid neoplasms. *X. Chen[1], F. Wang[1], Y. Zhang[1], M. Wang[2], W. Teng[1], X. Ma[1], L. Guo[1], J. Fang[1], Y. Zhang[2], P. Zhu[2], H. Liu[1].* 1) Hebei Yanda Lu Daopei Hospital, Langfang, Hebei, China; 2) Peking University First Hospital, Xicheng District, Beijing, China.

To gain a better understanding of the gene-fusion landscape in different types of hematologic malignancies and different age groups, 36 fusion genes were assessed by multiplex-nested RT-PCR in 6170 *de novo* acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), mixed phenotype acute leukemia (MPAL) or myeloproliferative neoplasms (MPN) patients. 34 distinct fusion genes were detected in 39.06% patients. Fusion genes were detected in 41.21% AML, 28.72% ALL, 22.39% MPAL and 79.96% MPN, respectively. The incidence of fusion genes was higher in pediatric AML than in adult AML. With the exception of infant AML, the incidence of fusion genes decreased with the growth of age. RUNX1-RUNX1T1, KMT2A-MLLT3, KMT2A-MLLT10, KMT2A-MLLT11 and FUS-ERG were more common in pediatric AML while KMT2A-PTD was more frequent in adult AML. The frequency of fusion genes in ALL was lower than in AML. Except for infant ALL, patients of 30-59 years had the highest frequency of fusion genes. BCR-ABL1 and SET-NUP214 were more common in adult ALL while TCF3-PBX1 and ETV6-RUNX1 were more common in pediatric ALL. The occurrence of fusion genes was more frequent in B-ALL than in T-ALL. BCR-ABL1, TCF3-PBX1, KMT2A-AFF1 and ETV6-RUNX1 were more common in B-ALL while KMT2A-MLLT4, SET-NUP214 and STIL-TAL1 were more common in T-ALL. This is a systemic analysis of fusion genes in the largest cohort of leukemia and MPN to date. The identification of gene-fusion spectrums will help to better evaluate the diagnostic and prognostic significance of fusion genes in hematologic malignancies.

2374W

Paternal inheritance of *BRCA* mutations in Algerian hereditary breast/ovarian cancer families. *F. Cherbal[1], H. Gaceb[1], C. Mehemmai[1], K. Yatta[1], R. Bakour[1], K. Boualga[2]. 1) Unit of Genetics, LMCB, Faculty of Biological Sciences, University of Sciences and Technology Houari Boumediene Algiers, Algeria; 2) Radiation Therapy Services, Anti Cancer Center, Blida, Algeria.*

   Background Breast cancer is the leading cause of cancer morbidity and mortality among Algerian women.To date, *BRCA* mutations can be transmitted from fathers to their daughters. To the best of our knowledge, there is no study on paternal transmission of *BRCA* mutations in Algerian population. From 2008 to 2016, 150 breast /ovarian cancer families were screened for complete *BRCA1* and *BRCA2* germline mutations (70 families) and for specific mutations detected in *BRCA1* and *BRCA2* genes in Algerian patients (80 families), respectively. **Methods** The patients and their families were referred from public hospitals and private medical clinics which provide oncology services throughout Algeria. Patients were selected on family history of breast and/or ovarian cancer, and male relatives for prostate cancer along three generations at any age. Complete screening of germline mutations in *BRCA* genes has been conducted by HRM-direct sequencing in 70 families and their relatives (86 individuals). 80 families and their relatives (94 individuals) have been screened by PCR- direct sequencing for 4 common mutations, detected in Algerian population (exons 3,5 and 11) in *BRCA1* gene: c.83_84delTG, c.181T>G, c.798_799delTT and c.2125_2156inA, and 2 common mutations in *BRCA2* gene (exons 10 and 14): c.1310_1313delAAGA and c.7235_7236insG, respectively. **Results** We noticed that paternal inheritance of *BRCA* mutations resulted in an earlier age of breast cancer in daughters (mean age 38.6 years). In 7 families with *BRCA* mutations, the parent of the origin of the mutation was the father. Four distinct pathogenic mutations in *BRCA1* gene have been detected in 6 families.The *BRCA2* mutation c.1310_1313delAAGA has been detected in one family. Eleven (11) individuals (9 daughters and 2 sons) inherited pathogenic *BRCA* mutation from fathers (9 affected with cancer).Two ovarian cancers have been identified in two families. Interestingly, in five families with *BRCA1* mutations, male relatives had family history of prostate cancer.Three fathers among seven have developed prostate cancer. **Conclusions** Daughters with paternally inherited *BRCA* gene pathogenic mutation have breast cancer diagnosed at earlier ages. Families with paternal inheritance of *BRCA1* mutation have male relatives with history of prostate cancer. Individuals belonging to families with paternal inheritance of *BRCA* mutations should be tested at earlier age. Further studies in larger cohort are needed to reaffirm our findings.

2375T

Novel acute myeloid leukemia amplicon panel for uniform detection of all target genes, Including *CEBPA*. *C. Clear[1], R. Hrdlickova[1], J. Nehyba[1], D. Fox[1], D. Obermoeller[1], M. Toloue[2]. 1) Bioo Scientific, Austin, TX; 2) Perkin-Elmer, Boston, MA.*

   Acute myeloid leukemia (AML) accounts for the largest group among adult acute leukemia patients, with an incidence of over 20,000 cases per year in the United States alone. AML is thought to occur from acquired somatic lesions in regulatory genes, which accumulate in hematopoietic progenitors throughout life. These mutations give rise to malignant clones, which outcompete or suppress normal hematopoiesis. NGS allows rapid and sensitive detection of mutations facilitating deep read coverage of all target regions. However, AML regions of interests include several complex and difficult regions not often covered in amplicon panels, such as GC-rich *CEBPA*. Even though this gene is extremely difficult to amplify, information about the sequence is needed for proper evaluation of AML genotypic status, as biallelic mutations in *CEBPA* have been revealed to be significant in the progression of AML. Therefore, we developed a panel that uniformly covers all regions of interest, with a simplified protocol. Whereas many panels fail to produce any coverage of the *CEBPA* gene, this panel has full mutation calling capabilities. *CEBPA* is composed of roughly 80% guanine and cytosine nucleotides which create numerous secondary structures, preventing amplification. An innovative combination of primer design and PCR conditions has been developed to overcome these difficulties, maintaining the high performance of all amplicons, even when the diversity of nucleotides in some regions is low. The developed panel covers 42 kb, comprising 105 coding exons and limited flanking intron regions of 28 genes. Libraries have 100% uniformity at a 10% of average read threshold, 95% on-target reads, and virtually no dimer contamination. Input is as little as 40 ng DNA from fresh and frozen tissue. High uniformity and high on-target reads allow maximal multiplexing, which is particularly important when identifying somatic mutations due to the necessity for deep coverage. Libraries generated are compatible with both Ion Torrent™ and Illumina® sequencing platforms. Performance of this panel has been validated by positive identification of described mutations in five key genes relevant to the genesis of the disorder. Samples containing mutations in the following genes, *CEBPA, NPM1, TET2, RUNX1, and DNMT3,* were analyzed on a MiSeq®, 600 cycle v3 kit.

2376F

Comparison of variant classification algorithms incorporating clinical and family history for breast and ovarian cancer. *J. Clifford, M.H. Black, A. Elliott, H.M. Lu, S. Li.* Bioinformatics and Computational Biology, Ambry Genetics, Aliso Viejo, CA.

Pathogenic mutations are more likely to occur in high-risk individuals while benign variants are unrelated to personal and family history. A proband's clinical and family history may be used to form a likelihood ratio (LR) of the probability that the individual is a mutation carrier vs. non-carrier. A variant LR is computed by estimating the joint probability of carrier status across all probands, which may be derived either from odds obtained from a logistic regression model (LRM) or observed proportions in a clinical population (history weighting algorithm (HWA)). While both methods are useful for variant classification, the relative merits of each under a variety of scenarios have not been directly compared. To demonstrate differences in performance and clinical utility of these approaches, we applied both methods to estimate the LRs of classified variants in *BRCA1* and *BRCA2*, using detailed clinical and family history data from 61,851 patients who underwent genetic testing at a single diagnostic laboratory in 2012-2015. To evaluate the performance of each approach under increasing numbers of independent variables, we tested gene-specific models with 6 and 20 binary variables, respectively. Discriminatory power for correctly classifying variants was then assessed using AUCs with 2-fold cross-validation. We also simulated synthetic variant LR distributions for a fixed number of mutation carriers/non-carriers; the number of carriers was incrementally increased to determine the minimum number of probands such that $\leq 5\%$ of the variant LR distribution for mutation carriers overlapped with the distribution for non-carriers. We observed similar performance, as measured by AUC, for both methods with 6 variables for *BRCA1* (LRM: 0.71, HWA: 0.70) and *BRCA2* (LRM: 0.61, HWA: 0.61). When the number of independent variables increased, LRM had higher AUC (*BRCA1*: 0.83, *BRCA2*: 0.70) than HWA (*BRCA1*: 0.74, *BRCA2*: 0.64). For the 20 variable LRM, our simulations indicated that 40 probands for *BRCA2* and 12 for *BRCA1* were the minimum number necessary to separate 95% of the carrier and non-carrier LR distributions. Variant LR log(e) thresholds were found to predict synthetic variants as benign (*BRCA1/2*:-4.8) and pathogenic (*BRCA1*:5.7, *BRCA2*:5.5) with 95% accuracy. These data show that when the number of variables is large, the LRM may outperform HWA, which may have important implications for variant assessment given the increasing availability of clinical and family history data.

2377W

A t(18;22)(q21;q11) involving *IGL/BCL2*, a rare event in in chronic lymphocytic leukemia. *A. Dowiak[1,2], J. Ewing[3], D. Zhao[1,2], E. Salazar[1,2], L. Thompson[3], B. Wyatt[3], K. Peterson[3], L. Batholomaus[3], A. Nguyen[1,2], R. Crippen[1,2], D. Serk[3], J. Reinartz[3,4], C.A. Tirado[2,3,4,5].* 1) University of California, Los Angeles, CA; 2) The International Circle of Genetics Studies, Los Angeles, CA; 3) Allina Health, Minneapolis, MN; 4) HPA, Minneapolis, MN; 5) University of Minnesota School of Medicine, Department of Laboratory Medicine and Pathology, Minneapolis, MN.

We report a 63 y.o. male who presented with splenogaly, pallor, mucucutaneous bleeding. Peripheral blood showed more than 5000 monoclonal B Lymphocytes/µl. Blood smears showed small, mature appearing lymphocytes with a narrow border of cytoplasm and a dense nucleus lacking discernable nucleoili, having partially aggregated chromatin. Larger atypical lymphocytes or prolymphocytes were present in 43% of the cells. Flow Cytometry confirmed the presence of CD19, CD20 and CD23 consistent with CLL. Chromosome analysis of the bone marrow showed an abnormal karyotype described as 46,XY,t(18;22)(q21;q11.2)[19]/46,XY[1]. FISH analysis on interphase nuclei revealed an abnormal clone with loss of D13S319 (13q14.3) in 68.0% of the cells examined. Deletion of chromosome 13 is the most common cytogenetic abnormality identified in CLL (approximately 50% of CLL). Recent studies suggest that deletion of chromosome 13q14 in 65% or more nuclei by FISH is associated with an intermediate to unfavorable prognosis in CLL. The t(18;22)(q21;q11.2) present in this case as well as the t(2;18)(p12;q21) is a variant of the t(14;18)(q32;q21) a rare abnormality in CLL. These abnormalities are found in less than 4% of CLL cases. They are usually present within the context of a complex karyotype in a subset of CLL, but can also be observed in cases of benign lymphocytosis . Herein, we report a t(18;22)(q21;q11.2) in a CLL patient as a sole cytogenetic abnormality. .

2378T

Differential expression of kallikreins as prognostic markers in recurrent prostate cancer. *E. Guzel[1], F.B. Buyukozer[2], O.F. Karatas[3], N.D. Dibek[2], S. Simsek[2], M.E. Ozgurses[2], E.S. Aslan[2], M. Ittmann[4].* 1) Istanbul University, Istanbul Faculty of Medicine, Istanbul, Turkey; 2) Biruni University, Department of Molecular Biology and Genetics, Istanbul, Turkey; 3) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 4) Department of Pathology and Immunology, Baylor College of Medicine, Houston, 77030, TX, USA.

In the western countries, prostate cancer (PCa) is currently one of the leading causes of cancer death among men over the age of 50 years. Clinical recurrence with increased serum prostate-specific antigen (PSA, also known as Kallikrein 3 or KLK3) level is observed in almost 30% of men within 5 years following a successful radical prostatectomy. Therefore, it is necessary to find out molecular contributors of PCa progression to develop novel diagnostic/ therapeutic tools. In this study, it is aimed to evaluate the expression profiles of Kallikrein family proteins in recurrent and non-recurrent prostate tumor specimens as well as in adjacent normal prostate tissues to propose novel prognostic markers in PCa. A total of 69 (35 recurrent, 34 non-recurrent) radical prostatectomy material and 19 normal prostate tissue samples were obtained from Baylor College of Medicine Prostate Cancer program and included into the study. Expressions of Kallikrein protein family members were evaluated using quantitative real-time PCR (qRT-PCR). For statistical analysis, Pearson's correlation test, receiver operating characteristics (ROC) analysis and Kaplan-Meier test were used. QRT-PCR results demonstrated differential expression of Kallikreins both in tumor vs. normal and recurrent vs. non-recurrent comparisons. In addition, Pearson's correlation, ROC and Kaplan-Meier analysis showed the power of Kallikreins for predicting PCa prognosis. Here, we report the deregulated expression of Kallikreins in tumor samples and recurrent PCa specimens, which suggest Kallikreins as valuable predictors of PCa progression.

2379F

A 34-gene, NGS assay for common hereditary cancers reveals a high percentage of variants of unknown significance in the *POLD1* and *POLE* genes. *D. Hodko, R. Nakles, R. Chen, C. Nery, I. Karbassi, F. Lacbawan.* Quest Diagnostics, Inc., 33608 Ortega Hwy, San Juan Capistrano, CA 92675-2042.

Multigene panels have become an important tool for risk assessment in hereditary cancers. While such panels provide the benefit of revealing pathogenic variants associated with a broad range of phenotypes, they also tend to reveal more variants of unknown significance (VUS). In genes other than *BRCA1* or *BRCA2*, *CHEK2* and *ATM* variants are typically found in the highest frequencies in previously published multigene panel studies. Often not included in these panels are the DNA polymerase genes, *POLD1* and *POLE*, known to be associated with a moderate risk for colorectal cancers. Here, we present a panel for the simultaneous sequencing of 34 cancer-susceptibility genes (including *POLD1* and *POLE*) and assess the frequency of variants identified among individuals with a personal or family history of cancers including breast, ovarian, colon, and skin. This panel combines next-generation DNA sequencing and confirmatory microarray analysis for comprehensive detection of both sequencing variants and genomic rearrangements. From initial panel screening results, we analyzed the first 350 consecutive individuals referred for hereditary cancer screening. 195 (55.7%) individuals carried a variant in one of the 34 genes analyzed. Pathogenic or likely pathogenic variants were discovered in 35 (10.0%) individuals. Among all pathogenic or likely pathogenic variants detected in the cohort, the greatest number occurred in BRCA2 (25.7%), *BRCA1* (20.0%), and *CHEK2* (11.4%). Out of 228 unique variants of unknown significance (VUS) detected, 2.2% occurred in *BRCA1* and 3.5% occurred in *BRCA2*. However, the data reveal a higher percentage of unique VUS occurring within *ATM* (11.8%), *POLE* (8.8%), *APC* (7.9%), and *POLD1* (7.5%). The greater percentage of unique VUS occurring in DNA polymerase genes, *POLD1* and *POLE*, may be attributable to analyzing the full coding sequences of each gene. Our analysis finds that only 1.8% of all observed VUS are missense variants located within the exonuclease (proofreading) domains (*POLD1* codons 311-541 and *POLE* codons 269-485), where the only known pathogenic variants in *POLD1* and *POLE* occur. In highlighting the *POLD1* and *POLE* genes, we show that the rate of novel variants and VUS among high-risk cancer families is greater in non-*BRCA1/2* genes within the panel. In agreement with other studies, our data support the view that multiplex hereditary cancer panels alter the focus of genetic counseling, with increased attention to variants of uncertain significance. .

2380W

Hereditary cancer panel results identify gaps in knowledge of cancer risks and limitations in current guidelines. *H. LaDuca[1], C. Espenschied[1], J.S. Dolinsky[1], L. Panos Smith[2], K. Fulk[3], M. Pronold[3], C. Horton[3], F.J. Couch[4,5], B. Tippin Davis[1]. 1) Department of Research and Development, Ambry Genetics, Aliso Viejo, CA; 2) Department of Genetic Specialists, Ambry Genetics, Aliso Viejo, CA; 3) Department of Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA; 4) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 5) Department of Health Sciences Research, Mayo Clinic, Rochester, MN.*

With increased utilization of hereditary cancer multigene panel tests (MGPT), independent efforts have been made to explore yield and utility for various testing indications. However, few systematic approaches have utilized this data to further our understanding of cancer risk associations and support modifications to current genetic testing and management guidelines. To identify where gaps may exist, we explored mutation frequency by cancer type in a large cohort of patients referred for hereditary cancer MGPT. Patients with a history of a single primary breast, colorectal (CRC, non-polyposis), ovarian, endometrial, pancreatic, or prostate cancer were selected from a larger cohort of approximately 200,000 cases referred for hereditary cancer MGPT at a single commercial laboratory. Pathogenic/likely pathogenic variant (PV) frequencies were calculated for 34 genes known to predispose to at least one of these six cancers. In general, results showed that the more commonly mutated genes for each cancer type have established risk estimates and associated management recommendations for the respective cancers, such as *CHEK2*, *BRCA1/2*, *ATM*, and *PALB2* in breast cancer and mismatch repair genes in CRC and endometrial cancer. Exceptions were most apparent for pancreatic and prostate cancers, suggesting that cancer-specific testing guidelines should be developed to encompass more commonly mutated genes beyond *BRCA1/2* such as *ATM*, *CDKN2A* and *PALB2* in pancreatic cancer (2.7%, 1.3% and 1.2%, respectively), and *HOXB13* and *ATM* in prostate cancer (3.2% and 2.1%, respectively). Other interesting findings were noted for *CHEK2*, *BRCA2* and *ATM*, where mutations were also frequent among cancer types not currently associated with these genes. For example, *BRCA2* PVs were detected in 1.0% of CRC and 0.7% of endometrial cancer patients, and *ATM* PVs were detected in 0.9% of CRC and ovarian cancer patients, even though these genes have not been associated with these cancers. *CHEK2* was among the top five most frequently mutated genes across all cancer types (includes p.I157T); however, risks for ovarian, endometrial and pancreatic cancers have not been well-defined for this gene, and *CHEK2* management recommendations are limited to breast and CRC surveillance. Research efforts aimed at generating precise cancer risk estimates and expanded testing and management recommendations for more commonly mutated genes will have the largest immediate impact for counseling patients and their families.

2381T

TaqMan dPCR liquid biopsy assays targeting the TERT promoter region. *M. Laig, K. Varma, V. Venkatesh. Thermo Fisher Scientific, 180 Oyster Point Blvd., South San Francisco, CA 94080.*

The enzyme Telomerase maintains telomeres at the ends of chromosomes. The Telomerase Reverse Transcriptase (TERT) gene codes for the enzyme's catalytic domain and is not expressed in normal somatic cells. As a consequence, normal cells acquire senescence by shortening of their telomeres during cell division and eventually undergo apoptosis. In contrast to normal somatic cells, expression of TERT is reinstated in cancer cells causing escape from senescence and apoptosis by maintaining the telomeres. It has recently been shown that mutations in the TERT promoter region play a key role in regulating and reinstating TERT expression. Up to 90% of cancers carry a mutation in the TERT promoter region. Mutations like C228T and C250T create new binding sites for the E26 transformation-specific (ETS) transcription factor that regulates TERT expression. Experimental evidence showed that the ETS factor GA-binding protein, alpha subunit (GABPA) binds to the de novo ETS motif and activates TERT transcription in cancer cells. We undertook a project designing TaqMan® dPCR Liquid Biopsy Assays addressing mutations in the TERT promoter. These assays are TaqMan SNP Genotyping assays that are optimized for use in digital PCR with the Applied Biosystems QuantStudio 3D. In digital PCR, partitioning the sample into many individual reaction wells facilitates detection and quantification of rare mutant alleles. TaqMan SNP Genotyping Assays reliably discriminate mutant and wild-type allele. This will enable easy and sensitive detection of TERT promoter mutations in cancer research samples. These assays are suitable for detection in liquid biopsy applications with cell free DNA (cfDNA). Assay design proved to be challenging due to high GC content and repetitive elements in this region of the TERT gene and required varying both the assay design and experimental cycling conditions. Template for wet-lab testing was synthetic plasmid carrying the mutation spiked into wild-type genomic DNA and wild-type genomic DNA control. The result was two TaqMan dPCR Liquid Biopsy Assays detecting the C228T and C250T mutations in the TERT promoter region. We showed that designing TaqMan dPCR Liquid Biopsy Assays for digital PCR is feasible for the challenging TERT promoter region.

2382F

Clearance of plasma *EGFR* mutations as a predictor of outcome following osimertinib treatment. *A. Markovets[1], C. Barrett[1], S. Vowler[2], G. Oxnard[3], J. Dry[1], K. Thress[1].* 1) Oncology, AstraZeneca, Waltham, MA; 2) Global Medicines Development, AstraZeneca, Cambridge, UK; 3) Dana-Farber Cancer Institute, Boston, MA.

Osimertinib is an *EGFR* tyrosine kinase inhibitor (TKI) designed to target activated *EGFR* harboring the T790M resistance mutation. While osimertinib provides high and durable response rates in non-small-cell lung cancer patients with T790M, prediction of response duration can be critical for disease monitoring and to guide subsequent treatment decisions. In this study, we investigated whether changes in the levels of *EGFR* mutations detectable in plasma using digital PCR (dPCR) (BEAMing by Sysmex/Inostics) are associated with clinical outcome on osimertinib. We performed computational analysis to assess fluctuation in *EGFR* mutation concentration levels in the circulating tumor DNA from longitudinal plasma samples of 143 pts from the AURA Phase I clinical trial. We demonstrated that clearance of plasma *EGFR* mutations after 6 weeks of osimertinib therapy associates with improved overall response rate and progression-free survival. We validated this finding in AURA Phase II and III study populations using a different dPCR platform Genestrat (BioDesix). For patients that lack such a "plasma response" further in-depth profiling is ongoing to identify causes of more rapid disease progression and potential resistance mechanisms. In summary, monitoring of plasma *EGFR* mutations after initiation of osimertinib therapy can predict clinical outcome, and may rapidly identify patients where deep tumor profiling could be used to characterize drug resistance and, potentially, inform future therapeutic combinations. .

2383W

Digital spatial profiling platform allows for spatially-resolved, multiplexed measurement of solid tumor protein distribution and abundance in FFPE tissue sections. *C. Merritt, G. Ong, K. Barker, J. Jung, I. Sprague, Y. Liang, S. Warren, P. Webster, D. Dunaway, J. Beechem.* Nanostring Technologies, Seattle, WA.

Characterization of the spatial distribution and abundance of proteins within tissues enables a better understanding of biological systems in many research areas, including immunology and oncology. However, it has proven difficult to perform such studies in a highly multiplexed manner. To address this unmet need, we have developed a novel optical-barcode based microscope and tissue-sampling platform designed to simultaneously analyze hundreds of proteins on a single FFPE section (Digital Spatial Profiling, DSP). DSP probes are not multiplex-limited by spectral resolution. Instead, "colors" are determined using barcode indexing oligos that are conjugated to antibodies. These indexing oligos are conjugated via UV-cleavable linkers and, following UV light exposure, are released and siphoned off the tissue surface via a microcapillary tip. This UV-cleavage is precisely controlled by a digital micromirror device that can illuminate discrete regions ranging from entire tissue microenvironments to single-cells. For quantification of signal, the photocleaved oligos are hybridized to NanoString barcodes, providing digital counts of the protein in each region-of-interest using NanoString nCounter instruments. Using this novel approach, we spatially resolve over 30 solid tumor targets, simultaneously, on Her2+ breast cancer samples and Diffuse Large B-Cell Lymphoma samples. Solid tumor targets were probed with antibodies that recognize phosphorylated or total protein content (e.g. Her2, Ki67, P-AKT, P-ERK). We demonstrate multiplexed detection from discrete regions within a tumor and the adjacent normal tissue, enabling systematic interrogation of immune activity in FFPE samples. We also show the ability to profile either whole tissue sections or tissue microarrays. Finally, we validate that indexing oligo conjugation and high multiplexing of antibodies do not interfere with specificity. The simplicity of the DSP platform allows high-resolution, high-multiplexed, spatially-resolved protein characterization in any lab capable of performing immunohistochemistry procedures, providing a potential method that can bridge the gap between translational research discovery and clinical applications. Continued work on the platform will expand the library of protein targets accessible for profiling and future assay development will demonstrate multiplexing up to 800 targets.

2384T

Clinical and molecular findings correlation in patients with hereditary cancer syndromes: Experience of a Colombian reference laboratory. *T. Pineda, S. Maradei, A. Zarante, Y. Gomez.* Genetics Institute, BIOTECGEN S.A., Bogotá, DC, Bogotá, Colombia.

Introduction: Hereditary cancer presents in individuals with a germline pathogenic variant in genes related to cell cycle. In average, they span about 5 to 15% of all human cancers. Gene panels have arisen as an interesting option to identify mutations in families with hereditary cancer syndromes, optimizing diagnosis, treatment and following up algoritms on individuals at risk, with strategies towards decreasing morbidity and mortality rates and proposing new targeted therapies. *Aim:* To describe the experience of a Reference Center in Colombia for the diagnosis of Hereditary Cancer and to provide new data about our population mutational profile. *Methods:* 200 patients with personal and/or family history of breast, ovarian or colon cancer were referred to our laboratory for genetic testing from 2015 through March 2017, all of whom received pre and post-test counseling by a medical geneticist. Genes to be tested on each case were determined by the referring physician. Genomic DNA obtained from each patient's sample was sequenced using Illumina technology. All clinically significant observations were confirmed and then interpreted for their potential pathogenicity. Overall, we tested 170 patients (85%) with personal history of breast, ovarian or colon cancer, with or without family history of cancer. Also 30 patients (15%) with no personal history of cancer but with a positive family history of any tumor suggesting a hereditary cancer syndrome were tested. *Results:* We identified pathogenic variants in at least one of the high risk related genes in 13,53% of individuals with personal history of breast, ovarian or colon cancer with or without family history. Variants of unknown significance (VOUS) were reported in 25,88% of these patients. In 16,67% of individuals with only family history of breast, ovarian or colon cancer, mutations were detected in at least one of the high risk related genes. Variants of unknown significance (VOUS) were reported in 10% of these patients. *Conclusion:* This study highlights the relevance of an adequate use of gene panels on hereditary cancer syndromes as an attractive, efficient and cost-effective tool to identify disease causing mutations in patients and families with a suspected hereditary cancer syndrome, and also, sets a new record on the definition of our population mutational profile.

2385F

Polymorphic SNPs in breast cancer using molecular studies in Indian population. *M. Rao[1,4], N. Shah[1,2], P. Shah[1,3], Y. Panchal[1], K. Katudia[1], H. Ray[1], N. Khatri[1], U. Bhatia[1], S. Shah[1], B. Shah[1].* 1) Supratech Micropath Laboratory and Research Institute, Ahmedabad, Gujarat, India; 2) Department of Pediatrics, Nassau University Medical Center,New York City, NY, USA; 3) Department of Medicine, Lahey Hospital and Medical Center, Boston, MA,USA; 4) Department of Zoology, School of Sciences, Gujarat University, Ahmedabad, Gujarat, India.

A large number of distinct mutations in the BRCA 1 and BRCA 2 genes have been detected. But the role of polymorphic SNPs and except other mutations among Indian women for Breast cancer is scarcely reported. In this cohort, 79 referral breast cancer(B.C) women during 2015-16 were screened for BRCA1 and BRCA2 and BRCA1/2 gene mutations using extracted DNA by Sanger DNA sequencing and Next Generation Sequencing(NGS). The sequence data were analyzed using standard Ion Torrent SuiteTM software. Only 35 women were found for breast cancer positive (35/79;50.72%). In These cases,BRCA1 gene mutation was positive for 9(26%) cases, followed by 15(43%) for BRCA2 and11(31%) cases for BRCA1/2. The germline mutations for BRCA1 were 39/82(47%) and BRCA2, 43/82(53%) from total of 82. The types of mutations were of 72 missense 72/82(88%),frameshift 5/82(6%), nonsense 4/82(5%) and inframe types 1/82(1%). The Novel mutations detected were 5(p.Pro163fs,p.Asn997fs,p.Ser1481fs and two splice site missense SNPs) in our cohort affecting protein truncation and malfunction in these both homo- and heterozygotic cases. Several missense polymorphic SNPs detected here(72) were in exons 10,11,14,16,9,2 and 3 of the two genes which probably were negative for B.C, but support for susceptibility to breast cancer markers. The age groups of <40 years had 11 patients with 21% and 41-60 years possessed 18 cases with 55% followed by >61 years old (6) with 24% mutations in BRCA1 and BRCA2 genes. The age group of 41-60 years old had high incidence of mutation types and were related to pathologic conditions. Thus these data implicate that early age of 20-60 years old is probably susceptible for germline mutations in our study. However, the role of high incidence of polymorphic SNPs in relation to the causation of breast cancer needs to be evaluated in our Indian population.

2386W

Frequency of pathogenic and rare variants of uncertain significance in cancer patients and control cohort. *K.M. Rocha, M. Lazar, S.A.M. Ezquina, G.L. Yamomoto, M.C. Varela, V.N. Takahashi, M. Aguena, M. Zatz, M.R. Passos-Bueno. Gen e Biol Evolutiva, Inst Biocien Univ Sao Paulo, São Paulo, SP, Brazil.*

From the past five years, our lab has been offering molecular diagnosis using targeted massively parallel sequencing of a panel containing genes associated with several diseases such as neuromuscular, skeletal dysplasias and neurodevelopmental disorders. This panel also contains a few cancer associated genes and, more recently, we started offering molecular diagnosis for cancer as well. Since our cancer cohort is much smaller than for the other diseases (107 x 1593 cases), in our experience, the interpretation of variants that have not been reported in ClinVar or disease-specific databases is much more complex for cancer. This is also challenging given that our population is admixed and not very well represented in public population variants databases. In order to generate a more consistent filter for reporting variants of uncertain significance (VUS) we sought to evaluate the frequency of variants found in cancer and control cohorts in 14 cancer associated genes. Variants classified as pathogenic or likely pathogenic in ClinVar were detected in 9.3% of cancer cases and in 1.2% of the control cohort. Considering not previously reported frameshift and stopgain rare variants (frequency less than 0.5%) that we might consider potentially pathogenic, we detected around 4.7% and 2.6% for cancer and control cohorts respectively. Assuming that the frequency ratio of known pathogenic variants in cancer and control cohorts should be maintained when analysing potentially pathogenic rare variants, our results indicate that it is necessary to increase the stringency of analysis parameters in order to avoid reporting potentially false positives. Supported by: FAPESP-CEPID, CNPq and FUSP. .

2387T

Hereditary cancer risk testing within a Colombian cohort reveals high incidence of Lynch syndrome. *J.A. Rugeles[1], A.Y. Zhou[2], A. Zimmer[2], J. van den Akker[2], A. Leon[2], S.E Aruachan[1], M.E Gonzalez[1]. 1) Cancer genetics, IMAT Oncomedica S.A, Monteria, Cordoba, Colombia; 2) Color Genomics, 1801 Murchison Dr #128, Burlingame, CA 94010.*

Background: To date, hereditary cancer genetics has been most well-studied within the Caucasian and European populations. However, little is known about the incidence of hereditary cancer mutations in other populations. Here, we describe the demographics and characteristics of 145 Colombian high risk breast cancer and colon cancer index cases and referred for a 30-gene panel for hereditary cancer risk (the Color Test). Of the 20 index cases who received a positive result, cascade family testing was pursued in 13 families. **Methods:** All patients were seen at IMAT Oncomedica (Monteria, Colombia) and given pre-test counseling by an in-house genetics specialist. Patients with a significant personal and/or family history of breast or colon cancer were referred for the Color Test. **Results:** In this cohort of 145 high risk breast or colon cancer cases, 20 pathogenic or likely pathogenic mutations were identified and reported. The mutation carrier rate in this cohort was 13.8% (20/145). Among the 96 patients who reported a personal history of breast cancer, the mutation carrier rate was 9.4% (9/96). Among the 19 patients who reported a personal history of colon cancer, the mutation carrier rate was 47.4% (9/19). Of the 9 mutations reported in the breast cancer cohort, 3 mutations were identified in *BRCA1* and *PALB2* respectively, 2 in *BRCA2*, and one in *CHEK2*. Surprisingly, in the colon cancer cohort, all 9 of the patients who received a positive result were found to have a mutation in a Lynch syndrome gene (5 *MSH2* carriers, 3 *MLH1* carriers, 1 *MLH1+MUTYH* concurrent pathogenic mutation carrier). **Conclusion:** These data indicate that there is a significant need for genetic testing within Colombia. Current evidence from the United States suggests that Lynch syndrome accounts for ~5% of all colorectal cancer cases. Thus, our finding that 47.4% of the colon cancer patients in our cohort were found to carry a Lynch syndrome mutation is an indication that Lynch syndrome may be significantly more prevalent in the Colombian population than the United States. Our finding that 9.4% of breast cancer patients carry a mutation is very similar to the rates reported in the United States, where ~10-15% of breast cancer patients are thought to carry a mutation. The number of patients collected for this study is still quite small and further recruitment and analysis is warranted to fully understand the true prevalence of hereditary cancer within the Colombian population.

2388F

Use of functionally classified 207 BRCA2 missense variants to calibrate sequence-based computational prediction models. *H. Shimelis, S. Hart, T. Hoskin, R. Moore, A. Thomas, V. Suman, F. Couch.* Mayo Clinic, Rochester, Rochester, MN.

Genetic testing of individuals with a family history of breast and/or ovarian cancer has led to the identification of many unique *BRCA2* and *BRCA1* missense variants of uncertain significance (VUS). While breast cancer risk conferred by truncating loss of function variants is known, risks conferred by many germline missense variants have not been established. Classification of missense variants using quantitative models such as the established *BRCA1* and *BRCA2* multifactorial Likelihood Model has proved challenging due to limited availability of family data for each VUS. In contrast, functional analysis of VUS using validated assays and *in silico* prediction models may be useful for classifying VUS in the absence of case-control or family data. We have used a homology directed repair (HDR) cell-based assay to characterize missense variants in the DNA binding domain (DBD) of *BRCA2*. The method has been validated using known pathogenic and known non-pathogenic *BRCA2* missense variants and has 100% sensitivity (95% confidence interval (CI): 75.3%–100%) and 100% specificity (95% CI: 81.5%–100%) for pathogenic *BRCA2* variants. A classifier of variant pathogenicity was also developed based on the mean and variances of the distributions of the HDR results of the known pathogenic and neutral variants (1). We subsequently conducted a comprehensive assessment of 207 *BRCA2* missense variants using the HDR assay and identified 71 deleterious variants with >99% probability of pathogenicity, 116 neutral variants with >99% probability of neutrality, and 20 with intermediate function. Functional assay results were compared to 40 *in silico* prediction algorithms, using default and functional data optimized thresholds. An ensemble prediction model of *in silico* prediction algorithms trained on *BRCA2* functional data was developed using random forest modelling (RF) and a naïve voting method (NVM). Both RF and NVM models yielded high area under the receiver operating characteristic curve (AUC) of 0.90, 95%CI: 0.84-0.95 and 0.89, 95%CI: 0.83-0.96, respectively. Taken together our results show that both validated functional assays alone or *in silico* prediction models trained on functional data provide a robust tool for clinical annotation of *BRCA2* VUS.

2389W

Concurrent multiple molecular alterations involving *ALK*, *RET*, *ROS1* and *MET* in non-small cell lung cancer. *Z. Tang, J. Zhang, H. Chen, W. Wang, L.J. Medeiros, G. Tang.* MD Anderson Cancer Center, Houston, TX.

Targetable activating alterations in cancer genes, such as *EGFR, KRAS, ALK, RET, ROS1* and *MET*, are usually considered as mutually exclusive in non-small cell lung cancer (NSCLC). Rare NSCLC cases with coexistent alterations of *EGFR/ALK*, *EGFR/RET* and *EGFR/ROS1* have been reported. In this study, the database of the Clinical Cytogenetics Laboratory at The University of Texas MD Anderson Cancer Center for NSCLC cases assessed by fluorescence *in situ* hybridization (FISH) for abnormalities of *ALK*, *RET*, *ROS1* or *MET*. Fifteen cases (3 men and 12 women; 14 Caucasians and 1 African American) with ages ranging from 43 to 81 years (median 60 years) with positive results of more than one of these genes were identified. The combination of alterations included *ALK/ROS1* (n=4), *ALK/MET* (n=3), *ALK/RET* (n=1), *RET/MET* (n=4), *RET/ROS1* (n=2), and *ROS1/MET* (n=1). The frequencies of involvement were similar for all 4 genes, 53% for both ALK and MET (n=8), 47% for both RET and ROS1 (n=7). Activating gene mutations were also detected by next-generation sequencing for *TP53* (n=6), *EGFR* (n=5), *KRAS* (n=3) and *STK11* (n=2). Nine patients reported a smoking history (8 heavy and 1 light) and 6 patients were non-smokers. These findings suggest the need for assessing a panel of genes in NSCLC. Since targetable agents are available for each of these activating alterations, treatment with more than one targeted agent may be beneficial for this rare group of patients.

2390T

Towards prediction of platinum treatment response in ovarian cancer using machine learning approaches. *A. Vladimirova, R. Goold, T. Klingler.* Station X, San Francisco, CA.

Ovarian cancer is the eighth most prevalent cancer in women in the US and the fifth most common cause of death. The standard of care for ovarian cancer consists of a combination of surgery and chemotherapy, typically platinum-tax-ane treatment. However, many patients develop platinum resistance, defined as a relapse within 6 months of starting platinum chemotherapy. Predictive methods for early evaluation of the potential for platinum resistance may benefit patients by identifying those that might be better served with alternative second-line therapies or by enrollment in relevant clinical trials. In this study we generate and evaluate predictive models of resistance and sensitivity to platinum drugs in ovarian cancer by using the GenePool™ genomics platform and the integrated TCGA (The Cancer Genome Atlas) ovarian cancer RNA-seq and clinical data, including manually-curated platinum status information for each patient, and complement these analyses with a battery of machine learning approaches. We utilize GenePool best practices RNA-seq work-flows on ovarian primary tumors to derive and prioritize genes most strongly associated with platinum response. First, using clinical information, we define two ovarian cancer cohorts, one platinum-sensitive and the other platinum-re-sistant. We then compare the expression levels of all genes in these cohorts and identify those that are most differentially expressed. The gene expression results are analyzed using a variety of classifier approaches such as logistic regression, support vector machines and deep learning techniques to create, optimize and evaluate predictive models of platinum sensitivity status. Our workflows leverage cross-validation, dimensionality reduction, a variety of performance metrics to evaluate our models, as well as visualization of results to facilitate interpretation. We demonstrate a combination of approaches to de-rive and validate predictive models of platinum response in ovarian cancer and illustrate the potential of similar approaches to benefit cancer patient care. .

2391F

Functional analysis of BRCAness in female cancers. *M.P.G. Vreeswijk[1], L.M. van Wijk[1], M. de Jonge[2], K.A.T. Naipal[6], T.G. Meijer[6], M. Meijers[1], N.T. ter Haar[2], N.S. Verkaik[6], R.A.E.M. Tollenaar[4], W.E. Mesker[4], A. Jager[7], C.D. de Kroon[5], J.R. Kroep[3], K.N. Gaarenstroom[5], D.C. van Gent[6], T. Bosse[2], H. Vriel-ing[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Nether-lands; 2) Pathology, Leiden University Medical Center, Leiden, Netherlands; 3) Medical Oncology, Leiden University Medical Center, Leiden, Netherlands; 4) Surgery, Leiden University Medical Center, Leiden, Netherlands; 5) Gynae-cology, Leiden University Medical Center, Leiden, Netherlands; 6) Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands; 7) Medical Oncology, Erasmus University Medical Center, Rotterdam, the Netherlands.

Background. *BRCA1-* and *BRCA2*-related breast and ovarian tumors respond very well to treatment with Poly (ADP-ribose) polymerase (PARP) inhibitors (PARPi) because of their deficiency in homologous recombination (HR). However, tumors may contain a deficiency in HR that is unrelated to ger-mline mutations in *BRCA1* or *BRCA2* (i.e. BRCAness). Assessment of the HR efficacy of tumors might therefore allow identification of an additional group of cancer patients that could benefit from PARPi treatment. **Experimental procedure.** Fresh tumor samples were collected from patients undergoing surgery or ascites/pleural fluid drainage. Samples were *ex vivo* irradiated with ionizing radiation, fixed after 2 hours incubation at 37°C and embedded in paraffin. The ability of replicating tumor cells to accumulate RAD51 protein at DNA double strand breaks (RAD51 foci) was used as functional read out for HR proficiency. **Results.** To date, HR-status has been determined in breast (n= 56), ovarian (n= 59) and endometrial (n=28) tumor samples. Impairment of HR was observed in 20% of breast and in 40% of both ovarian and endome-trial tumor samples. Apart from patients with a germline BRCA mutation, also a subset of tumors from non-BRCA mutation carriers showed a HR deficient phenotype. Apparently, germline BRCA testing is not sufficient to identify all HR deficient tumors. Genetic and epigenetic analysis of tumor DNA are ongo-ing. **Conclusion.** The functional analysis BRCAness in fresh tumor tissue is a promising new tool to identify HR deficient tumors. A significant proportion of breast, ovarian and endometrial tumors are defective for homologous recombi-nation. Functional analysis of HR in fresh tumor samples may therefore allow the identification of a much broader patient population that could benefit from treatments targeting deficiencies in HR.

2392W

Detection and quantitation of M-BCR and m-BCR fusion transcripts by pico-liter digital PCR. *J. Woolworth-Hirschhorn, T. Baker, Y. Wang, C.A. Schandl.* Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC.

The BCR-ABL1 fusion gene results from a reciprocal translocation involving the long arms of chromosomes 9 and 22. The major breakpoint (M-BCR) fusion is seen in 99% of CML cases while the minor breakpoint (m-BCR) is rarely identified. In adult Ph+ B-ALL patients, M-BCR is identified in about 40% of cases with m-BCR found in 60%. The clinical significance of minimal residual disease is known, but that of transcript type in CML and in Ph+-ALL is unclear. However, studies suggest prognostic differences and distinguishing between them may become clinically relevant. Thirty CML and three Ph+-ALL cases with previously analyzed BCR-ABL1 transcript quantification by an alternate method were identified for analysis including five cases with no transcript detected. Pico-liter digital droplet PCR (pddPCR) was used for absolute quantitation of BCR-ABL1 transcripts using one-step reverse-transcription and real-time PCR. The presence of the M-BCR (b2a2, b3a2) and m-BCR (e1a2) fusion transcripts were quantified separately in each patient sample in a single assay by consequent droplet fluorophore detection. Results were compared to those obtained by a real-time PCR assay unable to distinguish between the M-BCR and m-BCR fusion transcripts and, when available, a qualitative assay distinguishing M-BCR from m-BCR transcripts. Sixteen of thirty CML cases demonstrated the M-BCR transcript by pddPCR (11/16 with concurrent reference laboratory qualitative agreement), nine were dual-positive for M-BCR and m-BCR (3 of 9 identified as dual-positive by reference laboratory, 3 demonstrating only M-BCR; and 3 not tested), four were not detected, and one demonstrated m-BCR (in conflict with M-BCR reported). Two of three cases of Ph+ B-ALL were dual-positive by pddPCR (one in agreement and one not detected by reference method) while no transcripts were detected in the third at the collection time-point (in agreement with reference result). Two dual-positive cases were identified significantly earlier by pddPCR than reference methodology: a Ph+-ALL 4 months earlier and a CML 8 months earlier. This study suggests that pddPCR may be more sensitive and identify recurrence earlier in a patient's course. As additional outcome data are evaluated for patients with variable transcripts, it is likely that prognostic implications may emerge. Thus, consideration of an ultrasensitive assay able to differentiate between M-BCR and m-BCR such as pddPCR may become clinically imperative.

2393T

Frequent detection of chromothripsis in acute myeloid leukemia with complex karyotype and marker chromosomes. *J. Lee[1], C. Seol[1,2], J. Lee[3], E. Seo[1,2].* 1) Medical Genetics & Genomics Center , Universiy of Ulsan College of Medicine and Asan Medical Center , Seoul, South Korea; 2) Department of Laboratory Medicine, Universiy of Ulsan College of Medicine and Asan Medical Center , Seoul, South Korea; 3) Asan Institute for Life Sciences, Universiy of Ulsan College of Medicine and Asan Medical Center , Seoul, South Korea.

Background Genomic rearrangements can drive the development of cancer through different mechanisms including chromothripsis, a catastrophic genomic event, seen in 2-3% of all cancers. Recent studies have revealed that chromothripsis could also be relevant for hematologic malignancies and was detected in 1.2% of myelodysplastic syndromes. We investigated the frequency and characteristics of the chromothripsis in acute myeloid leukemia (AML) patients with complex karyotype and marker chromosomes using array comparative genomic hybridization (aCGH) and fluorescence in situ hybridization (FISH) analysis. **Methods** A total of eight patients diagnosed as AML with complex chromosomal abnormalities and marker chromosomes were studied. Genome-wide copy number alterations (CNAs) and chromothripsis were analyzed in BM samples using SurePrint G3 Human CGH 180K microarray (Agilent Technologies, USA). FISH was performed using BAC clones located within amplified regions. **Results** The analysis of copy number profiles derived from aCGH data identified chromothripsis showing multiple non-contiguous CNAs in three patients who were young males with less than 50 years. These patients had chromothripsis events involving different chromosomes: 5q, 9p, 14q and 21q. Complex karyotype of 3 patients showed a median number of 11.7 chromosomal abnormalities with 1 to 3 marker chromosomes. Important genomic changes found by aCGH but not detected by conventional metaphase cytogenetics were validated by interphase fluorescence *in situ* hybridization (FISH), in the case of large recurrent deletions and gains. Array CGH analysis revealed the origin of marker chromosomes: amplified oncogenes such as DAXX on 6p21.32 and ERG on 21q22.2. **Conclusion** Chromothripsis is a genetic abnormality in which tens to hundreds of clustered genomic rearrangements in a one-step catastrophic event. Our results identified that marker chromosomes reflect heavily rearranged chromosomes following chromothripsis. This study showed three AML patients with gene amplifications had the multiple rearrangements of chromosomal segments, chromothripsis at 5q, 9p, 14q and 21q. Amplified oncogenes and chromothripsis may contribute to the tumorigenesis and progression of AML. Molecular karyotyping such as array CGH and FISH is useful to investigate the origin of marker chromosomes and complex chromosomal rearrangements. .

2394F

Tetraploidy and near tetraploidy in acute myelocytic leukemia and myelodysplastic syndrome: A report of seven new cases. *J. Liu[1], D. Fortuna[2], U. Uppal[3], S. Nedumakel[1], J. Gong[3], Z. Wang[4], S. Peiper[2].* 1) Clinical Cytogenomics Laboratory, Department of Pathology, Anatomy, and Cell Biology Thomas Jefferson University, 1025 Walnut Street, Philadelphia, PA 19107; 2) Department of Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107; 3) Section of Hematopathology, Hematology and Flow Cytometry, Department of Pathology, Anatomy, and Cell Biology, Sidney Kimmel Medical College, Thomas Jefferson University, 125 South 11th Street, Philadelphia, PA 19107; 4) Molecular & Genomic Pathology Laboratory, Department of Surgery, Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, 1025 Walnut Street, Philadelphia, PA 19107.

Tetraploidy and near tetraploidy (T/NT) (81–103 chromosomes) karyogram with or without numerical and structural abnormalities is rare in acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS). Only single case or small case series have been reported to date. The clinical significance is not well understood mainly due to the rarity of the entity. Here we report the clinical, morphological, cytogenetics, molecular genetics and prognostic features in seven cases of T/NT-AML/MDS.   The seven cases were identified by the searching the database of the Clinical Cytogenetics Laboratory at The Thomas Jefferson University Hospital between 2009-2016. Available clinical, hematopathological and laboratory data were all reviewed. A total of seven patients (6 with AML and 1 with MDS) with a T/NT karyogram were identified, comprising ~0.47% of all AML and MDS patients during this seven-year interval at The Thomas Jefferson University Hospital. Four patients were male and three were female. Age ranged from 54–81 years (median, 72 years). T/NT karyogram was detected at initial diagnosis of AML/MDS in five patients, and acquired during the course of disease in two patients. Three patients had a non-complex and four had a complex T/NT karyogram. Other than chromosomal gains, del 7q/7 was the most common structural abnormality in our series (n = 4) followed by del 5q/5 (n=2) and del 17p (n=2). Recurrent translocations seen in AML/MDS which were also present in our series were i(17)(q10) and t(1;7). All patients exhibited medium to large sized blasts, frequently with irregular nuclear contours and prominent nucleoli as well as various levels of cytoplasmic vacuoles and/or inclusions. Auer rods were not common. Mutation tests for *FLT3* internal tandem duplication (ITD) and *IDH1* codon 132 on two patients were negative. Next generation sequencing (NGS) technology based hematological malignancy gene panel test was performed on one case and revealed a pathogenic mutation in the gene *SRSF2*.   In summary, T/NT karyogram is an infrequent abnormality in AML and MDS with distinctive morphologic features for blasts. Seven new cases were reported. The prognostic significance and possible mechanism of a T/NT karyogram are discussed. This is also believed to be the first time that NGS panel test was utilized to characterize this unique entity.

2395W

A t(3;8)(q26.2;q24) involving the *EVI1 (MECOM)* gene on 3q26 in a case of acute myeloid leukemia preceded by polycythemia vera. *K. Liu[1,5], C. Makar[1,5], A. Bond[1,5], K. Stieglbauer[2,3], S. Wheaton[2,3], J. Reinartz[2,3], K. Cunnien[2], K. Peterson[2], L. Bartholomaus[2], D. Serk[2], B. Wyatt[2], S. Khan[1,5], C.A. Tirado[1,2,3,4].* 1) The International Circle of Genetic Studies, Los Angeles, CA; 2) Allina Health, Minneapolis, MN; 3) HPHA, Minneapolis, MN; 4) The University of Minnesota School of Medicine, Department of Laboratory Medicine and Pathology, Minneapolis, MN; 5) The University of California, Los Angeles, CA.

We report a 68 year old male with a history of polycythemia vera diagnosed in 2000. He was initially treated with phlebotomy and did well for many years until he developed symptomatic splenomegaly and pruritus. The patient was treated with ruxolitinib in 2012 and his splenomegaly improved. In late 2015 and early 2016 the patient developed cytopenias, increasing hepatomegaly, and increased poikilocytosis with dacrocytes in peripheral blood samples, suggesting myelofibrosis. A bone marrow biopsy in June 2016 confirmed the presence of myelofibrosis and revealed a hypercellular bone marrow (80%) with increased reticulin fibrosis (MF2-3), 5% blasts, and normal 46,XY karyotype. The patient was not a candidate for bone marrow transplant at this time, and ruxolitinib was discontinued to determine if blood counts would improve. Unfortunately, cytopenias progressed, and follow-up bone marrow biopsy in November 2016 documented acute myeloid leukemia (post-polycythemic myelofibrosis with acute leukemic transformation) with 20-30% blasts in the bone marrow, dense reticulin fibrosis, focal collagen fibrosis (MF3), and pancytopenia with 14% blasts in the blood. Chromosome analysis revealed an abnormal male karyotype with a t(3;8)(q26.2;q23) in 13 of the 21 metaphase cells examined. The remaining 8 cells were cytogenetically normal. Molecular cytogenetic studies confirmed a rearrangement of *MECOM (EVI1)*; however, no *PVTI* rearrangement was detected. We are aware of a case report in the literature previously describing a t(3;8)(q26.2;q24) involving *MECOM* (*EVI1*) and *PVTI*; however, to the best of our knowledge the t(3;8)(q26.2;q23) has not been reported in acute myeloid leukemia (AML), de novo or therapy related-myelodysplastic syndrome (MDS), or MDS or myeloproliferaitve disorder progressing to AML. Rearrangements of chromosome 3 involving the *MECOM (EVI1)* gene on 3q26.2 are recurrent abnormalities in myeloid disorders and are usually associated with disease progression and a poor prognosis.   .

2396T

Identification of cancer-associated copy number variations through integrated genomic analysis. *M. Luo[1,2], B. Luo[1,3], J. Tang[1], A.G. Gleason[1], G.T. Akgumus[1], L.K. Conlin[1,2], M.M. Li[1,2].* 1) Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Fox Chase Cancer Center, Philadelphia, PA.

Copy number variations (CNV) and loss of heterozygosity (LOH) are common genomic alterations in cancer. These genomic alterations may be critical in cancer diagnosis, prognosis and treatment. CNVs can be detected through different technologies, including traditional cytogenetics, fluorescence in situ hybridization (FISH), chromosomal microarray analysis (CMA), and recently, next generation sequencing (NGS). In 2013, American College of Medical Genetics and Genomics (ACMG) published technical standards and guidelines for CMA including CNV detection and interpretation in neoplastic disorders. However, efficient analysis and accurate interpretation for CNVs still remains to be a challenge for clinical laboratories given the complexity and dynamics of the changes in tumor genomes. In our laboratory, different in-house tools have been created to integrate genotyping data, log R ratio, and gene content. These developments significantly increased our efficiency and sensitivity in the detection of disease-associated changes. They allow us to identify mosaic CNVs at approximately 5-10% allele frequencies and different levels of bone marrow engraftment after transplantation. These tools also assist discerning ploidy changes which are often important in tumor risk stratification. An algorithm to consolidate data from different technologies, including both traditional cytogenetics and sequencing analysis, as well as data from multiple specimens of the same patient, has also been implemented to improve the accuracy of data interpretation and diagnostic yield. The algorism facilitates the detection of reverse mutations, novel structural abnormalities, relevant germline changes, and other complex genomic alterations. Since 2008, we have studied more than 2,400 pediatric oncologic samples including both solid tumors and hematological malignancies. Clinically significant CNVs were identified in about 80% of the cases which provided solid genomic evidence for precision cancer care.

2397F

Characterization of interstitial deletions of 9q in ten AML cases. *Z. Qi[1], K.W. Wen[1], A. Ki[2], L. Lv[3], S. Prakash[1], S. Kogan[1], J. Yu[1].* 1) Department of Laboratory Medicine, UCSF, San Francisco, CA; 2) Clinical Cytogenetics Laboratory, UCSF Medical Center, San Francisco, CA; 3) Department of Oncology and Hematology, The Second Hospital of Jilin University, Changchun, Jilin, China.

Interstitial deletions of the long arm of chromosome 9 [del(9q)] are seen in 2-5% of acute myeloid leukemia (AML). However, precise breakpoints of del(9q) in AML remain mostly unknown, making it difficult to identify candidate pathogenic genes in the deleted regions and to further understand the pathogenic effect of the deletions. We present here molecular breakpoints of del(9q) in 10 AML cases, detected by single nucleotide polymorphism array. Our study revealed a cluster of proximal breakpoints within 9q13q21, a region enriched with segmental duplication and a "hotspot" for genomic rearrangements. But the distal breakpoints of the del(9q) varied greatly. The deleted 9q regions can be further divided into an approximately 14.4 Mb proximal region (chr9:70950015-85397699, GRCh37) and an approximately 24.0 Mb distal region (chr9:85397700-109427261). The former seems to be not neoplasm-related, since deletions of this region are also reported in individuals without hematological or other neoplasms; whereas deletions of the latter have been reported mostly in hematological neoplasms. It is noteworthy that a 7.3 Mb deletion within the distal region appears to occur in all cases with survival time less than 22 months in this study. Loss of function or reduced expression of the *GADD45G*, *MIRLET7A1*, *MIRLET7D*, *MIRLET7F1*, and *FANCC* genes in this 7.3 Mb region has been associated with neoplasms. The 10 del(9q) cases can be divided into four groups based on additional genetic changes and their response to the treatment, including three cases with t(8;21) or t(15;17) and good response to chemotherapy, two cases with biallelic *CEBPA* mutations that achieved and have remained in complete remission after bone marrow transplant, three cases with *FLT3*-internal tandem duplication and short survival time (<15 months), and two cases with other changes of unknown significance and short survival time (13 and 22 months). Taken together, our findings suggest that only the distal del(9q) appears to associate with AML. Del(9q) may not change the prognosis in cases with favorable AML mutations. The deletions, in particular the deletion of the 7.3 Mb region, may affect the survival time in other AML cases. Further investigation of the pathogenic effects of the del(9q) is warranted.

2398W

Cytogenetic characterization of Richter transformation in chronic lymphocytic leukemia. *G. Tang, Z. Tang.* Hematopathology, UT MD Anderson Cancer Center, Houston, TX.

Background: Richter's transformation (RT) is a rare disorder that occurs in < 10% of patients with chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL). The transformation refers to the development of either diffuse large B-cell lymphoma (DLBCL, most commonly) or classical Hodgkin lymphoma (rare) during the diseases course. Little information regarding karyotypic or fluorescence in situ hybridization (FISH) findings in RT is available in the literature. Methods: The cytogenetic findings from conventional karyotyping and/or FISH analysis for CLL panel (*ATM, CEP12, D13S319, LAMP1*, and *TP53*) in 32 CLL/SLL patients with transformation to DLBCLwere investigated. In 16 of these patients, the results of karyotyping or FISH analysis in a pre RT sample were also available. A comparison between the cytogenetic findings of the pre RT and the RT samples was performed. Results: 28 (88%) RT samples showed complex karyotype(>3 abnormalities), while 4 (12%) had either a normal diploid or non-complex karyotype. Abnormalities involving chromosome 17(-17 or del17p) which resulting in *TP53* deletion were identified in 16 (57%) samples with complex abnormalities; 9(32%) samples had either triploid or tetraploid clones; 2 (7%) samples had t(8;14)(q24.1;q32)/ *IGH-MYC* rearrangement. 25 samples also underwent FISH analysis, 12 (48%) had *TP53* deletion; 5 (20%) had *ATM* deletion and 1(4%) had both *TP53* and *ATM* deletion, 7 (28%) were negative for either deletion. 16 cases had karyotyping in a pre RT sample, 9 had a normal and 5 had an abnormal karyotype. In comparison with the subsequent RT samples, 7 developed new clones, 5 showed clonal evolution and 2 had no change in the karyotype. 13 samples had FISH testing on pre RT sample, 6 had a *TP53* deletion and 2 had *ATM* gene deletion. In the subsequent RT samples, 12 had FISH testing, 6 had a *TP53* deletion and 3 had *ATM* deletion; all cases with pre RT *TP53* or *ATM* gene deletion had the deletion in the RT sample, 2 pre RT samples that were negative for either deletion acquired a *TP53* or *ATM* deletion in the RT sample, while 3 remained negative.Conclusions: RT has a high frequency of complex karyotype, including -17 or del(17p), which are not commonly seen in CLL/SLL. Clonal evolution and/or the acquisition of new clones are frequent in RT, occurring in 12 of 16 (75%) tumors. *TP53* and *ATM* deletions were present in the pre RT and RT sample, supporting *TP53* or *ATM* deletion as a risk factor for the development of RT.

2399T

A plasma cell myeloma case with an abnormal clone showing a t(8;22)(q24.1;q11.2) within the context of a hyperdiploid complex karyotype.
C.A. Tirado[1,2,3], K. Liu[3], A.V. Dowiak[3], W. Wyat[1], S. Wheaton[2], K. Peterson[1], D. Serk[1], K. Cunnien[1], J.J. Reinartz[2]. 1) Pathology, Allina Health, Minneapolis, MN; 2) Hospital Pathology Associates (HPA), Minneapolis, MN; 3) The International Circle of Genetic Studies, Los Angeles, CA.

We report here a 74-year-old male who was seen for recurrent respiratory infections, fatigue, and weight loss in Nov 2016. A CBC revealed anemia, and a serum protein electrophoresis revealed a large IgG lambda monoclonal gammopathy (total IgG level of 6430) and depressed IgA and IgM levels. His urine immunofixation study showed trace IgG lambda protein with free kappa light chains. A PET/CT scan was negative for lytic lesions; calcium and creatinine were normal. Bone marrow biopsy showed 90% involvement by plasma cell myeloma (PCM)[90% plasma cells, 40% cellular bone marrow]. Cytogenetic analysis of the bone marrow showed complex karyotype as described as: 53,Y,add(X)(p22.1),del(1)(p13p22),+3,add(3)(p13),add(4)(p12),+6,del(6)(q13q25),t(8;22)(q24.1;q11.2),+9,+11,+15,+15,+21[7]/46,XY[13]. This particular pattern with deletion 1p and 6q, and a t(8;22)(q24;q11.2) within the context of a complex karyotype is seen in PCM. Fluorescence in situ hybridization analysis on CDC138 sample was positive for additional copies of CEP7 (centromere 7), CEP9 (centromere 9), CEP11 (centromere 11), and CEP15 (centromere 15) suggesting polysomy. FISH using the MYC Vysis break apart probe showed evidence of *MYC* rearrangement similar to the breakpoint site seen in Burkitt lymphoma with t(8;22)(q24;q11). FISH using the IGL break apart probe (Cytocell) showed evidence of 22q11.2 rearrangement . The signal pattern showed a residual green signal (BCR), a green signal on the derivative 8 and a red signal on the derivative 22 suggesting that the breakpoint at 22q11.2 in this patient was located downstream of the BCR region of the *IGL* gene. Variant Burkitt-type translocation, t(8;22)(q24;q11), is a very rare abnormality in PCM in this case is one of the few reported up to date. In these patients, *MYC* abnormalities appear late in the course of the disease and have an immature phenotype. A review of several cases in the literature suggest that this translocation leads the *MYC* gene under direct regulation of the enhancer of the partner gene, in our case, the *IGL* or a nearby gene, thereby causing high level transcription of *MYC*. This abnormality is usally present within a complex karyotype and is considered to be associated with tumor progression and poor prognosis. .

2400F

A pediatric B-ALL with doubled near-haploidy by chromosomal microarray and negative MRD at day 29 of induction: Case report. *J. Xu[1], S. Kuerbitz[2], L.B. Baughn[3], K. Eickholt[1], I. Chen[4], S.C. Reshmi[5], R.W. Novak[1], J. Hord[2].* 1) Pathology, Akron Children's Hospital, Akron, OH, USA; 2) Hematology/Oncology, Akron Children's Hospital, Akron, OH, USA; 3) Mayo Clinic, Rochester, MN, USA; 4) Pathology, The University of New Mexico, NM, USA; 5) Institute for Genomic Medicine, Pathology, Nationwide Children's Hospital, OH, USA.

Hypodiploidy (HD) and doubled HD occur in 1-2% and ~0.7% cases, respectively, of pediatric B-ALL. Cytogenomic variation, prognosis and optimal therapies for these cases are yet to be fully appreciated. We present a 13-year-old girl diagnosed with B-ALL, 94% blasts in bone marrow (BM) and negative central nervous system. FISH of 24h BM culture was negative for rearrangements of *BCR/ABL1, ETV6/RUNX1, MLL* and CEN 4 and 10. G-banding of direct and 24h cultures showed a hyperdiploidy with 49~52,XX-,+X,+X,+14,+14,del16q,+21,+21[cp9]/46,XX[16]. Chromosomal microarray (CMA) study indicated a pseudo-hyperdiploid clone from doubling of an original near-haploid (NH) clone with loss of all chromosomes except X, 14 and 21. There were several deletions, including deletions 5q34, 6p22.2 (involving several histone genes), and 16p13.3p13.2 (involving *CREBBP* gene) consisting with previous reports in NH ALL (Nat Genet. 2013; 45:242-252). There were mosaic and heterozygous deletions 16q13q23.2 and 16q23.3q24.3, likely subsequent to the doubling and suggesting clonal evolution. The patient was tested negative for *CRLF2* rearrangements, *JAK* fusions, multiplex RT-PCRs for known actionable fusions, and targeted *JAK/IL7R* sequencing alterations. Treatment included standard 5 drug induction chemotherapy with oral prednisone, intrathecal methotrexate, and intravenous vincristine, daunorubicin and pegaspariginase, which was tolerated without significant side-effects. Minimal residual disease (MRD) was negative at day 29 of induction.   This is one of few reported B-ALL cases with doubled NH characterized by CMA and raises two interesting points. 1) The G-banding/FISH of this case showed a hyperdiploidy with no evidence of HD, mix of HD/hyperdiploidy, or two copies of the same structurally rearranged chromosome, which would have triggered CMA testing for a pseudo-hyperdiploid. 2) HD was reportedly associated with a very poor prognosis in pediatric ALL. Although our patient is currently undergoing treatment and the outcome is unclear, this case report emphasizes the need for considering CMA for in B-ALL with non-classical hyperdiploidy for accurate risk stratification/treatment.

2401W

A novel *XPA* gene mutation (c.773delG, p.R258Lfs*11) in two siblings with Xeroderma Pigmentosum. *B. Balta[1], M. Erdogan[1], H. Gumus[2], A. Kiraz[1], R. Ertas[3].* 1) Department of Medical Genetics, Kayseri Training and Research Hospital, Kayseri, Turkey; 2) Department of Child Neurology,Erciyes University,Medical Faculty ,Kayseri, Turkey; 3) Department of Dermatology, Kayseri Training and Research Hospital, Kayseri, Turkey.

Xeroderma pigmentosum (XP) is a heterogeneous autosomal resessive disorder characterized by increased sunlight sensivity with the development of carcinomas such as basal cell carcinoma, squamous cell carcinoma, melanoma at an early age. Some patients have neurologic manifestation including acquired microcephaly, absent deep tendon reflexes and progressive senso-rineural hearing loss. The diagnosis of XP is based on clinical findings, family history and identification of pathogenic variants in *ERCC1, ERCC2, ERCC3, ERCC4, ERCC5, POLH, DDB2, XPA, or XPC*. Two brothers at the age of 5 and 4 were admitted our department with photosensivity, persistent erythema, mild intellectual disability and microcephaly. Parents have not consanguinity but are from same village. Cranial MRI shows mild ventricular dilatation in both lateral ventricles and 3.ventricle. By whole exome sequencing we detected a novel homozygous mutation in XPA gene, c.773delG, p.R258Lfs*11. The aforementioned mutation has not been previously reported.

2402T

Implementation of a multi-gene panel in selected Greek breast and ovarian cancer patients reveals loss-of-function mutations in multiple genes.
F. Fostira, I. Konstantopoulou, M.S. Papamentzelopoulou, D. Kalfakakou, P. Apostolou, I. Konstanta, A. Vagena, I.S. Vlachos, D. Yannoukakos. Molecular Diagnostics Laboratory, NCSR Demokritos, Athens, Attika, Greece.

   Introduction: NGS technologies, especially in the form of gene panels, have surely entered the clinical practice for cancer patient management. Among high-risk breast and/or ovarian cancer patients of Greek descent, although five *BRCA1* alleles with founder effect dominate the mutation spectra, additional susceptibility alleles in a number of genes are definitely present. **Materials and Methods:** In total, 853 breast and 408 ovarian cancer patients fulfilling NCCN genetic testing criteria have been analyzed by Trusight Cancer panel targeting 94 genes associated with cancer predisposition, in order to identify loss-of-function (LoF) mutations. **Results:** LoF mutations have been identified in 22% and 26% of the breast and ovarian cancer group respectively, in 29 genes involved in DNA repair. Of these, 87% and 89% lie in genes with known or suspected associations with breast and/or ovarian cancer predisposition, predominantly *BRCA1* and *BRCA2* (62%; 74%). Mutations in *CHEK2*, *PALB2* and *ATM* were more frequent in the breast cancer series and *RAD51C* in the ovarian cancer series. Of note is the remaining 12% of LoF alleles in non-associated genes to date, which can be either regarded as incidental findings, while putative causality for breast/ovarian cancer cannot be excluded. **Conclusions**: Among a high-risk group of breast and ovarian cancer patients of Greek descent, where there are strong founder effects, the mutational spectrum is highly heterogeneous with respect to both loci and alleles. This approach can shed light on associations with the post-BRCA genes and will provide connections of phenotypes with mutations in additional DNA repair genes.

2403F

Screening for rearrangements in *RB1* gene/ 13q14 through real-time PCR. *R.M. Freitas[1], P. Pereira Sena[2], M. Campos Junior[3], F. Regla Vargas[1,4].* 1) InstituiÇÃo Oswaldo Cruz, Rio de Janiero, Brazil; 2) Institut für Medizinische Genetik und Angewandte Genomik, Universitätsklinikum Tübingen, Alemanha; 3) Human Genetics Laboratory, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil; 4) Genetics and Molecular Department, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, Brazil.

   Patients with deletion of chromosomal region 13q that includes the *RB1* gene show retinoblastoma (RB) and variable clinical features. About 5-15% of the patients with RB are heterozygous for a gross deletion that includes the whole or substantial parts of *RB1* gene. Deletions had been identified by analysis of short tandem repeat (STR) *loci* within *RB1*, quantitative multiplex PCR (QM-PCR), multiplex ligation-dependent probe amplification (MLPA) and classical cytogenetic analysis. We have designed a method based in real-time PCR for search of deletions / duplications of *RB1* gene. The method was tested in control DNA and validated in a set of RB patients with partial or total *RB1* deletions. Additionally, the method was validated in one patient with trisomy 13 (Patau syndrome), who has three copies of *RB1* gene. The specificity, sensitivity and clinical utility of the assay were demonstrated in detecting allele-specific copy number variation, and can be useful for analysis relative copy number. We have selected, in addition to the *RB1* gene, two other genes (*SUCLA2* and *MED4*) that are adjacently located to *RB1* in chromosomal region 13q14.2. Genomic DNA was isolated from peripheral blood samples using standard salting out method. The amplified segments were analyzed by relative quantification, relative copy number method ($2^{-\Delta\Delta Ct}$) (Livak et al., 2001). Each plate contained an internal control (*ALB* gene) and a trisomic sample was also used to observe the duplication of the region, all in triplicate. The CT's (cycle threshold) values obtained were used to calculate the relative copy number of each sample. All reactions were performed with Sybr Green (Invitrogen®). The mean of CT's triplicates were calculated, and the $\Delta\Delta$CT and $2^{-\Delta\Delta Ct}$ of each sample, having as reference CT normal individual with the primer *ALB* control and the CT trisomic sample, were also estimated. Seven samples of retinoblastoma patients with partial or total *RB1* deletion detected by MLPA were used for validation. All samples were validated by real-time quantitative PCR. Seven retinoblastoma patients carried complete deletion of *RB1* gene were identified by the MLPA technique, among 66 retinoblastoma patients tested (Sena, 2013). These deletions comprise about 10.61% of the mutations identified in retinoblastoma (Price et al., 2014). Thus, the relative quantification real-time PCR technique to investigate deletions in the *RB1* gene becomes advantageous, since it is a fast, sensitive and low-cost technique.   .

2404W

Molecular insights into the missing heritability of familial ovarian cancer. *J. Stafford[1], G. Dyson[1], N. Levin[1,2], S. Chaudhry[1], C. Wernette[1], N. Petrucelli[1,2], M. Simon[1,2], M. Tainsky[1,2].* 1) Wayne State University , Detroit, MI; 2) Karmanos Cancer Institute, Detroit, MI.

While up to 25% of ovarian cancer (OVCA) cases are thought to be due to inherited factors, the majority of genetic risk remains unexplained. We sought to identify previously undescribed OVCA risk variants through the whole exome sequencing (WES) and candidate gene analysis of 48 European American women with ovarian cancer selected for high risk for genetic inheritance, yet negative for any known pathogenic variants in either *BRCA1* or *BRCA2*. Whole exome sequencing followed by candidate gene analysis and in silico SNP assessment was employed to identify suspect variants followed by validation using Sanger DNA sequencing. A SNP was considered suspicious if it was rare (MAF<.001 CEU), missense or frameshift, and predicted to be damaging (SIFT/PolyPhen) to a gene associated with hereditary cancer risk or involved in DNA repair or cell cycle control. Five pathogenic variants were detected including two frameshift mutations in *ATM*, two truncating variants in *RAD51D* and another pathogenic nonsense mutation in *FANCM* which predisposes to breast cancer but not featured on current testing panels. Numerous rare and predicted to be damaging variants of unknown significance (VUS) were detected in genes on current commercial testing panels, most prominently in *ATM* (n=6) and *PALB2* (n=5). The *BRCA2* variant p.K3326*, resulting in a 93 amino acid truncation yet considered benign, was overrepresented in our sample (odds ratio = 4.95, p = 0.01). This variant is has recently been shown to be associated with susceptibility to various cancers and coexisted in the germline with other pathogenic variants, suggesting a possible role as a modifier of genetic penetrance. Furthermore, we detected loss of function variants in non-panel genes involved in OVCA relevant pathways; DNA repair and cell cycle control, including *CHEK1, TP53I3, REC8, HMMR, RAD52, RAD1, POLK, POLQ,* and *MCM4*. Functional analysis upon knockdown of some of these genes led to significantly reduced levels of homology directed repair. Our findings suggest that the likely sources of the unexplained genetic etiology in OVCA are due to rare VUSs in panel genes, risk loci occurring in non-panel genes involved in DNA repair or cell cycle control, and polygenic risk inheritance. Definitive functional tests are necessary to discriminate variants of decreased function from benign polymorphisms and ideally, could reveal the impact of multiple low effect mutations as a means of personalized genetic risk evaluation.

2405T

Current laboratory practices in *BRCA1* and *BRCA2* testing: Results from an international survey of 86 laboratories. *A. Toland[1], S.L. Neuhausen[2], J.O. Culver[3], F. Couch[4], F.B.L. Hogervorst[5], D.M. Eccles[6], W. Foulkes[7], A.B. Spurdle[8], C. Houdayer[9], A.N. Monteiro[10], S. Plon[11], L. Brody[12], A. Forman[13], BIC Steering Committee.* 1) Internal Medicine/CBG, Ohio State Univ, Columbus, OH; 2) Dept Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA; 3) USC Norris Comprehensive Cancer Center, Los Angeles, CA; 4) Mayo Clinic, Rochester, MN; 5) Netherlands Cancer Institute, Amsterdam, The Netherlands; 6) University of Southampton, Southampton, UK; 7) McGill University, Montreal, Canada; 8) QIMR Berghofer Medical Research Institute, Herston, QLD, Australia; 9) Oncogenetics and INSERM U830, Institute Curie, Paris & Paris Descartes University, Paris, France; 10) Moffitt Cancer Center, Tampa, FL; 11) Baylor College of Medicine, Houston, TX; 12) National Human Genome Research Institute, Bethesda, MD; 13) Fox Chase Cancer Center, Philadelphia, PA.

Clinical testing for *BRCA1* and *BRCA2* mutations became available more than 20 years ago. Since then, the number of individuals undergoing testing and laboratories offering testing for hereditary breast and ovarian cancer (HBOC) have increased dramatically. With the expiration and overturning of the *BRCA1* and BRCA2 patents, limitations on which laboratories could offer commercial testing, both nationally and internationally, were lifted. These legal changes occurred approximately the same time as the widespread adoption of massively parallel sequencing (MPS) technologies. Although the combination of advances has significantly changed laboratory practice, little is known about how the new technologies are currently being used in HBOC testing. We sought to obtain a snapshot of current laboratory genetic testing practices for *BRCA1* and *BRCA2* across the world. We employed an online survey of 65 questions covering four areas: laboratory characteristics, technological details on methods used for testing, variant classifications, and customer-service information. Over 78 non-US based laboratories began the questionnaire with 48 laboratories from 6 continents fully completing the survey. The majority of participating laboratories were in Europe. A slightly modified questionnaire was developed for laboratories located in the United States (US). Eight of 27 US-labs invited to participate completed the survey. Survey results were tallied for US and non-US labs. Most non-US based laboratories (92%, 72/78) use MPS-based platforms for *BRCA* analysis with Sanger sequencing for validation of variants, and the remaining 7.6% use Sanger sequencing as their only platform. All eight US laboratories utilize MPS for identification of in/dels and single nucleotide variants. Multiplex ligation-dependent probe amplification is the most common method for identifying and/or confirming large rearrangements. Laboratories differed widely on: 1) criteria for minimum read depths; 2) amount of sequence from intronic and regulatory regions; 3) variant classification criteria and approaches; 4) number of genetic counselors on staff; and 5) deposition of identified variants into public databases. A wide range of approaches and standards were used to determine pathogenicity of *BRCA1/2* variants. Data from this study may be useful for national and international agencies to set recommendations for minimal quality standards for *BRCA1/2* clinical testing.

2406F

Hyb & Seq™: The next generation of simultaneous RNA and DNA detection in liquid biopsy without the library and sample preparation bottlenecks. *J. Beechem, M. Hoang, M. Walsh, M. McElwain, E. Piazza, M. Rhodes, D. Kim.* Nanostring Technologies, Inc., Seattle, WA.

Nucleic acid (NA) sequencing of circulating cell-free (cf) DNA and RNA or cfNA from blood is an exciting, noninvasive tool for personalized medicine research. However, clinical implementation is hindered by the need for complex sequencing library preparations that amplify the low levels of cfNA but also introduce sampling errors and artifacts. Hyb & Seq technology is a library-free, amplification-free, single molecule sequencing technique that uses cyclic nucleic acid hybridization of fluorescent molecular barcodes to sequence native single molecule targets. Here we demonstrate Hyb & Seq's simple workflow on plasma, simultaneously capturing and directly sequencing both DNA and RNA. Target-specific capture probes were mixed directly into plasma to hybridize with circulating cfNA molecules. Following a short incubation, excess probes were removed, and captured cfNA molecules were immobilized onto a flowcell, and each cfNA molecule directly sequenced using Hyb & Seq. Total processing time from plasma sample to start of sequencing was under 60 minutes, with hands-on time of less than 15 minutes. Experimental results showed that all 100 targets were captured onto the flow-cell using 5 mL – 10 mL of plasma without PCR amplification. Using a reference plasma sample with known mutations and allele frequencies, accurate detection was achieved down to ~1% mutation allele frequency. Finally, 100 targets were simultaneously sequenced and digitally quantified for relative cfDNA and cfRNA levels, and presence of the single nucleotide variant. The Hyb & Seq system demonstrates the unique ability to simultaneously capture and directly sequence cfNA using a rapid sample-to-answer sequencing process without amplification or enzymology, thus opening up the possibility of assessing complex biomarker signatures from liquid biopsy.

2407W

Optimizations in target enrichment and bioinformatics enable sensitive detection of copy number variations in targeted NGS. *J. van den Akker, G. Mishne, R. O'Connor, Z. Mahkovec, T. Sittler.* Color Genomics, Burlingame, CA.

While copy number variants (CNVs) account for a significant proportion of mutations, their detection in targeted next generation sequencing (NGS) assays has been historically challenging. However, we have developed sophisticated methods for the detection of CNVs (deletions and duplications) in our 30-gene NGS-based test for hereditary cancer risk (the Color Test). This has been accomplished by optimizations at multiple levels to boost detection sensitivity. First, in the enrichment step, we target not only coding sequences, but also extended flanking regions of small exons as well as the breakpoints of published structural variants. Next, CNVs are detected with multiple algorithms targeting different variant properties. Read depths are subject to custom bias correction and segmented by multiple methods by CNVkit (CBS & fused LAS-SO). Segments are annotated and processed by a multi-parameter algorithm trained on past data to call CNVs. In addition, CNVs are detected by an in-house developed split-read algorithm. In regions where reads are not consistently clipped, signals from paired-end reads with large insert sizes are used to assess CNV breakpoints. All variants classified as VUS, likely pathogenic, or pathogenic are confirmed by either aCGH, MLPA, or Sanger sequencing of a PCR fragment covering the breakpoints. Color has identified CNVs in over 220 samples to date, representing approximately 150 distinct variations. They were detected in 22 genes, with almost one-third occurring in *BRCA1*. Detailed breakpoint analysis revealed at least 44 distinct CNVs in *BRCA1* and 7 CNVs in *BRCA2*. Out of the 15 duplications detected in *BRCA1* and *BRCA2*, 12 were confirmed to be in tandem. This knowledge is required to predict the RNA sequence, which guides variant classification. While the clinical relevance of CNV detection in *BRCA1*, *BRCA2*, and the Lynch syndrome genes has been well established, data on other genes associated with hereditary cancer is scarce. Here, we highlight the CNVs we identified in *CHEK2* (n=27, 10 distinct CNVs), *ATM* (n=21, 11 distinct), and *RAD51C* (n=15, 11 distinct). Importantly, our assay can detect CNVs of all sizes, and more than 30% impacted only a single exon. Moreover, our split-read detection algorithm has detected 15 variants between 25-250 base pairs, a size range known to be difficult in most NGS assays. In conclusion, by combining multiple complementary approaches, CNVs of variable sizes can be detected across a multi-gene NGS panel.

2408T

Hereditary cancer risk testing of 11,570 individuals with a multi-gene panel. *J. Ji, A. Zhou, A. Zimmer, W. McFadden, J. Akker, R. O'Connor, K. Kim, D. Keen-Kim, A. Leon.* Color Genomics, Burlingame, CA.

Here, we describe the demographics and characteristics of 11,570 individuals who received the Color Test, a 30-gene next generation sequencing (NGS) based panel for hereditary cancer risk. The overall proportion of individuals who carried a pathogenic or likely pathogenic variant in this cohort, or mutation carrier rate, was 11.9% (1,377/11,570). We identified and reported 1,434 pathogenic or likely pathogenic variants in this cohort. A total of 49 (3.6%) individuals were found to carry two concurrent pathogenic variants in different genes, and one individual carried three concurrent pathogenic variants in different genes, indicating the importance of broader multi-gene panel testing. Additionally, 6 individuals carried two different pathogenic variants in the same gene. Many individuals in the cohort were referred to testing for a personal or family history of breast or ovarian cancer. Therefore, it was not surprising that *BRCA1* and *BRCA2* account for 31.5% (452/1,434) of all pathogenic variants identified. Among the 1,377 individuals who received a positive result, 201 (14.6%) carried a single pathogenic variant in *MUTYH*, while 4 individuals carried two different pathogenic variants in *MUTYH*. Additionally, 69 (5.0%) individuals carried the *CHEK2* c.470T>C (I157T) variant and 146 (10.6%) individuals the *APC* c.3920T>A (I1307K) variant, both of which are known to have lower penetrance and specific screening recommendations. The mutation carrier rate amongst individuals who reported no personal or family history of cancer was 8.3% (92/1,108). The majority of the cohort self-reported as ethnically Caucasian (51.9%, 6,001/11,570) or Ashkenazi (10.9%, 1,256/11,570), reinforcing a well-known need to increase genetic testing awareness and uptake within other underrepresented ethnic populations. Only 18.5% (2,140/11,570) of the cohort was male, even though the genes on the panel are relevant to both males and females, indicating a potential underlying bias for women to seek out genetic testing more prevalently than men. Finally, only 28.0% (386/1,377) of individuals who received positive results were under the age of 40, indicating the importance of increasing awareness and uptake of genetic testing within the younger population, at an age for which preventative care is more relevant. Taken together, these data reinforce the continued need to increase awareness and broaden access to genetic testing within the general population.

2409F

What's in a VUS rate? Simulated VUS calculations for hereditary cancer genes in a general population using population frequency data and ClinVar submissions. *K.E. Kaseniit, M. Judkins, R. Mar-Heyming, H.P. Kang, E.A. Evans, K. Ready, D. Muzzey.* Counsyl, Inc., South San Francisco, CA.

A common metric by which inherited-cancer tests are compared is the likelihood of a patient receiving a variant of uncertain significance ("VUS"). This "reported-VUS likelihood" or "VUS rate" metric has questionable value, however, because multiple factors influence it: the number of genes tested, the particular assayed regions (e.g., exons, introns), the laboratory variant classification criteria, and the patient ethnicity. By permuting these various determinants for panels of hereditary cancer genes that have specific medical management guidelines, we demonstrate that the reported-VUS likelihood value can vary greatly. We simulated the reported-VUS likelihood for a hypothetical laboratory running a test with three parameters under our control. The first parameter was the stringency in variant interpretation: we used weighted ClinVar classifications and applied different frequency cutoffs for benign (0.3-1%). Second, to represent the impact of an assay's particular region-of-interest, we varied the the number of bases sequenced into intronic regions. Finally, to evaluate the impact of ethnicity, we considered a US-population weighted patient cohort, as well as cohorts comprised of each ethnicity individually. All simulations modeled patients' variants using the ethnic-specific gnomAD allele frequencies. For a 27-gene panel, the simulated reported-VUS likelihood values differed by up to 7 percentage points based on the other panel parameters. Decreasing the panel size tended to restrict the range: a 6% spread for a hereditary breast cancer panel with nine genes, and a 3% spread for BRCA1 and BRCA2 only. With a non-stringent benign allele frequency cutoff of >0.3%, the VUS rate could differ by up to 2 percentage points for a 27-gene panel depending on whether the 20 intronic bases flanking exons were part of the assay. When stratified by ethnicity, the VUS rate for an African/African-American cohort in one scenario was 12 percentage points higher than the US-population-weighted rate. Our simulations do not reflect any true diagnostic laboratory, yet they show that there is a wide range of reported-VUS likelihoods based on the type of test being ordered, the classification criteria being used, the assay design, and the ethnic mix of the patient cohort. Our results indicate that the VUS rate in isolation is not a reliable measurement of quality, suggesting that multiple criteria should be considered when evaluating which genetic tests to offer patients.

2410W

Integration of calibrated functional assay data into *BRCA1* VUS evaluation. *B.A. Thompson[1,2], R. Bell[1], A. Thomas[3], B.E. Welm[4], J. Burn[5], S.V. Tavtigian[1].* 1) Department of Oncological Sciences, Huntsman Cancer Institute, Salt Lake City, UT; 2) Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Melbourne, Australia; 3) Department of Internal Medicine, Division of Genetic Epidemiology, University of Utah School of Medicine, Salt Lake City, Utah; 4) Department of Surgery, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah; 5) Institute of Genetic Medicine, International Centre for Life, Newcastle University, Newcastle upon Tyne, United Kingdom.

   Driven by massively parallel sequencing and allied technologies, the scale of genetic predisposition testing is on a dramatic uptrend. While many patients are found to carry clinically actionable pathogenic sequence variants, testing also reveals enormous numbers of Unclassified Variants (UV), or Variants of Uncertain Significance (VUS), most of which are rare missense substitutions. Following International Agency for Research on Cancer (IARC) variant classification guidelines, quantitative methods have been developed to integrate multiple data types for clinical UV evaluation in *BRCA1/2*; results from these analyses are recorded in the BRCA gene Ex-UV database (hci-exlovd.hci.utah.edu). In variant classification, the rate-limiting step is often accumulation of patient observational data. Recently, functional assays evaluating *BRCA1* RING domain (n=10) and C-terminal substitutions (n=242) have been calibrated, in principle enabling variant classification through a combination of sequence analysis-based predictions with functional assay results. This two-component classification was embedded in a decision tree with safeguards to avoid misclassification. For 38 previously classified variants (likely pathogenic, n=1; pathogenic, n=15; likely not pathogenic, n=2; not pathogenic, n=20) the two-component analysis resulted in sensitivity of 87.5% (95% CI: 62-98%), specificity of 100% (85-100%), and an error rate (classification of known neutral variants as pathogenic or vice versa) of 0.0%. Classification of a UV as likely pathogenic or likely neutral does not require certainty; the probabilistic definitions of the categories imply a tolerable error rate. Combining sequence analysis with functional assay data in two-component analysis lead to the classification of 141 *BRCA1* variants as likely pathogenic (n=39) or likely neutral (n=102). Moreover, once higher throughput functional assays achieve acceptable accuracy and have been calibrated, two-component analysis has the potential to dramatically accelerate UV classification.

2411T

Mate-pair sequencing provides advanced molecular characterization of genomic rearrangements in B-Chronic Lymphocytic Leukemia (CLL) and non-Hodgkins lymphoma. *S.S. Smoley[1], H.K. Kearney[1], R.B. Jenkins[1], N.E. Kay[2], K.K. Reichard[1], D.L. Van Dyke[1].* 1) Laboratory Medicine & Pathology, Mayo Clinic, Rochester, MN; 2) Internal Medicine, Mayo Clinic, Rochester, MN.

   Mate pair sequencing (MPseq) is a powerful new approach for the detection and molecular characterization of novel and recurrent chromosome rearrangements in neoplasia. We have explored MPseq on a series of 15 CLL and eight lymphoma cases that were not completely informative using other methods, e.g., FISH, conventional cytogenetics (karyotype), or chromosomal microarray (CMA). DNA was processed using the Illumina Nextera Mate Pair library preparation kit, multiplexed at 2 samples per lane and sequenced on the Illumina HiSeq. Data was aligned to the reference genome using BIMAv3 and abnormalities were identified using SVAtools, both of which are in-house developed bioinformatics tools. Of the 23 CLL and lymphoma cases evaluated, 76 abnormalities were detected by karyotype, FISH or CMA. Mate-pair identified 72 of these 76 abnormalities, identified IGH, IGL, RB1, and CDKN2A rearrangements and copy number changes not detected by the other methods, clarified the extent of 13q deletions in three cases, and in four cases recognized deletions and duplications associated with apparently balanced translocations that had been identified by chromosome analysis. The four abnormalities that MPseq did not detect were all low level clones below the resolution of MPseq testing (the sensitivity of MPseq is 10-25% depending on the abnormality).   Two illustrative examples: in a bone marrow aspirate from a CLL patient (case 18) with deletion 11q and 13q and IGH break-apart by FISH, karyotype analysis using CpG stimulation revealed an abnormal 2p and 14q, and an apparently independent clone with 11q deletion. MPseq confirmed loss of ATM and the 13q miRNAs MIR15A/MIR16-1 and identified a BCL11A/IGH translocation, which is an uncommon but recurrent immunoglobulin rearrangement in CLL. In a bone marrow aspirate from a patient (case 24) with a reason for referral of "suspected lymphoma" FISH identified 13q deletion, karyotype analysis identified 13q deletion and a t(18;22), and microarray identified heterozygous RB1 deletion, homozygous miRNA loss, and loss of IGL. Mate pair confirmed these abnormalities and also identified a juxtaposition of IGL 5' of BCL2 with resulting upregulation of the oncogene. These results illustrate the significant power of mate pair sequencing in comparison to FISH, karyotype, or CMA to identify clinically relevant gene fusions, as well as gains and losses of chromosomal segments in hematologic malignancies.

2412F

Loss-of-function *POLE* and *POLD1* variants may not be associated with early-onset colon cancer/polyposis. *E.K. Flynn, M.E. Roberts, L.R. Susswein, K. Retterer, J. Juusola, Z. Xu, R.T. Klein, K.S. Hruska.* GeneDx, Gaithersburg, MD.

   Statement of Purpose: The *POLE* and *POLD1* genes have been associated with Polymerase Proofreading-Associated Polyposis (PPAP), a highly penetrant autosomal dominant condition that increases the risk for early-onset colorectal cancer (CRC) and adenomatous colon polyps. It has been suggested that only missense variants within the exonuclease domain, and not any loss-of-function (LOF) variants, would be pathogenic for PPAP. Our aim was to assess whether *POLE* and *POLD1* LOF variants might be associated with an increased risk of early-onset colon cancer/polyposis as is seen in PPAP. **Methods:** We retrospectively reviewed the personal and family histories for 73 individuals from 52 families with a LOF *POLE* or *POLD1* variant identified by Exome Sequencing (WES), an unbiased population since testing was not ordered based on a cancer indication, and 14 individuals tested by a 32-gene hereditary cancer panel. "Loss-of-function" variant was defined as any variant predicted to cause protein truncation or nonsense-mediated decay (NMD), including nonsense and frameshift variants as well as canonical splice variants predicted to cause a large out of frame deletion.   **Results:** Among our 52 WES families and 14 inherited cancer panel probands, 48 unique LOF variants were identified (*POLE* = 33; *POLD1* = 15). On WES, 73 individuals from 52 families were found to carry a *POLE* (n = 54) or *POLD1* (n = 19) variant. Two families total (3.8%, 2/52) reported personal or family history of CRC. Of the inherited cancer referrals, 14 individuals were found to carry a *POLE* (n=7) or *POLD1* (n=7) variant. None of those <50 years of age (n=9) reported a personal history of colorectal cancer while 1/5 of those ≥50 years of age had a personal history. Additionally, none of the 14 individuals identified via our inherited cancer panel reported a personal history of colon polyps, and five (36%, 5/14) reported a family history of CRC.   **Conclusions:** In the WES cohort, only 3.8% (2/52) of families with a LOF *POLE* or *POLD1* variant reported a personal or family history of CRC. While our inherited cancer cohort is subject to ascertainment bias, as we expect referrals to have strong personal or family histories, this group lacks young-onset colon cancers and there are no reports of colon polyps. Based on these preliminary data, it appears that LOF *POLE* and *POLD1* variants may not predispose to early-onset colon cancer/polyposis.

2413W

Library-free, targeted sequencing of native genomic DNA and RNA from FFPE samples using Hyb & Seq™ technology – the Hybridization-based Single Molecule Sequencing System. *D. Kim, L. Manrao, R. Khafizov, M. Walsh, M. Korukonda, M. Gregory, M. Hoang, S. Woo, T. Rane, Y. Deng, M. Rhodes, J. Beechem.* Research and Development, NanoString Technologies, Seattle, WA.

   DNA sequencing is an enabling tool for personalized medicine research, but widespread implementation is hindered by complexities in sample preparation and sequence analysis. Hyb & Seq™ technology is a library-free, amplification-free, single-molecule sequencing technique that uses cyclic nucleic acid hybridization of fluorescent molecular barcodes onto native targets. Hybridization-based sequencing enables the simplest sample-to-answer workflow; both DNA and RNA are directly sequenced with almost no manipulation of the input material. The advantages of single molecule sequencing include simple error correction and digital counting to elucidate DNA copy numbers and RNA levels. Here we describe an end-to-end Hyb & Seq sequencing process that consists of: i) Sample preparation including rapid gene capture directly from formalin-fixed paraffin embedded (FFPE) tissue. ii) Library-free targeted sequencing of oncogenic mutations. iii) Data analysis using a Hyb & Seq assembly algorithm (ShortStack™) for variant calling. Targeted sequencing of oncogenic mutations was performed on NanoString's prototype Hyb & Seq system. **Highlights of Key Results:** • Total time from FFPE curls to start of sequencing was under 60 minutes, with total hands-on time of less than 15 minutes • Simultaneous capture of 100 DNA gene targets and 20 mRNA from one to three FFPE curls (10 microns thickness) was sufficient for sequencing with no PCR amplification and cDNA conversion • Dual strand capture and simultaneous sequencing of both strands of DNA, increased accuracy of variant detection by identifying single-stranded damage artifacts common in FFPE templates • Automated sequencing carried out on a microfluidic cartridge with runs exceeding 400 Hyb & Seq cycles • Sequencing accuracy reached 99.99% (QV40) when a base from a single molecule was read ≥ 5 times • All targeted variants including variant as low as 1% were detected Hyb & Seq's simplicity, flexibility, and accuracy offers an ideal sample-to-answer solution for the translational sequencing research lab.

2414T

The prevalence of mosaicism in common cancer susceptibility genes from 232,328 individuals undergoing sequential testing. *T. Slavin[1], B. Coffee[2], H. Cox[2], R. Bernhisel[2], K. Brown[2], J. Weitzel[1], S. Neuhausen[1], D. Mancini-Dinardo[2].* 1) Clinical Cancer Genomics, City of Hope, Duarte, CA; 2) Myriad Genetic Laboratories, Salt Lake City, UT.

   Background: Previous studies have suggested that age and personal cancer history are associated with an increased risk of mosaic next generation sequencing (NGS) findings. This has been attributed to aberrant mutations in hematologic cell lines termed clonal hematopoiesis. Here, we evaluated mosaic NGS findings from a large hereditary cancer clinical cohort to explore contributing factors. **Methods:** We assessed individuals who had multi-gene pan-cancer panel testing through a large genetic laboratory (Myriad Genetic Laboratories) and were found to carry mosaic variants (NGS read frequency of 10-30%) classified as either pathogenic, likely pathogenic, or uncertain. Logistic regression analysis was used to determine the risk of carrying a mosaic variant after adjusting for 1) personal cancer history (affected vs unaffected), 2) age (<50 vs ≥50 years), 3) germline mutation status, and 4) clinical indication for testing. Odds ratios (OR) and 95% confidence intervals (CI) are reported. **Results:** Overall, 348/232,328 (0.15%) individuals were identified as carrying a mosaic variant, most commonly in either *CHEK2* (29%), *TP53* (23%), or *ATM* (22%). Logistic regression analysis showed that age group (≥50 years) was associated with the greatest risk for carrying a mosaic variant (OR=3.8, 95%CI 2.8, 5.2; p<0.001). Individuals with personal history of cancer (OR=2.4, 95% CI 1.8, 3.2; p<0.001), or a germline pathogenic variant (OR=1.4, 95% CI 1.0, 2.0; p=0.045), also had a significantly higher risk of carrying a mosaic variant. **Conclusions:** Overall, we show that prior history of cancer, age, and germline mutation status were associated with increased risk of carrying a mosaic variant. As the mosaic variants were associated with advanced age, likely chemotherapy exposure, and in genes known to be associated with clonal hematopoiesis, we believe clonal hematopoiesis as the most likely cause of the majority of the mosaicism. It is likely that more cases would have been identified with inclusion of genes frequently associated with clonal hematopoiesis, such as *DNMT3A* or *TET2*. These findings suggest that individuals with germline DNA repair mutations have increased mosaicism; possibly related to underlying genomic instability. As clonal hematopoiesis has been associated with leukemia predisposition, these results may explain previous literature identifying an enrichment of germline DNA repair mutation carriers in therapy related leukemia cohorts.

2415F

Somatic driver mutations enhance survival prediction in familial chronic lymphoid leukemia. *W. Zhou[1,2], L. Goldin[2], M. Wang[1,2], M. McMaster[2], K. Jones[1,2], L. Burdette[1,2], S. Chanock[2], M. Yeager[1,2], M. Dean[2], N. Caporaso[2].* 1) Cancer Genomics Research Laboratory, Leidos Biomedical Research/ NCI, Bethesda, MD 20892; 2) Division of Cancer Epidemiology and Genetics (DCEG), NCI, 9609 Medical Center Drive MSC 9776 Bethesda, MD 20892.

   Chronic lymphoid leukemia (CLL) is the most common type of leukemia in adults. Traditionally, four common chromosomal aberrations (deletion of 11q, 13q14, 17p, duplication of 12) as well as the mutational status of the immunoglobulin heavy chain variable (IGHV) gene have been shown to stratify CLL patients into distinct prognostic groups. In the last few years, several tumor sequencing studies have revealed a number of recurrent mutations, some of which also contribute to prognosis. In this study, we performed SNP microarray data on 95 CLL subjects from 40 CLL families. Log R ratio and B allele frequency were used to assess copy number alterations as well as to calculate the proportion of cancer cells. Overall, 142 detectable somatic copy number alterations (SCNAs) from 54 subjects were detected (57%), including the four most common CLL-associated cytogenetic aberrations: losses of chromosome 13q14 (28.9%), 11q (5.6%), and 17p (2.1%); gain of chromosome 12 (4.2%). We found many chromosomal alterations were observed only in a portion of the cells (mosaicism). We performed whole exome-sequencing of whole blood samples from 98 CLL patients (including all 95 subjects with SNP array data). We used a somatic pipeline caller to detect somatic mutations and used other filtering methods to eliminate likely germline mutations. We focused on a panel of 52 genes that have been reported as having recurrent putative driver alterations in mature B-cell non-Hodgkin lymphomas. Overall, 69 somatic mutations (all nonsynonymous except 2 in an intron of the *NOTCH1* gene) within 29 genes were detected among 45 subjects (46%). The most frequently mutated genes were *TP53* (8.2%) and *NOTCH1* (8.2%). We found that patients having both a point mutation in a CLL driver gene and SCNA have worse survival (HR=3.17, 95%CI=0.97-10.35; P=0.056) than patients having either a point mutation (HR = 1.34, 95%CI=0.66-2.71; P = 0.42), as SCNAs (HR = 2.65, 95%CI=0.77-9.13; P = 0.12). *TP53* mutations were associated with poorest overall survival (HR=4.39, 95%CI=1.28-15.04; P=0.018). Our study demonstrates that combining SCNA and mutational data can add to predicting the pathway from an initial diagnosis to outcome in CLL.

2416W

Combined mutation and CNV detection by targeted next-generation sequencing in uveal melanoma. *A. de Klein[1], K. Smit[1,2], N. van Poppelen[1,2], J. Vaarwater[2], H. Dubbink[3], R. Verdijk[3], E. Kilic[2]. 1) Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 2) Ophthalmology, Erasmus MC, Rotterdam, Netherlands; 3) Pathology, Erasmus MC, Rotterdam, Netherlands.*

Uveal melanoma (UM) is a highly aggressive cancer in the eye. It is the most common type of primary eye cancer in adults and nearly 50% of UM patients die from liver metastasis. Chromosome and mutation status have been shown to correlate with the disease free survival. Loss of chromosome 3 and inactivating mutations in *BAP1*, which is located on chromosome 3, are strongly associated with high risk tumors that metastasize early. Other genes often involved in UM are *SF3B1* and *EIF1AX*, which are found to be mutated in intermediate- and low risk tumors, respectively. We developed a targeted sequencing method that can detect mutations in genes involved in UM and chromosomal anomalies of chromosome 1, 3 and 8. Whereas current UM-diagnostics involves several techniques for detection of copy number variations and somatic mutations, our targeted UM panel can detect losses and gains of chromosome 1, 3 and 8 and somatic mutations in the aforementioned genes in a single assay. By sequencing formalin-fixed paraffin-embedded and fresh UM-specimens, we showed that mutations and chromosome-status can reliably be obtained using targeted next generation sequencing (NGS).

2417T

Applying synthetic long reads to a custom Lynch Syndrome NGS panel to overcome pseudogene interference in *PMS2*, detect structural variations, and enable allelic phasing. *C. Kao[1], F. Mafra[1], M. Gonzalez[1], R. Pellegrino[1], S. Wenzel[2], K. Wimmer[2], F. Armelao[3], T. Ripperger[4], L. Fang[5], K. Wang[5], D. Kvitek[6], S. Garcia[7], A. Fehr[7], H. Hakonarson[8]. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Dept. for Medical Genetics, Molecular and Clinical Pharmacology Medical University, Innsbruck, Innsbruck, Austria; 3) Department of Gastroenterology, Ospedale Santa Chiara, APSS, Trento, Italy; 4) Institute of Human Genetics, Hannover Medical School, Hannover, Germany; 5) Department of Biomedical Informatics, Columbia University, NY; 6) Invitae, Inc., San Francisco, CA; 7) 10x Genomics, Inc., Pleasanton, CA; 8) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.*

Massively parallel sequencing technologies have transformed clinical diagnostic screening to make the process more accessible and mainstream for a variety of genetic disorders, including hereditary conditions such as Lynch syndrome (hereditary nonpolyposis colorectal cancer). However, these current methods use short reads (up to 200-400 nt) aligned back to a reference genome, which suffers from some inherent limitations. For example: 1) short reads alone are unable to phase variants to establish compound heterozygosity; 2) detection of structural variants, especially single allele (heterozygous) or copy neutral alterations (e.g. inversions, translocations) remain challenging; and 3) clinically relevant genes with highly similar paralogs are difficult to map/resolve with traditional short read NGS; this includes *PMS2,* one of the Lynch syndrome genes. A pseudogene, *PMS2CL*, located ~0.7Mb away is highly homologous to exons 9, 11-15 in *PMS2*. Pseudogene interference is especially problematic due to frequent *PMS2/PMS2CL* recombination events. Customized, non-NGS based assays have been developed to clinically screen for *PMS2* mutations, however these can be cumbersome and difficult to implement. We used a gel-bead-in-emulsion method with the Chromium instrument (10x Genomics, Inc.) to impart long-range information into traditional short read libraries. DNA from samples (extracted using standard bead/column-based techniques) were segregated into ~1M gel-beads-in-emulsions (GEMs), where mini-libraries are created and tagged with a unique, GEM-specific 16-mer barcode. All mini-libraries are then pooled and finished with Illumina-compatible adaptors and go directly to whole-genome sequencing or undergo target enrichment prior to whole-exome or panel-based sequencing. "Synthetic" long reads are assembled by grouping together short reads sharing the same barcode. Previously, we used linked whole exome sequencing to detect *PMS2* variants (point mutations and structural rearrangements) within the *PMS2CL* paralogous region in clinical "truth" samples. We present here a custom, linked-read panel targeting all Lynch syndrome genes (*MLH1, MSH2, MHS6, and PMS2)* using Agilent SureSelect baits implemented within a fully automated, high-throughput workflow and demonstrate its utility in phasing alleles, overcoming pseudogene interference, and detecting structural variants. Improvements in the informatics pipeline for phasing, variant calling, and visualization are also presented.

2418F

Tumor characteristics provide evidence for mismatch repair (MMR) variant pathogenicity. *S. Li, J. Clifford, D. Qian, Y. Tian, A. Elliott, H.M. Lu, M.H. Black. Ambry Genetics, 15 Argonaut, CA.*

Pathogenic mutations in mismatch repair (MMR) genes (*MLH1, MSH2, MSH6,* and *PMS2*) increase risk for Lynch syndrome and other cancers. MMR deficiency may result in microsatellite instability (MSI); tumor MSI testing and protein immunohistochemical (IHC) analysis for MMR protein expression are often used alone or together to evaluate risk for Lynch syndrome in clinical screening. It has been suggested that these quantifiable tumor characteristics may be useful in assessing germline MMR variant pathogenicity. We obtained research consent and clinical information for 77,071 patients who underwent germline multigene panel testing including MMR genes in 2012-2016. Among the 3,627 patients with available MSI or IHC information, we excluded 69 with conflicting MSI/IHC status and 238 carriers of non-MMR pathogenic variants. Variant pathogenicity was assessed according to ACMG guidelines. Among the remaining 3,320 individuals with MSI (n=1,055) or IHC (n=2,954) status (n=689 with both), we tested association between MMR carrier status and the combined MSI/IHC phenotype using Fisher's exact test, followed by estimation of MSI/IHC likelihood ratios (LR=%pathogenic variant carriers/%non-carriers). For each variant, we computed the tumor characteristic LR (TCLR) based on the estimated MSI/IHC LRs. While typically TCLR>1 for pathogenic and TCLR<1 for benign variants, we used the TCLR>10 or <0.1 as thresholds suggestive of pathogenic or benign classification, respectively. Carriers of MMR mutations were more likely to have abnormal MSI/IHC (OR=27.3, 95%CI: 16.2-49.0). Estimated LRs for MSI/IHC unstable/abnormal and stable/normal were 3.136 and 0.115, respectively. Overall, we identified 141 pathogenic/likely pathogenic, 217 benign/likely benign, and 223 VUS/unclassified MMR variants. Among pathogenic variants, 73.0% have TCLR>1 and none have TCLR<0.1; 89.9% of benign variants have TCLR<1 and none have TCLR>10. Moreover, 15% of VUS/unclassified variants have TCLR>1 and 66% TCLR<1, suggesting that many of these may be potentially classified. While no VUS/unclassified variants have TCLR>10, 17 (8%) have TCLR<0.1, providing stronger evidence for classification as benign/likely benign. These data show that MSI/IHC-based TCLRs provide evidence for or against pathogenicity of variants in MMR genes. Used independently or in conjunction with other evidence, the TCLR may inform variant classification, and thus have important implications for genetic testing and clinical management.

2419W

A case report of a rare germline SVA transposition event in *TP53*. *W. Cheng[1], R. Torene[1], J. Mester[1], Z. Zhang[1], P. McAndrew[1], P. Murphy[1], S. Solomon[1], C. Cina[2], R. Klein[1], K. Hruska[1]. 1) GeneDx, Gaithersburg, MD; 2) Kaiser Permanente Southern California, San Diego, CA.*

Germline mobile element insertions (MEIs) disrupting normal coding sequences have been reported in association with hereditary conditions. Although apparently rare, these variants may be under-reported to date. Here we present the first known germline SINE-VNTR-Alu (SVA) insertion in the coding region of *TP53*. A 48yo female presented for cancer genetics evaluation with a personal history of bilateral breast cancer, diagnosed at ages 24 and 26, and multiple non-melanoma skin cancers. Family history was notable for a sister with synovial sarcoma, diagnosed at 50, and maternal grandmother with breast cancer, diagnosed in her late 50s. To explore the possibility of a hereditary cancer syndrome, the patient was tested on a 23-gene hereditary cancel panel, including next generation sequencing (NGS) and concurrent deletion/duplication analysis. The NGS analysis pipeline detected a possible *TP53* variant in the coding region of exon 8. Three groups of sequencing reads were observed: reverse reads containing CCCTCT repeats, forward reads containing a poly-A tract, and reference sequences. Reads containing mismatched soft-clipped sequences flanked a 16-bp duplication; this finding suggested an MEI. NGS data then were analyzed with the 'Mobster' algorithm, which identified an SVA element with 56 supporting reads and confirmed the presence of a target site duplication (TSD). Unmapped paired reads in the region were used for assembly of the MEI sequence for characterization. This SVA element was found to be full length and belonged to the SVA-E subfamily based on the percent identity to the SINE-R derived region. Using patient-specific primers for PCR followed by Sanger sequencing, we were able to confirm the 5' breakpoint of the event. However, due to the large size and the GC-rich VNTR domain, we were unable to obtain a full-length Sanger sequence across the insertion. Considering the presence of the TSD, poly-A tail, and its high identity to a young SVA subfamily, the mechanism of insertion was most likely target-primed reverse transcription. This MEI was reported as a likely pathogenic variant based on ACMG/AMP variant interpretation guidelines. This insertion is predicted to cause loss of normal *TP53* function, consistent with the patient's clinical presentation and family history suggestive of Li-Fraumeni syndrome. Given that germline MEIs disrupting coding sequences are likely disease-causing, it is important to incorporate their detection into analysis pipelines.

2420T

Diagnostic yield and mutation spectrum of multigene panel testing for hypertrophic cardiomyopathy. *L. Qin, J. Wang, L. Hoang, T. Johnston, J. Dolinsky, T. Pesaran, C. Antolik. Ambry Genetics, Aliso Viejo, CA.*

Hypertrophic cardiomyopathy (HCM) is the most common inherited cardio-vascular disease with a worldwide prevalence of 1:500, and is the leading cause of sudden cardiac death in young people. Genetic etiology is suspected in up to 50% of HCM patients, and genetic testing is highly recommended for HCM patients and family members following the identification of a causative mutation in the proband. To gain insight into the diagnostic yield and mutation spectrum of HCM, a retrospective review was performed for 408 consecutive cases with a clinical suspicion of HCM who underwent multigene panel testing at our laboratory. A pathogenic mutation was identified in 27% (n=112) of these individuals. Eighty two pathogenic mutations in 15 genes were identified. The most frequently mutated genes were *MYBPC3* (48%)*, MYH7* (21%)*, TNNT2* (9%) and *MYL3* (5%). Further variant-specific analysis identified 38 individuals from 31 families who were positive for the mutation previously identified in their respective probands. The median age for mutation-positive individuals reported to have cardiomyopathy was 38 years (n=5) vs 14 years (n=21) for unaffected mutation-positive individuals, consistent with the age-related pen-etrance of HCM. This underscores the great utility of genetic testing in identi-fying pre-symptomatic at-risk individuals. In our study, the diagnostic yield is lower than the value reported in the literature, which may be explained by i) the presence of variable phenotypes in individuals who underwent testing on larger panels designed to detect multiple types of cardiomyopathy, ii) testing of unaffected individuals referred solely due to their family history, iii) differences in testing criteria for research studies compared to that for clinical testing, iv) the lack of clear phenotype information in some cases, or v) the strict variant interpretation guidelines used in a clinical diagnostic laboratory, which results in more frequent reporting of variants of uncertain significance (VUS). In the current cohort, VUS were detected in 46% of the individuals (n=189), including 34 probands who had pathogenic mutations. About one third of the VUS were in *TTN*, followed by *MYH6, MYH7,* and *MYBPC3*. VUS interpretation is chal-lenging and will benefit from continually emerging data, including population frequency estimates in larger and more ethnically diverse populations, familial co-segregation studies, and functional studies.

2421F

Prevalence of RASopathy gene mutations in patients who have had multi-gene panel testing for cardiomyopathy. *W. Zhang[1], S. Chadwell[2], M. Nicole[1], C. Prada[2]. 1) Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.*

Wenying Zhang, Sarah Chadwell, Nicole Moore, Carlos Prada   The RASopathies are a group of developmental disorders caused by pathogenic genetic variations in the RAS-MAPK gene pathway, including Cardiofaciocu-taneous syndrome (CFC), Costello syndrome, Neurofibromatosis 1, and the Noonan syndromes (Noonan syndrome, Legius syndrome, Noonan syndrome with Multiple Lentigines/LEOPARD syndrome, and Noonan syndrome-like disorder with loose anagen hair). Cardiomyopathy, specifically hypertrophic cardiomyopathy (HCM), is an established feature of CFC, Costello, Noonan syndrome, and Noonan syndrome with multiple lentigines/LEOPARD syn-drome. There may be individuals with milder forms of these RASopathies who have HCM as the only known manifestation of disease. Knowing these individuals have an underlying genetic cause for their symptoms could have implications for their health and/or the health of their offspring. We conduct-ed a retrospective study of all individuals who have had a cardiomyopathy gene panel ordered through Cincinnati Children's Hospital's Heart Institute Diagnostic Laboratory to determine the prevalence of pathogenic or likely pathogenic genetic changes in RAS pathway genes in individuals with cardio-myopathy. We evaluated variants in the following genes: *BRAF, CBL, HRAS, KRAS, MAP2K1, MAP2K2, NF1, NRAS, PTPN11, RAF1, SHOC2,* and *SOS1.* We reviewed 74 cases with cardiomyopathy, including 32 with HCM, 24 with dilated cardiomyopathy (DCM), 9 with both Left Ventricular Non-Compaction (LVNC) and DCM, 4 with LVNC only, 2 with Arrythmogenic Right Ventricular Cardiomyopathy (ARVC) and 3 with unspecified cardiomyopathy. We identified four patients (5.41%) with pathogenic or likely pathogenic variants in *HRAS, PTPN11,* and *RAF1* (2 individuals). Indication for testing for all 4 individuals was HCM; three of them were younger than 2 years old when first diagnosed with HCM. The remaining individual was 7 years old at the time of diagnosis. The prevalence of pathogenic or likely pathogenic RASopathy gene mutations in our HCM patient cohort is 12.5% (4/32). Based on these data, we conclude that the RASopathy genes should be included on multi-gene panels for cardio-myopathy to increase diagnostic yield for individuals with HCM.

2422W

Molecular aproach of targeted next generation sequencing of 68 genes involved in cardiac arrhythmias of 148 unrelated patients. *B. Turkgenc[1], S.G. Temel[2], O. Karadag[3], H.H. Aykan[4], F. Uysal[5], I. Yildirim Bastuhan[6], A. Sulu[7], S. Ugan Atik[8], B. Cinar[8], R. Dedeoglu[8], E. Gunay[9], M. Ramoglu[10], E. Cilsal[11], T. Mese[12], O. Ciftci[13], O. Hallioglu[14], F. Oztunc[8], T. Karagoz[4], O. Oz-demir[15], O. Baspinar[7], O.M. Bostan[5], F. Akalin[9], M. Kervanoglu[7], C. Ayabakan[16], Y. Alanay[17], A. Celiker[6], S.A. Ozer[18], M.C. Yakicier[19].* 1) Medical Biology and Genetics, Acibadem Diagnostic Center, ISTANBUL, Turkey; 2) Department of Histology and Embryology, Faculty of Medicine, Uludag University, Bursa, Turkey; 3) Department of Statistics, Hacettepe University, Ankara, Turkey; 4) Department of Pediatric Cardiology, Faculty of Medicine, Hacettepe University, Ankara, Turkey; 5) Department of Pediatric Cardiology, Faculty of Medicine, Uludag University, Bursa, Turkey; 6) Department of Pediatric Cardiology, Faculty of Medicine, Koç University, Istanbul, Turkey; 7) Department of Pediatric Cardiology, Faculty of Medicine, Gaziantep University, Gaziantep, Turkey; 8) Department of Pediatric Cardiology, Cerrahpasa Medical Faculty, Istanbul, Turkey; 9) Department of Pediatric Cardiology, Faculty of Medicine, Marmara University, Istanbul, Turkey; 10) Department of Pediatric Cardiology, Faculty of Medicine, Ankara University, Ankara, Turkey; 11) Pediatric Cardiology, Adana Numune Hospital, Adana, Turkey; 12) Pediatric Cardiology, Izmir Dr. Behcet Uz Pediatrics Hospital, Izmir, Turkey; 13) Department of Pediatric Cardiol-ogy, Kartal Kosuyolu Training and Research Hospital, Istanbul, Turkey; 14) Department of Pediatric Cardiology, Faculty of Medicine, Mersin University, Mersin, Turkey; 15) Department of Pediatric Cardiology, Kecioren Training and Research Hospital, Ankara, Turkey; 16) Department of Pediatric Cardiology, Baskent University Istanbul Hospital, Istanbul, Turkey; 17) Department of Pe-diatric Genetic, Acibadem Maslak Hospital, Istanbul, Turkey; 18) Department of Medical Biology and Genetics, Marmara University, Istanbul, Turkey; 19) Department of Molecular Biology and Genetic, Acibadem University, Faculty of Science, Istanbul, Turkey.

Arrhythmia is a problem associated with the rate or rhythm of heartbeat. Car-diac arrhythmias are one of the most common causes of sudden death in the world. Sudden death remains an important public health concern. The genetic background of these diseases is complex and heterogeneous with overlapping clinical symptoms that may not always be penetrant. Despite several genes have been reported in ion channel diseases, a large proportion of clinically diagnosed families remain without a recognized genetic cause of disease. In our study, we aimed to screen a targeted NGS panel comprising 68 genes known to be potentially related to cardiac arrhythmias including 104 LQTS, 2 JLNS, 8 SQTS, 9 Brugada syndrome, 10 catecholaminergic polymorphic ven-tricular tachycardia (CPVT), 5 Ventricular tachycardia, 4 Progressive familial heart block and 9 unexplained arrhythmia with complete coverage of the entire coding region of all the genes analyzed. The genetic study was performed by massive parallel sequencing (IonTorrent PGM) following Sanger sequencing for all suspicious variants and their segregations within the family. In 148 pa-tients tested we identified 319 variants of which 45 (14,11%) were pathogenic, 4 (%1,25) were potentially pathogenic, 249 (61,44%) were variant of unknown significance (53 of them-16,62% were novel) and 21 (6,58%) were probably benign. We identified pathogenic/potentially pathogenic variants in 30% (45) of probands and novel variants in 32% (47) of probands tested. In LQT patients, pathogenic/potentially pathogenic mutations were also observed in *RYR2, GPD1L, CASQ2, JUP, LMNA, MYBPC3* and *PRKAG* as well as in the reported OMIM genes. SQT patients had pathogenic mutations in *SCN5A, RYR2, DSP* and *ANK2* genes. Also, one of the CPVT patient had a *GPD1L* mutation. We identified 4 de-novo mutations in *KCNQ1, RYR2, CACNA1C* and *CALM2* responsible for LQT syndrome and 1 novel de-novo mutation in *SCN5A* relat-ed to Brugada Syndrome. Any suspicious variant has not be detected in the remaining 7% (10) of the patients. For future directions functional analysis of found variations are under estimated. This data suggests that our targeted car-diac panel is a gold standard approach to investigate the novel genetic risks and open new opportunities for prevention and therapy of lethal arrhythmias in the common pathologies. **Acknowledgements:** Inherited arrhythmia panel project is supported by grant from SANTEZ (0253. STZ.2013-2), Turkey.

2423T

Involvement of mtDNA variants in patients with cardiac manifestation. *H. Cui, J. Chonis, A. Balog, J. Semotok, K. Haude, D. Dorsainville, R.K. Bai, S. Suchy, G. Richard, D. Macaya.* GeneDx, Inc., Gaithersburg, MD.

Among hundreds of mitochondrial DNA (mtDNA) variants reported, many are associated with cardiac disease, with the prevalence of mtDNA-related cardiomyopathy estimated to be at least 1 in 10000-15000 in the general pop-ulation. Cardiac involvement has important impact on morbidity and mortality in patients with mtDNA disease. A total of 14 mtDNA genes encoding either mitochondrial tRNAs or structural subunits of the respiratory chain complexes were analyzed by our laboratory together with up to 77 nuclear genes in 3 different cardiomyopathy next-generation sequencing panels. In this retro-spective study, we report on the molecular findings of 578 patients referred for cardiomyopathy testing to evaluate the involvement of mtDNA variants in their etiology. In this cohort, 53 (9%) patients had pathogenic or likely pathogen-ic variants (PV/LPV) in mtDNA, 115 (20%) patients had PV/LPV in nuclear genes, and 7 (1.2%) patients had PV/LPV in both mitochondrial and nuclear genes. Half of the patients with positive findings in mtDNA were referred with a clinical diagnosis or suspected diagnosis of hypertrophic cardiomyopathy (HCM) or dilated cardiomyopathy (DCM). Among the 54 PV/LPV identified in mtDNA, 39 were located in mt-tRNA genes. Of those patients, 16 (41%) had HCM and 8 (23%) had DCM, suggesting mt-tRNA as a hotspot for variants associated with cardiomyopathy in the mitochondrial genome. Twelve patients had PV/LPV in the MT-TL1 gene encoding tRNA Leu (UUR), including the recurrent PV m.3243 A>G (tRNA Leu) in 8 patients, which is known to cause mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS) syndrome, making this the most common reported mtDNA variant. Another factor related to the clinical features of mtDNA-associated disorders is the tissue-specific mutation load (heteroplasmy). PV/LPV in 28 patients (52.8%) were found to be heteroplasmic in the submitted specimen type (pre-dominantly blood), including all patients with m.3243 A>G, ranging from 15% to 87%. For low level heteroplasmic variants (less than 25%), further testing in affected tissues (such as heart or skeletal muscle) was recommended to determine conclusively their clinical significance. In summary, mtDNA testing revealed positive results in at least 10% of patients with cardiomyopathy. Our data underscore that dual analyses of both mitochondrial and nuclear genes maximize the diagnostic yield for patients with these disorders.

2424F

A genetic backround in Czech patients with inherited cardiomyopathies.

P. Cibulkova[1], M. Dvorakova[1], P. Vanickova[1], H. Slavikova[1], A. Boday[1], N. Havranova[1], A. Taborska[2], K. Vesela[3], J. Januska[4]. 1) Laboratories AGEL, Novy Jicin, Czech Republic; 2) Ambulance of clinical genetic, Hospital and Comprehensive Cancer Centre Novy Jicin, Novy Jicin, Czech Republic; 3) Institute of Biology and Medical Genetics, 1st Faculty of Medicine, Charles University in Prague and General University Hospital, Praha, Czech Republic; 4) Department of Cardiology, Hospital Podlesi, Trinec, Czech Republic.

Objectives: Inherited cardiomyopathies - hyperthropic (HCM), dilated (DCM), arrythmogenic right ventricular (ARVC), restrictive and left ventricular non-compaction are common, genetically heterogenous, predominantly autosomal dominant disorders with variable penetrance and age of onset. In our study we focused on the association between cardiomyopathy phenotype and genetic variants in 78 genes encoding cardiac sarcomeric and non-sarcomeric proteins. A cohort of 255 Czech unrelated patients was investigated. **Methods:** Genomic DNA was isolated from peripheral blood leucocytes. The coding regions and splice junctions of the 78 genes were enriched by a custom designed probe library (targeted capture system Nimblegen) and sequenced by massive parallel sequencing on an Illumina platform. Capillary Sanger sequencing was used for a variants confirmation. All variants with possible splice effect were checked with RNA analysis. All detected missense variants were tested by five *in silico* prediction softwares for its effect on the protein function. Nonsense and frameshift variants were considered as pathogenic due to their influence on the protein function. **Results:** A disease causing pathogenic variant or likely pathogenic variant were identified in 106 cases out of 255 cohort (42%). One or more variants with unknown clinical significance (VUS) were identified in 19 cases (7%). Five patients with HCM are digenic and one is trigenic which correlates with their more severe form of inherited HCM. The majority of pathogenic variants were found in sarcomeric genes (98 pathogenic/likely pathogenic variants). Fourty five HCM phenotypes were caused by the pathogenic variant in *MyBPC3* gene, and the others by mutations in *MYH7*, *MYH6*, *CSRP3* and *TRIM63* genes. In the case of DCM patients pathogenic variants were found in *LMNA*, *DSP*, *MyBPC3*, *MYH7* and *TTN* genes. One ARVD case was caused by mutation in *RYR2* gene, two likely pathogenic variants were found in *DSP* gene. **Conclusion:** Germline pathogenic variants are the main cause of inherited cardiomyopathies. The massive parallel sequencing is a great tool for molecular diagnostic testing, cost- and time-effective, but it sometimes makes a clinical genetics counselling more complicated due to indentifying rare sequence variants with unknown clinical significance. In some cases segregation analysis are necessary, classification of variants and their pathogenic role should be interpreted very carefully (e.g. *TTN* gene missense variants).

2425W

Post-mortem cytogenomic study of Brazilian patients reveals the CNVs connection to complex congenital heart defects.

F.A.R. Madia[1], A.T. Dias[1], E.A. Zanardo[1], J.G. Damasceno[1], M. Rocha[2], A.M. Nascimento[1], T.V.M.M. Costa[1], S.N. Chehimi[1], G.M. Novo-Filho[1], M.M. Montenegro[1], Y.G. Oliveira[1], R. Schultz[3], F.T. Gonçalves[4], C. Fridman[4], C.A Kim[5], L.D. Kulikowski[1,5]. 1) Laboratório de Citogenômica LIM 03, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 2) Centro de Reprodução e Genética – Departamento de Saúde Coletiva – Faculdade de Medicina do ABC – Santo André-SP, Brazil; 3) Serviço de Anatomia Patológica, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 4) Departamento de Medicina Legal, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 5) Unidade de Genética do Instituto da Criança, Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil.

Heart development is a complex process and the molecular mechanisms remains to be clarified. About ~8% of congenital heart defects (CHD) are caused by chromosomal and single gene mutation disorders. Recent the involvement of copy number variations (CNVs) in the development of CHD was proposed. In this sense, we investigate the contribution of CNVs to the pathogenesis of complex CHD using molecular methods in 29 cases of stillbirth and new-born from Serviço de Verificação de Óbitos, HC-FMUSP. DNA samples from skin and heart tissues were evaluated using AmpFℓSTR® MiniFiler™ PCR Amplification Kit (Life Technologies™, California, USA) and Multiplex Ligation-dependent Probe Amplification (MLPA) with kits P036, P064, P070, P095 and P311 (MCR-Holland, Amsterdam, the Netherlands). A total of 10 cases showed genomic alterations: duplications (80%), deletions (10%), and mosaicism (10%). When comparing the most frequent abnormality, we observed that greatest of the cases were compatible with trisomy 18 syndrome (5 cases). Our results displayed duplication of LPIN2, SS18_ex4 and SS18_ex6 probes, associated with alterations the septal and valve defects, in 4 cases with trisomy 18 hypothesis. Also duplications of SIM2, NCAM2, USP25, HSPA13, TIAM1, APP_18 and TFF1_1 probes and D21S11 marker, associated with heart dysplasia, were found in two cases with trisomy 21 hypothesis. The first *Post-mortem* cytogenomic study of Brazilian patients with complex congenital heart defects showed efficiency for identifying pathogenic CNVs associated with complex CHD and also allowed the adequate genetic counseling. **Grants:** FAPESP: 09/53105-9 and FINEP-CT INFRA 0160/12 SP8.

2426T

Molecular review of Polish patients with Smith-Lemli-Opitz syndrome. *P. Halat, E. Ciara, D. Wicher, D. Jurkiewicz, D. Piekutowska-Abramczuk, M. Pelc, D. Siestrzykowska, B. Chałupczyńska, P. Kowalski, M. Krajewska-Walasek.* Department of Medical Genetics, The Children's Memorial Health Institute, Warsaw, Poland.

Smith-Lemli-Opitz syndrome (SLOS [MIM 270400]) is a rare metabolic disorder in which a defect in cholesterol biosynthesis results in multiple congenital anomalies and intellectual disabilities. The symptoms vary widely from mild to severe among affected SLOS individuals. This autosomal recessive disease is caused by pathogenic variants in the *DHCR7* gene [MIM 602858]. Sequence analysis of *DHCR7* gene detects approximately 96% of pathogenic variants. In approximately 4% of SLOS patients the second mutation remains unidentified. Worldwide the incidence of SLOS has been reported between 1:10,000 - 1:70,000 depending on the population being studied. SLOS appears to be the most frequent (1:10,000 - 1:30,000) among populations of Central Europe, especially in Polish population (1:2,300 - 1:3,937). In our study we examined a group of 112 Polish patients with SLOS by Sanger sequencing. Biallelic changes in *DHCR7* gene were revealed in 108 patients. Twenty five different pathogenic variants, of which five were frequent [c.452G>A, p.(W151*); c.470T>C, p.(L157P); c.964-1G>C, p.?; c.976G>T, p.(V326L); c.1054C>T, p.(R352W)] were identified in this group. They constitute 85% of all mutated alleles in Polish patients. The most common variants, c.452G>A and c.976G>T, constitute 68%. In our group 4 unrelated patients with a strong suspicion of SLOS had only one mutated allele [c.452G>A (n=3) or c.946-1G>C (n=1)]. In 3 families the child died before complete molecular diagnosis, therefore Sanger sequencing in their parents was performed. Analysis revealed that only one parent from each family was a heterozygous carrier of the molecular variant. We supplemented our molecular analysis with a dosage gene test to identify a disease-causing mutation on the second allele of *DHCR7* in undiagnosed probands. Performed MLPA analysis revealed lack of gross *DHCR7* deletions. Diagnosis of SLOS was confirmed at the molecular level with 96.4% compliance. Our report revealed lack of gross *DHCR7* deletions in examined Polish population. The absence of second mutation in the coding region of the *DHCR7* gene suggests that it may be located in the promoter or deep intron region. To help patients it is still necessary to continue research on SLOS. Grant reference: M21/16.

2427F

Reconciling newborn screening and genomic results to detect partial biotinidase deficiency: A BabySeq Project case report. *J.B. Murry[1,3], K. Machini[1,3], O. Ceyhan-Birsoy[1,2], M. Lebo[1,3,4], S. Fayer[5], C. Genetti[6], G.E. Van-Noy[6], T. Yu[4,6,7], P. Agrawal[4,6,8], R. Parad[4,9], I. Holm[4,6], A. McGuire[10], R. Green[4,5,11], A. Beggs[4,6], H.L. Rehm[1,3,4,11].* 1) Laboratory for Molecular Medicine, Cambridge, MA; 2) Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY; 3) Department of Pathology, Brigham & Women's Hospital, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 6) Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 7) Department of Neurology, Boston Children's Hospital, Boston, MA; 8) Division of Newborn Medicine, Boston Children's Hospital, Boston, MA; 9) Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, MA; 10) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 11) The Broad Institute of MIT and Harvard, Cambridge, MA.

The BabySeq Project is a randomized clinical trial assessing the impact of genomic sequencing (GS) on sick and healthy infants and their families. Newborns are recruited from ICUs at Boston Children's Hospital and Brigham and Women's Hospital (BWH) and the well baby nursery at BWH. All infants receive state-mandated newborn screening and family history, and half are randomized to receive GS. All babies in the sequencing group receive risk and carrier status for highly penetrant diseases that have onset or management during childhood as well as pharmacogenomic variants relevant to the pediatric population. Here, we describe GS in a newborn with compound heterozygous *BTD* gene variants. Pathogenic defects in the *BTD* gene lead to biotinidase deficiency (BTD) which may present with neurologic and cutaneous abnormalities. BTD is diagnosed based on clinical presentation and biotinidase enzyme activity. Those with partial BTD may develop symptoms upon exposure to metabolic stressors. The two identified variants included a well-established pathogenic variant (c.1612C>T, p.Arg538Cys), causing profound BTD in homozygosity. In addition, a novel splice variant (c.44+1G>A, p.?) was identified in the invariant splice donor region of intron 1, potentially predictive of loss of function. However, due to evidence of exon 1 being alternatively spliced in different tissues, this variant was initially classified as uncertain significance. Sanger sequencing of parental samples demonstrated biallelic inheritance and upon medical record review, it was determined that the infant failed NBS for BTD, though passed on a rescreen. These data led to reconsideration of the pathogenicity of the variant as likely pathogenic. Diagnostic enzyme testing, prompted by the return of the GS results, confirmed partial BTD (1.5 nmol/min/ml serum, reference range: >3.5, 1.47 +/- 0.41 partial deficiency) and the baby was started on biotin supplementation. This outcome highlights the complexity of interpreting novel truncating variants in genes where loss of function is associated with disease, but for which alternative splicing of multiple transcripts confounds variant interpretation, and is an example of complexities involved in interpreting and managing cases without significant medical precedent. In conclusion, this BabySeq case demonstrates the utility of combining genotype and phenotype to more effectively guide the interpretation of both GS and NBS and inform the treatment of a newborn.

2428W

One novel 2.43Kb deletion and one single nucleotide mutation of INSR gene in a Chinese neonate with Rabson-Mendenhall syndrome. *L. Yang, X. Chen, W. Zhou.* Children's Hospital of Fudan University, Shanghai, China.

Background: Insulin receptor gene (*INSR*) is responsible for Donohue syndrome (DS) and Rabson-Mendenhall syndrome (RMS). These two diseases are all characterized as insulin resistance. **Methods:** A Chinese neonate suffering from glucose homeostasis, hyperinsulinemia, dry skin, heavy hair, elevated testosterone and growth retardation was recruited. 2742 inherited disease-gene panel sequencing was performed to search for candidate point mutations, small insertions or deletions and copy number variants. **Results:** One pathogenic mutation (c.3355C>T, p.Arg1119Trp) and a novel 2.43Kb deletion (chr19:7150507-7152938) in *INSR* were found in this patient by Next Generation Sequencing test using ClearSeq. The patient is diagnosed as RMS. Follow-up Sanger sequencing and real-time quantitative PCR confirm the variants. We, therefore, supposed these variants were candidate mutations of this family. **Conclusions:** We report a novel 2.43Kb deletion in *INSR* gene and provide further proof to the power of Next Generation Sequencing in rare disease diagnosis. **Keywords:** Insulin receptor gene (*INSR*); Rabson-Mendenhall syndrome (RMS); Neonate; Mutation; Next generation sequencing.

2429T

Concurrent determination of *ABO RhD* blood types and the HIV-1 resistance marker *CCR5* deletion via rapid multiplex PCR and capillary electrophoresis-based genotyping. *E. Schreiber[1], S. Higdon[1], K. Varma[1], S. Jackson[1], T. Nong[2], J. Lee[2].* 1) Genetic Sciences Division, Thermo Fisher Scientific, South San Francisco, CA; 2) One Lambda TDX Division, Thermo Fisher Scientific, Canoga Park, CA.

For optimal outcomes when matching bone marrow donors with their recipients, it is preferable to use bone marrow of identical or compatible blood types. Bone marrow registries thus require high resolution HLA genotyping data to match donor specimens with their recipients. We developed a research assay to aid in these investigations, which utilizes buccal swab DNA from potential donors to determine the *ABO* and Rh-antigen genotypes. In addition, the assay detects a 32 bp deletion in the *CCR5* gene. Homozygous carriers of this deletion are resistant to HIV-1 infection, and thus could be valuable stem cell donors for HIV-infected recipients. This research assay consists of a multiplex-PCR reaction with 5 fluorescently-labeled and 12 allele-specific primers followed by capillary electrophoresis on the new Applied Biosystems SeqStudio™ Genetic Analyzer. Four targeted SNPs of the human *ABO* gene allow the determination of the A, A2, B, O1,2 and O3 alleles. Rh-antigen genotyping is determined by targeting a small deletion that differs between the *RhD* and *RhCE* genes. The peak pattern is analyzed with GeneMapper™ software, and the resulting peak/genotype table is translated into a genotype/phenotype report using a standard spreadsheet. We verified this research assay by analyzing a panel of DNA samples with known blood group antigens and *CCR5* gene types. The results of this verification were found to be 100% accurate. This easy to use, rapid research assay may prove useful for future development of bone marrow donor identification assays, and other areas of public health studies. For Research Use Only. Not for use in diagnostic procedures.

2430F

Chromosome microarray analysis as a first-tier test for cytogenetic abnormalities in patients with myelofibrosis. *J. Mazzeu[1], M.R. Paula-Júnior[1], J. Minuncio[2], R.S. Bonadio[1], A. Pic-Taylor[1], S.F. Oliveira[1], A. Nonino[2], C.C. Mascarenhas[3].* 1) Universidade de Brasilia, Brasilia, Brazil; 2) Secretaria de Estado de Saúde, Brasilia, Brazil; 3) Universidade Católica de Brasília, Brasilia, Brazil.

　Primary Myelofibrosis (PMF), the most infrequent and severe of Philadelphia–negative classical myeloproliferative neoplasms, is characterized by bone marrow fibrosis, extramedullary hematopoiesis and abnormal cytokine expression, with consequent hepatosplenomegaly and cytopenias. Myelofibrosis may also result from late progression of Polycythemia Vera (PV) or Essential Thrombocythemia (ET), being then termed as post-PV myelofibrosis or post-ET myelofibrosis. Cardinal features of myelofibrosis include extramedullary hematopoiesis, hepatosplenomegaly and cytopenias. More than 85 percent of PMF patients have a mutually exclusive mutation in one of the following three genes: *JAK2* (60-65%), *MPL* (5%), or *CALR* (20-25%). All of these, which are referred to as "driver" mutations, activate JAK-STAT pathway. Cytogenetic abnormalities detected by karyotype abnormalities can be found in 32 to 48% of patients and have a high correlation with the prognosis. Though several large chromosome abnormalities have been described in patients with myelofibrosis, most were detected by conventional karyotyping and the influence of minor cytogenetic abnormalities and loss of heterozygosity (LOH) on the course of the disease has not been explored. We studied 15 patients using CytoScan platform (750k or HD- Affymetrix). This platform allows the identification of copy number variations as well as segments of loss of heterozygosity. Chromosome microarray analysis showed abnormalities in 12/15 patients (80%). A total of 35 copy-neutral LOH segments were identified in 10/15 patients, while 10 deletions and 3 duplications were found in 5/15 and 1/15 respectively. Only three imbalances were large enough to be detected by conventional karyotyping and microarray analysis revealed that the rearrangement was more complex. Most aberrations identified were private in our sample with recurrence only of LOH 9p (3/15) and 13q deletion (2/15). Copy-neutral LOH segments are attributed to mitotic recombination in somatic cells and are increasingly being recognized in a variety of neoplasms. Based on the high prevalence of LOH segments identified in our cohort we propose chromosome microarray analysis as a first-tier test for cytogenetic abnormalities in patients with myelofibrosis. Grant support: FAPDF.

2431W

Mutation profiling of 16 candidate genes in de novo acute myeloid leukemia patients. *H.X. Liu, Y. Zhang, F. Wang, X. Chen, W. Teng, J.C. Fang, W.J. Liu, M.Y. Wang, P.X. Cao.* Pathology & Laboratory Medicine Division, Hebei Yanda Lu Daopei Hospital, Beijing, China.

　Object: Acute myeloid leukemia (AML) is a group of heterogeneous diseases. Along with the progress of molecular biology of hematological malignancies and the use of next generation sequencing technology, more and more molecular genetic markers were identified.Hematological malignances gene mutation profiling is no longer confined to the detection of only a few molecular targets, the current study and application of leukemia has entered the "mutaome" profile era. The initiation and progression of tumors are the result of mutational synergism of many different functional genes. Currently there are some studies that have found certain rules on the occurrence and mutation pattern of gene mutation. In this study, we analyzed 16 common mutated genes in 259 de novo AML patients and investigated the rules of mutation spectrum, offering an approach to further explore the molecular biological mechanism of tumorigenesis and find new therapeutic targets. **Methods**: 259 patients who were diagnosed as de novo AML were enrolled in. Mutations profiling of 16 candidate genes were performed in bone marrow samples by Sanger sequencing, which include ASXL1, CEBPA, DNMT3A, ETV6, FLT3(ITD and TKD), IDH1, IDH2, KIT, KRAS, NRAS, NPM1, PTPN11, PHF6, RUNX1, TET2 and TP53. **Results:** 76.83% (199/259) cases had at least one mutation, 31.66% (82/259) cases carried more than one. FLT3-ITD was the most common mutated gene (16.22%, 42/259), followed by CEBPA (15.06%, 39/259), NRAS (14.67%, 38/259) and NPM1(13.52%, 35/259). 97.14% of NPM1 and 29.63% of double-mutated CEBPA were found to co-occur with other mutations. Distinct patterns of co-occurrence were observed for different hotspot mutations within IDH2 gene: R140 mutations was associated with NPM1 and/or FLT3-ITD mutations, whereas R172 mutations co-occurred with DNMT3A mutations only. The concurrence of 86.57% of epigenetic regulation genes was observed, most of which co-occurred with NPM1 mutations. **Conclusions:** The results showed that there were certain rules in the mutation profiling and concurrence of AML patients, which was related to the function classification of genes. Defining the mutation spectrum and mutation pattern of AML will contribute to comprehensive assessment of patients and identification of new therapeutic targets.

2432T

Xq22.1 contiguous deletion syndrome as a diagnostic challenge: Detection of a 17 kb deletion ends 30-year diagnostic odyssey. *G. Raca[1,3], C. Fong[1], J. Ji[1,3], B. Liu[1], D. Estrine[1], C. Hinh[1], F. Clarito[1], S. Saitta[1,3], J.A. Church[2,3].* 1) Dept. of Pathology and Laboratory Medicine, Children's Hospital Los Angeles; 2) Division of Clinical Immunology and Allergy, Children's Hospital Los Angeles; 3) Keck USC School of Medicine, Los Angeles.

X-linked agammaglobulinemia (XLA) is caused by loss-of-function mutations in the *BTK* (Bruton's tyrosine kinase) gene in Xq22, and manifests in affected males with recurrent bacterial infections in the first two years of life. Mohr-Tranebjærg syndrome (MTS), an unrelated X-linked disorder, is caused by mutations in *TIMM8A* (Translocase of Inner Mitochondrial Membrane 8 Homolog A) and shows early childhood hearing loss, progressive dystonia in adolescence, vision impairment in young adulthood and dementia in the fourth decade. *BTK* and *TIMM8A* are 770 bp apart in Xq22.1 and a rare condition with features of both disorders is caused by Xq22.1 deletions encompassing both genes. We report a 39-year old male with clinical XLA diagnosed in infancy, and sensorineural hearing loss seen at age 2 and attributed to viral encephalitis. His XLA was managed with immunoglobulin replacement therapy and he was never hospitalized for serious infections. A new onset of frequent falls, unstable gait and loss of fine motor skills prompted neurological evaluation with no etiology identified. Genetic investigation for MTS was then pursued, with phenotype-driven analysis of a whole-genome chromosomal microarray (CMA) revealing a small deletion in Xq22.1. Although the deletion was only 17kb in size, it included the 3' end of *BTK* and all of *TIMM8A*. A custom PCR assay showed that the deletion included exons 15-19 of *BTK* and both exons of *TIMM8A*. To our knowledge this is the smallest deletion found for the Xq22.1 contiguous deletion syndrome. The *BTK* and *TIMM8A* gene regions are enriched in interspersed *Alu* elements and prone to small deletions from mispairing and rearrangements between *Alu* repeats (Arai et al., 2017). Genetic evaluation for XLA usually includes both sequencing and deletion/duplication assays, but focuses on the *BTK* gene. XLA patients with additional neurologic phenotypes may be referred for CMA analysis but most platforms lack sufficient resolution to identify small Xq22.1 contiguous deletions. Timely diagnosis of XLA and MTS allows appropriate clinical management of patients, anticipatory guidance for families and an opportunity to test maternal carrier status, enabling recurrence risk counseling. This case underscores that the diagnosis of an Xq22.1 contiguous deletion can be challenging; in addition to a high level of clinical suspicion, it requires close collaboration between the clinician and the diagnostic laboratory to identify an appropriate testing strategy.

2433F

HLA typing using capture based next generation sequencing. *S.K. Lai[1], P.L Chen[1,2,3,4].* 1) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan.

Human leucocyte antigen (HLA) genes are important in organ transplantation and immuno-related diseases. Current HLA typing methods, either by sequence-based typing (SBT) or sequence-specific oligonucleotide (SSO), may suffer from ambiguous typing results. The problem is caused by the phasing problem or the polymorphic sites lying outside the sequencing region. Using probe capture-based next generation sequencing (NGS) may resolve the phasing problem, and be able to sequence larger regions of genes including introns. Comparing to the traditional methods, NGS provides higher throughput, lower costs and possibility to cover more loci on demand. In this study, we established the probe capture-based NGS method for HLA typing. Probe sets for Class-I HLA gene (HLA-A, B, C) in exon 2-4 and Class-II HLA gene (DRB1, DPA1, DPB1, DQA1, DQB1, DRB3,4,5) in exon 2-3 were designed. Ten samples which were previously known to carry the most common HLA types in Taiwanese were used to validate this method. The results were 100% concordant to SBT method. Furthermore, our data showed that this method could resolve ambiguous HLA types. .

2434W

A patient with hereditary pyropoikilocytosis caused by a combination of a novel in-frame deletion and a common functional but non-pathogenic allele, α^{LELY}, in *SPTA1*. *T. Goto[1], T. Togawa[2], T. Ito[1], M. Kouwaki[1], H. Ogura[3], H. Kanno[3], S. Saitoh[2], N. Koyama[1].* 1) Department of Pediatrics, Toyohashi Municipal Hospital, Toyohashi, Japan; 2) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 3) Department of Transfusion Medicine and Cell Processing, Tokyo Women's Medical University, Tokyo, Japan.

Hereditary pyropoikilocytosis (HPP) and hereditary elliptocytosis (HE) represent genetically and phenotypically heterogeneous hemolytic anemia caused by mutations in *SPTA1*, *STPB*, and *EPB41* encoding the red blood cell (RBC) cytoskeleton proteins α-spectrin, β-spectrin, and protein 4.1R, respectively. HPP/HE caused by mutations in *SPTA1* or *SPTB* leads to qualitative defects of α- or β-spectrin. Here we describe a patient with HPP diagnosed by targeted next-generation sequencing (NGS). The proband was a Japanese male neonate born from nonconsanguineous parents. He presented indirect hyperbilirubinemia with hemolysis from 23 hours after birth. He underwent phototherapy intermittently for 9 days and he required RBC transfusion for hemolytic anemia on 58 days after birth. We analyzed his blood smear and revealed abnormal RBC morphology, poikilocytosis, including elliptocytes, spherocytes, and fragmented erythrocytes. We analyzed his parents' blood smear and only his father' smear showed elliptocytosis. His father had needed phototherapy for his neonatal jaundice, but no RBC transfusion was performed. We performed targeted NGS to the proband, using Agilent HaloPlex custom kit and Ilumina Miseq. The custom gene panel included 67 candidate genes, such as *SPTA1*, *SPTB*, and *G6PD*. Consequently, we identified 3 variants in *SPTA1* (NM_003126.2); c.800_802del in exon 6 (novel in-frame deletion), c.5572C>G in exon 40, and c.6531-12C>T in intron 45 in heterozygous state each. Sanger sequence using samples of his family members showed that his father and sister had only c.800_802del in heterozygous state while his mother had c.5572C>G and c.6531-12C>T in heterozygous state each. The doubly mutated allele carrying c.5572C>G and c.6531-12C>T is named α^{LELY}. Although α^{LELY} is found 20-30% of the population in Western country, the polymorphism results in partial skipping of exon 46 and a decrease in the amount of spectrin. Monoallelic and biallelic α^{LELY} are silent, while compound heterozygous for HE-causing *SPTA1* allele and α^{LELY} in trans cause a severe phenotype, HPP. Most of HE causing pathogenic variants were missense mutations. Only one in-frame deletion (c.1406_1408del) was reported in patients with HE and we considered that the c.800_802del was a novel pathogenic variant. This family represents a unique genotype-phenotype correlation; a pathogenic mutation in combination with a common functional but non-pathogenic allele in *SPTA1* could cause much severe phenotype.

2435T

Analysis of germ line predisposition in Chinese children with bone marrow failure. *N. Dai-jing, L. Hong-xing, C. Xue, W. Fang, Z. Hui-zheng, T. Wen.* Pathology & Laboratory Medicine Division, Hebei Yanda Lu Daopei Hospital, Langfang, China.

Purpose: The role of germ line predisposition plays in the pathogenesis of Bone Marrow Failure(BMF) has not been explored as yet, the aim of the study is to detect it in Chinese children with BMF. **Methods:** Bone marrow or peripheral blood samples of 110 children with BMF were collected. All coding exons and flanking sequences of BRCA2, FANCA, FANCC, FANCD2, FANCG, CTC1, DKC1, NHP2, NOP10, RTEL1, TERC, TERT, TINF2 and WRAP53 were amplified and sequenced by Next Generation Sequencing(NGS) technology. Clinical significance of mutations were predicted by SIFT and PolyPhen-2 simultaneously. **Results:** The median age was 10(1-13) years with male/female ratio of 1.56:1. All patients were confirmed Bone marrow hyperplasia or severe hyperplasia by Bone marrow aspiration smears. 96 patients were diagnosed as Aplastic anemia, 2 as low proliferative MDS, 10 as Fanconi anemia(FA) and 2 as Dyskeratosis congenita(DC). Mutations were observed in 58(52.73%)of them. The number of patients harboring mutations of each gene were 11, 15, 7, 6, 10, 4, 0, 2, 0, 10, 0, 4, 4 and 1 respectively. In the patients those with FANCA mutations, 3 were homozygous and 1 was compound heterozygous, those with FANCC mutations, 1 was homozygous, those with RTEL1 mutations, 1 was compound heterozygous, those with TERT mutations, 2 were compound heterozygous, and all the rest were heterozygous. All mutations were confirmed to be germ line derived. Compared with 71 healthy Chinese people we tested and 208 healthy Chinese people in 1000 genomes database, the total mutation rate and pathogenic mutation rate of the case group were statistically high(52.73% vs. 37.63%, p=0.007); (30.00% vs. 19.71%, p=0.029). **Conclusion:** A boy who carried FANCA c.1074_1075del/p. Y359PfsX49 monoallelic mutation manifested BMF at the age of 7 with positive chromosome breakage test, suggesting nonclassical Mendelian inheritance patterns in FA. Mutation rate and pathogenic mutation rate were significantly high in case group, suggesting an important role hereditary susceptibilities have played in the pathogenesis of BMF, haploinsufficiency may be the explaination to some extent. New genes related to Inherited bone marrow failure syndromes have being identified continuously. Tested genes here were very limited, we speculate expanding the spectrum may find more patients with hereditary predisposition, which may be useful to predict clinical outcome and optimize the diagnosis and treatment of these patients. .

2436F

The study of the applicability of three statistical methods in IGH immune repertoire analysis. *F. Wang, J.C. Fang, Y.L. Xu, M.M. Wang, T.T. Li, C.Y. Cao, Y. Zhang, H.X. Liu.* Molecular Medicine Lab., Hebei Yanda Ludaopei Hospital, Beijing, China.

   Objective To explorer the applicability of rarefaction, Shannon-Wiener Index and Simpson Index in immunoglobulin heavy chain(IGH) immune repertoire analysis. **Methods** We selected 15 B cell tumor patients at diagnosis and 9 healthy individuals who were tested IGH clonal rearrangement using capillary electrophoresis(CE). Then extracted DNA from bone marrow or peripheral blood sampled from them. We amplified IGH V region segments using LymphoTrack® IGH Assay-PGM kit, and performed next generation sequence on Ion Torrent PGM. The FASTQ data sets were performed quality control and comparison analysis using MiXCR software, and were analyzed rarefaction, calculated Shannon-Wiener Index and Simpson Index, and drawed rarefaction curve using VDJtools software. Then analyzed the differences of Shannon-Wiener Index and Simpson Index between groups using Mann-Whitney U test. **Results** We grouped these samples into 2 groups(B cell tumor patients and healthy individuals) based on the positive and negative CE results. Rarefaction curve visualized the diversity of IGH V region segments in 2 groups. Compared to negative samples, the curve of positive samples faster reached the platform period when the sample sizes were same, and the diversity did not increase following the growing of the sample size. The median value of divesity in the positive samples was 45(;15-100);, which was lower than the negative samples that was 370(;219-580);(;P=0.000);, indicating that the IGH V region segments diversity in positive samples was reduced compared to negative samples. Shannon-Wiener Index and Simpson Index could quantify the diversity, and the median values were 1.639(;1.091-3.980);vs 236.069(;142.975-488.816);, 0.127(;0.023-0.614);vs 0.995(;0.991-0.998);between positive and negative samples respectively. The two index there were both statistically difference(;P=0.00);. **Conclusion** In this study, we could visualized and quantified the IGH immune repertoire diversity using rarefaction, Shannon-Wiener Index and Simpson Index.

2437W

47,XY,+21/46, XX chimera identified in an infant with ambiguous genitalia without Down syndrome features. *C. Charalsawadi, S. Jaruratanasirikul, P. Limprasert.* Department of Pathology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand.

   Chimerism resulting from a fusion of two different zygotes is a very rare genetic finding in humans and is mostly identified by the coexistence of XX and XY chromosomal complements in an individual. Most reported cases exhibit a chi 46,XX/46,XY karyotype and ambiguous genitalia while only a few cases exhibited an abnormal lineage coexisting with a normal one. To the best of our knowledge, only three, non-twin cases carrying both trisomy 21 and a normal karyotype have been reported. These include two cases with a chi 47,XY,+21/46,XX karyotype and a single case with a chi 47,XX,+21/46,XY karyotype. A DNA polymorphism study to identify the mechanism underlying chimerism was performed in two cases. Herein, we report an additional case of a chi 47,XY,+21/46,XX karyotype. Physical examination at 1 month of age revealed a small phallus, an opening of the urogenital slit, labioscrotal swellings without rugae formation, and a palpable gonad in the right inguinal area. Besides ambiguous genitalia, the infant had no features of Down syndrome. Growth and developmental milestones were within normal ranges. Cytogenetic analyses found approximately 17% of a 47,XY,+21 lineage for both G-banding karyotype analysis of 100 metaphase cells and FISH analysis of 1000 interphase cells obtained from uncultured peripheral blood. Short tandem repeat analysis using AmpFLSTR™ Identifiler™ Plus PCR Amplification Kit and ABI3130 Genetic Analyzer revealed 3 alleles at 7 different loci, in which 5 loci indicated a double paternal contribution and 2 loci indicated a double maternal contribution to the patient. This finding suggests the chimerism in our patient arose from either early fusion of dizygotic twins from double fertilization, or double fertilization of an ovum and a polar body by two different sperm with subsequent fusion of the zygotes.

2438T

Novel approach using NGS assay for detection of mutations of the *CYP21A2* **gene.** *G. Bennett[1], C. Clear[2], R. Hrdlickova[2], J. Nehyba[2], D. Fox[2], D. Obermoeller[2], M. Toloue[1].* 1) PerkinElmer, Boston, MA; 2) Bioo Scientific, Austin, TX.

Congenital Adrenal Hyperplasia (CAH) is a disease associated with mutations in the cytochrome p450 (CYP) 21A2 (*CYP21A2*) gene, leading to impaired function and reduced levels of the enzyme 21-hydroxylase (21OHD). CAH is the most common autosomal recessive disorder, with frequencies of carriers ranging between 4-10% in the population. The *CYP21A2* gene has a nonfunctional pseudogene, *CYP21A1P*, located 30 kb upstream of *CYP21A2*. *CYP21A2* and *CYP21A1P* are located in the HLA class III region on chromosome 6p21.3. The nucleotide homology between these two genes in their exonic regions is 98%. Two genetic events create non-functional chimaeras between *CYP21A2* and *CYP21A1P* – gene conversion and rearrangement. Gene conversion replaces specific sequences in *CYP21A2* with homologous sequences from *CYP21A1P*. Genomic rearrangement forms chimaeras between both genes by a 30 kb deletion of the neighboring sequence and bringing the two loci together. Both events generate a different set of products which will impact the phenotype. Unfortunately, it is impossible to detect these chimaeras by classical next generation sequencing (NGS) approaches. Congenital Adrenal Hyperplasia Amplicon Panel allows the rapid detection of these mutations using a novel, simplified protocol. The workflow starts with a long range PCR which will amplify the target gene, as well as rearranged chimaera if present, by using as little as 20 ng DNA from fresh and frozen tissue. A portion of this enriched template is taken and placed in a second PCR to generate amplicons that cover all coding exons, exon-intron boundaries, and a portion of the promoter. Further, tags are bound to each amplicon to differentiate between product resulting from gene conversion and the product from gene rearrangement during analysis. The final panel has 100% uniformity at 20% average amplicon read threshold, 92% on-target reads, and <1% dimer formation. Using the NEXTflex® barcode system, 384 samples can be multiplexed in a single Illumina® MiSeq® run. Additionally, the panel is compatible with both Ion Torrent™ and Illumina® sequencing platforms. The entire *CYP21A2* gene and its chimaera with *CYP21A1P* are each amplified in 3.2 kb comprising 10 exons and a partial promoter region. This approach is unique in the way that it will detect complex genetic changes in the *CYP21A2* and *CYP21A1P* loci by taking advantage of precise and economical NGS technology.

2439F

NEBNext Direct CFTR enrichment panel: Interrogation of genetic variants by Illumina sequencing. *S. Adams[1], K. Patel[1], A. Emerman[1], S. Bowman[1], C. Elfe[1], T. Davis[2], S. Russello[2], A. Barry[2], C. Hendrickson[1].* 1) Directed Genomics, Ipswich, MA; 2) New England BioLabs, Ipswich, MA.

Prompt and accurate diagnosis of cystic fibrosis (CF) is vital for achieving optimal therapies and outcomes. The screening and detection of variants in the cystic fibrosis transmembrane conductance regulator (CFTR) are critical for the effective management of this disease. More than 2000 CFTR mutations have been reported worldwide and refined sequencing technologies may provide improved genotype–phenotype correlations and better-informed genetic counseling. Here, we introduce the NEBNext Direct CFTR target enrichment panel for the interrogation of genetic variants in CFTR. The NEBNext Direct® technology uses a novel approach to selectively enrich nucleic acid targets ranging from a single gene to several hundred genes without sacrificing specificity. The approach rapidly hybridizes both strands of genomic DNA with biotinylated baits, captures the targets on streptavidin beads, enzymatically removes off-target sequence, and directly converts captured molecules into Illumina-compatible sequencing libraries in a single day protocol. Unlike alternative hybridization methods, NEBNext Direct does not require upfront library preparation. Instead, target molecules acquire an 8 base pair sample ID and a 12 base pair Unique Molecular Identifier (UMI) as part of the targeting protocol. The UMI tags each individual molecule prior to the final PCR amplification and enables identification of PCR duplicates. The CFTR panel consists of a single pool of baits targeting both strands of DNA across 23 exons and 7 intronic sites, for a total covered territory of 6.7 kb. The panel provides uniform coverage of the CFTR region and allows for germline variant detection with 10 ng of input DNA.

2440W

Novel pathogenic variants are routinely detected even in extensively-sequenced genes, such as CFTR. *N. Faulkner, C. Perreault-Micale, M. Zhu, K. Robinson.* Good Start Genetics, Cambridge, MA.

   Cystic fibrosis (CF), caused by pathogenic variants in the *CFTR* gene, is one of the most common autosomal recessive disorders. Carrier screening for CF is recommended for all couples considering a pregnancy, regardless of ethnicity. However, many carrier screening laboratories only offer testing for limited panels of the most prevalent pathogenic variants, resulting in CF screening tests with lower detection rates in non-Caucasian populations. Using a clinical NGS-based assay, we sequenced the coding region and intron-exon boundaries of *CFTR* in a large pan-ethnic population of individuals referred for carrier screening. Sequencing results were filtered for known pathogenic variants, as well as previously unreported ("novel") variants predicted to be pathogenic based on their truncating effect (nonsense, frameshift, and canonical splice site variants). Despite the fact that *CFTR* has been extensively studied with >2000 variants identified to date, we detected 34 different novel truncating *CFTR* variants in our patients. One novel variant was found in two individuals; all others were detected only once. Self-reported ethnicities of the 35 carriers included 11 Asian, 8 African American, 4 Caucasian, 2 Hispanic, 1 South East Asian, and 9 individuals who did not provide their ethnicity, indicating that the majority of novel variant carriers were not Caucasian. Five of the variants (c.51delC, c.3743C>G (p.Ser1248X), c.874_875delGA, c.1943delA, and c.1807delG) have been subsequently reported in CF patients, further affirming their pathogenic status. Of note, we did not detect any mutational hot spots as novel variants were distributed throughout the gene. It is remarkable that, in a well-studied gene like *CFTR*, we identified 35 carriers of pathogenic variants that had not been previously reported in the literature. Even the most comprehensive targeted panel of known pathogenic variants would have missed these carriers. Novel truncating variants were preferentially detected in non-Caucasian individuals, highlighting both the need for better characterization of the variant spectrum in these populations and the importance of offering them a truly pan-ethnic carrier screening test. Additionally, screening for large (exon-level) deletions and duplications across the *CFTR* gene can further enhance detection rates by identifying carriers of both rare and novel changes. We recently added del/dup analysis to our CF carrier screening test and will share our experience.

2441T

Mutations in the *DHX37* gene identified by whole-exome sequencing (WES) are a novel cause of the embryonic testicular regression syndrome (ETRS). *T.E. da Silva[1], A.M. Lerário[1], C.E. Keegan[2], M.Y. Nishi[1], F.M. Ferreira[3], B.M. Yashar[2], E.M.F Costa[1], B.B. Mendonca[1], S. Domenice[1].* 1) Disciplina de Endocrinologia e Metabologia, Laboratório de Hormônios e Genética Molecular/ LIM42, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil; 2) Department of Pediatrics and Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, USA; 3) Laboratório de Imunologia, Instituto do Coração, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil.

   WES have been an efficient tool in the genetic investigation of disorders of sex development (DSD). Affected and non-affected members of two unrelated Brazilian families with familial ERTS (Family 1 (F1)-2 affected brothers and the unaffected parents; Family 2 (F2) - affected maternal uncle and his nephew) and one Chinese-American family with sporadic ETRS were studied by WES using a HiSeq 2500 platform. A novel single heterozygous variant, p.Arg-308Gln in the DHX37 was identified in the affected members of the two families and in the Chinese-American child. This variant was confirmed by Sanger in all affected membrs and in the F1 father and the F2 mother. p.Arg308Gln variant was considered damaging in six predictor sites (Mutation Taster, Mutation Assessor, SIFT, Polyphen-2, CADD, GERP++) and was absent in EXAC, 1000GENOME, ESP6500, dbSNP population databases. Thirty-five additional patients with sporadic 46,XY DSD due to abnormalities of gonadal development (5 with ETRS and 30 with gonadal dysgenesis) were screened for this variant by Sanger methods. One out of 15 sporadic ETRS patients had the same heterozygous variant. *DHX37* (12q24.31) encodes a RNA helicase protein that belongs to the DEAH (Asp-Glu-Ala-His) family. Male patients with 12q24.31 deletions and genital abnormalities have been described in the literature, reinforcing the hypothesis that the DHX37 could to be involved in the 46,XY DSD etiology. Absence of the p.Arg308Gln DHX37 variant founder effect between the two Brazilian families was confirmed by hapltypes analyses. Epigenetic analysis performed by Infinium Methylation EPIC 850K BeadChip demonstrated hypomethylated promoter regions of *DHX37, FGF9, DMRT1* in the F1 father when compared with sons. The identification of deletirous variant in DHX37 in 6 patients with embryonic testicular regression syndrome provides strong genetic evidence that this gene is a novel candidate gene for 46,XY DSD due to abnormalities of gonadal development.

2442F

Diagnostic utility of whole-exome sequencing for chronic kidney disease. *E. Groopman[1], S. Cameron-Christie[2], J. Fleckner[2], R. March[2], D. Goldstein[3], A. Gharavi[1].* 1) Department of Medicine, Columbia University, New York, NY; 2) Centre for Genomics Research, Astrazeneca UK LTD; 3) Institute for Genomic Medicine, Columbia University, New York, NY.

   Background: Whole-exome sequencing (WES) has recently been introduced into clinical diagnostics, but its value for adult constitutional disorders requires further evaluation. We are assessing the diagnostic yield of WES in a large cohort of adults with all-cause Chronic Kidney Disease (CKD) or End-Stage Renal Disease (ESRD) recruited at Columbia University (N=1920) or enrolled in the AURORA study (N=1141). **Methods:** To date, the exomes of 765 patients have been analyzed using the American College of Medical Genetics (ACMG) guidelines for clinical sequence interpretation. **Results:** 97.4% of 765 patients were adults and 51.0% were non-Caucasian; 54% had a glomerulopathy, 7% had congenital defects, and 17.5% had CKD of unknown etiology. In total, 82 (10.7%) patients had a diagnostic variant for a genetic form of nephropathy. Among these diagnosed cases, 19.5% had presented with "CKD of unknown origin" and 41.5% noted no family history of nephropathy. In 47.6% of diagnosed cases, the molecular diagnosis confirmed the clinical diagnosis (e.g., Alport syndrome); in 52.4% it clarified the diagnosis (e.g., *UMOD* mutation in a patient with tubulointerstitial nephropathy) or provided an alternative diagnosis (e.g., Dent disease in a case of suspected glomerulopathy). Diagnostic variants were mainly found in genes for glomerulopathies (65.9%), followed by those for cystic disease (11.0%), tubulointerstitial disease (9.8%), other nephropathies (9.8%), and congenital anomalies (3.7%). In addition, 6 patients (0.8%) had a secondary, pathogenic variant in one of the 59 ACMG actionable genes. In the majority of cases, the results impacted clinical decision-making, including elements such as initiation of targeted surveillance, family counseling, selection of transplant donors, and modification of pharmacotherapy.**Conclusions:** In a large all-cause CKD cohort, WES gave a molecular diagnosis for 1 in 10 patients, and in 52.4% of positive cases, pinpointed an etiology not detected using traditional diagnostics, impacting clinical care. The completion of this study will provide comprehensive information about the clinical utility and the impact of genetic testing for nephropathy across broad demographic subgroups and etiologic subtypes.

2443W

De novo unbalanced insertional translocation, der(X)ins(X;5) (q?13;q12.3q13.1) in an adult female patient with developmental delay and ovarian insufficiency identified by DNA MicrorrayCGH and FISH. *M. Pitch[1], M. Guo[2], M. Shaver[2], A. Schroeder[3], M.Anwar Iqbal[2].* 1) University of Rochester School of Medicine & Dentistry, Rochester, NY; 2) Department of Pathology and Laboratory Medicine, University of Rochester, Rochester NY; 3) Division of Genetics, University of Rochester, Rochester, NY.

   Insertional translocations (IT's) involving the X chromosome are a relatively rare entity. Insertional rearrangements on the X chromosome may cause disruption of genes responsible for sexual maturity or by chromosome effect such as inhibition of meiotic pairing or skewed X inactivation (Prueitt 2002). We report a 21 year old patient with a history of intellectual disability, developmental delay who presented to clinic with amenorrhea. Previous karyotype testing at age 18 were reported as normal. The patient was referred for array comparative genomic hybridization (aCGH) and whole exome sequencing (WES). The WES findings did not report any pathogenic variants. Interestingly, a 3.3Mb duplication of 5q12.3q13.1 segment was revealed by aCGH. Fluorescence in-situ hybridization (FISH) with probe RP11-841D3 (5q12.3) identified the duplication 5q12.3q13.1 segment inserted within the long arm of chromosome X near the vicinity of XIST region (Xq13.2). Parental studies did not identify a carrier translocation by FISH. Follow up X chromosome inactivation studies in the patient lymphocytes reported highly skewed non-random X inactivation (97.3%). Our results demonstrate that the clinical symptoms in the patient are potentially due to the disruption of the XIST gene leading to skewed X-inactivation.

2444T

Pathogenic variants and variants of uncertain significance in autosomal dominant polycystic kidney disease (ADPKD) causative genes are commonly found in early-onset PKD patients: 2.5-year experience in a CAP/CLIA diagnostic laboratory. *W. Chen, D. Allingham-Hawkins, C. Iwaszczenko.* Clinical DNA Testing, PreventionGenetics, Marshfield, WI.

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common human Mendelian disorders with a prevalence of 1:400 to 1:1000. ADPKD is typically a late-onset disorder with nearly complete penetrance. An individual who carries a germline pathogenic variant in one of the three known ADPKD causative genes (*PKD1, PKD2* and *GANAB*) is expected to develop renal cysts detectable by ultrasound by age 30 or later. It has been well documented in literature that biallelic *PKD1* variants (including hypomorphic alleles) or multiple alleles from different PKD causative genes can manifest as a much more severe early-onset condition prenatally or in newborns. Here we present the genetic findings of autosomal dominant PKD genes in early-onset PKD patients tested in our CAP/CLIA diagnostic laboratory in the past 2.5 years. Of 97 unrelated patients no older than two years at age (including seven prenatal cases) referred for our PKD panel testing (including *PKD1, PKD2* and *GANAB*, as well as *HNF1B* and *PKHD1*), 40 cases (41%) were found to have at least one pathogenic variant or variant of uncertain significance in at least one of the autosomal dominant genes (*PKD1, PKD2, HNF1B* and *GANAB*). Notably, early-onset PKD in 18 patients (18.5% of total cases) can be possibly explained by biallelic *PKD1* variants while five cases (5.1% of total cases) had notable variants from different PKD causative genes including the autosomal recessive PKD (ARPKD) gene *PKHD1*. In 13 patients (13.7% of total cases), we found only one pathogenic variant or variant of uncertain significance (i.e. no other notable variants were found). Interestingly, we found a heterozygous pathogenic or highly suspected pathogenic *HNF1B* variant in remaining four patients (4.1% of total cases). In summary, our experience with diagnostic testing of very young children (including prenatal testing) with symptoms of PKD supports the hypothesis that biallelic *PKD1* variants and multiple alleles from different PKD causative genes can result in exceptionally early-onset PKD. Our data supports the clinical utility of testing autosomal dominant PKD genes in patients with early-onset PKD.

2445F

Diagnostic strategy in segmentation defect of the vertebrae: A retrospective study of 73 patients. *M. Lefebvre[1,2], A. Dieux-Coselier[3], G. Baujat[4], E. Schaefer[5], J. Saint-Onges[2], A. Bazin[6], L. Piinson[7], T. Attié-Bitach[4], C. Baumann[8], M. Fradin[9], G. Pierquin[10], S. Julia[11], C. Quelin[9], B. Doray[12], S. Berg[12], C. Vincent-Delorme[3], L. Lambert[13], N. Bachmann[14], D. Lacombe[15], B. Isidore[16], S. Odent[9], B. Demeer[17], A. Selicorni[18], A. Goldenberg[19], C. Thauvin-Robinet[1,2], A. Verloes[8], V. Cormier-Daire[4], J.B. Riviere[2], L. Faivre[1,2], J. Thevenon[1,2,20].* 1) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Est, FHU-TRANSLAD, CHU Dijon, France; 2) GAD EA4271 « Génétique des Anomalies du Développement » (GAD), Université de Bourgogne, Dijon, France; 3) Service de Génétique, CHRU Jeanne de Flandre, Lille, France; 4) Département de Génétique, Hôpital Necker Enfants Malades, AP-HP, Paris, France; 5) Service de Génétique, CHU Strasbourg, France; 6) Service de diagnostic anténatal, Hôpital René Dubos, Pontoise, France; 7) Service de Génétique, CHU Montpellier, France; 8) Département de Génétique, Hôpital Robert Debré, AP-HP, Paris, France; 9) Service de Génétique, CHU Rennes, France; 10) Service de Génétique, CHU Sart Tilman, Liège, Belgique; 11) Service de Génétique, CHU Toulouse, France; 12) Service de Génétique, Hôpital Félix Guyon, CHU La Réunion, France; 13) Service de Génétique, CHU de Nancy, France; 14) Center for Human Genetics, Bioscientia, Ingelheim, Germany; 15) Service de Génétique, CHU Bordeaux, France; 16) Service de Génétique, CHU Nantes, France; 17) Service de Génétique, CHU d'Amiens, France; 18) Azienda Ospedaliera San Gerardo, Monza, Italia; 19) Service de Génétique, CHU Rouen, France; 20) Service de Génétique et procréation, CHU Grenoble-Alpes, Grenoble, France.

Segmentation defect of the vertebrae (SDV) are commonly found while exploring malformative syndromes. Multiple SDV (M-SDV) are defined by the association of fused and/or abnormally formed vertebra and represent a rare clinical entity. The molecular bases of M-SDV are not fully elucidated due to the wide range of phenotypes and classification issues. The genes involved are in the Notch signaling pathway which is a key system in somitogenesis. Here we report on the genetic screening in a diagnostic cohort of patients with M-SDV and further delineate the genotype-phenotype correlations in order to establish a decision tree. The clinical and molecular data of 73 patients were included. A targeted sequencing of the 5 known spondylocostal dysostosis (SCD) causing genes (*DLL3, MESP2, LFNG, HES7* and *TBX6*) was performed in the 48 first patients and a whole exome sequencing (WES) in 28 patients A targeted sequencing approach diagnosed 10 patients with autosomal recessive variants in the 5 known genes: 4 patients with *TBX6* variations, 2 with *LFNG* variations, 2 with *DLL3* variations, one with *MESP2* and one with *HES7* variations. Among the 4 patients presenting with biallelic *TBX6* variation, two had M-SDV and a 16p11.2 deletions encompassing *TBX6* associated with a published common haplotype. WES diagnosed two patients and allowed differential diagnosis of Spondylo-carpo-tarsal fusion syndrome (FLNB) and Klippel-Feil syndrome (MEOX1). The diagnostic rate of the gene panel including the 5 genes of SCD was 10/73 (13.7%) in the global cohort, and 8/10 (80%) in the subgroup fulfilling the SCD criteria. We also established genotype phenotype correlation in patients presenting with biallelic *TBX6* variations, ranging from regional M-SDV to SCD. The molecular bases of SDV are complex and a complex genetic architecture should be investigated. The diagnostic work-up of a M-SDV should be guided by the clinical description. Because of their genetic complexity, we would suggest a targeted sequencing of the five disease causing genes for SCD. In the subgroup of patients presenting with asymmetric thoracic SDV, no variation was identified and those patients should be included in research project considering non mendelian inheritance.

2446W

Small 17p13.3 duplication including *BHLHA9* in a Brazilian family with incomplete penetrance of split-hand/foot malformation. *W.A.R. Baratela[1], H. Hijazi[2], C.M.B. Carvalho[2], G.L. Yamamoto[1,3], S.S. Costa[3], C. Rosenberg[3], C.A. Kim[1], D.R. Bertola[1,3], J.R. Lupski[2,4,5].* 1) Unidade de Genética, Instituto da Criança - Hospital das ClÃnicas - Universidade de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Department of Molecular and Human Genetics, BCM, Houston, TX; 3) Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, Instituto Biociências – Universidade de São Paulo, SP, Brazil; 4) Department of Pediatrics, BCM, Houston, TX; 5) Texas Children's Hospital, Houston, TX.

Split-hand/foot malformation (SHFM) with long bone deficiency (SHFLD) is a rare limb developmental abnormality typically involving the central rays of the autopod associated with long bone deficiencies, especially the tibia or the femur. Recurrence in familial cases is compatible with an autosomal dominant inheritance, with incomplete penetrance. Genomic imbalances, especially microduplications (254 to 527 kb) in 17p13.3, but also the presence of triplications and quadruplications, encompassing *BHLHA9*, a gene that encodes a basic helix-loop-helix transcription factor, have been found in several families. Although these copy-number gains are present in the majority of the reported cases, suggesting a gene dosage effect, the penetrance is approximately 50%, indicating that other genetic and/or epigenetic factors are required for phenotypic expression. We report a familial case with three individuals (the proband, his father and his paternal uncle) presenting SHFM with absent tibia, therefore SHFLD. High-density aCGH was performed in these individuals, as well as in three non-affected family members. A microduplication of ~ 62 Kb in 17p13.3 has been identified in all affected individuals, as well as in the proband's paternal grandmother and his paternal cousin. Exome sequencing performed in the proband's father failed to show any rare pathogenic variant in *BHLHA9* or other gene associated with SHFM. In the present family, all affected individuals were males, giving further support to previous descriptions involving male sex bias. In an attempt to identify a possible second genetic factor that could contribute to the high variable phenotypic expression and incomplete penetrance, exome sequencing was performed in one affected individual, but no other good candidate was found. Thus, further studies are required to understand the complete genetic mechanism responsible for SHFLD. FAPESP (2013/08028-1; 2015/21783-9), CNPq (302605/2013-4).

2447T

Novel pathogenic variants in craniosynostosis genes identified by NGS. *E. König[1], K. Platzer[2], I. Spier[3], S. Spranger[4], A. Tzschach[5], B. Zirn[6], T. Schweitzer[7], W. Kress[1], E. Klopocki[1].* 1) Insitute of Human Genetics, University of Würzburg, Würzburg, Germany; 2) Institute of Human Genetics, University of Leipzig, Leipzig, Germany; 3) Institute of Human Genetics, University of Bonn, Bonn, Germany; 4) Praxis fuer Humangenetik, Bremen, Germany; 5) Institute of Clinical Genetics, Technische Universität Dresden, Dresden, Germany; 6) Genetikum, Stuttgart, Germany; 7) Department of Pediatric Neurosurgery, University of Würzburg, Würzburg, Germany.

Craniosynostosis, the premature fusion of one or more cranial sutures, is a frequent craniofacial malformation affecting 1 in 2500 newborns. Premature ossification of the cranial sutures can occur either as isolated malformation or as part of a syndrome. So far genetic causes have been identified mainly for syndromic craniosynostoses, i.e. mutations in *FGFR2*, *FGFR3*, *TWIST1*, and *EFNB1*. High phenotypic variability and clinical overlap of the different syndromes often hamper straight forward genetic diagnostics, thus, necessitating conventional sequencing of more than one gene. Furthermore, in more than 50% of cases the underlying genetic cause remains unknown. We compiled a next generation sequencing (NGS) gene panel comprising 68 genes including known and candidate craniosynostosis genes of the syndromic and isolated type as well as genes associated with bone development. Target enrichment was performed by the Nextera Rapid Capture (Illumina) as well as by SureSelect[QXT] (Agilent). Sequencing was done on the Illumina MiSeq platform. Sequencing data were analyzed with the GensearchNGS software (PhenoSystems). Performance of the NGS gene panel was validated by sequencing five control patients with known mutations. All of these mutations were detected correctly. Subsequently, we sequenced genomic DNA of 54 patients with syndromic as well as isolated craniosynostosis. Hot spot sequencing of *FGFR1-3* and *TWIST1* was performed previously for most of the cases. Until now sequencing of craniosynostosis patients with our NGS gene panel enabled the detection of likely pathogenic variants in 32% of the cases. We identified 8 already known mutations in *FGFR1, FGFR2, FGFR3,* and *TWIST1*. In addition, we detected 9 novel, potentially disease causing variants in the genes *ERF*, *MEGF8*, *MSX2*, and *TCF12*. We want to highlight a novel, likely pathogenic variant in *MEGF8*, the Carpenter syndrome 2 gene: a homozygous variant (c.828G>A) predicted to affect the splice site in two children of consanguineous parents. Our patients show an overlapping phenotype with Carpenter syndrome 2 (MIM #616294). An *in-vitro* minigene assay showed that the c.828G>A variant leads to skipping of exon 5 which is predicted to result in a frameshift and subsequently in a premature stop codon. In summary, our NGS panel approach provides a useful tool in determining the underlying genetic cause of craniosynostosis in a time and cost efficient manner especially in patients with an unclear clinical diagnosis.

2448F

Detection of mosaic copy-number variation from whole-exome sequencing using XHMM and custom SNP approach. *A. Sorlin[1,2], E. Tisserant[1,2], J. Thevenon[1,2,3], Y. Duffourd[1,2], P. Kuentz[1,2,4], V. Carmignac[1,2], V. Cormier-Daire[5], C. Michot[5], V. Malan[6], MP. Beaujard[6], F. Morice-Picard[7], S. Naudion[7], C. Rooryck-Thambo[8], C. Vincent-Delorme[9], T. Smol[10], E. Boudry-Labis[10], S. Hadj Rabia[11], A. Phan[12], MP. Cordier[13], M. Till[13], J. St-Onge[1,2,14], C. Thauvin-Robinet[1,2,3], AL. Mosca Boidron[1,2,16], L. Faivre[1,2,3], JB. Rivière[1,2,14,15], P. Vabres[1,2,17], P. Callier[1,2,16].* 1) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement, CHU Dijon Bourgogne, Dijon, France; 2) Equipe d'Accueil 4271, Génétique des Anomalies du Développement, Université Bourgogne Franche-Comté, Dijon, France; 3) Service de Pédiatrie 1 et de Génétique Médicale, CHU Dijon Bourgogne, Dijon, France; 4) Génétique Biologique Histologie, CHRU de Besançon, Besançon, France; 5) AP-HP, Hôpital Necker-Enfants malades, Genetics Departement, Centre of Reference for Skeletal Dysplasia, INSERM UMR 1163, Institut Imagine, University Paris Descartes-Sorbonne Paris Cité, Paris, France; 6) Service d'Histologie-Embryologie-Cytogénétique, Hôpital Universitaire Necker-Enfants Malades, Paris, France; 7) Service de génétique médicale, CHU de Bordeaux-GH Pellegrin, Bordeaux, France; 8) Laboratoire de génétique moléculaire, CHU de Bordeaux-GH Pellegrin, Bordeaux, France; 9) Service de génétique clinique, CHRU de Lille - Hôpital Jeanne de Flandre, Lille France; 10) Laboratoire de Génétique médicale, CHRU de Lille - Hôpital Jeanne de Flandre, Lille France; 11) Service de dermatologie, Hôpital Universitaire Necker-Enfants Malades, Paris, France; 12) Pediatric Dermatology Department, Hôpital Femme Mère Enfant, Hospices Civils de Lyon, , Lyon F-69000, France; 13) Service de génétique, GH Est-Hôpital Femme Mère Enfant, Hospices Civils de Lyon, Lyon, France; 14) Child Health and Human Development Program, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; 15) Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, Quebec, Canada; 16) Laboratoire de génétique moléculaire et de Cytogénétique, Plateau Technique de Biologie, CHU Dijon Bourgogne, Dijon, France; 17) Service de Dermatologie, CHU Dijon Bourgogne, Dijon, France.

Whole exome sequencing (WES) emerged as a powerful tool for deciphering the molecular basis of developmental disorder, either single nucleotides variants (SNVs), indels, or recently copy number variants (CNVs). In mosaic development disorders, it has allowed detection of post-zygotic point mutations (mosaic SNVs) in various genes, mainly in the RAS-MAPK and PI3K-MTOR pathways. However, detection of mosaic CNVs still relies on conventional cytogenetic studies, such as array-CGH and array-SNP. We sought to develop an all-in-one strategy for patients with mosaic disorders, using trio-WES on lesional skin from patient and blood from his parents. We combined a conventional SNV detection pipeline with a CNV detection approach, based on both a read depth approach (using XHMM) and a custom SNP-based approach on inherited variants. In a cohort of 57 patients with various mosaic disorders involving the skin, pathogenic SNVs were detected in 22 patients (38.5%) in already known or previously uninvolved genes. In a subgroup of 19 patients with hypomelanosis of Ito (linear mosaic depigmentation), we were able to detect 6 mosaic CNVs (10,5%): 3 complete mosaic trisomy (for chromosomes 7, 12 and 15) and 3 mosaic CNVs of smaller size (3q26.1-q29 and 6p22.3-p25.3 gains, 13q12.11-q13.3 loss) increasing our diagnostic rate to just under 50%. Blood and skin karyotypes and array-CGH were previously negative, due either to the absence of mosaic cells in blood or to their elimination from cultured fibroblasts. All variants were confirmed using array-CGH, array-SNP and FISH on skin DNA. Among the 3 complete mosaic trisomies, the data on single nucleotides polymorphism inheritance and b-allele frequency provided information on the parental origin of extra chromosomes and clues to understand the underlying mechanism, such as meiotic non-disjunction. Hence, we have confirmed that chromosomal mosaicism is frequently associated with mosaic pigmentation disorders (6/19, 32% in our cohort). The use of an appropriate fresh tissue sample is essential, as CNVs were not detected in blood or cultured fibroblasts. Our combined approach showed a good efficiency to provide information both on SNVs and CNVs in a single one-step assay, offering new perspectives in the study of mosaic developmental disorder.

2449W

Diagnosing connective tissue disorders by clinical exome sequencing. *H. Cheng[1], X. Wang[1,2], M. Leduc[1,2], M. Walkiewicz[1,2], W. Bi[1,2], P. Liu[1,2], R. Xiao[1,2], L. Meng[1,2], F. Xia[1,2], Y. Yang[1,2], Z. Chen[1,2].* 1) Baylor Genetics, Houston, TX 77021; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030.

Connective tissue disorders (CTDs) mainly manifest in skin, eye, musculoskeletal, cardiovascular and neuron systems. Diagnosis by targeted gene approaches can be challenging due to late disease onset, variable phenotype expressivity, clinical and genetic heterogeneity and overlapping phenotypic features. This study retrospectively evaluated CTDs diagnosed by clinical exome sequencing in our lab including Ehlers–Danlos (EDS) syndrome related (including Cutis laxa), Osteogenesis Imperfecta (OI) related (including Hypophosphatasia, Cole-Carpenter syndrome), Marfan syndrome (MS) related (including Loeys-Dietz syndrome, Familial thoracic aortic aneurysms), and a newly described genetic disorder related to defects in the *ABL1* gene. Of the ~3000 positive clinical exome cases identified in our laboratory, 97 individuals were diagnosed with CTDs mentioned above. Of the CTD positive cases, 26 individuals were diagnosed with EDS related disorders. The genes most frequently identified were *COL5A1* (11.5%) and *PLOD1* (11.5%). Causative variant in *TNXB* was identified in one individual affected with hypermobility type EDS. Two individuals affected with vascular type EDS had mutations in *COL3A1*. Out of the 19 cases affected with OI related disorders, the genes frequently identified were *COL1A1* (26.3%) and *COL1A2* (15.8%). Causative variant in *COL1A2* was identified in one case with EDS features instead of OI. Out of the 46 individuals affected with MS related disorders, the genes frequently mutated were *FBN1* (21.7%) and *MED12* (13.0%). *De novo* variants in *SKI* were identified in two individuals with Shprintzen-Goldberg syndrome. All the *MED12* (X linked Lujan syndrome) variants were *de novo*; the variant identified in a female patient was reported in manifesting female carriers. In one family, two molecular diagnoses of Loeys-Dietz (*TGFB3*) and Stickler (*COL2A1*) syndromes were established, in which multiple family members were affected with either one or both diseases. Intriguingly, 8 individuals in 5 unrelated families were first reported in our laboratory to be affected by a new AD genetic disorder caused by gain of function variants in *ABL1*. The phenotypic features of the *ABL1*-related disorder resemble those in patients with Marfan syndrome. In summary, clinical exome sequencing in our laboratory identified a wide spectrum of genotypes in individuals with connective tissue disorders; in addition, it also enabled the discovery of a new disease gene related to CTDs.

2450T

Mendelian disorders of cornification in Iran: Spectrum of clinical and genetic characteristics in a cohort of 189 consanguineous families including a new entity of ichthyosis follicularis with profound bilateral deafness and leukonychia totalis. *L. Youssefian[1,2], H. Vahidnezhad[2,3], A. Saeidian[1], S. Sotoude[4], H. Mahmoudi[5], A. Touati[1], M. Barzgar[6], M. Daneshpazhooh[5], N. Aghazadeh[5], A. Azizpour[5], O. Zargari[7], M. Abiri[2,8], P. Mansouri[9], E. Londin[10], P. Fortina[11], S. Zeinali[2,8], T. White[12], J. Uitto[1].* 1) Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA;; 2) Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Iran; 3) Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran; 4) Department of Dermatology, Children's Medical Center, Center of Excellence, Tehran University of Medical Sciences, Tehran, Iran; 5) Department of Dermatology, Razi Hospital, Tehran University of Medical Sciences, Tehran, Iran; 6) Skin Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 7) Dana Clinic, Rasht, Iran; 8) Kawsar Human Genetics Research Center, Tehran, Iran; 9) Skin and Stem Cell Research Center, Tehran University of Medical Sciences, Tehran, Iran; 10) Computational Medicine Center, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania; 11) Sidney Kimmel Cancer Center, Department of Cancer Biology, Thomas Jefferson University, Philadelphia, Pennsylvania, USA; 12) Department of Physiology and Biophysics, Stony Brook University, Stony Brook, NY 11794, USA.

Heritable forms of ichthyoses, also referred to as generalized Mendelian disorders of cornification, are clinically diagnosed as dry and scaly skin, and histopathology reveals thickening of stratum corneum at the outer layer of the epidermis. In this study, we have developed a gene-targeted next generation sequencing (NGS) array consisting of 38 genes reported in association with ichthyosis phenotypes and sequenced total of 189 Iranian patients with clinical diagnosis of non-syndromic (autosomal recessive congenital) ichthyosis (ARCI), and syndromic forms of ichthyosIs. The designed library contained 38 genes (*ABCA12, ABHD5, AGPS, ALDH3A2, ALOX12B, ALOXE3, AP1S1, ARSE, CERS3, CLDN1, CYP4F22, EBP, ELOVL4, GJB2, GJB3, GJB4, GJB6, KRT1, KRT10, KRT2, KRT9, LIPN, LOR, NIPAL4, PEX7, PHYH, PNPLA1, PNPLA2, POMP, SLC27A4, SNAP29, SPINK5, ST14, STS, TGM1, TGM5, VPS33B, ZMPSTE24*) divided into 351 targets covered by 558 amplicon probes with an estimated coverage of 99%. NGS panel detected pathogenic and likely pathogenic variants in 155 out 189 families (82% detection rate). In 110 families, mutations were found in 8 associated ARCI genes: (*ABCA12*: 12.5%; *ALOX12B*: 9%, *ALOXE3*: 3%, *CERS3*: 8%, *CYP4F22*: 6.5%, *NIPAL4*: 4.5%, *PNPLA1*: 18%, and *TGM1*: 37%). Most common mutated gene in ARCI was *TGM1* (37%). Surprisingly, about 26% of all ARCI mutations were found in *PNPLA1* and *CERS3*; the reported variants in these genes are extremely rare. In 45 families, mutations were found in 18 genes associated with different syndromes in which ichthyosis is one of their clinical manifestations. We also characterized a new type of entity of ichthyosis follicularis with profound bilateral deafness and leukonychia totalis which segregate in 4 patients in a large multiplex family. All patients had compound heterozygous mutations (a missense and a deletion) in *GJB2*, and the parents of the proband were heterozygous. In Xenopus oocytes expressing the missense variant, membrane currents were elevated ~7 fold higher than wild-type *GJB2* expressing cells denoting increased hemichannel activity in mutant cells in comparison to control cells. Co-expression of both missense and wild-type *GJB2* reduced the magnitude of the current to about half between expressions of either protein separately, similar to of other *GJB2* associated skin disorders. Collectively, identification of specific mutations in a large cohort of ichthyosis patients allowed us to examine phenotype/genotype correlations.

2451F

Multigene next generation sequencing panel identifies pathogenic variants in patients with unknown subtype of epidermolysis bullosa: Subclassification with prognostic implications. *A. Saeidian[1], H. Vahidnezhad[1,2], L. Youssefian[1], A. Touati[1], S. Sotoudeh[3], M. Abiri[2,4], M. Barzgar[5], N. Aghazadeh[6], H. Mamoudi[6], S. Norouzzadeh[4], M. Hamid[2], H. Bagherian[4], S. Zeinali[4], P. Fortina[7], J. Uitto[1].* 1) Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA; 2) Molecular Medicine Department, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran;; 3) Department of Dermatology, Children's Medical Center, Center of Excellence, Tehran University of Medical Sciences, Tehran, Iran;; 4) Kawsar Human Genetics Research Center, Tehran, Iran;; 5) Skin Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran;; 6) Department of Dermatology, Razi Hospital, Tehran University of Medical Sciences, Tehran, Iran;; 7) Department of Cancer Biology, Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, Pennsylvania, USA;.

Purpose: Epidermolysis bullosa (EB), the prototype of heritable blistering diseases, is caused by mutations in genes expressed in cutaneous basement membrane zone. We have now evaluated the molecular basis of EB in 91 families, many of them of unknown subtype. **Methods:** A next generation sequencing panel (NGS) covering 21 distinct genes, including 18 EB-associated genes, was developed, and mutation profiles, together with clinical features, were used to classify the patients into distinct clinical subcategories. **Results:** A total of 72 pathogenic/likely pathogenic variants in 68 families were identified in 11 candidate genes, most of them (75%) being homozygous consistent with considerable consanguinity in the cohort. Half of the mutations (48.6%) were previously unreported. Refined analysis of the types of mutations with addition of variants of unknown significance suspected of causing EB allowed subclassification of patients into different subcategories of EB in 76 of 91 families, with prognostic implications. A genetically challenging case with homozygous loss-of-function pathogenic mutations in two different EB-associated genes resulting in different subtypes was identified. **Conclusion:** Utilization of next generation sequencing panel of EB-associated genes allowed diagnostic subclassification in the neonates particularly in families of unknown subtype, with prognostication of the overall long-term outcome of the disease.

2452W

Exome sequencing in 170 patients with diverse ataxia-related phenotypes identifies the genetic basis of disease in over 50%. *A. Knight Johnson[1], M. Sun[1], V. Nelakuditi[1], D. Fischer[1], K. Arndt[1], L. Ma[1], D. del Gaudio[1], M. Burmeister[3], V. Shakkottai[4], K. Boycott[5], J. Warman Chardon[6], D. Waggoner[1], C. Gomez[2], S. Das[1].* 1) Human Genetics, University of Chicago Genetic Services, Chicago, IL; 2) Department of Neurology, The University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 4) Neurology Department, University of Michigan Health System, Ann Arbor, MI; 5) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, ON; 6) Neurogenetics, Children's Hospital of Eastern Ontario, Ottawa, ON.

Ataxias are a group of neurological disorders that demonstrate significant clinical and genetic heterogeneity, and as such, exome sequencing provides an effective approach for the genetic diagnosis of this condition. We examined the utility of an exome sequencing-based approach for the molecular diagnosis of patients with a wide range of ataxia-related phenotypes. We studied 170 patients with ataxia of unknown genetic etiology referred from 63 different genetics or neurology clinics primarily within the United States and Canada, in whom prior trinucleotide repeat expansions had either been excluded or not implicated. Patients ranged in age from 2 years to 88 years. Exome-based target capture was performed using Agilent SureSelect technology and samples were sequenced on a NextSeq Illumina instrument, with an average coverage of 96% of all coding exons at minimum 30X read depth. Our analysis was focused on variants within 441 curated genes associated with ataxia and ataxia-like phenotypes as either an isolated finding, or as part of a more complex syndromic presentation. Combined, pathogenic variants and suspected diagnostic variants were identified in 88 of the 170 patients, providing a positive molecular diagnostic rate of 52%. Pathogenic and suspected diagnostic variants were observed in 57% of patients ≤30 years old (45/79), and 47% of patients >30 years old (43/91). Overall, forty-six different genes were implicated in the ataxia phenotype in our cohort. Pathogenic or suspected diagnostic variants in *SPG7*, *SYNE1*, *ADCK3*, *CACNA1A*, *ATP1A3* and *SPTBN2* together accounted for 41% of the positive cases. 33% of patients in whom a molecular diagnosis was made had a positive family history. Our strategy of analyzing a large set of ataxia-related genes in all referred patients provided a high diagnostic yield, and was particularly useful for making a molecular diagnosis in patients who had an atypical presentation for a particular disorder. In several cases the results from our cohort broadened the known clinical spectrum of genetic ataxia-related disorders. Our cohort represents the general population of patients with ataxia seen in neurology and genetics clinics, regardless of age or the presence or absence of additional clinical features. Our finding of pathogenic or suspected diagnostic variants in 52% of our cohort therefore represents the likely diagnostic yield of exome sequencing for unselected patients with ataxia referred for clinical exome sequencing.

2453T

Incidentaloma in neurogenetics: Pathogenic variant in NSD1 in a patient with spinocerebellar ataxia. *H.M. Velasco, D.A. Ramirez.* Genetic Institute, National University, Bogota Cundinamarca, Colombia.

Background: Genetic studies of late-onset sporadic ataxias (> 40 years) are not routinely indicated. In patients for whom the most common acquired causes are ruled out, diagnostic algorithms such as triplet expansions in genes related to spinocerebellar ataxia (SCA) can be used, and in the unresolved cases, next-generation sequencing (NGS), such as Whole Exome Secuencing (WES), are available for definitive diagnosis. **Case report:** Female patient with usual facial phenotype and anthropometry, who developed ataxia at 45 years of age, with no relevant family history, and initial clinical approach that ruled out common etiologies. Genetic assessment was performed at age 47 yielding negative studies for SCAs through triplet expansion (SCA 1,2,3,6). WES was performed, reporting a heterozygous pathogenic variant c.248delA (p.N83MfsX4) in the NSD1 gene (related to Sotos Syndrome), which was not associated with ataxia and is not related to the patient's phenotype. The variant was confirmed by complete sequencing of NSD1 in two different laboratories through Sanger sequencing. **Conclusion:** Findings obtained through NGS generally offer valuable results which serve to make medical decisions. However, finding incidental results that lead to define new clinical and bioethical actions is also possible. Is necessary confirm and establish the biological importance of this type genetic incidentaloma. This particular finding in NSD1 could demonstrate that there are clinically silent pathogenic variants or that these variants could be related to genopatogenic phenomena in ataxia.

2454F

Intracellular FmRPpolyG-HSP70 complex: Possible use as biochemical marker of FXTAS. *G. Bonapace[1], R. Gullace[1], D. Concolino[1], G. Arabia[2], A. Quattrone[2,3], R. Procopio[3], G. Iannello[2,3], M. Gagliardi[3], G. Annesi[3].* 1) Pediatrics, University Magna Graecia, Catanzaro, Italy; 2) Institute of Neurology, Department of Medical Sciences, University "Magna Graecia" of Catanzaro; 3) Institute of Molecular Bioimaging and Physiology, National Research Council, Section of Germaneto, Catanzaro, Italy.

   Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset neurodegenerative disorder that affects about 25% of carriers of small, non-coding CGG-repeat expansions (55–200 repeats) in the premutation range within the fragile X gene (*FMR1*). Main clinical features of FXTAS include intention tremor, cerebellar ataxia, and Parkinsonism. The pathogenesis of FXTAS has been demonstrated to involve different overlapping molecular and biochemical mechanisms, with a common feature due to an increased expression of FMR1 gene Recently, great emphasis on toxic aggregates, produced by a RAN translation process on the 5' expanded CGG region of the gene, has been given . These aggregates contain a fragile mental retardation protein with a polyglycin stretch on the aminoterminal ends, (FmRp-polyG) and so far, have been isolated and characterized only in drosophila models, mouse model and post mortem FXTAS patients brain, but never in FXTAS living patients**.** Unfortunately, the role of these aggregates in the pathogenesis of FXTAS is still not well understood. Furthermore, the syndrome is frequently misdiagnosed because of the early diagnosis in adult carriers remains difficult due to the lack of specific biochemical markers. The same situation hampers the possibility to predict the clinical outcome in children carriers of premutation on the *FMR1* gene, negative for the Fraxa syndrome, but at risk for developing FXTAS. Here we demonstrate, for the first time, using immunohistochemistry (IHC) and Western Blot procedure (WB), the presence *in vivo*, of toxic aggregates produced by the expanded polyG FMRP protein in fibroblasts from a carrier of a premutated FMR1 allele associated to clinical signs of FXTAS. This finding is very promising and offer the possibility to use the interaction between polyG-FMRP and HSP70 as novel early diagnostic biomarker of FXTAS in living patients. Even though the validation of this procedure needs larger research to study its predictive value for the clinical outcome of FMR1 premutation carriers without signs of FXTAS, we recommend considering its use for patients showing neurological signs of FXTAS associated to a well defined CGG premutated allele. .

2455W

The AAGAAAG duplication at nucleotides 2023-2029 of SCN8A gene of EIEE13 (early infantile epileptic encephalopathy-13) presenting with no epilepsy, but variable expression of intellectual disability, ADD/ADHD and autism in the same family. *M. Hajianpour[1], D. Lewis[2].* 1) Pediatrics/Medical Genetics, East Tennessee State University, Quillen College of Medicine, Johnson City, TN; 2) Pediatrics, Holston Medical Groups, Kingsport, TN.

   Proband is a 15-year-old male with a history of developmental delay, speech and language deficit, intellectual disability (ID) with FSIQ of 40, and ADHD. He was born full-term, by vaginal delivery to a 24-year-old, G1, P0 mother. No prenatal exposure to teratogens. BW; 8 lbs 1 oz, and BL; 21". He had mild motor delay. He did not talk in sentences until 3 years of age. He is in 9th grade special education but functions well below this level. He has autistic behavior and ADHD. The EEG and MRI for an episode of shivering were normal. He has had hypermetropia since age 5 years. Molecular studies for ID (a panel of 140 genes by AmbryGenetics) showed a pathogenic mutation of *SCN8A* **Gene on Chromosome 12q13 (paternally inherited): c.2023_2029dupAAGAAAG.** The test also showed two variants of uncertain significance (**VUSs**) of **VPS13B Gene: p.T1271S, p.S2596F (maternally inherited)**.   Family history is remarkable for LD in father (currently, has his own business), and 2 out of 4 brothers, one of whom has also ADD. Another brother has ADD only, and youngest brother (4 yo) is apparently normal. Testing for SCN8A mutation and VUSs on brothers are in progress.   Alterations in SCN8A gene are typically inherited in an autosomal dominant fashion in association with early infantile epileptic encephalopathy-13 (EIEE13); characterized by frequent tonic seizures or spasms beginning in infancy with specific EEG findings Additional features may include mild cerebral atrophy, progressive microcephaly, interrupted myelination on brain imaging (which are not present in proband and his father), regression of speech and language skills, developmental delay, intellectual disability, hypotonia, difficulties with coordination and balance, and autistic behaviors. Out of all these features, only DD/ID, motor coordination problems and autistic behavior are present in proband, and only history of LD in his father (who is a carrier of the same SCN8A alteration). Mother (with 2 VUSs) had LD requiring special education. She may also have a coagulation disorder with tendency to bleeding (as well as one of her sons). Contribution of maternally inherited 2 VUSs to the phenotype in proband is possible.   Trudeau et al. (2006) identified heterozygosity for a 2-bp deletion in SCN8A gene, in 4 members of a family; 3 had cognitive and behavioral deficits/ ADHD, while 1 had cerebellar atrophy, ataxia, and mental retardation, indication variable expression of the phenotype in members of the same family.

2456T

Further investigation of variants discovered in an early onset dementia cohort: Additional family member sequencing. *S.A. Bucks[1], C.J. Hale[2,3], D.S. Hanna[2,3], M.B. Dauer[2,3], L. Mathew[2,3], D. Sarroza[2,3], S. LaDuke[2,3], M.A. Weaver[2,3], T. Tran[2,3], M. Rumbaugh[1], D. Tsuang[1,4], S. Jayadev[1], T.D. Bird[1,4,6], M.O. Dorschner[2,3,5].* 1) Neurology, University of Washington, Seattle, WA; 2) Pathology, University of Washington, Seattle, WA; 3) UW Medicine Center for Precision Diagnostics, University of Washington, Seattle, WA; 4) Department of Medicine (Medical Genetics), University of Washington, Seattle, WA; 5) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 6) Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle, Washington.

Dementia affects an estimated 47 million people worldwide. It is caused by, or associated with, several neurological disorders, including Alzheimer's disease (AD), Parkinson's disease (PD), and frontotemporal dementia (FTD). Identifying the precise genetic cause of dementia contributes to accurate diagnosis, inclusion into appropriate clinical trials, potential discovery of novel gene targets for therapeutic development, and more informative genetic counseling for family members. To determine the genetic cause of dementia in patients, we previously developed a clinical exome-based panel using an optimized IDT Exome v1.0 capture system for library enrichment followed by Illumina sequencing. The panel included 38 genes associated with dementia phenotypes, including genes in AD, PD, FTD, and amyotrophic lateral sclerosis (ALS). The clinical panel was developed by experts engaged in neurodegeneration research and clinical practice at the University of Washington. We previously identified pathogenic or likely pathogenic genetic variants in 14 out of 145 cases (10%) of early onset and/or familial dementia, including in the genes *MAPT, NOTCH3, PSEN1,* and *PSEN2*. We also identified 33 variants of uncertain significance (VUS) in 29 cases. To further clarify the pathogenic nature of these VUSs, we then used Sanger sequencing to analyze samples from affected or unaffected family members of probands from the initial cohort. After Sanger sequencing the relatives of 12 probands with dementia, we identified three candidate variants that were also present in several affected individuals: *NPC1*, c.1778delA, p.Asn593Thrfs*7; *PSEN1*, c.322C>T, p.Arg108Trp; and *PSEN1*, c.122_128delGACGGAG, p.Arg41Thrfs*31. Remaining families that showed no clear aggregation of the proband's VUS with the dementia phenotype are undergoing whole exome analysis to identify candidate variants that may contribute to the family's dementia. Exome sequence data for members of a single family will be compared to eliminate private, rare benign polymorphisms and enrich for variants carried by more than one affected individual. Retrospective analysis of our procedure (dementia panel, Sanger sequencing in relatives, and exome analysis) should provide insight into which steps are necessary to determine the genetic cause of dementia in most patients suspected of having single gene drivers of dementing disease.

2457F

Comparison of the diagnostic yield of multi-gene panels for neuromuscular disorders. *A. Gruber, K. Williams, J. Schroeder.* PreventionGenetics, Marshfield, WI.

Neuromuscular disorders (NMDs) are a clinically, pathologically, and genetically heterogeneous group of diseases that impair muscle function or the peripheral nervous system. With the reduction in NextGeneration Sequencing (NGS) costs, genetic testing has become a cost efficient approach for NMD clinical diagnosis. A patient cohort of 950 patients with a wide range of NMDs was evaluated for pathogenic variants using NGS. Thirteen NMD panels were retrospectively analyzed and ranged from comprehensive (124 genes) to phenotype-specific (6 genes). A positive result was reported in patients with one or more pathogenic or likely pathogenic variants based on ACMG variant interpretation guidelines. The highest performing panel was a custom muscular dystrophy panel (57 genes, 10 patients) in which all patients were thoroughly evaluated with muscle biopsy results from a single provider. For this panel, 60% of patients were positive, 40% were inconclusive, and there were on average 2.3 VUSs/patient. The nemaline myopathy panel (10 genes, 58 patients) had a 47% positive rate, 22% inconclusive rate and 0.7 VUS/patient. Other panels which include limb-girdle muscular dystrophy (24 genes, 236 patients), congenital muscular dystrophy (27 genes, 104 patients), and congenital myasthenic syndrome (15 genes, 265 patients) were positive in 20%-30% of cases and averaged less than one VUS/patient. Two recently released panels, including a comprehensive neuromuscular panel (124 genes, 5 patients) and comprehensive neuropathy panel (71 genes, 7 patients), were positive in 20%-30% of cases but had 2.5-3 VUS/patient. In conclusion, larger multigene panels have become a cost effective tool in clinical diagnosis of NMDs. Smaller phenotype based panels are less likely to result in variants of uncertain significance, which are a burden for providers. Due to the clinical and genetic heterogeneity of NMDs, large multigene panels are now commonly offered as a diagnostic test. However, they do come at the cost of additional VUSs being identified due to the large number of genes being analyzed.

2458W

Identification of a novel *de novo* nonsense mutation of the *NSD1* gene in monozygotic twins discordant for Sotos syndrome. *J. Han, S. Her.* Pediatric Neurology, Daejeon St. Mary's Hospital , Daejeon, South Korea.

   Sotos syndrome is a congenital overgrowth disorder characterized by facial gestalt, excessively rapid growth, acromegalic features and a non-progressive cerebral disorder with intellectual disability. We used diagnostic exome sequencing (DES) to identify a heterozygous *de novo* mutation of the *NSD1* gene in monozygotic twins with Sotos syndrome. DES revealed a novel nonsense mutation c.2596G>T (*p.Glu866**) of the *NSD1* gene in the proband, the first of monozygotic twins. Sanger sequencing analysis of the proband and his family members showed that this nonsense mutation was present in the proband and his sibling, but was absent in their parents, indicating that it occurred with *de novo* origin. Interestingly, the identical male twins showed somewhat different clinical, cognitive and behavioural phenotypes. Abnormal clinical manifestations including seizures, scoliosis, enlarged ventricles, and attention-deficit/hyperactivity disorder (ADHD) were found in the proband (first twin), but not in the sibling (second twin). This finding expands the phenotypic spectrum associated with variable expression of the Sotos syndrome caused by *NSD1* mutation, and it adds further support for postconceptual mutation, epigenetic change and/or an environmental factor involved in the cause of the Sotos syndrome. We demonstrate that DES is a valuable approach to identify a genetic cause in sporadic cases of Sotos syndrome due to inherited *de novo* mutations.

2459T

Novel myopathic phenotype due to a newly detected stop-loss mutation in *MYH7* gene. *K. Sumegi[1,2], Z. Banfai[1,2], K. Hadzsiev[1,2], E. Pal[3], E. Kovesdi[1,2], B. Melegh[1,2].* 1) Department of Medical Genetics, University of Pecs, Pecs, Hungary; 2) Szentagothai Research Centre, University of Pecs, Pecs, Hungary; 3) Neurology Clinic, University of Pecs, Pecs, Hungary.

  Purpose: Defects of the slow myosin heavy chain (MyHCI) encoding *MYH7* gene result predominantly in cardiomyopathies and skeletal myopathies including distal myopathy and skeletal muscle abnormalities caused by thick filament deposition in the sarcomeres. However, according to the reports from recent years, phenotypic characteristics of myopathies arising from *MYH7* gene defects have a rather wide spectrum. Pathogenesis of the C-terminal mutations of *MYH7* gene are less known, only two article deals with the phenotypic impact of the elongated mature protein product due to termination signal loss. Here we present a male patient with an unusual phenotypic variant of early-onset and predominant involvement of neck muscles, with muscle biopsy indicating myopathy and sarcoplasmic storage material. **Methods:** Our investigation consisted of standard clinical examination procedures, including electromyography, electroneuronography, MRI, histopathological (Hematoxylin-Eosin, modified Gomori and NADH staining, electron microscopy) and laboratory examinations. Our molecular genetic investigation included the diagnostic Sanger sequencing of *MYH7* gene. We also used various prediction software (Mutation Taster, Human Splice Finder 3.0, Netgene2) in order to characterize the clinical significance of the detected mutations. We also applied the Parcoil2 software in order to investigate probable structural changes in the rod region of the protein due to the identified mutation. **Results:** Sequencing of *MYH7* revealed three novel point mutations at the 3'-end of the gene, one of them (c.5807A/T) causing the loss of termination signal and elongation of the mature protein. Parcoil2 result showed a probability score approaching 1 for the likelihood of coiled-coil formation of the elongated region. The elongated protein likely disrupts the functions of the sarcomere by multiple functional abnormalities, resulting in severe myopathy of certain axial muscles and leading to the defect of thick filament degradation, causing protein deposition and accumulation in the sarcomeres. **Conclusion:** The number of currently identified mutations causing termination signal loss in *MYH7* are few, their impact on the structure of MyHCI is unique and do not show a uniform phenotype. Examination of *MYH7* mutations is essential to improve our knowledge about congenital myopathies, classification of clinical patterns and the diagnosis of currently unidentified cases.

2460F

Whole exome sequencing: An effective and comprehensive genetic testing approach for leukodystrophy. *F. Zou, K. Retterer, J. Scuffins, D. McKnight.* GeneDX, Gaithersburg, MD.

Leukodystrophies are a group of rare genetic disorders caused by abnormal development or destruction of the white matter of the central nervous system (CNS) with or without peripheral nervous system involvement. The diagnosis of leukodystrophies is challenging due to heterogeneous phenotypes, although early diagnosis is critical for prognosis and family planning and in some cases to take advantage of available treatments to slow disease progression. The objective of this study was to establish the positive diagnostic rate (PDR) for leukodystrophies by whole exome sequencing (WES) and summarize a variety of underlying genetic etiologies. In this retrospective study, we reviewed the results of clinical WES testing of 541 patients with leukodystrophy reported as a clinical feature. A positive test outcome was defined as presence of one or two pathogenic or likely pathogenic variants in a single gene, depending on the mode of inheritance. In 33% (176/541) of cases, WES yielded a positive result, 44% (242/541) of cases were reported as non-diagnostic (inconclusive), and 23% (123/541) of cases were negative. Approximately 51% (90/176) of patients had an autosomal recessive (AR) disorder, 41% (72/176) had an autosomal dominant (AD) disorder, and 8% (14/176) had an X-linked (XL) disorder. WES-trio testing with concurrent sequencing and analysis of probands and both parents revealed that the majority of patients diagnosed with an AD disorder (81%, 58/72) had *de novo* variants. The PDR for WES-trios was 34% (136/399), which was significantly higher than the 23% PDR (20/86, p < 0.05) for proband-only WES. Overall, pathogenic or likely pathogenic variants were reported in 133 different genes and most frequently seen in well-known leukodystrophy genes including TUBB4A (3.4%, 6/176), RNASEH2B (2.8%, 5/176), GFAP (2.8%, 5/176) and EIF2B5 (1.7%, 3/176) and POLR3A (1.7%, 3/176). More interestingly, over 8% of these genes were only described in connection with a specific disease within the last two years. In summary, WES is an effective diagnostic tool that can identify the underlying genetic etiology of leukodystrophy in up to 34% of patients. The WES-trio approach which is more efficient than proband-only WES is recommended to enable segregation analysis and aid in variant interpretation. These results, along with the growing number of newly described leukodystrophy genes, emphasize the importance of a comprehensive and flexible genetic testing approach for leukodystrophy.

2461W

Copy number analysis using next-generation sequencing: Comprehensive genetic testing and its application to neuromuscular and epilepsy panels. *A. Entezam, J. Westbrook, C. Tan, S. Yang, R. Truty, N. Patel, P. Taylor, M. Rabideau, H. White, T. Winder, S. Aradhya.* Invitae, San Francisco, CA.

Objective The diagnosis of Mendelian disorders is routinely performed through gene sequence or chromosomal copy number variant (CNV) analysis. Although exon-level CNVs contribute significantly to disease burden, their testing has traditionally been limited to a few loci. We evaluated a next-generation sequencing (NGS) method that simultaneously evaluates sequence changes and gene-level CNVs in hundreds of genes. Our data show that a high frequency of gene and exonic CNVs are associated with neurological disorders. Methods We investigated subsets of 300 genes curated for neuromuscular disorders in 2316 unrelated individuals and up to 186 genes curated for epilepsy in 1131 unrelated individuals. Validated coverage-based CNV detection algorithms and custom algorithms designed to flag split-read signals were applied to all samples. Results Among individuals with neuromuscular disorders, 247 had pathogenic variants detected with CNV analysis. These variants represented 39% of all neuromuscular diagnostic findings. The majority of pathogenic CNVs were in PMP22 or DMD, but they were also identified in several other genes not traditionally tested for CNVs. Among individuals with epilepsy, 44 had whole-gene or exonic CNVs in 23 genes, representing 14% of the pathogenic variants in epilepsy genes. Conclusions Exonic CNVs explain a substantial (~10%) proportion of pathogenic variants across several disorders and occur in a broader variety of genes than previously appreciated. This NGS-based sequence and CNV detection method can be used routinely in germline genetic testing in neurological disorders to identify both common and rare events that together result in a high diagnostic yield. .

2462T

Combining repeat expansion testing with phenotype based NGS panels provides significant diagnostic benefit. *H.A. Marton, D.P. Kolev, W.A. Langley, P.L. Nagy.* MNG Laboratories, Atlanta, GA., US.

Phenotype-based Next Generation Sequencing (NGS) panels are a powerful tool for clinical diagnostics and have made variant calling cheaper, faster, and more accessible. The availability of NGS panels has increased over the past few years to cover a range of conditions including inherited and rare diseases. A number of diseases assessed by NGS panels are known to have multiple genotypic causes, ranging from single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) to repeat expansions. While NGS panels can precisely identify SNPs and CNVs, due to the current limitations of sequencing technology, NGS cannot accurately detect large repeat expansions. To date, more than 20 diseases have been found to be caused by repeat expansions, including Huntington's, Friedreich's Ataxia (FRDA), myotonic dystrophy, spinocerebellar ataxia (SCA), amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), Fragile X, and others. Research has shown that repeat expansions are responsible for as many as 1 in 3 cases of familial ALS and 1 in 4 cases of familial FTD. For other diseases, such as the rarer SCAs, the prevalence in the general population is currently unknown. By including repeat expansion testing with phenotype-based NGS panels, we can increase the clinical diagnostic sensitivity of the NGS panels. We currently perform repeat expansion testing for FRDA, C9ORF72, HTT, eleven SCAs, and Fragile X. Initial findings suggest that repeat expansion testing in conjunction with our neuromuscular and movement disorder panels increases positive reports by 5-20%.

2463F

Diagnostic yield for neurological and neuromuscular disorder testing via high-depth multi-gene panel analysis with integrated sequence and copy number detection. *T. Winder, A. Entezam, S.P. Taylor, H. White, M. Rabideau, S. Yang, C. Tan, J. Westbrook, S. Aradhya.* Invitae Corporation, San Francisco, CA.

Objective The objective of this study was to evaluate diagnostic yield using systematically curated multi-gene panels and simultaneous sequence and exonic copy number variant (CNV) detection. Methods Using high-depth next-generation sequencing (NGS), we investigated subsets of approximately 300 genes curated into ten disease-specific, multi-gene panels in 2,316 unrelated individuals diagnosed with neurological or neuromuscular disorders. Each gene in every panel was evaluated with our validated custom-built algorithms that use depth-of-coverage information and split-read detection to identify CNVs, small and large indels, and single nucleotide changes. Results We observed clinically reportable variants in 727 individuals. A definitive molecular diagnosis was reached for 27.9% of the 2,316 patients, and another significant percentage had results that would likely reach clinical significance with additional evidence such as observation of de novo occurrence or compound heterozygosity in trans. Whole-gene or exonic CNVs composed 39% of all diagnostic findings, two-thirds of which were identified in PMP22 or DMD. The highest observed diagnostic yield was obtained in DMD testing (65.6%), in which 66% of the findings were CNVs. The most frequently observed pathogenic variants were PMP22 CNVs. Diagnostic yields for muscular dystrophy, myopathy, neuropathy, and comprehensive neuromuscular disorders were 43.7%, 14.6%, 19.5%, and 19.1%, respectively. Conclusions Pre-curated multi-gene panels and our custom NGS-based system enable simultaneous sequence and CNV calling and, owing to high yield, are appropriate for diagnosing neurological conditions before exome sequencing.

2464W

Millder-Dieker syndrome: Clinical, radiological, and molecular characterization. *A. Cortes, N. Laza, A. Latorre, C. Silvera, P. Garavito.* Universidad del Norte, Barranquilla, Atlantico, Colombia.

Background: Miller-Dieker syndrome (MDS) (OMIM 247200) is caused by a microdeletion in chromosome band 17p13.3, that involves the genes *PAFAH1B1 (LIS1)*, *YWHAE, HIC1, PRPF8*, among others. MDS has a prevalence of 1/100.000 live births. These patients present brain anomalies, like convulsions and cognitive decline; also, facial dysmorphisms, like prominence in the frontal and occipital regions, small nose, anteverted nostrils, low set ears, prominent lips, micrognathia, and heart and genitourinary anomalies. The purpose of this case report is to present the clinical manifestations suggestive of a patient with MDS and its molecular studies for confirmation. Case report: 8-month-old female patient referred from Neuropediatrics to the Genetic clinic because of epilepsy, hypotonia, and global developmental delay; she was the first pregnancy full term of parents age-appropriate. On physical examination, she presented generalized hypotonia, fixed gaze, anteverted nostrils, carp-shaped mouth, micrognathia, large ears with mild posterior rotation, long neck, and cutaneous syndactyly between second and third toes. Brain MRI reported loss of cortical gyrus, lissencephaly, cortical thickening, and CT brain scan did not show brain calcification. Chromosomal Microarray Analysis demonstrated a 657kb microdeletion of 17p13.3. This region included 4 OMIM genes (*RFA1, R4RL1, DPH1, HIC1, and SMG6*). Discussion: MDS is a low prevalence disease that should be considered in patients with lissencephaly associated with facial dysmorphism and consider in the differential diagnosis the isolated lissencephaly and the microcephalic osteodysplastic primordial dwarfism. Our case did not show the typical brain calcifications described in the literature.

2465T

Clinical application of whole exome sequencing in patients with uncertain neurological disorders. *Y. Lee[1], J.S. Lee[2,3], B.C. Lim[2], Y. Yoo[1], T.K. Yoo[1], S. Lee[1], J. Seo[1], M. Lee[1], J.M. Ko[2], M. Choi[1,2], J.H. Chae[2].* 1) Department of biomedical sciences, Seoul National University College of Medicine, Seoul, Republic of Korea; 2) Department of Pediatrics, Seoul National University College of Medicine, Seoul National University Children's Hospital, Seoul, Republic of Korea; 3) Department of Pediatrics, Gachon University Gil Medical Center, Incheon, Republic of Korea.

Inherent complexity of brain developmental processes inevitably generates patients with neurological defects of varying clinical outputs that frequently challenge conventional diagnostic criteria. Therefore, accurate diagnosis of the patients with unconventional or complicated pediatric neurological defects demands well-coordinated combination of robust genetic analytic ability and delicate clinical evaluation. Using a trio-based whole exome sequencing approach, we analyzed 310 patients with a broad spectrum of neurological disorders without any pathogenic variants revealed by the conventional methods. Overall, we identified disease-causing variants in 54.5% of the probands and 64.5% of them were previously reported as pathogenic. We also discovered 60 potentially pathogenic variants that require further functional assessments (19.4%). These genes containing novel variants are also enriched with the neuron part gene ontology classes (adjusted $P = 0.0007$), and coexpression network analysis using developing human brain transcriptome data revealed that these genes are more tightly connected to the known disease associated genes than expected by chance (the best $P = 0.0003$), providing support that these novel genes might have roles in human brain development. Furthermore, we introduce a few clinical vignettes that well illustrate potential diagnostic pitfalls that one could have encountered without this approach. Our results highlight the heterogeneity of clinical and genetic profiles of pediatric neurology patients and propose utility of whole exome sequencing for improved patient care and discovery of novel pathophysiology mechanisms.

2466F

Molecular diagnosis of Colombian patients with myopathies through next generation sequencing panel. *R. Garcia-Robles[1], P.A. Ayala-Ramirez[1], T. Pineda[2], M. Garcia-Acero[1], M. Guerra[1], F. Suarez-Obando[1]. 1) Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Biotecgen, Bogota, Colombia.*

Introduction: Myopathies are a group of disorders characterized by muscle weakness, with a high clinical overlap that makes the diagnosis tough to establish among all the phenotypes. Recently, some new tests based on NGS (next generation sequencing) have been used to detect the molecular alteration and allow an adequate diagnosis of patients. Additionally, the molecular characterization of the disease is necessary for genetic counseling, prognosis, and management. Methodology: Since January of 2014 to December of 2016 we assessed 50 patients with clinical suspicion of muscular dystrophy using the Invitae Comprehensive Muscular Dystrophy Panel. This panel analyzes sequence changes and exonic deletions/duplications in 55 genes associated with muscular dystrophies. The panel includes: *ANO5, B3GALNT2, B4GAT1, CAPN3, CAV3, CHKB, COL6A1, COL6A2, COL6A3, DAG1, DES, DMD, DNA-JB6, DPM1, DPM2, DPM3, DYSF, EMD, FHL1, FKRP, FKTN, GAA, GMPPB, ISPD, ITGA7, LAMA2, LARGE, LMNA, MYOT, PLEC, PNPLA2, POMGNT1, POMGNT2, POMK, POMT1, POMT2, SGCA, SGCB, SGCD, SGCG, SYNE1, TCAP, TMEM43, TMEM5, TNPO3, TOR1AIP1, TRAPPC11, TRIM32, TTN.* Results: We analyzed 51 patients clinically diagnosis with myopathy/muscular dystrophy (range of age: 2 to 61 years). In 4 patients the diagnosis was not be established. The 3 most frequent diagnoses were: Duchene/Becker muscular dystrophy (58%), Titinopathy (8%) and Bethlem myopathy related to COL6A1 (6%). Conclusions: Molecular diagnosis based on NGS panels increase the possibility of precise diagnose of muscular diseases, (myopathies and muscular dystrophies), where the phenotype is highly overlapped. The increased use of NGS, in the diagnostic of specific groups of diseases, allows the improvement of its clinical utility, clarify its indications, and limitations, and gives to the physician, a highly useful tool for the improvement of genetic counseling and treatments based on accurate diagnosis. We used the Invitae Comprehensive Muscular Dystrophy Panel in 50 individuals with clinical suspicion of muscular dystrophy, reaching specific diagnostic rate in 92% of the cases.

2467W

A novel approach distinguishing the *SMN1* and *SMN2* genes in spinal muscular atrophy (SMA) using a linked-read NGS custom panel. *R. Pellegrino[1], C. Kao[1], M. Gonzalez[1], F. Mafra[1], J. Garifallou[1], C. Kaminski[1], L. Tian[1], S. Garcia[2], O. Thomas Mueller[3]. 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) 10XGenomics, Pleasanton, CA; 3) Department of Pathology and Laboratory Medicine Johns Hopkins All Children's Hospital, FL.*

SMA types I-IV are caused by loss of both copies of the survival motor neuron 1 (*SMN1*) gene in 5q13. Expression of the SMN protein is essential for motor neuron survival. Most functional SMN protein is produced by *SMN1*, with a small contribution from a paralog, *SMN2*. Resolving *SMN1/SMN2* by sequencing is challenging using short read NGS due the unstable high homology regions containing a variety of pseudogenes and repetitive elements subject to "*de novo*" rearrangements. Sequencing methods reliant on short reads have difficulty defining large structural variations and repetitive gene sequences. Advances in microfluidics technology and precision reagent delivery allow long-range information (short reads) to be rescued and preserved through the use of the 10x Genomics Chromium platform. Each input DNA fragment (~40-200kb) is partitioned into a gel-bead in emulsion (GEM). This approach can be extremely powerful to define breakpoints, large rearrangements and "dark-matter" regions such as the one involving the SMA genes. **METHODS** We sequenced subjects with clinically diagnosis of SMA, using the Chromium technology (10X Genomics) to generate barcoded reads that identify segments that are immediately contiguous to each other in the genome. High Molecular Weight (HMW) DNA was isolated and QC performed with Genomic TapeStation to verify the size of the DNA (where all samples achieved the mandatory 60Kb). In the microfluidic Genome Chip, a library of Genome Gel Beads is combined with an optimal amount of HMW template gDNA in Master Mix and partitioning oil to create Gel Bead-In-Emulsions (GEMs). The barcoded libraries then underwent a special custom baits design (target enrichment) using an expanded prototype OneSeq bait set (Agilent SureSelect) containing "bridging baits" spaced within intronic and intergenic regions to preserve long-range information following capture and Illumina sequencing **RESULTS** We use linked read technology coupled to whole exome sequencing to differentiate between single nucleotide variants, indels, and larger structural variants residing in the *SMN1* vs *SMN2*. This was done through accurately phasing of the compound mutations in the complex region for all cases.

2468T

Detection rate of chromosomal microarray in individuals with ADD/ADHD. *A.L. Baxter, A. Ortega, C. Bracken, E.R. Harward.* Lineagen, Inc., Salt Lake City, UT.

Chromosomal microarray (CMA) is a guideline recommended first-tier genetic test for individuals with autism spectrum disorder (ASD), developmental delay (DD), intellectual disability (ID), or multiple congenital anomalies (MCA). It is often used for assessing the underlying genetic cause in other childhood disorders, although the utility is not as well established. Attention deficit disorder (ADD) and attention deficit hyperactivity disorder (ADHD) affect 11% of children in the United States and are frequently part of the phenotype of individuals tested by CMA. Additionally, over 150 genetic conditions have ADD/ADHD as a documented clinical feature. However, few studies have assessed the utility of genetic testing for individuals with ADD/ADHD. We retrospectively reviewed the ICD codes and medical records submitted for individuals who had high-resolution CMA through our laboratory between July 2012 and April 2017 to identify those with ADD/ADHD. A total of 1,891 individuals with ADD/ADHD were identified and, of those, 8.4% (159) had an abnormal result. There was a subset of 165 individuals with ADD/ADHD who had other clinical features but did not have ASD, DD, ID, or MCA (typical first-tier clinical indications); 7.3% (12) of that group had an abnormal result. Fifteen individuals had only ADD/ADHD listed as the clinical indication for testing and 13.3% (2) of them had an abnormal result. These findings demonstrate that ADD/ADHD has a similar CMA detection rate as typical first-tier indications. Therefore, CMA should be considered in the evaluation of an individual with ADD/ADHD.

2469F

Repeat expansions that cause central nervous system disorders present challenges for long-read sequencing technologies. *M.T.W. Ebbert, S.L. Farrugia, K. Jansen-West, M. Prudencio, M. van Blitterswijk, R. Rademakers, L. Petrucelli, J.D. Fryer.* Neuroscience, Mayo Clinic, Jacksonville, FL.

Background. Structural variants are the greatest source of variation in the human genome, affecting an estimated 13% of an individual's genome, and long-read sequencing technologies such as PacBio and Oxford Nanopore Technologies (ONT) provide an exciting opportunity to understand how an individual's genome structure affects her health. Many types of structural variants exist, including insertions and deletions (INDELs), and translocations, amongst others, and researchers are discovering how structural mutations cause several diseases, including cancer, autism, and schizophrenia. Structural mutations like INDELs are commonly reported, but repeat expansions are a special class that has received little attention. Repeat expansions are an insertion where a given base nucleotide sequence (e.g., CAG) is expanded to be repeated hundreds to thousands of times, for unknown reasons. Numerous CNS diseases, like Huntington's disease, spinocerebellar ataxias (SCA's), frontotemporal dementia (FTD), and Amyotrophic Lateral Sclerosis (ALS) are caused by repeat expansions, many of which are GC-rich, making them difficult to sequence. We tested whether PacBio and ONT sequencing platforms can accurately sequence repeat expansions for SCA36 (GGCCTG in *NOP56*), Fuch's disease (TGC in *TCF4*), and ALS/FTD (GGGGCC in C9orf72). **Methods.** We cloned specified repeats into 5 plasmids of known size: (1) SCA36 (66 repeats); (2) Fuch's disease (220 repeats); (3) ALS/FTD (423 repeats); (4) ALS/FTD (774 repeats); and (5) a control plasmid with no repeats. All plasmids were digested with a single-cut restriction enzyme and sequenced on both the PacBio RS-II and the ONT's MinION. Plasmids were verified for correct size on a gel. **Results.** All repeat plasmids presented significant challenges to both sequencing technologies, while both platforms performed seamlessly on the control plasmid. Both platforms sequenced the primary plasmid backbones without issue until they reached the given repeat region, where coverage dropped dramatically. Particularly interesting was the number of reads where the sequencers appeared to read "through" the repeat, but were missing significant portions of the repeat sequence. i.e., reads contained sequence on both sides of the repeat region, but the number of repeats present in the final sequence was significantly lower than the true number of repeats. **Conclusion.** While the repeats presented significant challenges, these can be overcome through informatic improvements.

2470W

Attacking a VUS from multiple angles: An integrated and functional approach for reclassifying variants of uncertain significance. *R.N.T. Lassiter, J. Wang, E.C. Weltmer, J. Tinker, H. Newman, T. Carter, S. Ichikawa, A. Chamberlin, I. Petrik, R. Karam, T. Pesaran.* Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA.

Neurodevelopmental disorders (NDDs) involve a wide range of symptoms and severity. Individuals may present with multiple indications including epilepsy, autism, intellectual disability, and developmental delay. The nonspecific nature of NDDs can lead to significant diagnostic challenges. Multi-gene panel testing (MGPT) can help to stratify NDDs and identify causative variants. Though large MGPT allows for the molecular diagnosis of many possible rare diseases, they can also result in numerous variants of uncertain significance (VUS). High VUS rates can be confusing and discouraging to both patients and clinicians. Here we describe three effective approaches for reducing VUS rates in neurology MGPT. Combining parental co-segregation studies, protein structural analysis, and RNA functional splicing assays with additional strong lines of evidence, led to a significant decrease in VUS rates. Parental co-segregation studies were performed following initial results of the proband's neurology MGPT. Parental samples were tested for all VUS detected in proband MGPT in autosomal dominant (AD) and X-linked (XL) genes. Of the 200 VUS submitted for parental co-segregation studies, approximately 5% (10/200) were upgraded to variant likely pathogenic (VLP), 35% (70/200) were downgraded to variant likely benign VLB, and 60% (120/200) remained VUS. These data show that testing of parental samples is informative in ~40% of VUS. Protein structural analysis can be leveraged to identify and explain the effects of a variant on protein function. The structure based assessment strategy utilized for classification consists of several complementary tools (Rosetta, FoldX, ELM, etc.) to address the distinct structural impact of a variant. Of the 49 VUS with a structural assessment, 96% contributed to reclassification. Approximately 78% (38/49) were upgraded to likely pathogenic VLP, 18% (9/49) were downgraded to likely benign VLB, and 4% (2/49) remained VUS. RNA functional studies, designed to identify alternative splicing, resulted in reclassification from VUS to likely benign VLB in 100% (3/3) of splicing cases. These data show that a comprehensive approach to variant reclassification is an effective method for reducing VUS rates in neurology MGPT.

2471T

Parental variant study is informative for variant classification in significant number of neurodevelopment genes. *E.C. Weltmer, R. Lassiter, S. Ichikawa, T. Pesaran, J. Tinker, H. Newman, T. Carter, N. Ghahramani, L. Guidugli, J. Wang.* Ambry Genetics, Aliso Viejo, CA.

Parental variant study (PVS) can provide powerful segregation data for variant classification. However, in genes with reduced penetrance, variable expressivity, or a non-specific associated phenotype, common in neurodevelopmental disorders, PVS may be less powerful. We sought to identify characteristics common to genes for which PVS was informative for variant classification. Retrospective analysis of 100 cases for which PVS was completed July 2016-March 2017 for variants of unknown significance (VUS) found on multigene panels for neurodevelopmental disorders. Parental samples were received after the proband report was issued. VUS in autosomal dominant (AD) and X-linked (XL) genes were eligible for PVS. Variants were classified using a 5-tier system based on established algorithms. Genes in which PVS was/was not informative were compared. 179 unique VUS were found in 84 AD or XL genes. Seven VUS (4%) were upgraded to likely pathogenic (VLP), 68/179 (38%) downgraded to likely benign (VLB), 104 (58%) remained VUS. The 7 upgraded VUS were found in 5 genes: *ATP1A2, CHD2, CREBBP, KCNQ2, PCDH19*; all due to *de novo* event in proband. 52/68 (76%) downgraded VUS were due to informative PVS (found in asymptomatic parent); these 52 VUS were found in 33 genes. Of these 33 genes, 5 (15%) had ≥3 unique VUS downgraded per gene: *CHD8, EHMT1, FOXP1, KCNT1, PACS1*; 28 had 1-2 downgraded VUS per gene. PVS was informative for 38/104 (38%) VUS in 29 genes but VUS were not reclassified due to lack of additional supportive evidence (parental inheritance alone insufficient). PVS was not informative for 44/104 (42%) VUS found in 15 genes with multiple inheritance pattern, incomplete penetrance, variable expressivity, or non-specific disease phenotype (e.g., developmental delay, epilepsy). PVS was also not informative for 14/104 (13%) VUS inherited from a symptomatic parent, and 8 (8%) with only one parent tested. PVS was informative for 45% (38/84) of genes with VUS and 19% (38/196) of all genes analyzed. PVS was most likely to be informative for genes associated with a well described syndrome, e.g., CHARGE, Kleefstra, Rubinstein-Taybi syndrome, or severe, early-onset phenotype, e.g., infantile epileptic encephalopathy. Understanding the types of genes for which PVS is likely to be informative may be useful for clinicians coordinating parental testing. Further understanding of the genes themselves may allow for effective variant classification using evidence not limited to PVS.

2472F

Neurogenetic disease study in Mali reveals novel mutations and suggests new disease-causing genes. *G. Landoure[1,2,3], O. Samassékou[3], B.G. Burnett[4], L. Cissé[1], D. Dramane[5], S.P. Coulibaly[3], T. Coulibaly[1], S. Diallo[6], S.H. Diallo[7], K. Bagayoko[1], N. Bocoum[1], S. Diarra[1], H.O. Ba[8], K-L. Chen[2], S. Coulibaly[9], T. Coulibaly[1], M. Karambé[1], M. Keita[10], B. Koumaré[9], Y. Maiga[7], M. Ouologuem[11], A.B. Schindler[2], A. Simaga[12], A.S. Sissoko[1], H. Stanescu[13], C.J. Sumner[14], A. Touré[15], M. Traoré[16], R. Kleta[13], K.H. Fischbeck[2], The H3Africa Consortium.* 1) Service de Neurologie, CHU du Point G, Bamako, Bamako, Mali; 2) Neurogenetics Branch, NINDS, NIH, Bethesda, MD; 3) Faculté de Médecine et d'Odontostomatologie, USTTB, Bamako, Mali; 4) Department of Anatomy, Physiology and Genetics, USUHS, Bethesda, MD; 5) Service de Médecine, Hôpital Mère-Enfant "Le Luxembourg", Bamako, Mali; 6) Service de Médecine, Hôpital Fousseyni Dao, Kayes, Mali; 7) Service de Neurologie, CHU de Gabriel Touré, Bamako, Mali; 8) Service de Cardiologie, CHU de Gabriel Touré, Bamako, Mali; 9) Service de Psychiatrie, CHU du Point "G", Bamako, Mali; 10) Service d'ORL, CHU de Gabriel Touré, Bamako, Mali; 11) Service de Médecine, Hôpital du Mali, Bamako, Mali; 12) Institut d'Ophtalmologie Tropicale de l'Afrique, Bamako, Mali; 13) Centre for Nephrology, UCL, London, UK; 14) Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD; 15) Service de Pédiatrie, CHU de Gabriel Touré, Bamako, Mali; 16) Service de Cytogénétique et de Biologie Reproductive, INRSP, Bamako, Mali.

   Hereditary neurological diseases (HNDs) are debilitating and often untreatable disorders caused by mutations in many known genes; however the genetic basis of many others is still not established. In addition, despite its genetic diversity, studies of hereditary neurological disorders in the African population have been limited. We have aimed to characterize families with HND and to identify the underlying genetic defects. Patients have undergone neurological examination, and laboratory evaluation was performed to consolidate our diagnoses or exclude common causes. For genetic analysis, candidate gene analysis was done in some families and exome sequencing coupled with linkage analysis was performed in other cases. To date, 587 participants from 141 families have been enrolled. The mean age was 26.9 years and sex ratio was 1.18 (male:female). Consanguinity was reported by 40% of families, and a recessive inheritance pattern was found in 34%. Major clinical entities included ataxia, muscle disease, hereditary spastic paraplegia (HSP), Charcot-Marie-Tooth disease (CMT), Huntington's disease (HD) and progressive myoclonic epilepsy (PME). Several families displayed unusual features. Genetic diagnosis was established for 37 families. Among them were four novel mutations in three and one families with HSP and CMT, respectively. Interestingly, many of these diseases were not previously described in Africa, especially in sub-Saharan Africa. While the majority were described broadly elsewhere, one disease was identified in only one Chinese family and another in an Arab family in Israel. Exome sequencing and disease gene panel testing have excluded known disease-causing genes in several families; suggesting new variants or entities. Linkage analysis has identified regions of interest in some families and exome sequencing variants in the regions of interest are being investigated. Our study established for the first time the molecular diagnosis of HNDs in sub-Saharan Africa and expanded their genetic epidemiology. Our results open the way for the discovery of new disease-causing genes in this and other populations. **Keywords:** neurogenetics, genetic diversity, novel mutations, Mali.

2473W

Characterization of copy number variations of genomic regions containing long noncoding RNA in children with neurological phenotypes. *S.C. Smith[1], V. Srivastava[2], J. Saunders[3], N. Miller[3], L.D. Cooley[1,2], E.A. Repnikova[1,2,3].* 1) Department of Pathology/Genetics, Children's Mercy Hospital, Kansas City, MO; 2) University of Missouri-Kansas City School of Medicine, Kansas City, MO; 3) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO.

   Children with congenital malformations and neurological disorders comprise 5-10% of the general pediatric population and represent the largest group referred for genetic evaluation. Microarray analysis provides a diagnostic yield of ~10-15% for patients with such conditions. While the diagnostic yield from whole exome and genome sequencing is significantly higher at ~25-34%, a significant proportion of affected children is still undiagnosed. A definitive genetic diagnosis remains elusive for many children because their phenotypic features are not specific and the majority of them do not have identifiable mutations or copy number variations (CNV) of the known clinically significant genes. In recent years, comprehensive analysis using RNA deep sequencing showed that a large number of long non-coding RNAs (lncRNAs) display neuronal-specific expression, suggesting their role in neurological disorders. In this study we screened for CNVs containing lncRNA in children referred for clinical microarray in order to identify new diagnostic and therapeutic targets for further research. In a review of 4647 patients referred to our laboratory for microarray analysis for various clinical indications during 2008-2012, we identified 50 individuals (35 males/15 females) with deletion (33 patients) or duplication (17 patients) of a genomic region containing only lncRNA(s). No genetic diagnosis was established for these patients. The average size of a deletion or duplication was 249.5 kb and 266.3 kb, respectively. Of those 50 patients, 48 (~96%) presented with various neurological disorders (ex. autism, developmental delay, epilepsy) compared with 58% of patients overall who had neurological phenotypic features for the same period of time. We found that 11 CNVs containing lncRNA were recurrent and 22 were unique. Interestingly, 16 CNVs containing lncRNAs in our dataset were not recorded in the benign CNVs per the Database of Genomic Variants (DGV), and 18 are present in the DGV, but at a frequency of 0.06% or lower. The highest frequency of such CNVs was observed for chromosomes 5, 8, 21 and X. Our initial investigation seems to support the hypothesis that lncRNA may be involved in neuronal development in humans. Further studies of CNVs containing lncRNAs, analysis of lncRNA tissue-specific expression and the genes they regulate will help to clarify how lncRNAs may influence disease susceptibility.

2474T

Analysis of total RNA in the cerebrospinal fluid environment. *S.L. Farrugia, M.T.W. Ebbert, K. Jansen-West, M. Prudencio, R. Rademakers, M. van Blitterswijk, L. Petrucelli, J.D. Fryer.* Neuroscience, Mayo Clinic, Jacksonville, FL.

Cerebrospinal fluid (CSF) shields the brain from physical trauma and pathogens while also aiding in cerebral circulation to supply the brain with blood. CSF has shown great potential as a predictive and diagnostic tool for a wide range of diseases and conditions such as multiple sclerosis, Alzheimer's Disease, and Parkinson's Disease. Metabolites in CSF have been examined for their role in depression and several behavioral disorders. Although metabolites and proteins in CSF have been analyzed extensively with conventional methods, these experiments have not accounted for all of the valuable information present in the RNA. MicroRNAs in CSF have received some research attention due to their potential role as a biomarker in brain diseases, however, the rest of the RNA population has barely been investigated. Analysis of total RNA in CSF can provide a plethora of information vital to understanding CSF function and various illnesses. We are collecting CSF from lumbar punctures of patients with various health diagnoses, including healthy controls. Total RNA is isolated, polyadenylated, and sequenced. The reads are then base called and aligned with the Burrows-Wheeler Aligner (BWA). Reads that are not able to be aligned to the human genome are tested for alignment with microbial species (fungi, bacteria, and virus). The study of total RNA in CSF will provide a more complete picture of the cerebral environment, including the "normal" RNA population in healthy individuals. Additionally, any bacterial RNA discovered can prompt the characterization of the "CSF microbiome." In future studies, CSF of diseased patients can be compared more extensively with control patients using unbiased or targeted RNA analysis to search for meaningful differences between the two groups. The categorization and evaluation of total RNA in CSF is essential to utilizing the full capabilities of CSF as a biomarker.

2475F

Optimized, modular, target-enrichment gene panels for the detection of genetic variants associated with neurological disorders. *A.B. Emerman[1], K.M. Patel[1], S.K. Bowman[1], S.M. Adams[1], C.D. Elfe[1], T.B. Davis[2], S. Russello[2], A. Barry[2], C.L. Hendrickson[1].* 1) Directed Genomics, Ipswich, MA; 2) New England Biolabs, Ipswich, MA.

Next-generation sequencing (NGS) is increasingly being used to better understand the genetic basis of inherited neuropathies and neuromuscular disorders, as well as for diagnostic testing for these conditions. The enrichment of targets of interest prior to NGS reduces sequencing costs while focusing data collection on the desired regions, resulting in greater coverage of relevant regions of the genome relative to whole genome sequencing. The diversity of neurological disorders necessitates customized gene panels for focused diagnostic tests or genomic studies; however, custom gene panels are often costly and time-intensive to develop de novo. To reduce cost and improve turnaround time for custom gene panels for neurological studies and diagnostics, we have predesigned bait sets specific to over one hundred genes associated with these disorders for use with the NEBNext Direct target enrichment kit. These bait sets are a part of a larger collection of predesigned content that has been optimized for specificity and uniform coverage, and subsets of genes from this collection can be selected and pooled into customized panels with a rapid turnaround time. The NEBNext Direct technology converts targets of interest into sequencer-ready libraries in a 1-day protocol while removing off-target sequences and incorporating unique molecular identifiers (UMIs) and sample indexes. Here we demonstrate that high-performing custom panels can be rapidly generated to target diverse neurological disorders using the predesigned bait sets. Subsets of genes from the predesigned neurological gene content were combined into custom panels to target specific neurological disorders. Each unique panel maintained high specificity and uniformity of coverage of the selected targets regardless of panel size or content. Additionally, we demonstrate that genetic variants within the selected targets were confidently detected with 10 ng of input DNA.

2476W

Autism spectrum disorder: A study of series of cases with genetic alterations. *J.L. Mussolini, A.M. Freitas, D. Brunoni.* Universidade Presbiteriana Mackenzie, São Paulo, Brazil.

The Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with communication and social interaction impairments, also repetitive and restricted behaviors. Affecting about 1% of the population, ASD has a heterogeneous clinical picture, besides several syndromes and psychopathies in comorbidity. Between the causal factors, it's known that the genetic causes represent up to 20% being that the copy number variations (CNVs) in microdeletions and microduplications forms are the principal ones. These cytogenomic alterations involve dozens of genes. The objective of this present work is to describe a convenience sample found in 6 ASD patients. The patients were investigated in the Clínicas TEA of Universidade Presbiteriana Mackenzie and the tests were performed in commercial laboratories. Five patients were diagnosed by microarray and one by exome sequencing. The genomic coordinate interpretation was performed with usual bioinformatics tools, like Genome Browser. The chromosomic regions involved were: 3p26.3; 6q; 7q31.3; 7q36.3; 9p21.1; 17q22 and the possible relevant genes in the manifestation of the clinical picture were: *B4GALT1*, *PTPRN2*, *CNTN4*, *CNTN6*, *OXTR*, *GRM7*, *SETD5*, *BZRAP1*, *SYNGAP1* e *IMMP2L*. These results were already been published in other patients with ASD. Hundreds of ASD patient samples are mentioned in the international literature. Individuals who present genetic alterations similar to the studied patients have in common the intellectual disability. Thus, this clinical feature should be the main indication to investigate the genetic cause comorbidity in patients with ASD. Both microarray and exome are of great use and importance in diagnosis due to their high resolution. It is clear that the case study should be expanded using the genetic tests already available clinically. This knowledge will improve the understanding of the changes caused by genetic alterations in individuals with ASD.

2477T

An inherited distal 16p11.2 deletion demonstrates association with rhizomelic shortening, variable expressivity, and incomplete penetrance for psychiatric illness: A case report. *E. Morris[1,2], B. McGillivray[1], A. Lehman[1], J. Austin[1,2].* 1) Department of Medical Genetics, University of British Columbia, Vancouver, BC; 2) Department of Psychiatry, University of British Columbia, Vancouver, BC.

Background: The most commonly reported 16p copy number variant (CNV) - 16p11.2 deletion syndrome (16pDS) (~600 kb; 29.5 -30.1 Mb) - is associated with intellectual disabilities (ID)/developmental delays (DD), schizophrenia (SZ), congenital anomalies, dysmorphic features and obesity. Non-overlapping "distal" 16p11.2 deletions (~200 kb; 28.7 -28.9 Mb) are less commonly reported, but have a variable phenotype similar to that of 16pDS. We report here on a parent/child dyad who share a distal 16p11.2 deletion. **Methods:** Patient 1 (index) was assessed (for indications including seizures, DD, and short limbs) at ages 2, 17, and 36, when she had clinical chromosomal microarray (CMA) (Affymetrix Cytoscan HD array, clinical thresholds: 200 kb (deletions), 400 kb (duplications), Chromosome Analysis suite (v2.0.0 195)). Fluorescence *in situ* hybridization (FISH) was used to confirm the deletion and determine origin. **Results: Patient 1** was born at term to non-consanguineous parents. A 46XX karyotype was confirmed at age 2 and 17 (testing indications: myoclonic seizures, DD, rhizomelic shortening of upper limbs). At age 17, she was also noted to have small hands (<3rd %ile) and feet (-3.5 SD), large posteriorly rotated ears (+3 SD), brachycephaly, high nasal bridge, and low posterior hairline. She had developed psychosis at 13, and received diagnoses of bipolar disorder type I, ADHD, possible ASD and ID. At 37, she is obese with impaired fasting glucose, and ongoing psychiatric symptoms. CMA revealed a 238Kb deletion at 16p11.2 (28 819 028 – 29 056 973; hg19)) confirmed by FISH as being paternally inherited. **Patient 2** is the 69 year-old father of Patient 1. He completed high school despite difficulties, and is not obese but has diabetes mellitus type II. He has rhizomelic shortening of both upper limbs but no history of psychiatric illness. **Discussion**: This report supports emerging evidence that this CNV is associated with psychiatric symptoms, including bipolar disorder, and contributes to the generation of a more thorough clinical description of the phenotypic range of distal 16p11.2 deletion, which we suggest, includes rhizomelic limb shortening.

2478F

Diagnostic yield of chromosomal microarray in a community sample of adults with schizophrenia. *C. Lowther[1], G. Costain[2], J. Waserman[3], K. Boyd[4], A. Noor[5], M. Speevak[5], D.J. Stavropoulos[5], D. Merico[6], J. Wei[6], A.C. Lionel[6], C.R. Marshall[6,7], S.W. Scherer[6,8], A.S. Bassett[9,10].* 1) University of Toronto, Toronto, Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Humber River Regional Hospital, Toronto, Ontario, Canada; 4) Department of Psychiatry & Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada; 5) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; 6) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 8) Molecular Genetics and McLaughlin Centre, University of Toronto, Ontario, Canada; 9) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 10) Toronto General Hospital Research Institute, University Health Network, Toronto, Ontario, Canada.

Schizophrenia (SCZ) is a severe psychiatric disorder associated with significant impairments in intellectual functioning. Rare copy number variations (CNVs) have been established to play an important role in the aetiology of SCZ. Additionally, CNVs associated with SCZ have been shown to negatively affect IQ in population-based controls without any major neuropsychiatric disorder. However, very few studies have examined the impact of rare CNVs on IQ in SCZ. The aim of this study was to perform genome-wide microarray testing in a community ascertained cohort of adults with SCZ to determine the diagnostic yield of rare CNVs across a range of IQ levels. We recruited 546 adults of European ancestry with SCZ from six community psychiatric clinics across Central and Eastern Canada. Each individual was assigned to one of three IQ sub-groups (average IQ, borderline intellectual functioning, ID) based on standardized IQ tests and/or educational attainment. We used rigorous methods to detect genome-wide rare CNVs from high resolution SNP microarrays and compared the burden of rare CNVs >100 kb deemed to be clinically relevant by a laboratory cytogeneticist between each of the IQ sub-groups. There were 39/546 (7.1%; 95% CI 5.2-9.7%) SCZ subjects with at least one pathogenic CNV detected. However, the yield of pathogenic CNVs was not uniformly distributed across the SCZ cohort: individuals with SCZ and average IQ had the lowest yield of pathogenic CNVs (n=9/325; 2.8%), followed by borderline intellectual functioning (n=9/130; 6.9%), and ID (n=15/62; 24.2%). In contrast, there was no significant difference in the yield of CNVs classified as variants of unknown significance between any of the IQ sub-groups. Many of the pathogenic CNVs identified in this study have been previously associated with SCZ, including 1q21.1 duplications, 2p16.3 deletions overlapping *NRXN1*, 2q13 deletions/duplications, 15q11.2-q13.1 duplications, 16p11.2 deletion/duplications, and 22q11.2 deletions. We also identified several CNVs that have not been previously associated with SCZ, including a 3q27.1-q27.2 deletion, 6q26-q27 deletion, 13q14.13-q14.3 deletion, Xp11.23-p11.22 duplication, and an Xp22.33-p22.2 deletion. Finally, we identified a novel 280 kb 3p26.1 deletion overlapping two genes, *ITPR1* and *SUMF1*, which has not been previously reported in the literature. The results of this study have important implications for clinical practice and may help inform future rare variant studies of SCZ. .

2479W

Clinical utility of exome sequencing in individuals with large homozygous regions. *A. Prasad[1], M.A Sdano[2], R.J. Vanzo[1], M.A. Serrano[1], C.H. Hensel[1], E.R. Wassman[1].* 1) Lineagen, Salt lake City, UT; 2) Department of Biochemistry, University of Utah, Salt Lake City.

Chromosomal microarray analysis (CMA) is recommended as the first-tier clinical diagnostic test for individuals with developmental disabilities. CMA platforms with single nucleotide polymorphism probes can detect large homozygous regions within the genome, which represent potential risk for recessively inherited disorders. Clinical laboratories typically proceed with targeted sequencing by selecting candidate genes within the homozygous regions based on a patient's phenotype. However, heterozygous variants associated with autosomal dominant or X-linked conditions could also explain at least part of the patient's phenotype. The purpose of this study is to evaluate heterozygous variants in a cohort of individuals with large homozygous regions. We performed whole exome sequencing (WES) in 53 patients referred to our clinical laboratory for a variety of phenotypes including autism spectrum disorder (ASD), developmental delay, epilepsy, intellectual disability or microcephaly where homozygosity was detected by CMA. We have previously examined homozygous variants and identified pathogenic or likely pathogenic variants in 6/53 cases (11%). These variants, in *GJB2*, *TPP1*, *SLC25A15*, *TYR*, *PCCB*, and *NDUFV2,* are implicated in a variety of diseases. In this study, we investigated heterozygous likely protein disrupting variants across the whole exome and identified an additional 8/53 (15%) cases with pathogenic or likely pathogenic variants. We found one of the patient to have a homozygous pathogenic variant in *TPP1* as well as a likely pathogenic hemizygous variant in the *OTC* gene. In one of the patients with strabismus, dysmorphic ear shape, language and developmental delays and ASD, we identified a pathogenic missense variant in the *SOS1* gene which is reported to cause Noonan syndrome. In two unrelated patients with muscle weakness, we identified a likely pathogenic hemizygous variant in *PNPLA4* which has been shown to cause mitochondrial respiratory chain complex deficiencies. The other likely pathogenic variants were in the *CADM1*, *HBB*, *SFTPC* and *ASMT* genes. These findings have important medical management implications including screening and treatment options. Our study highlights the clinical utility of WES in individuals whose CMA uncovers homozygosity. Importantly, our results demonstrate that some of the causative mutations do not lie in regions of homozygosity, suggesting that the appropriate follow up test is WES rather than a targeted approach.

2480T

A framework to identify contributing genes in patients with Phelan-Mc-Dermid syndrome. *A.C. Tabet[1,2,3,4], T. Rolland (first Co-author)[2,3,4], M. Ducloy[2,3,4], J. Levy[1], J. Buratti[2,3,4], A. Mathieu[2,3,4], C. Leblond[2,3,4], B. Keren[5], C. Mignot[6], A. Jacquette[7], E. Pipiras[8], S. Chantot-Bastaraud[9], D. Héron[10], C. Le Caignec[11], L. El Khattabi[12], P. Vago[13], AL. Mosca-Boidron[14], C. Missirian[15], D. Sanlaville[16], C. Coutton[17], C. Rooryck[18], J. Puechberty[19], J. Chiesa[20], J. Lespinasse[21], C. Dubourg[22], A. Toutain[23], R. Toro[2,3,4], F. Amsellem[24], R. Delorme[2,3,4,24], T. Bourgeron[2,3,4].*
1) AP-HP, Department of Genetic, Cytogenetic Unit, Robert Debre Hospital, Paris, France; 2) Human Genetics and Cognitive Functions, Institut Pasteur, Paris, France; 3) CNRS UMR 3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 4) Université Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; 5) Cytogenetics Unit, Pitié Salpetrière Hospital, APHP, Paris, France; 6) Neurogenetics Unit, Pitié Salpetrière Hospital, APHP, Paris, France; 7) Clinical Genetics Unit, Pitié Salpetrière Hospital, APHP, Paris, France; 8) Cytogenetics Unit, Jean Verdier Hospital, APHP, Bondy, France; 9) Cytogenetics Unit, Trousseau Hospital, APHP, Paris, France; 10) Clinical Genetics Unit, Trousseau Hospital, APHP, Paris, France; 11) Clinical Genetics Unit, Nantes Hospital, Nantes, France; 12) Cytogenetics Unit, Cochin Hospital, APHP, Paris, France; 13) Genetics Unit, CHU Estaing, Clermont-Ferrand, France; 14) Cytogenetics Unit, Dijon Hospital, Dijon, France; 15) Genetics Unit, La Timone Hospital, Marseille, France; 16) Cytogenetics Unit, Lyon Civil Hospital, Lyon, France; 17) Cytogenetics Unit, Grenoble Hospital, Grenoble, France; 18) Genetics Unit, Bordeaux Hospital, Bordeaux, France; 19) Genetics Unit, Montpellier Hospital, Montpellier, France; 20) Genetics Unit, CHRU Nimes, Nimes, France; 21) Cytogenetics Unit, Chambéry-Hôtel- Dieu Hospital, Chambery, France; 22) Genetics Unit, CHU Rennes, Rennes, France; 23) Genetics Unit, Bretonneau Hospital, Tours, France; 24) Department of Child and Adolescent Psychiatry, Robert Debre Hospital, APHP, Paris, France.

Phelan-McDermid syndrome (PMS) is characterized by a variety of clinical symptoms with heterogeneous degrees of severity, including intellectual disability (ID), absent or delayed speech, and autism spectrum disorders (ASD). It results from a deletion of the distal part of chromosome 22q13 that in most cases includes the *SHANK3* gene. *SHANK3* is considered a major gene for PMS, but the factors that modulate the severity of the syndrome remain largely unknown. In this study, we investigated 85 patients with different 22q13 rearrangements (78 deletions and 7 duplications). We first explored the clinical features associated with PMS, and provide evidence for frequent corpus callosum abnormalities in 28% of 35 patients with brain imaging data. We then mapped several candidate genomic regions at the 22q13 region associated with high risk of clinical features, and suggest a second locus at 22q13 associated with absence of speech. Finally, in some cases, we identified additional clinically relevant copy-number variants (CNVs) at loci associated with ASD, such as16p11.2 and 15q11q13, which could modulate the severity of the syndrome. We also report the first inherited *SHANK3* deletion transmitted to five affected daughters by a mother without ID nor ASD, suggesting that some individuals could compensate for such mutations. In summary, we shed light on the genotype-phenotype relationship of patients with PMS, a step towards the identification of compensatory mechanisms for a better prognosis and possibly treatments of patients with neurodevelopmental disorders.

2481F

Post-zygotic copy number variations in autism spectrum disorder discordant monozygotic twins. *D.L. Nuñez[1], C. Ospina[2], F. Paredes[2], M.C. Lattig[1].*
1) Universidad de Los Andes, Colombia; 2) Liga Colombiana de Autismo.

Copy Number Variations (CNVs) are deletions and duplications which change the number of copies of chromosome regions. CNVs may contribute greatly to genomic variation in human population, and also provide a potential risk to develop a wide kind of disorders such as Autism Spectrum Disorder (ASD). Pathogenic and normal CNVs are formed in germinal and post-zygotic cells. Specifically, meiotic formation of CNVs is mediated by mistakes in homologous recombination while post-zygotic CNVs are formed by mechanism related with DNA replication (Campbell, 2015). A minority of CNVs associated to genetic disorders are inherited, suggesting that most pathological CNVs are developed by germinal or early post-zygotic processes. Indeed, the most critical period for somatic changes is the early embryonic development due to high cell division rate (Iourov, 2010). Current studies with monozygotic (MZ) twins, genetically identical at conception have shown CNVs discordances in elder age (Koldby, 2016). Moreover, although post-zygotic mutations are poorly characterized in neurodevelopmental disorders, recent evidence suggest the role of multi-tissue mosaic mutations in individuals with ASD (Freed D., 2016). We report a Colombian family with both parents and 3 sons: an older healthy brother and a couple of phenotypically discordant MZ twins (one MZ twin has clinical diagnosis of ASD while the other MZ twin does not). The clinical evaluation was completed with the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observational Schedule (ADOS) tests. DNA from blood samples of MZ twins were evaluated by SNP 6.0 microarray of Affymetrix, and the .CEL files were analyzed by PennCNV sotfware. We found a small 46Kb deletion present only in the ASD twin. The deletion includes the complete exon 2 of the GPHN gene (GENE ID 10243). This gene encodes a neuronal assembly protein named Gephyrin that interacts with subunits of $GABA_A$ receptors regulating its clustering and diffusion in postsynaptic membrane (Dejanovic, 2015). Indeed, exonic deletions in GPHN have been implicated in risk for schizophrenia, autism and epilepsy (Balan, 2017). Moreover, the reduction in its expression levels has been related with Alzheimer disease (Puthiyedth, 2016). This result supports previous observations that somatic CNVs could be also responsible of neurodevelopmental disorders such as ASD.

2482W

Chromosomal microarray analysis of pediatric patients with autism spectrum disorders and intellectual disabilities. *A.R. Patel[1], D.D. Jhala[2], S.S. Chettiar[1].* 1) Department of Genetics, GeneXplore Diagnostics and Research Center Pvt. Ltd., Ahmedabad, Gujarat, India; 2) Human Genetic Center, Department of Zoology, Gujarat University, Navrangpura, Ahmedabad, Gujarat, India.

Autism specrum disorder (ASD) and Intellectual disabilities (ID) are complex and phenotypically heterogeneous neurodevelopmental disorders characterized by significant deficit in cognitive and adaptive skills, debuting during the developmental period. Chromosomal microarray (CMA) offers superior sensitivity and high resolution survey of the entire genome for identification of submicroscopic copy number variants (CNV). It also contributes to understand the etiology of disease which makes it the first tier genetic testing for patients with autism spectrum disorder (ASD) and Intellectual disabilities (ID). In the present study, we studied 67 children (43 Males and 24 Females; Age ranges from: 2.0 yrs.-14yrs.) diagnosed with Autism or Autism spectrum disorder or intellectual disabilities. array CGH was performed on oligonucleotide microarray platform. Data of the array was analyzed using cytogenomics software (CytoGenomics v4.0.3, Agilent technologies). Statistical analysis was performed using SPSS.   Results of the microarray analysis showed 28 clinically relevant(Likely Pathogenic) chromosomal imbalances in 14 patients leads to an overall diagnostic yield of 20.89%(14/67). 11 clinically significant variations were detected in patients with Autism or Autism spectrum disorders. Among which Microdeletion of 5,741kbp at chr15q11.2-q13.1 was detected in three patients. 17 chromosomal imbalances were detected in the children with Intellectual disabilities. The range of the imbalances in ID patients ranges from 712kbp to 10463 kbp. While, both groups contained more males than females, a significantly higher percentage of males were present in the ASD group. The likelihood of an abnormal microarray result increased with the number of clinical abnormalities. Microarray analysis will likely become the diagnostic genetic test of choice as its potential implication in Autism spectrum disorders and in idiopathic Intellectual disabilities patients.

2483T

Whole genome sequencing of neurodevelopmental disorders in Japanese. *C. Abe-Hatano[1,2], A. Iida[3], K. Ishikawa[3], Y. Momozawa[4], S. Kosugi[5], I. Nishino[3], K. Inoue[1], Y. Kamatani[5,6], M. Kubo[4], Y. Goto[1,3].* 1) Department of Mental Retardation and Birth Defect Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Japan; 2) Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Medical Genome Center, National Center of Neurology and Psychiatry, Kodaira, Japan; 4) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Neurodevelopmental disorders (NDDs) are clinically and genetically heterozygous diseases characterized by autism spectrum disorder (ASD), intellectual disability/developmental delay (ID/DD), attention-deficit/hyperactivity, motor and tic disorders, and language communication disorders. Severe NDDs and/or major congenital malformations affect 2-5% among global population, but the etiology remains unclear. We employed whole genome sequencing (WGS) to identify the causative genes in Japanese patients with NDDs. Thirty six patients (20 males and 16 females) with NDDs from 36 unrelated families were included in this study. The common clinical features among all patients were included ID/DD. We performed WGS in all patients and their parents and used to identify the causative genes by the candidate gene approach. The candidate variants were evaluated by *in silico* analysis programs. Finally, we confirmed the variants by Sanger sequencing. We found a total of seven single nucleotide variants (SNVs) in seven known NDDs genes, *MID* (1 SNV), *TCF4* (1), *SPAST* (1), *TSPYL2* (1), *SLC6A8* (1), *SATB2* (1), and *SMS* (1). We found two genes with *de novo* single nucleotide variants in two families and one gene with autosomal dominant inheritance. The remaining four was occurred in four families with X-linked recessive NDDs. Six SNVs were considered to be novel variants by comparing our data with the variants deposited in the public databases. The remaining two SNVs were previously reported. One of these SNVs, c.1738C>T (p.R580W) in *TCF4*, was the causative variant in Pitt-Hopkins syndrome, which is characterized by severe developmental delay and breathing anomalies. Another variant, c.1496G>A (p.R499H) in *SPAST* was the same variant mutated in autosomal dominant spastic paraplegia. We also identified a 10-base deletion of *ADNP* in one patient with stereotypic behavior and ventricular septal defect. Now we are investigating the genotype-phenotype correlations of these patients, and the WGS data from the remaining 28 patients. This is the first report of WGS of NDDs in Japan. A total of seven SNVs and one small deletion in eight patients were identified. These data might contribute to a more precise understanding of correlation between genotypes and phenotypes.

2484F

Diagnostic yield of chromosomal microarray analysis in patients with intellectual disability and developmental delay. *A.C. Ceylan[1,2], S. Citli[1], H.B. Eerdem[3,4], I. Sahin[3,4], E. Acar Arslan[5].* 1) Department of Medical Genetics, Trabzon Kanuni Training And Research Hospital, Trabzon, Turkey; 2) Department of Medical Genetics, Yıldırım Beyazıt University, Ankara Atatürk Training and Research Hospital, Ankara, Turkey; 3) Department of Medical Genetics, Atatürk University, Medical Faculty, Erzurum, Turkey; 4) Department of Medical Genetics, Ankara Dışkapı Training and Research Hospital, Ankara, Turkey; 5) Department of Child Neurology, Karadeniz Technical University, School of Medicine, Trabzon, Turkey.

   Intellectual disability (ID) and developmental delay (DD) are clinically heterogeneous neurodevelopmental disorders seen in approximately 1 % of children. Chromosomal microarray (CMA) is a first-tier test in the evaluation of individuals with ID and DD with the diagnostic yield varying based on population examined, ranging from 5-25% in recent literature. In this study, we present the diagnostic rates for copy number variations (CNVs) detected by CMA in a cohort of 125 Turkish patients with ID and DD. Affymetrix CytoScan Optima® chips were used to perform CMA in 125 patients with ID and DD at Trabzon Kanuni Training and Research Hospital. Data analysis was performed using Chromosome Analysis Suite3.1 (ChAS) software. The analysis and interpretation of the obtained results were performed by using public genomic databases, such as UCSC, OMIM, DGV, DECIPHER, CLINGEN. In 24 individuals (from 23 families), CMA analysis revealed likely pathogenic CNVs including 6 microduplications (2p25p24, 22q11.2 [2 cases], 2p25, 14q32, and 15q11.2), 14 microdeletions (1p36.3, 1q21 [2 cases], 2q36, 6p25.1p23, 6q25q27, 14q32, 15q11.2, 16q24 [KBG syndrome], 18p11.3, 18q22, 18q21q23, 19p23, and 22q11.2) and 2 triplications identified in same locus (22q11.2) in two patients. All identified CNVs were interpreted as having a pathogenic significance in concordance with the clinical findings in the patients examined. Interestingly, in 2 out of 24 individuals multiple CNVs have been identified including 2 different deletions (6q13q14 and 8q21.3) in one patient and 1 deletion (6q25) and 2 duplications (6q24.1 and 6q26) in another patient. The diagnostic yield of CMA in our study was 19.2% which is consistent with the results of previous studies (5-25%). The length of 18 CNVs were below 5 Mb and could not be detected with conventional karyotyping. The length of 7 CNVs were also between 5-10 Mb. Interestingly siblings, whose parents were consanguineous and carrying 22q11.2 duplication, were diagnosed with 22q11.2 triplication. 23 of 27 CNVs were revealed de novo. Although 8 out of 23 families had consanguineous marriages, our analyses revealed de novo variations in these families rather than homozygous mutations. CMA is a first-tier test in the evaluation of individuals with ID and DD which provides opportunities to discover new ID-DD associated syndromes and uncover the genetic background of many syndromes by revealing genetic heterogeneity and identifying new loci for novel candidate genes.

2485W

Unravelling structural chromosomal rearrangements by whole genome sequencing: Results of the ANI project, a French collaborative study including 55 patients with intellectual disability and/or congenital malformations. *D. Sanlaville[1], F. Diguet[1], P.A. Rollat-Farnier[1], J. Amiel[2], M.A. Belaud-Rotureau[3], B. Benzacken[4], P. Callier[5], P. Collignon[6], B. Demmer[7], M. Doco-Fenzy[8], L. Faivre[9], B. Gilbert-Dussardier[10], A.M. Guerrot[11], B. Keren[12], V. Kremer[13], A. Lebbar[14], C. Le Caignec[15], J. Lespinasse[16], V. Malan[17], C. Missirian[18], C. Perbel-Richard[19], V. Paquis[20], M.F. Portnoi[21], J. Puechberty[22], C. Rooryck-Thambo[23], V. Satre[24], A.C. Tabet[25], R. Touraine[26], A. Toutain[27], C. Schluth-Bolard[1].* 1) Service de Génétique, Hospices Civils de Lyon; Centre de Recherche en Neurosciences de Lyon, INSERM U1028; CNRS UMR5292; UCBL1; GENDEV Team, Lyon, France; 2) Service de Génétique Médicale, Hôpital Necker-Enfants Malades, Paris, France; 3) Laboratoire de Cytogénétique et de Biologie Cellulaire, CHU Pontchaillou, Rennes, France; 4) Laboratoire de Cytogénétique, Hôpital Jean Verdier, Bondy, France; 5) Laboratoire de Cytogénétique, CHU Dijon, France; 6) Service de Génétique Médicale, CHI Toulon, France; 7) Centre d'activité de génétique clinique, Pôle pédiatrie, Hôpital Nord, CHU Amiens, France; 8) Service de Génétique, CHU Reims, France; 9) Centre de référence anomalies du développement et syndromes malformatifs, CHU de Dijon, France; 10) Génétique Médicale, CHU La Milétrie, Poitiers; EA 3808 Université de Poitiers, France; 11) Unité de Génétique Clinique, CHU Rouen, Franc; 12) Département de Génétique, GH Pitié-Salpêtrière, Paris, France; 13) Laboratoire de Cytogénétique, CHU Strasbourg, France; 14) Laboratoire de Cytogénétique Constitutionnelle, Hôpital Cochin, Paris; 15) Service de génétique médicale, Institut de biologie – CHU Nantes, France; 16) Laboratoire de Génétique chromosomique, CH Général, Chambéry, France; 17) Service de Cytogénétique, Hôpital Necker Enfants Malades, Paris, France; 18) AP-HM, Département de Génétique Médicale, Laboratoire de Génétique Chromosomique; 19) Service de Cytogénétique Médicale, CHU Estaing, Clermont-Ferrand, France; 20) Service de Génétique Médicale, CHU de Nice, France; 21) Département de Génétique Médicale, Hôpital d'Enfants Armand Trousseau, Paris, France; 22) Service de génétique, Hôpital Arnaud de Villeneuve, Montpellier, France; 23) Service de Génétique Médicale, CHU de Bordeaux, France; 24) Laboratoire de Génétique Chromosomique, Hôpital Couple Enfant, CHU Grenoble, France; 25) Laboratoire de Cytogénétique, Hôpital Robert Debré, Paris, France; 26) Service de Génétique Clinique, Chromosomique et Moléculaire, CHU Hôpital Nord, Saint-Etienne, France; 27) Service de Génétique, CHRU Tours, France.

   Introduction: *Apparently balanced chromosomal rearrangements* (ABCR) *associated with abnormal phenotype* are rare events, but may be challenging for genetic counseling. Abnormal phenotype may be explained either by cryptic genomic imbalances detectable by array-CGH or by gene disruption or position effect. However, breakpoint cloning using conventional methods is laborious and not performed routinely. Recently, Whole Genome Sequencing (WGS) proved to be a powerful and rapid technique to characterize ABCR breakpoints at the molecular level. **Material and methods:** The ANI project is a French collaborative study that aims at characterizing ABCR in patients presenting with intellectual disability and/or congenital anomalies. We included 55 patients (41 reciprocal translocations, 4 inversions, 2 insertions, 8 complex chromosomal rearrangements). Array-CGH showed no pathogenic imbalance. *Breakpoints were characterized by paired-end WGS* and confirmed by Sanger sequencing. Expression studies of disrupted and neighboring genes were performed on blood cells. **Results:** *Breakpoints were characterized for 49/55 patients (89%)*. The rearrangements showed unexpected complexity, since 211 breakpoints were identified against 119 breakpoints according to karyotype and included 6 cases of chromoanagenesis. 59% (29/49) of rearrangements resulted from non-homologous end-joining (NHEJ) mechanism, 27% (13/49) may involve micro-homology mediated mechanisms and 14% (7/49) may involve a combination of both mechanisms. Ninety-four breakpoints disrupted a gene (44%). In *17/49 patients, WGS allowed a diagnosi*s, either by gene disruption (12) or by position effect (5), thus a diagnostic rate of 35%. **Conclusion:** These results revealed the diversity and complexity of chromosomal rearrangements at the molecular level. It also show the *diagnostic relevance of WGS approach in ABCR*. This work is supported by ANR and DGOS (PRTS 2013 grants).

2486T

Using parental report questionnaires to identify developmental delay in a 22q11.2 deletion specialty clinic. *K. Coleman[1,2], L. Kobrynski[3], C. Chris Gunter[1,2,3], O. Ousley[1,4].* 1) Marcus Autism Center, Children's Healthcare of Atlanta, 1920 Briarcliff Rd NE, Atlanta, GA 30329; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322; 3) Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322; 4) Emory Autism Center, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA 30322.

22q11.2 deletion syndrome (22q11.2DS) occurs in approximately 1:4000 live births and encompasses significant phenotypic heterogeneity. In addition to providing complex medical care, a multidisciplinary team is responsible for assessing these patients for a variety of developmental delays and making appropriate referrals. In any setting, this team could benefit from establishment of reliability for parental-report questionnaires. We conducted a medical chart review for children seen within the Emory-CHOA (Children's Healthcare of Atlanta) 22q Specialty Clinic. During each medical visit parents completed developmental questionnaires appropriate to their child's age, including 1) Communication and Symbolic Behavior Scales Developmental Profile Infant Toddler Checklist (**CSBS-DP-ITC**) for ages 6 to 24 months (n=87, mean age=15.0 months; SD=5.2); 2) Child Development Inventory Profile (**CDIP**) for 25 to 72 months (n=86, mean age=47.4 months; SD=14.2); or 3) Adaptive Behavior Assessment System, Second edition (**ABAS-II**) for 6 to 21 years (n=101, mean age=12.3 years; SD=4.1). We determined that approximately half of the children showed evidence of developmental delay across communication, social, and conceptual domains for each age group. For the **CSBS-DP-ITC**, referrals for further services were indicated for each composite score: Social (39.1%), Speech (54.0%), Symbolic (51.7%), and Total (52.9%). For the **CDIP**, children scored 2 SD below the mean at the following rates for each domain score: Social (60.5%), Self-help (68.6%), Gross Motor (60.5%), Fine Motor (51.2%), Expressive Language (67.4%), and Language Comprehension (62.8%). For the **ABAS-II**, children and adolescents scored 2 SD below the mean at the following rates for each composite score: Conceptual (44.5%), Social (29.7%), Practical (51.5%), and General Adaptive (51.5%). In addition, for a subset of children who were assessed at multiple time points, we found evidence of sustained developmental delay. We conclude that parental report questionnaires provide quantitative measurements of various developmental skills, allowing medical professionals to initiate referrals for developmental delay and assisting them in determining if a comprehensive speech/language or neuropsychological assessment is warranted in children with 22q11.2DS. Notably, despite the high rates of developmental delay found, many children also obtained age-appropriate scores, highlighting the range of outcomes for children with 22q11.2DS.

2487F

Increased yield of clinically relevant candidates in the UK 100,000 Genomes Project using the Fabric Genomics platform. *M. Babcock[1], C. Son Rigby[1], M. Falcioni[1], A. Guo[1], E. Kofman[1], A. P. Fejes[1], J. Grigonis[1], E. S. Kiruluta[1], M. Yandell[2], M. G. Reese[1].* 1) Fabric Genomics, Oakland, CA; 2) Department of Human Genetics, University of Utah, Salt Lake City, UT.

The 100,000 Genomes Project, spearheaded by Genomics England (GeL), is a United Kingdom National Health Service sponsored study aimed at identifying disease-causing genetic variants in patients and families with rare genetic diseases and cancer using a whole genome sequencing (WGS) approach. For this study, clinical history was used to recruit patients into specific disease categories, each of which were associated with gene panels curated in the GeL PanelApp tool. Fabric Genomics™, a clinical interpretation partner for the 100,000 Genomes Project, has analyzed over 600 clinical cases using Opal™ Clinical. The variant filtering and prioritization protocols utilized for case analysis include GeL's variant tiering methodology, ClinVar, and Fabric Genomics' proprietary variant and gene ranking algorithms VAAST (Variant Annotation, Analysis and Selection Tool) and Phevor (Phenotype Driven Variant Ontological Re-ranking Tool). We report results showing that by applying VAAST and Phevor we increase the clinical candidate yield compared to using the GeL tiering system alone. We identified candidate causal genes/variants in 44.7% of the cases. In 19.6% of these cases (8.7% overall) candidates were only obtained by using the VAAST/Phevor top 20 ranked genes/variants.

2488W

Next generation sequencing based on long range PCR: A reliable, expeditious, cost effective genetic testing strategy for lysosomal storage diseases. *M.C. Vanaja[1], P. Ranganath[1,2], A. Bhattacherjee[1], A. B Dalal[1].* 1) Diagnostics, Centre for DNA Fingerprinting and Diagnositcs, Hyderabad, Telangana, India; 2) Department of Medical Genetics, Nizams Institute of Medical Sciences, Hyderabad.

Lysosomal storage diseases (LSDs) are rare inherited metabolic disorders resulting from defective lysosomal function. There are over 50 different disorders which affect different parts of the body including the skeleton, brain, heart and the central nervous system. Many factors make LSD diagnosis difficult, including phenotype and penetrance variability, shared signs and symptoms, and problems inherent to biochemical diagnosis. Even though enzyme analysis is the most effective method for definitive diagnosis, molecular genetic testing can refine the enzymatic diagnosis. Further accurate prenatal diagnosis can be done using molecular genetic testing. Molecular genetic testing has conventionally been done using the Sanger sequencing technique. While Sanger sequencing is very useful for sequence analysis of small genes, when applied for large genes it becomes time consuming, costly and laborious, requiring multiple PCR reactions for generating amplicons for sequencing. We have designed a strategy based on next generation sequencing (NGS) for the simultaneous testing of different lysosomal genes, using Long Range PCR technique. Primers were designed for each gene to amplify 5-10kb fragments containing the exons and intronic regions. Long PCR was set up for the patients DNA who were diagnosed to be affected with the respective LSDs based on clinical features and enzyme assay for respective genes. Libraries were constructed by mixing the amplified products from gene fragments of **NEU1**, **SMPD1**, **IDUA, ARSA**, **NPC1**, **NPC2**, **GBA**, **GAA**, **GLB1**, **GNPT-AB**, **and GALNS.** The constructed libraries were sequenced on an Illumina MiSeq platform. The Quality analysis of the fastq file generated was done by FASTQC, followed by data alignment by Burrow's Wheeler algorithm, Variant calling by GATK pipeline and Variant Annotation by Annovar. The identified sequence variants were validated by Sanger sequencing. Initial standardization was done using 40 patient samples with known variants in 9 genes and we found 100% concordance in the identified sequence variants by NGS. Later, 40 samples with unknown variants were included and the identified 38 likely pathogenic sequence variants which were validated by Sanger sequencing. We conclude that Long-range PCR combined with next-generation sequencing is a reliable, expeditious, cost effective choice for mutational analysis of lysosomal storage diseases and can be applied as a method of choice in low income countries for cheaper molecular diagnostics. .

2489T

StarSeq, an innovative method based on NGS for accurate detection of punctual mutations and copy number variants in children with intellectual disability and obesity. *A. Bonnefond[123], L. Montagne[1234], M. Derhourhi[123], E. Durand[123], I. Rabearivelo[123], O. Sand[123], B. Delobel[4], P. Froguel[1235].* 1) CNRS UMR8199, Lille, France; 2) Univ. Lille, Lille, France; 3) Institut Pasteur de Lille, Lille, France; 4) Hôpital Saint-Vincent de Paul, Lille, France; 5) Imperial College, London, UK.

Although targeted or exomic next generation sequencing (NGS) is now widely used for the detection of punctual coding mutations, DNA microarrays (e.g. CGH/SNP microarrays) are still the current gold standard methods for the detection of copy number variants (CNVs) in the molecular diagnosis of genetic disorders. Indeed, several studies demonstrated reproducibility issues in CNV calling based on whole-exome sequencing. We have developed a groundbreaking, NGS-based, one-step tool (StarSeq) for an accurate and cost effective detection of both punctual mutations and CNVs. We assessed its usefulness and accuracy for the molecular diagnosis of patients with monofactorial diseases. We used a custom capture (NimbleGen SeqCap EZ Choice XL) which included probes regularly spaced in the entire genome, in combination with a capture targeting the whole exome (NimbleGen SeqCap EZ MedExome Enrichment Kit). The resulting probes covered >80 Mb of the genome. Probes were placed every 10 kb apart in the regions known to be affected by CNVs (based on the literature and public database), and probes were placed every 25 kb in the rest of the genome. We tested this capture in a cohort of 120 obese children with intellectual disability and their parents, whose 50% had already been analyzed by DNA microarrays (CGH Agilent 60k v2) for CNV detection. The sequencing was performed through Illumina HiSeq 4000, using 100 bp paired-end reads, and a mean sequencing depth higher than 85× was achieved for each individual. We detected all the CNVs previously identified by DNA microarrays and we found additional CNVs which were not found by DNA microarrays in our patients, probably due to the better resolution of StarSeq. In parallel, we detected novel punctual mutations with the same accuracy as a standard exome sequencing. We were able to diagnose more than half of the patients, establishing causality to either punctual mutations and/or CNVs. For instance, in a young boy with obesity, macrocephalia and large abdominal lipoma (without intellectual disability), we surprisingly detected a chr16p11.2 deletion along with a heterozygous nonsense mutation (p.R130*) in *PTEN*. We demonstrated that our custom capture in combination with NGS enables an accurate, inexpensive detection of CNVs and punctual mutations in only one step. StarSeq dramatically improves genetic diagnosis and thus has the potential to replace standard exome sequencing and DNA arrays as this method is cost-effective and fast.

2490F

Genetic causes of intellectual disability in 102 consanguineous families from Jordan. *T. Froukh[1], O. Nafie[2], J. Sommerfeld[3], S. AlHait[4], M. Sturm[3], A. Baraghiti[1], P. Koch[3], J. Hanselmann[3], B. Kootz[3], J. Gohlke[3], T. Haack[3], W. Al-Ameri[5], R. Abou Jamra6[6], A. Alfrook[4], M. Hamdallah[7], A. Riess[3], P. Bauer[3], O. Riess[3], R. Buchert[3].* 1) Philadelphia University, Amman, Jordan; 2) Mutah University, Amman, Jordan; 3) University of Tübingen, Tübingen, Germany; 4) Private Clinic, Amman, Jordan; 5) Ibn Alhytham Hospital, Amman, Jordan; 6) University Medical Center Leipzig, Leipzig, Germany; 7) Farah Hospital, Amman, Jordan.

Intellectual disability (ID), defined by an intelligence quotient of less than 70, occurs in approximately 1 to 3% of the population and tends to be higher in low-income countries and in inbred communities. Despite high rates of consanguineous marriages and likely enrichment for recessive forms of ID, the genetic bases of ID in Jordan are largely unstudied. In this study, we used whole exome sequencing (WES) to identify the genetic causes of ID in 102 consanguineous families from Jordan with one or more intellectually disabled progeny. While sequencing and variants filtration are progressing, likely pathogenic mutations were identified in 15 families and strong candidates were identified in other 20 families. Variants in six genes not previously implicated with ID were assigned as possible causative to ID in six families, including: APBA1, EGFL6, DENND5A, NEK5, SALL3, and PKP3. In addition, two *de novo* CNV deletions were detected in two families, one with 2.8Mb on 4q21.22-21.23, previously recorded as pathogenic and the other with 1.5Mb on 7q11.23, previously associated with Williams Beuren Syndrome. This study demonstrates the power of examining consanguineous families to identify genes not previously implicated with ID and motivates larger gene discovery endeavors in the Jordan population.

2491W

Identification of copy number variations from whole-exome sequencing using eXome Hidden Markov Model (XHMM): A French experience. *E. Tisserant[1], J. Thevenon[1,2], A. Bruel[1], A. Sorlin[1], M. Assoun[1], N. Marle[3], V. Carmignac[1], S. Nambot[1,2], M. Lefebvre[1,2], A. Vitobello[1], D. Lehalle[1,2], F. Tranmauthen[1,2], C. Philippe[1,3], P. Kuentz[1], M. Polleau[3], J. Thibaud[1], C. Poe[1], C. Thauvin[1,2], L. Faivre[1,2], AL. Mosca-Boidron[1,3], Y. Duffourd[1], P. Callier[1,3].* 1) Equipe GAD, EA4271, UMR 1231, Dijon, France; 2) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est, Centre Hospitalier Universitaire Dijon, 21079 Dijon, France; 3) Laboratoire de Génétique Chromosomique et Moléculaire, Plateau technique de biologie, Centre Hospitalier Universitaire Dijon, 21079 Dijon, France.

Whole-Exome Sequencing (WES) is becoming a standard application for the detection of gene mutations responsible for human disease, especially SNVs and indels. To date, most studies of CNVs were diagnosed by array-CGH or array-SNP with a lower detection limit of about 30kb. The challenge is now to detect both SNVs and CNVs using an exome-wide approach with a single test. Several algorithms have been developed to call CNV from WES data (Exome CNV, CONTRA, ExomeCopy, ExomeDepth, ConiFer, XHMM). The XHMM program is optimized for the identification of rare variants in large exome data sets, in particular exonic CNVs smaller than 30kb. We used XHMM to screen for exome CNVs in data from 350 patients with developmental disorders for whom exome sequencing was performed in a research or diagnostic setting with normal array-CGH 4x180K Agilent design used in our lab. We identified pathogenic CNVs, ranging in size from 685pb to 11kb (all confirmed by qPCR), in five patients (1.4%). Among these, four patients showed clinically relevant deletions in neurodevelopmental disorders (4.5 kb homozygous deletions of *PPT1*; 5kb for *PIGN*; 637pb for *TCF4*; 2.2kb homozygous *CLCN2* deletion), one patient had Cohen syndrome (11kb homozygous *VPS13b* deletions). This study shows that the XHMM program detected five CNVs from whole-exome data. This study underlines the interest of whole-exome sequencing in genetic disorders to identify SNVs and CNVs especially those smaller than 30kb not detected by array-CGH.

2492T

MIDAS Project status report: Trio whole exome sequencing in patients with intellectual disability. *Y. Dinçer[1,2], J. Schulz[1], M.Y. Cohen[1], D. Wahl[1,3], S. Wilson[1], S.H. Eck[1], V. Mall[2], I. Rost[1], HG. Klein[1].* 1) Center for Human Genetics and Laboratory Diagnostics (AHC), Martinsried, Germany; 2) Chair of Social Pediatrics, Technical University of Munich, Munich, Germany; 3) Practice for Genetic Counselling and Psychotherapy, Augsburg, Germany.

   Introduction: This report presents the first 16 of 50 patients with intellectual disability (ID) enrolled in the Multiple Integration and Data Annotation Study (MIDAS). Due to the extensive clinical and genetic heterogeneity, the causes of ID remain mainly unknown. Whole exome sequencing (WES) was shown to be an effective diagnostic strategy for the detection of causative mutations in patients with rare ID; however, NGS approaches are still limited by time-consuming human interpretation of genomic data. To overcome this bottleneck, MIDAS aims to accelerate NGS data analysis and to enhance the validity of the results by computer-based variant prioritization using clinical data of the patient. For this purpose, a central software system for data integration in a diagnostic laboratory based on genotype-phenotype correlations will be developed within MIDAS.   **Methods**: By now, 16 patients (8 males, 8 females) with unexplained ID and their unaffected parents (trios) were analyzed in MIDAS. All patients were evaluated by clinical geneticists and phenotypic features were recorded in a standardized way. Before enrollment, the patients had undergone extensive diagnostic testing, including karyotyping, genome-wide array comparative genomic hybridization (aCGH), and Fragile-X analysis. A subset of the patients was analyzed by targeted gene tests specific for the suspected disorder, but all these tests did not led to a diagnosis. We sequenced the patients as well as their healthy parents and used the WES data to perform a trio analysis afterwards. Identified variants were evaluated using population (dbSNP, ExAC, gnomAD) and disease databases (HGMD, OMIM, ClinVar). Identified variants were evaluated by molecular geneticists and clinicians in the context of patients' clinical presentation.   **Results**: We were able to identify four *de novo* loss-of-function mutations as well as one autosomal recessive inherited mutation in a consanguineous family. So far, the diagnostic yield for patients with ID enrolled in MIDAS and analyzed by Trio WES is 31 %. All causative *de novo* mutations were located in genes (*ADNP*, *ARID1B*, *GATAD2B*, *KDM5C*) associated to known ID syndromes.   **Discussion**: Whole exome sequencing is an effective diagnostic strategy for the detection of causative mutations in patients with rare diseases. Nevertheless, due to large amount of genetic data by NGS approaches in diagnostics, powerful computational algorithms are necessary to accelerate NGS data interpretation.

2493F

Copy number reanalysis: The hidden contribution of *MED13L* to intellectual disability. *L.K. Conlin[1], J.R. Murrell[1], R. Rajagopalan[1], V. Jayaraman[1], E.J. Romasko[1], B. Devkota[1], P. Vaidiswaran[1], P. Jayaraman[1], S. Biswas[1], M. Sarmady[1], A. Santani[1], B.L. Krock[1], M.C. Dulik[1], E. Zackai[2], I.D. Krantz[2], M. Luo[1], N.B. Spinner[1].* 1) Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA.

   While exome sequencing (ES) as a clinical diagnostic tool has advanced greatly in recent years, identification of copy number variation (CNV) using ES is not widely used clinically due to low sensitivity and high false-positive rates. In order to optimize CNV detection, we used ES data from 160 probands and compared the results to their clinical chromosomal SNP microarray data. As part of this validation, we identified pathogenic CNVs in haploinsufficient genes *MBD5, ANKRD11, SYNGAP1,* and *KCNH2*, heterozygous deletions *in trans* with a sequencing variant in *SPATA5* and *TRAPPC9,* and a likely pathogenic duplication of *FGF12*. Surprisingly, deletions in *MED13L* were identified in three probands. For all but two of these findings, the CNVs were also detected by the array, but the CNV had not been reported clinically as the disease-gene association was not known or not well understood at the time of original testing, and the CNVs were below the reporting criteria (20 probes and 200kb) for variants of uncertain significance. In particular, variants in *MED13L* were first described in a small cohort of patients with transposition of the great arteries in 2003; however, in 2013, reports began to emerge linking *MED13L* to both syndromic and non-syndromic intellectual disability (ID). The frequency of *MED13L* CNVs prompted us to re-evaluate array data from over 12,000 patients who received clinical testing since 2008. We identified 5 additional *MED13L* deletions, only one of which had been reported as a variant of uncertain significance at the time of testing. In total, 7 unique deletions in *MED13L* were identified in 9 individuals, including two brothers who inherited the deletion from a mosaic parent. The deletions ranged in size from 10kb to 214kb, with the majority of deletions involving only one or two exons. Overall, exonic deletions of *MED13L* were found in ~0.1% of individuals with array testing who presented with ID or developmental delay, compared to ~2% in the ES cohort. The differences may be due to the poor coverage of this gene in earlier array designs, thus the contribution of deletions of this gene to ID may be underappreciated. While reanalysis of ES data has highlighted the impact of new knowledge regarding gene-disease associations on improved diagnostic yield, our experience also shows the importance of reanalysis of CNV data, both as a supplement to ES sequencing as well as increasing diagnostic yields of chromosomal microarray testing.

2494W

Trio whole genome sequencing for undiagnosed patients with moderate or severe intellectual disability. *B. Cogné[1], T. Besnard[1], S. Kury[1], AS. Denommé-Pichon[2,3], S. Schmitt[1], X. Latypova[1], A. Toutain[4], P. Parent[5], B. Gilbert-Dussardier[6,7], S. Odent[8,9], A. Boland[11], V. Meyer[11], O. Pichon[1], C. Le Caignec[1], B. Isidor[1], R. Redon[10], D. Bonneau[2], JF. Deleuze[11], S. Bézieau[1].* 1) Service de Génétique Médicale, CHU Nantes, Nantes CEDEX 1, 44093, France; 2) Département de Biochimie et Génétique, CHU Angers, Angers, 49100, France; INSERM UMR 1083 - CNRS 6214, Angers, 49100, France; 3) Service de génétique, CHU de Nancy, Vandœuvre-lès-Nancy, 54500, France; INSERM UMR 954, Vandœuvre-lès-Nancy, 54500, France; 4) Service de Génétique, CHU Tours, Tours, 37044, France; 5) Service de Pédiatrie et Génétique Médicale, CHRU Brest, Brest, 29609, France; 6) Service de Génétique, CHU Poitiers, BP577, Poitiers, 86021, France; 7) EA 3808 Université Poitiers, France; 8) Service de Génétique Clinique, CHU Rennes, Rennes, 35033, France; 9) CNRS UMR 6290, Université Rennes 1, Rennes, 35033, France; 10) INSERM, UMR 1087, l'Institut du Thorax, CHU Nantes, Nantes, 44007, France; 11) Commissariat à l'Energie Atomique/Direction des Sciences du Vivant/Institut de Génomique, Centre National de Génotypage, 91057 Evry, France.

Trio Whole Exome Sequencing (WES) has proven to be a valuable strategy to identify molecular basis of syndromic or non-syndromic intellectual disabilities (ID). As part of the *HUGODIMS* project, a trio-based WES analysis was performed in 69 patients affected with moderate to severe ID without familial history of ID and normal array-CGH and fragile X testing. A definitive diagnostic or a strong candidate gene was identified for 69% of the cases (48/69). In patients without diagnosis, Whole Genome Sequencing (WGS) is able to increase coding genes coverage together with detecting structural variations (SVs) and giving an insight into noncoding variants. Up to now, however, WGS implementation for diagnostic or research purposes is still challenging due to computational and analysis issues. As a pilot study, we performed WGS in 7 patients negative after trio WES and their healthy parents on Illumina HiSeqX5 using TruSeq PCR-free. Single nucleotide variants and short indels were jointly called across the 21 samples using GATK HaplotypeCaller. For SVs, we used both read depth and split/paired-reads approaches: i) to detect copy number variants (CNVs), we generated a read count matrix over 500bp non-overlaping intervals covering the entire genome and a read-depth approach using the R algorithm cn.MOPS (Klambauer et al. 2012) was applied, calling CNVs from 3 consecutive intervals, ii) split and paired-reads analysis were performed with MANTA (Chen et al. 2016) on the 21 WGS jointly, it allowed the identification of all SVs classes (CNVs, insertions, inversions and translocations). We then developed Python scripts to identify *de novo events* and to annotate SVs with several external databases (RefSeq, OMIM, segmental duplications, DGV gold standard etc). We also extracted rare coding variants in genes covered by a SV to identify compound heterozygous events. As preliminary results, we identified two *de novo* missenses variants in the same exon of a strong ID candidate gene not covered in the previous WES analysis, a *de novo* 2Mb inversion disrupting a known ID gene and a *de novo* balanced reciprocal translocation. Confirmation of these results using another technique is underway. We will therefore report the first results of this pilot study and we will discuss actual challenges subsisting with WGS analysis.

2495T

Reanalysis of whole exome and genome data leads to new diagnoses in children with intellectual disability and developmental delay. *C.R. Finnila[1], S. Hiatt[1], M. Amaral[1], K. Bowling[1], M. Thompson[1], G. Barsh[1], E.M. Bebin[3], K. Brothers[2], K. East[1], D. Gray[1], W. Kelley[1], N. Lamb[1], J. Lawlor[1], E. Lose[1], R.M. Myers[1], C. Rich[2], S. Simmons[3], G.M. Cooper[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Louisville, Louisville, KY; 3) University of Alabama at Birmingham, Birmingham, AL.

As a part of the CSER Consortium, we are actively sequencing the genomes of children with developmental delay/intellectual disability (DD/ID) and of their unaffected parents. To date, we have enrolled 463 families (1308 individuals) and have sequenced exomes from 368 individuals (120 families) and whole genomes from 940 individuals (343 families). Of the 480 DD/ID affected children sequenced and analyzed, 59% did not receive any findings (pathogenic, likely pathogenic or VUS) related to their disease. In an attempt to identify new diagnoses in probands where there was not one previously identified, our team has elected to perform reanalysis of sequence data. With more groups performing routine exome and genome sequencing, more sequence data is available describing the genetic makeup of both healthy controls and affected individuals. The rapid accumulation of data poses both a challenge and an advantage to analysis, as available data on a given gene changes relatively quickly and integration of such data into our analysis pipeline must therefore occur on a routine basis. The increased amount of new information that is now available prompted our reanalysis efforts. To begin the reanalysis process, we annotate additional data pertaining to variants or genes (gnomAD data, ClinVar updates, publication data, and updates to impact predictions such as CADD and RVIS) then re-filter and manually curate variants. Through reanalysis efforts, we have identified 19 diagnostic variants in 22 children, representing 6% of previously undiagnosed cases. These variants were not returned at the time of original analysis due to inadequate support for pathogenicity of the variant (14/19) or the variant was originally scored as a VUS, but new evidence suggests upgrading to likely pathogenic/pathogenic (5/19). Additionally, both our first-line analysis and reanalysis efforts have greatly benefited from collaborations initiated via GeneMatcher. We have submitted 50 genes to date, and approximately 25% have resulted in an active collaboration with the goal of publication. Our findings demonstrate that reanalysis of exome/genome sequence data, as well as data sharing with clinicians and researchers, leads to increased diagnosis for those with DD/ID and contributes greatly to the scientific and medical knowledgebase.

2496F

The clinical application of chromosomal microarray in in the diagnosis of children with developmental delay/intellectual disability in Korea: A single tertiary center experience. *Y. Kim[1], L. Yoon[1], Y. Shin[2], Y. Lee[3], S. Nam[3].* 1) Department of Pediatrics, Pusan National University Hospital, Pusan National University School of Medicine and Biochemical Research Institute, Busan, Korea; 2) Department of Rehabilitation, School of Medicine, Pusan National University Hospital, Pusan National University School of Medicine and Biochemical Research Institute, Busan, Korea; 3) Department of Pediatrics, Pusan National University Yangsan Hospital, Pusan National University School of Medicine and Biochemical Research Institute, Yangsan, Korea.

Background: Chromosomal microarray (CMA) is the first-tier genetic test for children with developmental delay (DD), intellectual disability (ID), autistic spectrum disorders (ASD), and multiple congenital anomalies in many countries. In this study, we report our experiences with the use of CMA in Korean children with DD/ID. Methods: We performed CMA in a cohort of 143 children with DD/ID from Jan 2010 to Apr 2017. We retrospectively reviewed their medical records. The CMA was conducted with CytoScan 750K array (Affymetrix, USA) with an average resolution of 100 kb. Results: Eighty-one patients (56.6%) had non-syndromic DD/ID. Among 62 patients (43.3%) with syndromic DD/ID, there were 52 patients (36.4%) with dysmorphisms, 10 patients (7.0%) with epilepsy, and 6 patients (4.2%) with ASD. Among total 103 copy number variations (CNVs) identified in 73 patients (51.0%), 30 CNVs (21.0%) were pathogenic. Genetic loss of CNVs were more common in pathogenic CNVs than benign CNVs and variants of unknown significance (P=.045). Pathogenic CNVs showed bigger average size of CNVs (P<.001) and more genes between CNVs (P <0.001) than benign CNVs and variants of unknown significance. The diagnostic yields were 11.1 % in non-syndromic DD/ID and 27.4% in syndromic DD/ID. Microdeletion and duplication syndromes were diagnosed, including Phelan-Mcdermid syndrome, DiGeorge syndrome, steroid sulfatase deficiency, 16p11.2-p12.2 microdeletion syndrome, 15q13.3 microdeletion syndrome, 1q21.1 recurrent microdeletion, 6q25 microdeletion syndrome, Wolf Hirschhorn syndrome, Miler-Dieker syndrome, 7q11.23 duplication syndrome, and 16p13.11 recurrent microduplication. Conclusion: This study provides additional evidence of the clinical usefulness of CMA in the diagnostic genetic testing for children with DD/ID in Korea.   .

2497W

Case study: Identification of a pathogenic microdeletion using exome data. *M.N. Luong, W. Scaringe, C.Y. Lee, M. Li, Y. Shen, S.P. Strom, H. Gao.* Fulgent Genetics, Temple City, CA., Select a Country.

Recent advances in bioinformatics have enable the identification of Copy Number Variants (CNV) using targeted next generation sequencing. In this case study, we utilized the CNVexon™ algorithm to detect CNVs using exome data in an individual with developmental delay. A multi-gene deletion was detected, and confirmed to be a microdeletion of 2q12.3-2q13. Evaluation of the identified microdeletion using publicly available CNV databases resulted in a classification of pathogenic, as it has been previously reported in two individuals with developmental delay (ISCA, Kaminsky et al., 2011) and is absent from the general population. This novel large genomic deletion was previously detected by chromosomal microarray analysis, indicating that CNV analysis by exome sequencing is a good alternative to traditional microarray analysis for detecting and evaluating copy number variants.

2498T

Genetic evaluation of patients with intellectual disability (ID) using chromosomal microarray and next-generation sequencing at the "ID clinic".
K. Takano[1,2], T. Kosho[1,2], K. Wakui[1,2], E. Nishi[3], A. Unzaki[2,4], M. Ishikawa[2], E. Kise[2], T. Yamaguchi[1,2], R. Kawamura[1,2], M. Motobayashi[5], T. Fukuyama[6], N. Fueki[7], S. Hirabayashi[8], Y. Inaba[6,8], T. Kaname[9,10], S. Hata[10,11], Y. Matsubara[10,12], Y. Fukushima[1,2,4]. 1) Department of Medical genetics, Shinshu University School of Medicine, Matsumoto, Japan; 2) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Japan; 3) Division of Medical Genetics, Nagano Children's Hospital, Azumino, Japan; 4) Problem-Solving Oriented Training Program for Advanced Medical Personnel:NGSD Project, Japan; 5) Division of Neonatology, Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; 6) Division of Neurology, Nagano Children's Hospital, Azumino, Japan; 7) Division of Rehabilitation Medicine, Shinano Iryofukushi Center, Shimosuwa, Japan; 8) Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; 9) Department of Genome Medicine, National Center for Child Health and Development , Tokyo, Japan; 10) Initiative on Rare and Undiagnosed Diseases in Pediatrics, Japan; 11) Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development , Tokyo, Japan; 12) National Research Institute for Child Health and Development , Tokyo, Japan.

Intellectual disability (ID) results from significant limitations in both intellectual functioning and adaptive behavior and starts before the age of 18. ID is one of the most frequent developmental disorders in children, and its prevalence is 1–3% in the general population. However, more than half of ID was of unknown etiology due to its clinical and genetic heterogeneities. Recently, genome-wide approaches, such as chromosomal microarray (CMA) and next-generation sequencing (NGS) have improved the diagnostic yield of ID. The "ID clinic" was established at the Center for Medical Genetics, Shinshu University Hospital in April 2014. We provide clinical diagnosis, systematic genetic evaluation, and genetic counseling to patients with ID or global developmental delay (GDD). To date, 109 patients have visited the "ID clinic". In the first-tier genetic testing, CMA detected 6 pathogenic copy number variants. In the second-tier genetic testing, targeted NGS using the "ID panel" (49 or 80 ID-related genes) identified 15 pathogenic variants. In the third-tier genetic testing including medical exome sequencing (TruSight One Sequencing Panel, Illumina), whole-exome sequencing or specific genetic testing, genetic causes were identified in 7 patients. The total diagnostic yield was 25.7% (28/109). Genetic evaluation in the "ID clinic" is thought to be useful in providing definitive diagnosis, expected clinical course, and recurrence risk to patients and their family.

2499F

Targeted next-generation sequencing of 75 genes in Japanese patients with intellectual disability and multiple congenital anomalies of unknown etiology. *D.T. Uehara[1], K. Tanimoto[2], J. Inazawa[1,3].* 1) Department of Molecular Cytogenetics, Tokyo Medical and Dental University, Tokyo, Tokyo, Japan; 2) Genome Laboratory, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan.

Intellectual disability (ID) is a genetically heterogeneous condition affecting 1-3% of the population, often associated with other clinical findings broadly referred to as multiple congenital anomalies (MCA). The genetic diagnosis of a large subset of cases remains unknown, and identifying their causes is a core step for the consideration of targeted therapies. Aiming at detecting pathogenic single nucleotide variants leading to undiagnosed ID/MCA in Japanese patients, previously negative for genomic rearrangements, we have employed high-throughput sequencing using a custom panel composed of 75 genes. In this panel, 61 genes are the most frequently mutated known genes causing ID/MCA, selected on the basis of published studies in large cohorts. Most of those genes are causative for autosomal dominant forms of ID, consistent with the sporadic nature of ID/MCA. The remaining 14 genes were included as candidate genes that emerged from our previous microarray screenings, and were selected based on the higher probability of exhibiting haploinsufficiency. To date, we have screened 48 patients, in which likely causative variants were found in 25% of the cases. Notably, we detected pathogenic variants in genes recently implicated with ID/MCA, namely *DDX3X*, *PPP2R1A*, *PPP2R5D* and *WDR45*. Among candidate genes, we identified a missense mutation (c.536C>T, p.Ala179Val) in *GAP43*, a gene highly expressed in brain that had been previously pointed as the candidate to explain the phenotype of patients carrying microdeletions at 3q13.31. Due to the great cost-effectiveness, targeted sequencing as a first approach is ideal for the exclusion of mutations in known genes prior to the application of broader strategies such as whole-exome sequencing.

2500W

Diagnostic testing using capture-based NGS reveals a high rate of mosaicism in genes associated with neurodevelopmental disorders. *D. McKnight, W. Wang, A. Lai, K. Retterer, R. Brandon, G. Richard, A. Lindy.* GeneDx, Gaithersburg, MD.

Historically, mosaicism was challenging to detect using conventional sequencing methods and has likely represented an under-reported cause of genetic disease. Capture-based NGS with high read-depth, mosaic-aware variant calling, and reduced random sequencing error provides a more sensitive and quantitative approach for the identification of mosaicism. We conducted a retrospective study to identify genes with a high rate of mosaicism tested by NGS. The study included only likely pathogenic or pathogenic variants (L/PV) in 188 genes associated with autosomal dominant or X-linked disorders and a high positive rate (>9 reported L/PV). These genes were assigned to phenotypic groups based on the major disease association and rate of mosaicism was calculated by group. Mosaic L/PVs were observed in 9-39% of reads in AD genes, 10-25% of reads in XL genes in females, and 13-88% of reads in XL genes in males. In a two year time period, 1.4% (111/7859) of L/PVs were classified as mosaic. Mosaic variants were identified in 59/188 genes and accounted for 4.8% (111/2314) of all L/PVs in these genes. The highest rate of mosaicism was observed in genes associated with neurodevelopmental disorders (NDD); specifically 8.4% (17/202) of L/PVs were mosaic in brain malformation syndromes (BMF) and 3.2% (57/1778) in epilepsy and intellectual disability disorders (Epi/ID). Notably, some BMF genes with high rates of mosaicism were PAFAH1B1 (15.8%, 3/19), FLNA (13.6%, 3/22), and TUBA1A (12.0%, 3/25), and Epi/ID genes included SCN2A (7.8%, 6/77), PCDH19 (6.3%, 3/48), and MECP2 (4.3%, 4/93). In contrast, only 1/2378 and 1/658 L/PVs were mosaic in genes associated with cardiomyopathy/arrhythmia and neuromuscular disorders, respectively. Parental mosaicism was observed in 32 cases; mosaic L/PVs were observed in 2-35% of reads including 8 parents who had mosaic L/PVs in <5% of reads. Clinical information was available for 63% (20/32) of parents and 75% were unaffected, whereas 25% were either similarly or more mildly affected than their child. Collectively, these data emphasize that mosaic L/PVs are more common than previously reported in some genes associated with NDD. Diagnostic testing for patients with NDD should include capture-based NGS with high read-depth and mosaic-aware variant calling. Additionally, parental testing by NGS or a comparable quantitative method, should be considered when the child has a L/PV in a gene with a high mosaic rate.

2501T

Recurrent telomere captures as the mechanism producing uniquely complex 1p UPD mosaicism. *P.L. Pearson[1], A. Santos[1], F. Campagnari[3], A.C.V. Krepischi[1], A.M. Vianna-Morgante[1], M.L.R. Câmara[2], R.C.E.A. Brasil[2], L. Vieira[1], C. Rosenberg[1].* 1) Genetics and Evolutionary Biology, University of São Paulo, Sao Paulo SP, São Paulo, Brazil; 2) Center of Odontological Assistance, Faculty of Odontology, São Paulo State University, Araçatuba, SP, Brazil; 3) Deoxi Biotecnologia, Araçatuba, São Paulo, Brazil.

A complex UPD mosaicism of chromosome 1p was detected by SNP microarray analysis in a patient presenting a 4 Mb 1p36 terminal deletion with associated clinical features**.** The microarray pattern revealed a progressive distal reduction of heterozygosity frequency from the expected 50% towards the 1p36 terminal deletion, indicating the presence of mosaic uniparental disomy (UPD). UPD involving terminal deletions has been observed as a result of telomere capture, with the reduction of heterozygosity remaining constant along the entire captured segment. In our case, the unique SNP array pattern observed can be explained by overlap of differently sized UPD segments producing a stepped pattern of heterozygosity resulting from recurrent telomere captures from the normal homologue in independent cell lineages. Comparisons between blood and buccal mucosa, the only tissues available from our patient, indicated four independent telomere captures involving blood cells, and an additional fifth capture encompassing the entire 1p in buccal mucosa. The putative timing of these events suggests that the first four segments arose in a common precursor of blood and buccal mucosa, with a further UPD segment arising in buccal mucosa following embryonic differentiation of the two tissues. This latter segment contributed approximately 50% of the total UPD observed in buccal mucosa and facilitated measuring the UPD contribution of the segments with more distal breakpoints. The contribution of individual segments to the total frequency of UPD mosaicism in both tissues was approximately linearly additive. The most likely mechanism for this complex mosaicism is a chain of events initiated by occurrence of a 1pter deletion in either a gamete or a very early embryonic cell division. The deletion was "rescued" by several early, independent telomere captures, involving varying lengths of the normal 1p homologue: this resulted in a pattern of overlapping segments of partial 1p UPD, with the most distal segments exhibiting the most reduction in heterozygosity, and the most proximal, the least. Although this is the first description of multiple independent telomere capture events of a single deletion lesion, other cases probably exist but have remained undetected due to inadequate microarray procedures. High SNP density appears to be one of the most important requirements.

2502F

Genetic spectrum of limb-girdle muscular dystrophy in Taiwan. *Y.L. Lin[1], C.C. Yang[2], N.C. Lee[1], C.H. Huang[3], Y.H. Chien[1], T.M. Ko[3], W.L. Hwu[1].* 1) Medical Genetics and Pediatrics, National Taiwan University Hospital, Taipei, Taiwan; 2) Neurology, National Taiwan University Hospital. Taipei, Taiwan; 3) GenePhile BioScience Laboratory, Ko's Obstetrics and Gynecology, Taipei, Taiwan.

    Background Limb-girdle muscular dystrophy (LGMD) is a genetically and clinically heterogeneous group of muscular dystrophies. Individuals with LGMD generally show weakness and wasting of the muscles in the arms and legs, and gradually worsen over time. The muscles most affected are usually the proximal muscles. The age of onset, severity, and features vary among different subtypes of LGMD, sometimes even within the same family. The various types of LGMD are caused by mutations in different genes. At least 31 genes have been found to be associated with LGMD which is far too much to screen each gene one-by-one. The NGS approach could save both time and cost, as well as labor. **Materials and methods** We designed a targeted panel for LGMD containing 32 muscular dystrophy related genes, from which exome-captured libraries were built. Our goal is to describe the spectrum of mutations and genes related to LGMD patients in Taiwan using NGS panel we designed, identify factors associated with diagnostic findings, and evaluate the assessments of genetic variants. **Results** Thirteen LGMD patients were enrolled in the study. We were able to identify likely pathogenic (LP) or pathogenic (P) variants in 8 patients (61.5%) according to ACMG guideline. Variants were most commonly observed in *DYSF* gene (25%), then in *TRAPPC11* (12.5%) and *CAPN3* (12.5%), all of which are associated with type 2 LGMD, specifically, 2B, 2S, 2A, respectively. Two variants were found in genes associated with type 1 LGMD, *LMNA* and *HNRNPDL*, particularly 1B and 1G, respectively. The rest of the variants were associated with other myopathies, such as cardiac and distal myopathy, and Duchenne muscular dystrophy (DMD)/Becker muscular dystrophy. **Conclusion** Targeted panel helps us to molecular diagnosis of LGMD. In our cohort, type 2 LGMD is the most common form of LGMD. This benefits the management, clinical care and genetic counseling in these patients.

2503W

A novel intronic mutation in MTM1 detected by RNA analysis in a case of X-linked myotubular myopathy. *A.H. AlHashim[1,2], H. Gonorazky[2], K. Amburgey[2], S. Das[3], J. Dowling[2].* 1) pediatric neurology, king Fahad Medical City, Riyadh, Saudi Arabia; 2) Division of Neurology and Program for Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON Canada; 3) Department of Human Genetics, University of Chicago, Chicago, IL USA.

    X-linked myotubular myopathy (XLMTM) is a rare neuromuscular condition that presents with neonatal hypotonia and weakness and is associated with severe morbidities (including wheelchair, feeding tube, and ventilator dependence) and early death. It is defined by muscle biopsy features, including central nuclei, abnormal oxidative stain distribution, and type I fiber hypotrophy. Mutations in myotubularin (*MTM1*) account for all genetically solved cases of XLMTM, but have not been discovered in all individuals with characteristic clinical and biopsy features. Of note, there are some forms of autosomal centronuclear myopathy that can resemble XLMTM, such as those associated with mutations in *BIN1*, *DNM2, RYR1*, and *SPEG*, though rarely are such cases a completely phenocopy of XLMTM. In this study, we present a case that illustrates the importance of considering non-coding mutations as a cause of XLMTM and illustrate the utility of RNA analysis in individuals with a phenotype suggestive of a particular genetic diagnosis.

2504T

Development of a unified *DMPK* and *CNBP* PCR workflow for determining repeat expansions relevant to myotonic dystrophies. *J. Wisotsky, J. Kemppainen, G. Latham, B. Hall.* Asuragen, Austin, TX.

Myotonic Dystrophy type 1 and 2 (DM1 and DM2) are the most common adult onset muscular dystrophies. DM1 is caused by a trinucleotide expansion of CTG in the 3' UTR of myotonic dystrophy protein kinase (*DMPK*). Clinical phenotypes manifest in patients who have >50 repeats with classic disease occurring in those with >100 repeats. DM2 is caused by a tetranucleotide expansion of CCTG in intron 1 of the cellular nucleic acid-binding protein (*CNBP*). Pathogenicity is observed in those with >75 repeats. In both cases, expansions often include hundreds to thousands of repeat units. DM1 and DM2 molecular diagnosis relies on a complex combination of PCR and Southern blot analysis. Here, we describe a unified PCR/ capillary electrophoresis (CE) assay that interrogates repeat loci in both *DMPK* and *CNBP* and generates accurate sizing up to 200 (*DMPK*) or 75 (*CNBP*) repeats. AmplideX®PCR reagents were optimized for the repeat-primed amplification of both *DMPK* and *CNBP* repeats using distinct dye-tagged primers. Amplicons were sized on a 3500xL Genetic Analyzer (Thermo Fisher). A total of 40 DNA samples were evaluated: 20 previously-characterized genomic DNA samples (Coriell) with *DMPK* or *CNBP* expansions and 20 specimens collected from healthy volunteers. A subset of samples were also amplified by gene-specific PCR and evaluated by Sanger sequencing as an orthogonal comparison to identify interruptions within *CNBP*. The PCR/CE assay achieved accurate repeat sizing for both *DMPK* (≤200) and *CNBP* (≤75). Furthermore, samples with >200 CTGs in *DMPK* were identified in agreement with reference results. *CNBP* alleles with <25 repeats contained interruptions identified by both CE and sequencing, whereas in expanded alleles fewer interruptions were observed. Current methods fail to identify expansions in *DMPK* and *CNBP* using a streamlined assay format. We report a novel PCR based workflow that combines optimized reagents and unifies *DMPK* and *CNBP* molecular analyses. This technology can advance clinical research by enabling rapid, simple, and accurate profiling of repeat expansions in individuals with phenotypes consistent with DM1 or DM2.

2505F

Clinical correlations of a streamlined molecular assay based on AmplideX® PCR/CE technology that determines repeat size for both normal and expanded alleles in *DMPK* for myotonic dystrophy 1. *B. Hall[1], P.J. Snyder[2], J. Kemppainen[1], T.W. Prior[2], G.L. Latham[1].* 1) Research and Development, Asuragen, Austin, TX; 2) Department of Pathology, Ohio State University, Columbus, OH.

Myotonic dystrophy type 1 (DM1) is an autosomal dominant disease characterized by more than 50 CTG repeats in the 3' UTR of *DMPK*. In clinical labs, DM1 testing requires a combination of PCR and Southern blot (SB) analysis because PCR cannot reliably amplify over 100 repeats. Here, we describe results using a novel PCR technology that permits sizing of normal and expanded alleles in a single tube to identify up to 200 repeats using capillary electrophoresis (CE). The assay can identify longer repeats that can be further sized by SB analysis if desired. Genomic DNA samples were procured from 100 presumed healthy whole blood donors (Asuragen) and 100 DM1 affected patients (Ohio State University) with different repeat sizes (50-2,000) and degrees of mosaicism identified previously by SB. Sample gDNAs were amplified using prototype AmplideX® DMPK PCR/CE reagents (Asuragen) in laboratories at Asuragen or Ohio State. FAM-labeled amplicons were resolved on 3130 or 3500 xL Genetic Analyzers (Thermo Fisher). CTG-specific and gene-specific priming were optimized with other PCR components in a single reaction to produce complementary data signatures from repeat-primed and full-length amplicons. These signatures confirmed the genotypes of normal and expanded alleles without the need for separate PCR and SB, resulting in a faster and simpler assay and analysis workflow. Prototype PCR reagents could resolve more than 200 CTG repeats from as little as 20 ng gDNA and generated repeatable results. The assay was sensitive to mosaicism down to 2.5% mass fraction. Assay performance and clinical correlations are currently being assessed using samples from patients with mild to severe phenotypes over a range of expansions and different degrees of mosaicism. Current methods fail to amplify moderate-to-large repeat expansions in DM1, and allele drop-outs may be indistinguishable from frequently-encountered homozygous samples. We report the first single-tube, long-read PCR technology that can resolve zygosity, identify expansions to over 200 repeats, and flag larger expansions. This approach may significantly reduce the need for DM1 SB analysis.

2506W

NGS-based diagnostics at Newborn Screening Ontario. *L. Racacho[1,2], D.E. Bulman[1,2,3,4].* 1) Newborn Screening Ontario, Ottawa, ON; 2) Children's Hospital of Eastern Ontario, Ottawa, ON; 3) Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON; 4) Department of Pediatrics, Faculty of Medicine, University of Ottawa, Ottawa, ON.

Newborn Screening Ontario (NSO) is a public health program that screens approximately 140000 babies annually in the province of Ontario for 30 treatable disorders. NSO also provides repatriated diagnostic tests with massively parallel sequencing (MPS) to target 76 newborn screening genes and 406 nuclear encoded mitochondrial genes. We have implemented SeqCap EZ (Roche) library preparation for MPS on the MiSeq (Illumina) to deliver this service. In addition, a hybrid cloud informatics solution was adopted to quickly meet our needs without major restructuring of our information systems' ecosystem. Our first year of delivering MPS-based diagnostics has yielded 76 positive, 148 negative and 67 inconclusive findings. Of 291 patients reported thus far, 25 originated from screen positive cases. We discuss the challenges of delivering MPS-based diagnostics at NSO and present the distribution and classification of the sequence variants that we have identified to date.

2507T

Wide range of maternal heteroplasmy for inherited pathogenic mtATP6 variants. *J. Thompson[1], E. White[1], C. Gault[1], R. Batorsky[1], E. Lopez[1], T. Ross[1], D. Campagna[2], M. Fleming[2].* 1) Res & Development, Claritas Genomics, Cambridge, MA; 2) Dept of Pathology, Boston Children's Hospital, Boston, MA.

Pathogenic variants in mitochondrial DNA can cause a wide variety of diseases. Understanding the impact of these variants can be difficult because the variants may be heteroplasmic and vary across tissues within an individual and over time. Three individuals with congenital sideroblastic anemia from different families were identified by Sanger sequencing with the identical pathogenic variant in mtATP6 at an allele frequency estimated to be >80%. Sanger sequencing showed evidence of the allele in one of the proband's mothers, but not in the two other mothers or one female sibling. Given that Sanger sequencing is insensitive to low-level heteroplasmy, we employed next generation sequencing technology to determine whether the variants might be present at low levels in other family members. However, when short-read NGS technologies are used to determine low-level heteroplasmy, they are subject to potential contamination by the many mtDNA homologs found throughout the nuclear genome. We used two independent primer sets for long range PCR of mtDNA to eliminate the nuclear DNA issue. Each set of primers yielded two amplification products that covered the whole mtDNA genome. The independent amplifications also served to minimize any potential issues with PCR-induced errors. Use of these primers verified that the pathogenic variant was present in the maternal samples (1.9% to 37.2%) while not present in unaffected family members (<0.2%).

2508F

Development of a strategy for the genetic diagnosis of mitochondrial diseases in paediatric population. *G. Barcia[1], Z. Assouline[1], M. Rio[1], M. Maryse[1], S. Hanein[2], C. Fourrage[1], B. Ruzzenente[2], M. Metodiev[2], C. Bole[2], P. Nitschke[2], A. Munnich[1,2], J.P. Bonnefont[1,2], J. Steffann[1,2], A. Rotig[2].* 1) Department of genetics, Necker-Enfants Malades Hospital, APHP, Paris Descartes University, Paris, France; 2) INSERM U1163, Imagine Institute, Paris Descartes University, Paris, France.

Purpose: Mitochondrial diseases (MD) are a clinically and genetically heterogeneous group of disorders caused by mutations of nuclear DNA (nDNA) or mitochondrial DNA (mtDNA) and resulting in mitochondrial respiratory chain dysfunction. The global incidence of MD is estimated 1/4000 live birth, corresponding to 200 new patients/year in France. Clinically, childhood MD are characterized by a more severe phenotype than adult MD and by a frequent brain involvement. **Methods:** In 2015-2016, we performed whole exome sequencing (WES) or targeted next generation sequencing (TNGS) in 240 consecutive patients with a diagnosis of MD according to 5 criteria, namely clinical, biochemical, metabolic, histological and neuroradiological. We aimed to evaluate 1) in which cases is more appropriate to choose TNGS or WES as a first diagnostic step and 2) which of our inclusion criteria is the best predictor of a molecular diagnosis. **Results:** We identified pathogenic variants in 45% of patients, respectively 32% by TNGS and 47% by WES. In patients with at least 4 inclusion criteria the diagnostic yield of TNGS was 63% compared to 71% of WES. In 8 patients, WES allowed the detection of mutations in genes not involved in mitochondrial function. However, TNGS, allowed the detection of exonic deletions or duplications in 8 patients (3 with a negative WES result). In these series, increased lactic acid, abnormal respiratory chain assembly and altered respiratory chain enzyme activity were the better predictor of a positive results with causal mutations identified respectively in 58%, 62% and 45% of patients. **Conclusions:** We established a molecular diagnosis in 45% of patients in a cohort of 240 patients affected by childhood MD. In the MD diagnostic landscape, TNGS and WES are complementary approaches, but in patients with several criteria evocating a MD, in particular increased lactic acid or abnormal respiratory chain function, it is appropriate to choose TNGS as a first diagnostic step.

2509W

Complexities of mitochondrial gene testing. *C. Kaiwar[1], E.W. Klee[2], K.D. Kennelly[3], P.S. Atwal[4].* 1) Center for Individualized Medicine, Mayo Clinic, Scottsdale, AZ; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 3) Department of Neurology, Mayo Clinic, Jacksonville, FL; 4) Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL.

Single large deletions in the mitochondrial DNA are associated with progressive external ophthalmoplegia (PEO), Kearns–Sayre syndrome (KSS) and Pearson syndrome. Patients can present with a spectrum of symptoms ranging from early onset KSS with severe muscle involvement or a less severe, later onset, predominantly myopathic PEO. The size of the deletion as well as the level of heteroplasmy is negatively correlated with the phenotypic severity of the disease. We report a 39 year old woman with congenital myopathy (congenitally dislocated hips and scoliosis requiring surgery), who recently developed numbness in extremities and worsening balance. Family history was significant on the maternal side for myopathy and neuropathy of variable presentations in different members. Physical examination revealed a high arched palate, wide–based gait, weakness worse distally, absent reflexes and markedly reduced distal sensation with gait imbalance. EMG revealed bilateral median neuropathies at the wrists and patchy distal neurogenic and more diffuse myopathic changes. MRI of the head without contrast was normal. Laboratory levels for lactate and creatine kinase were normal and free and total carnitine levels were low. Muscle biopsy from two different sites showed severe denervation atrophy with extensive reinnervation. Electron transport chain analyses were reported as normal. Whole exome sequencing on a blood sample did not reveal any variants. Mitochondrial analysis showed a Variant of Uncertain clinical Significance (VUS) in *MT-CYB*, m.14841 A>G, p.Asn32Ser that was homoplasmic in the proband. A mitochondrial genome panel, including nuclear encoded mitochondrial genes, was ordered on skin fibroblasts as well as the biopsied vastus lateralis. A novel deletion m. 5975_16074del10100 was identified in the skin fibroblasts and confirmed by Southern blot. This deletion was not identified in the muscle specimen, although its presence at levels below detection cannot be excluded. The muscle specimen did have smaller deletions of multiple sizes, present at a low level of heteroplasmy, confirmed by long range PCR. These mitochondrial aberrations, varying between tissues, demonstrate that a high level of suspicion from the clinician and testing of multiple tissues is key to confirming mitochondrial phenotypes that are otherwise undiagnosed or often misdiagnosed.

2510T

Mitochondrial genome sequencing in phenotype-based panels and exome sequencing increases test sensitivity. *M.A. Reott, N.A. Rouse, R.R. Kelly, K. Patterson, K.H. Stovall, W.A. Langley, P.L. Nagy.* MNG Laboratories, Atlanta, GA.

Next Generation Sequencing (NGS) allows rapid variant analysis for identifying disease causing variants. Exclusion of mitochondrial sequencing analysis in NGS disease screening panels limits their power by ignoring the potential of pathogenic mitochondrial variants. Initially thought to be rare, mitochondrial genetic diseases represent important and common sources of disorders. Recent genetic epidemiological studies quantifying the most common pathogenic mtDNA variants have shown the incidence of clinical mitochondrial diseases is about 1 in 5000. Additionally, a survey of newborn cord bloods revealed that 1 in 200 infants harbored common pathogenic mtDNA variants. Coupled with variant assessment and curation database software, the addition of mtDNA sequencing to phenotype-based panels and exomes increases the ability to discover variants of interest that aid in the determination of patient treatment. Upon including mtDNA sequencing with the panels, we have positively identified pathogenic or potentially pathogenic variants in patient samples having no definitively pathogenic nuclear genome variants. The effect of this inclusion has been particularly beneficial in NGS panels associated with neurological disorders. The increased sensitivity of the NGS panels and the ease of ordering a single test instead of two provides added value for both clinicians and their patients.

2511F

Expert specification of the ACMG/AMP variant interpretation guidelines and application of the ClinGen gene curation framework to genetic hearing loss. *A. Abou Tayoun[1,2], A. Oza[3], S. Hemphill[3], B. Cushman[3], M. DiStefano[3], J. Shen[3,4], A. Chapin[5], A. Giersch[4], M. Kenna[6], A. Pandya[7], J. Alexander[8], H. Duzkale[9], W. Zhang[9], L. Schimmenti[10], K. Booth[11], H. Azaiez[11], H. Mutai[12], T. Matsunaga[12], L. Downie[13], H. Kremer[14], MA. Moreno-Pelayo[15], M. Morin[15], I. del Castillo[15], C. Morton[4], H. Rehm[3,4,16], S. Amr[3,4] On behalf of the ClinGen Hearing Loss Working Group.* 1) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 3) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA; 4) Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, USA; 5) ARUP Laboratories, Salt lake City, UT, USA; 6) Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; 7) Division of genetics and Metabolism, University of North Carolina School of Medicine, Chapel Hill, NC, USA; 8) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 9) Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH, USA; 10) Department of Pediatrics, Mayo Clinic, Rochester, MN, USA; 11) Molecular Otolaryngology and Renal Research laboratories, Department of Otolaryngology, University of Iowa Hospital and Clinics, Iowa City, Iowa, USA; 12) Division of Hearing and Balance Research, National Institute of Sensory Organs, National Hospital Organization Tokyo Medical Center, Tokyo, Japan; 13) 3Murdoch Children's Research Institute, Royal Children's Hospital, University of Melbourne, Victoria, Australia; 14) Department of Otorhinolaryngology and Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboudumc, Nijmegen, The Netherlands; 15) Genetics Department, Hospital Ramón y Cajal-IRYCIS-CIBERER, Madrid, SPAIN; 16) The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

The recently published ACMG/AMP variant interpretation guidelines (Richards *et al* 2015), along with the more recent Clinical Genome Resource (ClinGen) framework for assessing the clinical validity of gene-disease associations (Strande *et al* 2017), represent major milestones in genomic interpretation. Applying these frameworks, ClinGen is building an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research (www.clinicalgenome.org). The ClinGen Hereditary Hearing Loss Working Group (HL WG) was assembled in June 2016 from a team of experts including over 40 genetic counselors, researchers, geneticists, and clinicians from 12 U.S. and 14 international academic institutions on five continents. The aim of this HL WG is to create a comprehensive, standardized knowledge base of genes and variants relevant to HL which affects 1 in 300 infants, half of whom may have a genetic etiology due to one of over 400 described genes (Abou Tayoun *et al* 2016). This HL WG leveraged the unique clinical and research expertise of its members as well as both published and unpublished data to characterize key attributes of HL genes. Using this information, the HL WG has specified most of the ACMG variant interpretation guidelines relevant to HL including allele frequency cutoffs for benign and likely benign classification, functional assays supporting pathogenicity (or not), criteria for loss-of-function as a disease mechanism, and gene-specific phenotypes. The group continues to specify the remaining ACMG/AMP rules in concert with the ClinGen Sequence Variant Interpretation (SVI) Working Group. Using ClinGen's gene-disease clinical validity classification framework, over 93 gene-disease associations currently under expert review have been curated resulting in 48 strong/definitive, 25 moderate, and 20 limited/conflicting/no evidence classifications, and more are underway. Over 11,108 interpreted variants in 99 hearing loss genes have been submitted to ClinVar (May 2017), of which 653 variants (5%) across 68 genes have significant conflicting classifications. An ongoing goal of the standardization is to apply specific rules established by this HL WG to evaluate those variants, and to reach expert panel (3 star) status in ClinVar in order to resolve discrepancies in variant interpretation and to facilitate patient management and care. Funding: NIH U41 HG006834.

2512W

A comprehensive resource and guideline for the development and validation of exome-based panels for clinical laboratories. *R. Niazi[1], M. Gonzalez[1], M. Sarmady[1,2], A. Abou Tayoun[1,2].* 1) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

Exome-based panels (exome slices) are becoming the preferred diagnostic strategy for genetically heterogeneous disorders. The advantages of this approach include 1) Enabling frequent updates to gene lists without the need for re-designing 2) Reflexing to exome analysis bioinformatically without requiring additional sequencing 3) Streamlining lab operation by using established exome kits and protocols. However, exome-slices are subject to the same pitfalls of exome captures including potential lack of adequate coverage for all the relevant genes. Additionally, more bioinformatics analysis is needed during test development and validation to identify gaps in coverage and/or regions of high homology and to design ancillary assays. Currently there is no uniform standard for the development and validation of exome slices in a clinical setting. Here, we present guidelines, resources, and principles to design and validate exome slices in a clinical lab. This includes criteria for gene selection, a resource for gene technical assessment, a guideline for establishing the test's performance characteristics, and a framework for variant filtration based on the disease's relevant genetic attributes. We have categorized the most relevant and useful public resources for each of the above. However, we identified a gap in the online resources and therefore built a web-based application called ExomeSlicer, an algorithm that leverages next generation sequencing (NGS) exon-level quality metrics from 166 clinical exome samples to characterize technically challenging regions such as GC-rich, repeat, and/or high homology regions. ExomeSlicer allows users to visually inspect such regions (and any other) in genes of interest and provides scores, based on coverage and mapping quality metrics, to guide the decision for ancillary assays development. We applied all of the above to develop an epilepsy clinical exome slice. 100 clinically valid genes were selected for which more than 10 problematic regions, requiring ancillary assays, were identified by ExomeSlicer. These included repeat (e.g. *ARX* exon 2), GC-rich (e.g. *SYNGAP1* exon 1) and high homology (e.g. *CHRNA7*) regions. Finally, our overall validation demonstrate >99% sensitivity and specificity for variant detection, and 100% accuracy for filtering clinically significant variants. In summary, we provide a comprehensive resource and guideline for successful development and validation of clinical exome-based panels.

2513T

Exome sequencing leads to the identification of two novel MYO15A mutations in a German family with autosomal recessive non-syndromic hearing loss. *R. Birkenhager, S. Arndt, A. Aschendorff, R. Laszig.* Department of Otorhinolaryngology, Head and Neck Surgery, University Medical Center Freiburg, Killianstrasse 5, D-79106 Freiburg, Germany.

Hearing impairment is the most prevalent sensorineural disorder and also the most common birth defect. Approximately one of every 1000 newborns is born profoundly deaf and one of every 300 newborns has a permanent mild-to-profound congenital hearing impairment. More than half of pre-lingual hearing impairment cases occur due to genetic factors. Currently, over 120 genes and over 180 loci have been identified for non-syndromic hearing impairment. Autosomal recessive non-syndromic hearing loss (ARNSHL) is a genetically heterogeneous sensorineural disorder, with prelingual hearing loss and absence of other clinical manifestations. Based on the clinical diagnosis it is not possible to recognize in which gene mutations are present. This is only feasible in exceptional cases (7/180). The aim of this study is to identify the pathogenic gene in a non-consanguineous German family with 13 children, 8 affected with prelingual ARNSHL. Hearing testing BERA/Electrocochleography and radiological a high-resolution CT scan was made. Mutational analysis of two affected family members was performed using direct sequencing of the coding exon and intron transitions of the genes *GJB2* and *GJB6*, including deletion analysis. For investigation of autosomal recessive nonsyndromic hearing loss genes, whole exome sequencing was performed, with the "INVIEW HUMAN EXOME" platform; array Agilent Genomics SureSelectXT All Exon V5. No mutations could be identified in the DNFB1 gene locus, containing the genes *GJB2* and *GJB6*. A further targeted analysis of other genes was not possible; therefore complete exome sequencing took place. All known genes for hearing impairment were analyzed. Only in the gene *MYO15A* two heterozygous pathogenic variations were detected. A deletion c.3141delC, p.Pro1047fsX1124 and a splice-side mutation c.IVS3+1G>A, the mutations were confirmed by sanger sequencing for all 14 family members. The protein encoded by *MYO15A* (MIM 602666), is critical for the differentiation and elongation of the stereocilia and has important roles in actin organization in hair cells. Myosin *MYO15A* is localized at the tip of the stereocilia of cochlear and vestibular hair cells. Diagnostic testing for this gene is not routinely offered due to its large size. The study demonstrated that exome sequencing is a powerful molecular diagnostic strategy for ARNSHL, an extremely heterogeneous genetic disorder.

2514F

Genome and exome sequencing application in clinical diagnostics for deafness: Not quite there yet. *H. Azaiez[1], K.T. Booth[1,2], G. Yu[3], F. Bu[1], D. Walls[1], E.A. Black-Ziegelbein[1], R.J. Smith[1,2,3].* 1) Otolaryngology, University of Iowa, Molecular Otolaryngology & Renal Research Laboratory, Iowa City, IA; 2) Department of Molecular and Cellular Biology, University of Iowa, Iowa City, Iowa; 3) Iowa Institute of Human Genetics, University in Iowa City, Iowa.

The elucidation of the genetic etiology of human disorders has been revolutionized with the advent of targeted genomic enrichment and massively parallel sequencing (TGE+MPS). In diseases which display great genetic heterogeneity, such as deafness, TGE+MPS has been shown to be a proven method both for research applications and in clinical diagnostic settings. Several platforms differing by the genomic regions they cover are available such as whole exome sequencing (WES), whole genome sequencing (WGS) and disease-specific targeted gene panels (TGP). In a comparative study to assess the coverage of known deafness-causing genes and its impact on diagnosis accuracy between the different platforms, we have evaluated the data derived from sequencing 100 samples using a deafness-specific TGP we have developed; OtoSCOPE and WES. Additionally, 10 samples also underwent WGS. Bioinformatic analysis was completed on a custom pipeline and was limited to the regions that cover genes and mutations involved in deafness. The average target depth of coverage for deafness genes on OtoSCOPE was >400 reads, whereas it only reached 70 and 31 reads for WES and WGS respectively. Consequently, coverage at 10, 20 and 30X was consistently higher with OtoSCOPE than WES and WGS with 96.5% of regions of interest covered at 30X in OtoSCOPE versus 53% and 47% with WES and WGS respectively. This drop in coverage correlated with a decline in base and variant quality, an increase in uncertainty of zygosity status and, most notably, a lower detection rate of genomic variants. A comparison of variant lists after filtering for quality and minor allele frequency (≤2%) revealed that OtoSCOPE detected variants that were missed in either WES or WGS. Of those variants, some were of clinical significance and were deemed to be the causative variants for the deafness phenotype in patients. Detection of Copy Number Variations (CNVs) also differed between the different platforms with OtoSCOPE showing the most robust and accurate detection rate especially for heterozygote CNVs. In conclusion, several factors need to be considered when choosing and implementing a clinical diagnostic test, chiefs amongst them, are sensitivity, accuracy, turn-around-time and cost. We showed that for deafness, targeted gene panels such as OtoSCOPE offer higher diagnostic yields at lower cost when compared to WES and WGS making them the most sensible choice for comprehensive genetic testing for deafness.

2515W

Targeted exome sequencing as a molecular diagnostic tool for syndromic hearing loss. *Y.S. Lima[1], M.A. Chiabai[2], J. Shen[3], M.S. Córdoba[4], B.R. Versiani[4], R.A. Benício[4], R. Pogue[2], A. Pic-Taylor[1], J.F. Mazzeu[5], S.F. Oliveira[1].* 1) Institute of Biology,University of Brasília, Brasília, Brasília, Brazil; 2) Catholic University of Brasília, Brasília, Brazil; 3) Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) University Hospital of Brasília, University of Brasília, Brasília, Brazil; 5) Faculty of Medicine, University of Brasília, Brazil.

Target exome sequencing based-panels are being used successfully as a diagnostic tool for genetic diseases. In this work, we propose a panel designed for syndromic hearing loss. It is estimate that approximately 50-60% of the world's hearing loss cases have a genetic etiology and 30% of these are represented by syndromic hearing loss, which includes more than 400 syndromes. The customized Ion Ampliseq panel is composed of exons related to 51 genes associated with syndromic hearing loss, added to *GJB2*, covering 98.95% of the exons. 26 patients with a clinical diagnosis of a syndromic hearing loss with no previous molecular diagnosis were analyzed: 17 with a clinical diagnosis of a syndrome with described genes (included in the panel) and nine with syndromes with undescribed genes (Goldenhar syndrome) or without clinical diagnosis. Potentially causative mutations were found in nine patients with clinical diagnosis of syndromes with described genes and none with undescribed genes, whereas we were able to diagnose six patients, reaching a diagnostic rate of 23% (6/26). From the nine different mutations identified, three are novel mutations, related to Waardenburg, Treacher Collins and CHARGE syndromes. We noticed that all the patients diagnosed by the panel had a previous clinical diagnosis, suggesting that the panel has more effectiveness in diagnosing this group of patients. Considering only this group, the diagnostic rate would increase to 35% (6/17). Regarding patients with syndromes with described genes that we were not able to find any potentially causative mutation, we hypothesize that another gene not included in the panel can be involved or the causative mutation can be in the few regions not covered by the panel. Concerning to patients with Goldenhar syndrome or without clinical diagnosis, we think that the enormous genetic heterogeneity and variety of rare described syndromes not covered by the panel can influence the difficult to find the genetic etiology for these patients.The panel here developed demonstrates a good effectiveness in molecular diagnostic when compared to others in the literature, especially for clinical diagnosed patients with a syndrome with described genes. .

2516T

Kagami Ogata Syndrome caused by a 14q32 microdeletion that did not encompass MEG3 DMR. *W.T. Keng[1], K.H Teh[2], F. Mansor[2], Azli. Ismail[3], Z. Zakaria[3].* 1) Genetic Department, Hospital Kuala Lumpur, Kuala Lumpur, Malaysia; 2) Paediatric Dept, Hospital Sultan Abdul Halim, Kedah, Malaysia; 3) Unit Hematologi, CaRC, Institute of Medical Research, Malaysia.

   Kagami Ogata Syndrome patients have small thorax and developmental delay. Some have abdominal wall defect or contractures. Radiologically, the ribs have characteristic coat hanger configuration. The pregnancy is often associated with polyhydramnios and large placenta. This syndrome is caused by paternal UPD 14, epimutation or microdeletion on maternal chromosome 14. There are 2 differentially methylated regions [DMR] on 14q32 region, DLK1-MEG3 intergenic differentially methylated region (IG-DMR) and MEG3-DMR. The microdeletions usually involved MEG3-DMR. We describe here a patient who presented with respiratory distress at birth, small thorax and rib cage with coat hanger appearance. Chromosome 14q32methylation analysis showed normal MEG3 DMR methylation. Array CGH detected a 126.8 kb deletion at 14q32 which deleted several imprinted genes including part of MEG3 gene but not the MEG3 DMR region. There have been 5 previous reported cases of Kagami Ogata syndrome caused by microdeletion that did not involve DMR region on chromosome 14q32. Those cases and our case showed that deletion of imprinted genes without involvement of DMR is enough to cause Kagami Ogata Syndrome and methylation study on its own is not enough to rule out this syndrome.

2517F

***De novo* variant in *SOS2* with a concurrent *SHOX* deletion: Report of a patient with short stature, dysmorphic features and heart defect.** *D. Lyalin[1], L. Grote[2], S. Amudhavalli[2], J. Jenkins[2], C. Saunders[1], L. Zhang[1], E. Repnikova[1].* 1) Department of Pathology and Laboratory Medicine, Children's Mercy Hospital, University of Missouri-Kansas City School of Medicine, Kansas City, Missouri; 2) Division of Clinical Genetics, Children's Mercy Hospital, University of Missouri-Kansas City School of Medicine, Kansas City, Missouri.

   Noonan syndrome (NS) is a common autosomal dominant disorder characterized by short stature, congenital heart disease, and facial dysmorphism with an incidence rate of ~1/2500 live births. NS is one of the RASopathies, a genetically heterogeneous group of multisystemic disorders caused by the dysregulation of the RAS/MAPK pathway. To date, 75-80% of NS can be explained by heterozygous pathogenic sequence variants in the known genes. Patients with *SOS2*-related NS exhibit a wide range of clinical findings, including typical facial features of NS, cardiac anomalies, learning difficulties, and ectodermal involvement including curly hair, sparse eyebrows, and keratosis pilaris. All the missense variants in *SOS2* reported in NS patients are located in the DH domain and could result in gain-of-function mutations, leading to upregulation of the RAS/MAPK pathway.   Here we report a heterozygous *de novo* pathogenic variant, c.1127C>G (p.Thr376Ser) in exon 9 of *SOS2* identified in a 7 y.o. male with NS features. In addition, this patient has a paternally inherited ~372 kb deletion involving *SHOX*. This patient's features include short stature with a muscular build, dysmorphic features (sparse eyebrows, anteverted ears with overfolded upper helices, a broad nasal base, high arched palate), pectus carinatum, stretchy/loose skin over abdominal wall, keratosis pilaris, hyperopic astigmatism, pelviectasis, and septal hypertrophy. Analysis of the clinical features of 12 previously reported patients with *SOS2* missense variants revealed that all patients had striking ectodermal involvement and facial dysmorphisms typical of NS. In contrast, short stature, which is observed in our patient, was noted in only 2/12 patients. Our patient is the first with this variant to have septal hypertrophy, although it was previously noted in two patients with a different *SOS2* variant.   This patient was found to have two distinct genetic disorders: NS and *SHOX* deficiency. His short stature is likely due to the *SHOX* deletion and the *SOS2* variant explains the dysmorphic features and cardiac anomaly. Also, this is the first report of a patient with an *SOS2* variant and a concurrent *SHOX* deletion. Both, microarray and sequencing analyses were important for establishing the diagnosis in this patient as well as appropriate genetic counseling.

2518W

Defects in cohesin components STAG1 and STAG2 expand the locus heterogeneity of "cohesinopathies". *B. Yuan[1,2], J. Neira[1,3], T. Santiago-Sim[1,2], J. Rosenfeld[1,2], W. Jin[2], M.P. Adam[4,5], J. Dean[6], C.T. Fong[7], L. Hudgins[8], S. Madan-Khetarpal[9], M. McGuire[1,2], L. Slattery[8], L. Martin[7], L. Ramsdell[4,5], D.D.D. Study[10], W. Bi[1,2], S.W. Cheung[1,2], A. Breman[1,2], J. Smith[1,2], C. Eng[1,2], Y. Yang[1,2], R. Xiao[1,2], P. Liu[1,2].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Baylor Genetics, Houston, TX 77021, USA; 3) Texas Children's Hospital, Houston, TX 77030, USA; 4) Seattle Children's Hospital, Seattle, WA 98105, USA; 5) Department of Pediatrics, University of Washington School of Medicine, Seattle, WA 98105, USA; 6) Clinical Genetics Service, NHS Grampian, Aberdeen , AB25 2ZA, Scotland; 7) Department of Pediatrics, University of Rochester Medical Center, Rochester, NY 14642, USA; 8) Division of Medical Genetics, Stanford University, Stanford, CA 94305, USA; 9) Children's Hospital of Pittsburgh of UPMC, Pittsburgh PA 15224, USA; 10) The Deciphering Developmental Disorders (DDD) Study.

    The cohesin complex is critical in various cellular functions, including sister chromatid cohesion, recombination-mediated DNA repair, chromosomal looping and architecture, and gene transcriptional regulation. Cohesin is a multi-subunit protein complex composed of four structural components, including SMC1A, SMC3, RAD21 and STAG1/2, and numerous regulatory factors, such as NIPBL, MAU2, PDS5A/B, WAPL, and HDAC8, which are in charge of the precise progression of the cohesion cycle. Defects of the cohesion pathway have been associated with developmental disorders collectively described as "cohesinopathies", notably Cornelia de Lange Syndrome (CdLS) that are caused by loss-of-function mutations in the genes encoding SMC1A, SMC3, RAD21, NIPBL and HDAC8. We investigated subjects referred for clinical whole exome sequencing (WES) with various genetic disorders and identified pathogenic or likely pathogenic variants in the known CdLS genes including 5 in *NIPBL*, 14 in *SMC1A*, 4 in *SMC3*, 2 in *RAD21* and 8 in *HDAC8*. Interestingly, by expanding the analysis to genes involved in the cohesion pathway in addition to the well-established CdLS genes, we identified two *de novo* heterozygous likely pathogenic single nucleotide variants (SNVs) in *STAG1* and three *de novo* SNVs in *STAG2*. When we examined regions encompassing these two genes in our copy number variants (CNVs) database of subjects who were referred for chromosome microarray analysis (CMA) due to neurodevelopmental disorders, we identified two *de novo* deletions and one deletion of unknown inheritance for *STAG1* and one deletion of unknown inheritance for *STAG2*. *In silico* analysis suggested haploinsufficiency as the likely mechanism for disease-contribution. Patients identified in this study did not present with the classical distinctive facial features of CdLS, making the diagnostic process more challenging. Clinical profiling suggested overlapping features with CdLS, including developmental delay, short stature, microcephaly, micrognathia, limb abnormalities and male genital hypoplasia among others. This study therefore contributes to genetic and phenotypic expansions of "cohesinopathies". Moreover, this study demonstrates the efficacy of genome-wide assays, such as WES, in new gene discovery by allowing investigation of genes in the same pathway or with functions related to the known disease-contributing genes.

2519T

Constitutive supernumerary marker chromosomes are the chromothripsis remnant of the supernumerary chromosome present in trisomic embryos. *N. Kurtas[1], L. Leonardelli[2], L. Xumerle[2], M. Delledonne[2], A. Brusco[3], K. Chrzaowska[4], A. Schinzel[5], S. Guerneri[6], E. Manolakos[7], S. Giglio[8], T. Liehr[9], O. Zuffardi[1].* 1) Molecular Medicine, University of Pavia, Pavia, Pavia, Italy; 2) Department of Biotechnology, University of Verona, Verona, Italy; 3) Department of Medical Sciences, University of Turin, Torino, Italy; 4) Department of Medical Genetics, The Children's Memorial Health Institute, Warsaw, Poland; 5) Institute of Medical Genetics, University of Zurich, Zurich, Switzerland,; 6) Laboratory of Medical Genetics, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; 7) Laboratory of Genetics, Access to genome P.C., Thessaloniki, Greece; 8) Biomedical Experimental and Clinical Sciences "Mario Serio", University of Florence, Firenze, Italy; 9) Institute of Human Genetics, Jena University Hospital, Jena, Germany.

    We hypothesize that most *de novo* constitutional non-recurrent supernumerary marker chromosomes (sSMC) are the final outcome, via a chromothripsis event, of the partial trisomy rescue of supernumerary chromosomes present in trisomic embryos. The hypothesis is based on the following facts, fully coincident with the known biological characteristics of human trisomies: (i) sSMCs are associated with maternal age, (ii) most sSMCs are in mosaic with a normal cell line, (iii) in some sSMCs maternal UPD for the chromosome by which the marker derives was reported. Moreover, a few sSMCs appear to be constituted by non-contiguous portions of the original chromosome, a clear indication of a previous chromotripsis event. To investigate this hypothesis, we collected DNA from 20 cases of non-recurrent de novo sSMCs, already defined by array-CGH, and from their parents. We performed whole genome paired-end sequencing (WGS) in the first five cases, chosen among those in which the array-CGH had shown their constitution by non-contiguous portions. We also analysed the trios by microsatellites spread along the entire chromosome by which the sSMC originated. As a result, in four cases, sSMC2a, sSMC2b, sSMC7, sSMC17, sSMC18, a disordered assembly of chromosomal segments was highlighted. In two of them, sSMC7 and sSMC2b, we could demonstrate that the marker was even more complex, revealing some regions that were represented at different level of copy number gain (duplicated, triplicated, or even amplified) as if multiple types of sSMC were present in the different blood cells. In the fifth case, sSMC18, the non-contiguous regions remained in the reference order. In each case the novel order of the chromosomal segments was confirmed by breakpoints (bps) PCR amplification and cloning. In all bps the fusion signature showed repair mechanisms such as non-homologous end joining and microhomology mediated processes. Microsatellite and SNPs analysis in the five trios indicated a maternal origin of the sSMC18 and sSMC17 with biparental origin of the related homologous chromosomes, and a paternal one in the remaining sSMC7, sSMC2a, and sSMC2b, with maternal hetero/isodisomy of their related chromosomes. These data demonstrate both a link between numerical and structural anomalies and that early lethal trisomies may leave a dramatic legacy in the postnatal life. .

2520F

Association of 22q11.2 duplication with two variants in *FREM2* gene in a patient with an unusual phenotype. *M.I. Melaragno[1], M.Z. Souza[1], A.G. Dantas[1], M.F.F. Soares[2], E.J. Bhoj[3], D. Li[3], R.P. Silva[3], V.A. Meloni[1].* 1) Genetics Division, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Department of Imaging Diagnosis, Universidade Federal de São Paulo, Brazil; 3) Center for Applied Genomics – Childrens Hospital of Philadelphia, USA.

We report on a 6-month-old male patient, the fourth child of a non-consanguineous couple, who presents a unique phenotype with facial asymmetry, ocular hypertelorism, malformed ears, macrostomia, right-hand radial deviation, proximal implantation of a hypoplastic fifth finger in right hand, corporal asymmetry, and complex congenital heart disease. Skeleton X-ray showed cranial asymmetry, accentuation of the dorsal kyphosis and lumbar thoracic scoliosis, butterfly vertebrae anomaly in T11 and T12, subluxation of the right elbow, right radial agenesis, and shortening and bowing of the ulna. 3-D helicoidal tomography showed ribs fusion and hemivertebrae scoliosis. The cardiac evaluation revealed *cor triatriatum sinistrum*, atrial and ventricular septal defects, and pulmonary hypertension, surgically corrected. MLPA with P250 kit revealed a 22q duplication inherited from his father who presents a normal phenotype. Chromosomal microarray showed a 2.8 Mb duplication: arr[hg19] 22q11.21(18640729_21465659)×3 pat. Exome-sequencing showed two variants in *FREM2* (FRAS1 related extracellular matrix protein 2) gene, each one inherited from a parent: c.1837C>T:p.R613C, and c.2770_2772del:p.924_924del (GRCh37). Some clinical features of our patient could be explained by the 22q11.2 duplication such as facial dysmorphisms and congenital heart defects, features not found in his father. As expected, 22q11.2 duplications are frequently inherited and associated with variable phenotypes. A patient recently reported in the literature presents a smaller 22q11.2 duplication associated with a mutation in *SALL4*, a gene responsible for Duane-radial syndrome. This patient had multiple skeletal anomalies, but no Duane anomaly. Our patient presents a worst, but similar skeletal anomalies, and also no ophthalmologic anomaly. Exome sequencing in our patient revealed no rare variants in *SALL4* (Spalt Like Transcription Factor 4) gene. However, two heterozygous variants were found in *FREM2* gene, whose mutations are associated with the Fraser syndrome, an autosomal recessive disease. Our patient is a compound heterozygote for the gene *FREM2*, but his phenotype is not compatible with Fraser syndrome, although we cannot exclude the possibility of a phenotype derived from the association of two genetic alterations. Thus, the unusual phenotype in our patient remains unexplained and the diagnosis, a challenge for geneticists. Financial support: FAPESP, Brazil.

2521W

Case report of an unusual situation in which mother and daughter have two different chromosomal abnormalities. *J. Neri[1,2], E. Gil[2].* 1) Medicina, Universidade Potiguar, Laureate International Universities, Natal, RN, Brazil; 2) DNA Center,Natal, RN, Brazil.

V.F.O.A., 20 years old, female, young parents at the time of pregnancy and childbirth. She did not cry at birth, stayed one day in incubator with oxygen halo and more six days for phototherapy and observation. She is accompanied by specialists and is in stimulation therapies from the first month of life, but had significant motor retardation, moderate-severe intellectual disability and never developed speech. Physical examination revealed macrocrania, hypertelorism, alternating strabismus, ogival palate, discrete splenomegaly, genu recurvatum, intermittent hypotonia. As a complementary evaluation, she has several nuclear magnetic resonance imaging of brain describing multiple malformations, among them, cerebral atrophy, tapering of corpus callosum and important ventricular dilatation; several altered electroencephalograms; continuous thrombocytopenia. During evolution, she had repeated urinary tract infections and otitis, controlled with continued use of Nitrofurantoin, and due to the development of severe scoliosis, surgery was performed for spinal fixation. Her karyotype describes mosaic of normal cells and cells with additional material on a chromosome 16 [46,XX/46,XX,der(16)add(16)(p13.3)?], which aCGH revealed to be part of short arm of a chromosome 7 (7p22.3-p15.3). The karyotype of the father and the older sister were normal. In the karyotype of the mother, robertsonian translocation involving chromosomes 13 and 14 was observed [45,XX,rob(13;14)(q10;q10)]. The mother does not present any clinical alterations detected so far and negated miscarriages.

2522T

Breakpoint mapping in five Brazilian cases of distal 5p deletion: Influence of copy number variable regions and haploinsufficiency to clinical phenotype. *S.N. Chehimi[1], E.A. Zanardo[1], J.R.M. Ceroni[2], F.A.R. Madia[1], G.M. Novo-Filho[1], T.V.M.M. Costa[1], J.G. Damasceno[1], M.M. Montenegro[1], A.M. Nascimento[1], Y.G. Oliveira[1], C.A. Kim[2], L.D. Kulikowski[1].* 1) Laboratório de Citogenômica - Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brasil; 2) Unidade de Genética, Instituto da Criança - Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brasil.

Recent cytogenomic studies demonstrated the contribution of copy number variations (CNVs) to the phenotype variability among individuals with deletions syndromes. We used Illumina BeadChip array (San Diego, CA, USA) to map the breakpoints in five patients with 5p distal deletion (four females and one male). Four of the patients had a previous diagnosis of 5p- Syndrome by G-banded karyotype and one of them was confirmed after performing array. Array analysis revealed a 5p *de novo* deletions ranging from 17.43 Mb to 25.63 Mb (5p15.33-5p14.1) in four patients and unexpected showed a 5p interstitial deletion for one of the patients. The interstitial deletion present a breakpoint start at 4,788,892 and end at 22,219,836, while the four other deletions present start at 25,328 and with different breakpoints ends (18,022,107; 18,921,988; 22,039,679; 22,219,836; 25,658,882), respectively. Apart from the 5p deletion, we found a total of 33 benign CNVs (ranging from 5 to 13 CNVs in each patient). We observed that the patient with mild to moderate intellectual disability present the larger 5p deletion (25,6 Mb) while the other patients present moderate to severe cognitive impairment. Our analysis also showed haploinsufficiency for several different genes, including *SEMA5A*, *CTNND2*, *MARCH6* and *TERT* that have already been associated with the main clinical phenotype of the 5p deletion syndrome, even though *TERT* gene was not deleted in the patient with the interstitial deletion. This study highlights the importance of mapping the breakpoints in deletions syndromes in order to study the phenotypic consequences for both clinical and research applications. Also, the study of haploinsufficiency in the era of gene therapy and CRISPR-Cas9 technology can provide future alternative treatments and healthcare to patients with the 5p deletion syndrome and other genetic disorders.

2523F

Mosaicism and chimerism: Mechanistic answers for chromosomally distinct cell populations. *R. Rowsey, L. Baughn, H. Kearney, U. Aypar, D. Van Dyke, N. Hoppman, E. Thorland.* Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Constitutional blood specimens that contain an admixture of cell populations are uncommon and can represent either mosaicism or chimerism. The distinction is that mitotic mosaicism derives from a single fertilization, whereas chimerism derives from separate zygotes. Distinguishing between the two requires techniques other than G-banded karyotyping, such as a SNP-microarray. With an understanding of SNP patterns, this testing can provide mechanistic answers to the origins of the multiple cell lines. Before the routine use of SNP-based arrays, most recognized cases of chimerism were those in which XX and XY cell lines were present, and therefore were unlikely to be a result of mosaicism. Today, examination of the allele patterns on a SNP-microarray will easily demonstrate two non-identical cell lines, regardless of sex chromosome complement. The allele patterns also reveal if the two cell lines present are related. For example, in a case of twin chimerism, we observed several regions (approximately 25%) across the genome with a normal allele pattern, implying both cell lines have identical alleles in those regions, whereas in a mixture of unrelated cell lines share no identical allelic regions, and therefore a noisy allelic pattern throughout the genome. Perhaps the most challenging mosaic/chimeric cases are those that contain multiple abnormal cell lines, as it becomes difficult to understand the mechanism and to predict the clinical implications. We observed two patients with a novel combination of cell lines that highlights these challenges. They each had, two abnormal cell populations: one XX cell line with an unbalanced Robertsonian translocation (der(13;13) in one patient and der(21;21) in the other), and one cell line with monosomy X and no translocation. Utilizing SNP-microarray on the patient with the der(21;21), several mechanistic factors were inferred. First, we observed a normal allele pattern for most of the genome, consistent with mosaicism. Second, the allele pattern on chromosome 21 confirmed the der(21;21) was an isochromosome, as commonly observed in cases of der(21;21). Finally, the array showed X chromosome heterozygosity, confirming one X was lost to form the 45,X cell line, effectively ruling out the possibility of Y chromosome material being present. These example cases will demonstrate that the academic puzzles leading to a mechanistic understanding of mosaic/chimeric cases can also provide clinically relevant information.

2524W

Comparison of diagnostic yield between clinical exome sequencing and whole exome sequencing. *Y. Niu[1], H. Yen[2], Y. Xue[1,2].* 1) Fulgent Genetics Inc, 2250 N. Druid Hills Road Suite 110, Atlanta, GA 30329; 2) Fulgent Genetics Inc, 4978 Santa Anita Ave., Suite 205, Temple City, CA 91780.

Background: Next generation sequencing enables exome sequencing rapidly become a common molecular diagnostic test for individuals with genetic disorders. Whole exome sequencing (WES) captures all coding regions in the genome, namely ~20,000 genes which may or may not have a disease association. To better facilitate clinical utility of the technology, clinical exome sequencing (CES) was created which captures genes associated with human diseases (~5,000 genes). The advantages for CES are easier interpretation of variants found in known disease-associated genes and better sequencing coverage. However gene discovery happens at a rapid pace and new genes are associated with diseases on monthly basis if not weekly. Therefore only whole exome sequencing make analyzing these newly discovered genes possible. **Methods and materials:** A hundred consecutive cases of each CES and WES were investigated. The following items were compared: diagnostic yield, testing indication, patient age and gender, number of reported variants. Diagnostic finding was defined as one pathogenic (P) or likely pathogenic (LP) variant identified in phenotype match autosomal dominant disease or two P/LP variants identified in phenotype match autosomal recessive disease. In addition, we examined if the gene is in CES gene list when diagnostic finding was identified in WES test. **Results:** There is no age and gender difference of patients tested for CES or WES. As literature reported, the main testing indications for both CES and WES are intellectual disability, developmental delay, epilepsy, neuromuscular symptoms, dysmorphic features, etc. The diagnostic yields for CES and WES are similar with CES being 28% and WES being 25%. The numbers of reported variants are also similar. For WES cases with diagnostic finding, 84% (21/25) of cases have the genes in CES capture set meaning these cases could be diagnosed by CES as well. There are 4 WES cases that genes with diagnostic findings are only in WES capture set. These genes have been associated to diseases within the last three years. **Conclusion:** Both CES and WES are effective genetic testing approaches for individuals with symptoms that suggest extreme heterogeneity. Analyzing more genes in cases of WES may not necessarily lead to higher diagnostic yield. However, WES has the benefit of capturing newly disease-associated genes that are missed by CES.

2525T

Familial interstitial deletion 1(q43q44) due to maternal complex balanced insertional translocation (IT) and inversion in 3p. *Y. Hadid, Z. leibovitz, H. Shapira, M. Ziv, J. Levitaz, C. Furman, R. Broneshter, R. Nabwani, S. Atalla, H. Bar El, A. Pavlenco, A. Shalata.* Bnai Zion, Haifa, Israel.

Balanced interchromosomal insertional translocation (IT) is a rare chromosomal rearrangement that occurs when a part of one chromosome inserts into a non-homologous chromosome. A 41 year old mother was examined at 24th week of her ninth pregnancy. The examination demonstrated: small fetal head circumference, enlarged colpocephalic lateral ventricles and narrowed, widely separated vertically oriented frontal horns, and near complete agenesis of the corpus callosum. The hemispheres were small with widening of the subarachnoid space. Otherwise, the fetal anatomical scan and growth were normal. Sonographic evaluation after one week demonstrates the same pathologic finding and the pregnancy was terminated. The family history reports a normal large inbreeding family compound of healthy parents of Jewish origin with seven healthy siblings and one affected child with a 1q43q44 8.1Mb deletion (arr[hg19] 1q43q44(240,676,914-248,844,657)x1). Using a SNP Microarray from TOP fetal blood, the same deletion was detected as previously reported in their affected daughter. G- banding karyotyping and FISH studies (using ASI analysis software), revealed a maternal balanced insertion of the band 1q43q44 into the band 3p11 and a paracentric inversion between band 3p11 and 3p23. In conclusion when a CNV is detected in a patient and is not found by array in parental genomes, the deletion or gain is considered to be De-novo. However, to rule out a balanced IT/ cryptic changes in the parents a further investigation (e.g. G-banding karyotype and FISH) are crucial, especially when the finding is repeated in the family.

2526F

Investigating complex structural variants using third generation genome sequencing and mapping technologies. *R. Rajagopalan[1,2,3], D.A. McEldrew[1], P. Vaidiswaran[4], M. Zaykaner[1], V. Jayaraman[1], J. McCaffrey[3], J. Sibert[3], E. Young[3], S.J. Pastor[3], M. Xiao[3], L.K. Conlin[1].* 1) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia PA; 2) Dept. of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia PA; 3) School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia PA; 4) Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia PA.

Structural variations (SVs) are genomic aberrations greater than 1kb and include deletions, duplications, inversions, translocations and other complex rearrangements. While chromosomal microarray (CMA) is the gold standard diagnostic test for copy number variants (CNV), cytogenetics is still required to detect structural rearrangements. While whole-genome short-read sequencing (WGS) is poised to replace CMA as a standard diagnostic test, it suffers from short read lengths, mappability issues, and reference bias for accurately detecting complex rearrangements. Here, we set out to investigate the utility of third generation technologies such as optical genome mapping (>70x), 10X linked-read sequencing (30x) and mate-pair sequencing (30x) compared to the current standards such as WGS (30x), SNP array and karyotyping for two patients with constitutional chromothripsis-like events involving chromosomes 1 and 22. In order to understand the significance of the SVs in these individuals, both copy number and precise breakpoint determination are required for clinical interpretation. We identified hundreds of structural variations per sample from the WGS data using Delly (read-pair, split-read), Pindel (split-read) and CNVnator (read-depth). SVs identified from WGS were enriched for false-positive variants originating from the complex regions of the genome, and many breakpoints were not identified. For chr1q, we identified 23 novel adjacencies using mate-pairs (>3 fragments), 24 breakpoints using 10X linked-reads (>9 barcodes), 74 insertions, deletions and intra-chromosomal translocations using optical maps. Mate-pair sequencing resolved the most of the breakpoint regions associated with CNVs from SNP array and the optical mapping identified more novel insertions compared to the other methods. However the default pipeline for the optical maps do not provide the origins of the insertions. All three technologies identified mutually exclusive breakpoints and novel adjacencies unseen by the array. In contrast, all three technologies failed to resolve all the breakpoints on chr22. This may be due to the insertions in the acrocentric p-arm, or that this individual only had one large, derivative chr22. Our data suggest that the third-generation mapping and sequencing technologies show great promise for investigating complex SVs as compared to the short-read sequencing; however, no single technology could resolve the complete picture with existing analysis tools.

2527W

Bohring-Optiz syndrome caused by an ASXL1 mutation inherited from a germline mosaic mother. *D. Copenheaver[1], S. Bale[1], M. Deardorff[2], E. Bedoukian[2].* 1) GeneDx, Gaithersburg, MD; 2) Children's Hospital of Philadelphia, PA.

Introduction: Bohring-Opitz syndrome is a rare congenital disorder characterized by intrauterine growth retardation, poor feeding, mental retardation, trigonocephaly, prominent metopic suture, exophthalmos, nevus flammeus of the face, upslanting palpebral fissures, hirsutism, and flexion of the elbows and wrists. There have been approximately 60 reported cases of Bohring-Opitz syndrome and all have been reported to be due to de novo ASXL1 gene mutations (1). **Case Report:** We report a case of a 3 year old female who presented with intrauterine growth retardation, failure to thrive, microcephaly, seizures, choanal atresia, atrial septal defect, chronic respiratory failure, a facial hemangioma and facial dysmorphism. Previous testing included normal microarray and CDH7 sequencing. Both parents are cognitively normal without congenital malformations. The mother, 34 years old, has a history of Crohn's disease, immune thrombocytopenic purpura and degenerative disc disease. There are three normal siblings and no pregnancy losses. **Results:** Trio whole exome sequencing revealed a heterozygous pathogenic nonsense variant in ASXL1 (p.R965*; c.2893C>T). The sequencing data demonstrated that the variant was inherited maternally but indicating that this maternal variant was underrepresented in comparison to the normal allele. These results suggested mosaicism for the variant in the mother. Additional testing on maternal buccal cell DNA was also consistent with mosaicism. **Conclusion:** This is the first report of Bohring-Opitz syndrome caused by a mutation inherited from an unaffected, somatic mosaic parent with presumed germline mosaicism. This phenomenon has been reported for other traditionally de novo dominant disorders like CHARGE syndrome and Cornelia de Lange syndrome (2)(3). This case demonstrates that haploinsufficiency of ASLX1 is not lethal in the germline, and emphasizes the need for accurate low but non-negative recurrence risk counseling for families with children with Bohring-Opitz syndrome. 1.Hoischen, A., et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. Nature Genet. 43: 729-731, 2011. 2.Jongmans, M.C., et al. Familial CHARGE syndrome and the CHD7 gene: a recurrent missense mutation, intrafamilial recurrence and variability. Am J Med Genet A. 146A: 43-50, 2008. 3.Slavin, T.P., et al. Germline mosaicism in Cornelia de Lange syndrome. Am J Med Genet A. 158A(6): 1481-5. 2012.

2528T

Concurrent exome and copy number variation (CNV) analyses enable more precise diagnoses and shorten diagnostic testing time in patients with pediatric disorders. *A.V. Dharmadhikari[1], A. Braxton[1,2], S. Narayanan[1,2], P. Ward[1,2], J. Zhang[2], W. He[1], F. Vetrini[1], J. Beuten[1], X. Ge[2], M. Tokita[2], T. Santiago-Sim[2], J. Smith[1,2], P. Stankiewicz[1,2], S.W. Cheung[1,2], C. Bacino[1,2], A. Patel[3], A.M. Breman[1,2], X. Wang[1,2], M. Leduc[1,2], P. Liu[1,2], M. Walkiewicz[1,2], R. Xiao[1,2], F. Xia[1,2], J.R. Lupski[2], C. Eng[1,2], Y. Yang[1,2], W. Bi[1,2].* 1) Baylor Genetics, Houston, TX; 2) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY.

Exome sequencing detects single nucleotide variants (SNVs) and small indels in the coding region of the genome and has been increasingly applied in clinical genomics for molecular diagnosis. However, using exome data to detect CNVs is still in its early stage. Chromosomal microarray analysis (CMA) on the other hand has been an established method in the clinical lab to detect CNVs for various clinical indications including intellectual disability and multiple congenital anomalies. Here, we describe the findings from concurrent CMA and exome testing (WESCMA) performed at Baylor Genetics in order to assess the diagnostic effectiveness of detecting SNVs and CNVs simultaneously. We analyzed 440 consecutive patients who underwent WESCMA testing between September 2013 and March 2017. CMA testing was performed using custom designed Agilent microarrays with exonic coverage of >4,800 disease genes and 'backbone' genome-wide coverage of one interrogating oligonucleotide/30kb. Each sample was also concurrently analyzed by a HumanExome Beadchip SNP array for quality-control of exome data. A molecular diagnosis was established in 126/440 (29%) patients and consisted of pathogenic SNVs identified in 106/440(24%) patients by exome and pathogenic CNVs identified in 24/440 (5%) patients by CMA. One patient had a homozygous deletion involving *TNNT1* and *TNNI3*, which was detected by CMA and also by NGS read depth. Examples of pathogenic CNVs include also a *de novo* ~2Mb 1p36 deletion associated with monosomy 1p36 syndrome, monosomy X in a patient with Turner syndrome, and a 7q11.23 deletion in the Williams-Beuren syndrome region. Notably, in 4/126 (3%) patients with a molecular diagnosis, both diagnostic SNVs and CNVs were detected. One patient with end-stage renal disease had a heterozygous pathogenic variant in the *WDR19* gene and an intragenic deletion of ~4 Kb including part of the *WDR19* gene on 4p14. The other three patients have a dual molecular diagnosis with pathogenic CNV and SNV mapping to different loci. Of the 24 patients with causative CNVs detected by CMA, the SNP array analyses performed for WES quality control detected 15/24 (62%) CNVs. The positives from the diagnostic CMA assay can also be used as controls to validate CNV deletions from exome sequencing data. In conclusion, the approach of concurrent detection of point changes, small indels and large CNVs provided more precise diagnoses, and shortened the diagnosis time by avoiding sequential studies.

2529F

Diagnostic exome sequencing (DES) coupled with rules-based candidate gene analysis identifies a causative *RRAS* lesion in a patient with a novel RASopathy. *C.B. Mroske[1], J.M. Hunter[1], K.L. Helbig[1], D.N. Shinde[1], S.J. Hipp[2], S.L. Whiteway[2], J.M. Berg[2].* 1) Clinical Genomics, Ambry Genetics, Aliso Viejo, CA; 2) Pediatrics, Brooke Army Medical Center, San Antonio, TX.

Germline alterations in the RAS/ MAPK signaling network underlie a group of disorders collectively referred to as RASopathies. RASopathies, which include diseases like Noonan Syndrome, are characterized by a wide spectrum of variable symptoms such as cardiac and growth defects, facial dysmorphism, cognitive deficits, and predisposition to certain malignancies. Flex *et al.* recently reported two unrelated patients in whom germline alterations in the *RRAS* gene, which is a member of the RAS/ MAPK network hitherto unassociated with RASopathy, constituted the molecular etiology of symptoms that were nevertheless consistent with RASopathy. These patients, who were both female, harbored the respective germline *RRAS* alterations, p.G39DUP and p.V55M. Herein, we report a 7 year-old male who is unrelated to the patients reported by Flex *et al.* and who displays symptomology consistent with RASopathy. By employing family-trio DES coupled with rules-based candidate gene analysis, we show that our proband's causative lesion is the previously reported *de novo*, germline *RRAS* alteration, p.G39DUP. Like the previous patient with this alteration, our proband displays motor delay and has a thick lower lip. Although our proband has not been diagnosed with a specific hematologic malignancy, he does exhibit splenomegaly and thrombocytopenia, both of which are associated with JMML. Notably, the aforementioned patient reported by Flex *et al*. developed AML, which was suspected to be secondary to JMML. Additional RASopathy features manifested by our proband that were not described in the previously reported *RRAS* patients include speech delay; VSD; scoliosis with pelvic tilt; craniosynostosis; joint hypermobility; hydrocele; and chronic fatigue. In summary, DES coupled with rules-based candidate gene analysis allowed us to efficiently determine the molecular lesion underlying our proband's symptoms. Not only do our data recapitulate the findings of Flex, 2014, they also expand the number and types of symptoms observed in patients with germline *RRAS* alterations. We propose that *RRAS* be added to the list of genes that are commonly sequenced in individuals suspected of having a RASopathy. We recommend that patients with *RRAS* alterations be monitored carefully for the emergence of hematological malignancy. The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of the Army, the Department of Defense, or the U.S. Government.

2530W

Clinical evaluation with Holm´s criteria for Prader-Willi syndrome in a cohort of 15 Mexican pediatric patients. *R. Lara-Enríquez[1], C. García-Delgado[1], A.B. Morales-Jiménez[1], L.B. Munóz-Martínez[1], A. Cervantes[2], V.F. Morán-Barroso[1].* 1) Genetics Department, Hospital Infantil de México, Federico Gómez, México City, México City, México; 2) Genetics Service, Hospital General de México Dr. Eduardo Liceaga. Facultad de Medicina, Universidad Nacional Autónoma de México, México City, México.

   Introduction: Prader-Willi Syndrome (PWS, MIM #176270) is an imprinting disorder caused by the absence of paternally expressed genes in 15q11.2-q13. PWS has a prevalence of 1 per 10,000 to 30,000 live births, and is characterized by neonatal hypotonia, hypogonadism, facial dysmorphic characteristics, psychomotor retardation, obesity, and intellectual disability. A paternal deletion is identified in 70% cases; 20 to 25% are due to maternal uniparental disomy, 5% are due to abnormalities in the imprinting control center, and a few cases are caused by chromosome rearrangements. **Objective:** To describe the type and frequency of clinical features in PWS Mexican patients with FISH analysis for del(15)(q11.2-q13). **Methodology:** From 2007 to 2016, 15 PWS were reviewed by direct clinical evaluation and/or medical chart review using Holm´s criteria (HC). Cytogenetic GTG banding and FISH analyses using the *Prader Willi/Angelman Region Probe* – Tricolor Probe (Vysis®) were performed in peripheral blood. Three cases had other molecular studies. **Results:** The patients were 11 males (73%) and 4 females (27%) from 1 to 14 years old; 14 (93%) had a GTG karyotype analysis and only in one of them a del(15)(q11.2q13) was identified. All the patients had a FISH analysis (5 were performed in other Institution), in 10 (67%) patients a del(15)(q11.2q13) was identified. Patients were allocated in two different groups: Group I (GI) 10 FISH-positive patients with clinical diagnosis accordingly to HC; Group II (GII) 5 FISH-negative patients with clinical PWS diagnosis. Two patients of Group II had a methylation test. 30% of patients under 3 years old did not comply to HC, eventhough they had a PWS molecular diagnosis. Four patients of the GI were under 3 years old and only one of them had less than HC 6 points (pts). All patients over 3 years old had 8 pt except in one case that had 6.5 pts. Regarding GII, 2 patients had 8 pts for HC, and the rest were above 7 pts. **Discussion**: We confirm a higher association between specific symptoms such as feeding problems, characteristic facies, hypopigmentation, with paternal del(15)(q11.2q13). PWS was more frequently diagnosed in males, we considere that it may be associated to clinical manifestation of hypogonadism were more evident in them. This study suggests that the reported frequency of del(15)(q11.2q13) in PWS Mexican patients is similar to the reported frequency. HC were useful for the diagnosis of PWS in Mexican patients.

2531T

Diversity of JAG1 mutations in Japanese patients with Alagille syndrome. *T. Togawa, K. Ohashi, T. Tokio, K. Ito, T. Endo, S. Saitoh.* Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan.

   Alagille syndrome (ALGS; OMIM 118450) is an autosomal-dominant multisystem disorder affecting the liver, heart, face, eyes, skeleton, and other organs. The prevalence is estimated to be 1 in 70,000 live birth. ALGS is caused by mutations in one of two genes in the Notch signaling pathway, *JAG1* or, rarely, *NOTCH2*. *JAG1* is located in 20p12 and consists of 26 exons. Mutations in *JAG1* can be identified in approximately 90% of clinically diagnosed patients with ALGS. To date, over 500 *JAG1* mutations have been reported from Western countries, while *JAG1* mutations of Japanese patients with ALGS have not been provided in large series. We aimed to elucidate the diversity of *JAG1* mutations in Japanese patients with ALGS. We carried out molecular genetic analyses with targeted next-generation sequencing (NGS) and/or multiplex ligation probe amplification (MLPA) covering the entire exons of *JAG1*. NGS analysis was performed with the Ion Torrent Personal Genome Machine system (Thermo Fisher Scientific) and MLPA was performed with the SALSA MLPA probemix P184-C2 JAG1 (MRC-Holland). We detected 42 pathogenic variants in 42 unrelated patients who were diagnosed or highly suspected as ALGS. Thirty-five of 42 pathogenic variants were confirmed by the targeted NGS analysis and the other 7 were by MLPA. Regarding the mutation type, twenty-one single nucleotide variants (8 missense (19%), 10 nonsense (24%), 3 splice site (7%)), 9 small deletions (21%), 5 duplications (12%), 4 whole-exon deletions (10%), and 3 single exon deletions (7%) were detected. This prevalence of *JAG1* mutations by mutation type was similar to that reported in the previous report (Leonard LD, et al. Eur J Hum Genet, 2014). Twenty-eight (65%) of all the mutations were presented in 7 exons of *JAG1*; exon 2, 4, 6, 10, 17, 23, and 24. In contrast, there was no mutation in 13 exons; exon 1, 7, 8, 11, 14, 15, 16, 19, 20, 21, 22, 25, and 26. Distribution of mutations in *JAG1* was also similar to that of the previous study from Western country (Warthern DM, et al. Hum Mutat, 2006). By our 23 segregation analyses using parents' samples, we confirmed 17 (74%) as de novo, 6 (26%) as of maternal origin, and no paternal origin. These results could provide the characteristics of *JAG1* mutations in the patients with ALGS in Japan.

2532F

Profile and characterization of genetic variants identified in 208 patients with Kabuki syndrome assists in pathogenicity determination. *C.F. Li[1], E. Wei[2], C. Baldwin[2], J.M. Milunsky[2].* 1) Calgary Laboratory Services, Dept. of Pathology and Laboratory Medicine, University of Calgary, Calgary, AB, Canada; 2) Center for Human Genetics, Cambridge, MA, USA.

Kabuki syndrome is characterized by distinctive facial features, multiple congenital anomalies, and intellectual disability. It is caused by mutations in the lysine methyltransferase 2D (*KMT2D*) or lysine demethylase 6A (*KDM6A*) gene, which are involved in normal histone methylation. Patients with clinical features of Kabuki syndrome along with their parents if available were tested using Sanger sequencing and MLPA for both genes. Genetic variants found were classified into pathogenic, likely pathogenic, uncertain significance, likely benign and benign according to the American College of Medical Genetics (ACMG) standards and guidelines for interpretation of sequence variants. There were 179 genetic variants of *KMT2D* found in 203 patients and 8 healthy parents from 194 families, and 4 genetic variants of *KDM6A* found in 5 patients and 1 healthy parent from 4 families. The pathogenic and likely pathogenic variants are due to frameshift, nonsense mutation, splicing changes, and missense mutation that substitutes an amino acid with different properties. The variants of uncertain significance are due to missense mutation that substitutes an amino acid with different properties, missense mutation that substitutes an amino acid with same properties, variant in an intron or silent substitution not reported, in-frame deletion or duplication, variant in an intron or silent substitution near an acceptor splice site not affecting splicing, and variant in an intron that appears to reduce splicing. The likely benign and benign variants are due to missense variants that substitute an amino acid with the same properties, missense variants that substitute an amino acid with different properties found in an individual without phenotype or having a co-occurring pathogenic variant found at different site or a variant in an amino acid that is not conserved, in-frame deletion, duplication or insertion, variant in an intron or synonymous variant found in individual without phenotype. Hence, frameshift, nonsense mutation and a change in splicing were only found in pathogenic or likely pathogenic variants, while missense variants that substitute an amino acid with the same properties was only found in benign, likely benign, or variants of uncertain significance. In addition to the ACMG criteria for interpretation of genetic variants, analysis of the substituted amino acid properties may further assist in the classification of the missense variants found.

2533W

Molecular diagnosis of tuberous sclerosis complex by next generation sequencing in pediatric patients from Mexico. *M.E. Reyna-Fabián[1], N.L. Hernández-Martínez[1], M.A. Alcántara-Ortigoza[1], J.T. Ayala-Sumuano[2], A. Varela-Echavarría[3], A.E. González del Ángel[1].* 1) Genetics, Instituto Nacional de Pediatría, Mexico City, Mexico; 2) Universidad Politécnica de Huatusco. Huatusco, Veracruz, Mexico; 3) Instituto de Neurobiología, Universidad Nacional Autónoma de México. Querétaro, México.

The tuberous sclerosis complex (TSC) is caused by mutations in the tumor suppressor genes *TSC1* and *TSC2*. To date, more than 1,500 disease-causing mutations have been described in these genes, where conventional molecular biology techniques as PCR and Sanger sequencing have helped in the identification of small mutations in ~80% of individuals who meet definitive diagnosis; however, the absence of an identification of the responsible genotype in the remaining cases difficult confirmatory diagnosis, genetic counseling and treatment. The Next Generation Sequencing (NGS) has emerged as an alternative approach to identify pathogenic variants in either coding and noncoding gene regions, besides variants in somatic mosaicism (~6% of the TSC cases) and small and large gene deletions/duplications (present in 17% of the TSC cases). In this study, we included five TSC pediatric patients that fulfill definitive diagnosis according to the most recent criteria and in which no mutation was previously been identified by Single Stranded Conformational Polymorphism and PCR-Sanger sequencing. The coding and noncoding regions from both *TSC1* and *TSC2* genes was subjected to NGS by an Illumina platform and analyzed for pathogenic variants using the Galaxy suite. The results lead us to identify disease-causing non-sense variants in three cases, two of those in *TSC1* (c.2102C>T, p.Gln701*; c.2227C>T, p.Gln743*) and one in *TSC2* (c.2448dup, p.Asp817*). According to the search in ExAC, LOVD, dbSNP and HGMD databases, two out of those three variants have previously been reported in TSC cases, whereas the other one resulted in a novel variant (p.Asp817*). The Sanger sequencing performed in probands and their parent's confirmed two cases as unique and one familial case that carries the p.Gln743* pathogenic variant. The next step is to evaluate the presence of mosaicism in the other two patients in which no pathogenic variant was found. The results show that deep sequencing is an effective strategy for variant detection in those cases where conventional diagnostic methods were not effective. Identifying the disease causing genotype with new methodologies as NGS allows to extent the molecular diagnosis to other familial members in order to discard minimal expression of the disease and to suggest prenatal diagnosis.

2534T

Application of next generation sequencing in NICU experiences from a 1239-patient pilot study. *W. Zhou, L. Yang, H. Wang, B. Wu.* Division of Neonatology, Children's Hospital of Fudan University, Shanghai, China.

OBJECTIVE: To evaluate the clinical utility of using the focused medical exome as a first-tier screening /diagnostic tool for patients with suspected genetic diseases in NICU. METHODS: A total of 1239 patients enrolled into levels 3 and 4 NICU with suspected genetic conditions were offered for a ~2700-gene focused medical exome. The vast majority of patients were tested as singletons, no selections were made based on the type of genetic etiology, family history or prior testing. The study include 723 (58.4%) female and 516 (41.6%) male patients. RESULTS: Pathogenic or likely pathogenic variants in disease genes associated with patients' clinical phenotype were identified in 110/1239 (8.9%) patients. In addition, 133 (10.7%) patients were found to have variants of uncertain significance (VUS) in disease genes that were highly or partially associated with patients' clinical presentations. These include 19 patients with 1 pathogenic variant/1 VUS, 27 patients with 1VUS/1VUS in genes for autosomal recessive conditions and 87 patients with variants in genes for autosomal dominant or X-linked conditions. The primary abnormalities for these neonates include cardiovascular, neurological, metabolic, gastrointestinal, hematological , respiratory , allergy/immunologic/infectious disorders, dermatological, intrauterine growth retardation, urogenital and craniofacial malformations. Recurrent positive findings include ABCC8 gene for congenital hyperinsulinism, G6PD for Favism, BTK for Agammaglobulinemia, X-linked 1, KCNQ2 for neonatal seizures, COL7A1 for Epidermolysis bullosa dystrophica, MUT for Methylmalonic aciduria. For 35 patients who had pathogenic/likely pathogenic or even VUS findings in ABCC8, G6PD (G6PD deficiency), F7 (Factor 7 deficiency) and F8 (8 deficiency) genes, since the molecular findings were highly related to the patients' phenotype, immediate medical intervention was implemented right after the preliminary molecular findings were made. CONCLUSION: Our results successfully demonstrate the focused medical exome is a powerful tool for first tier screening /diagnosis in NICU. It complements the metabolic disease-focused new born screening, detects broad spectrum of conditions and enables early intervention and decision making for a significant portion of NICU patients. The results of this pilot study enables us to promote the same strategy for future NICU care in China.

2535F

Rapid Paediatric Sequencing (RaPS): Implementation of a framework for rapid genetic diagnosis in critically ill children using whole-genome sequencing. *H. Williams[1], L. Boukhibar[1], E. Clement[2], S. Drury[3], W. Jones[2], L. Ocaka[1], A. Gagunashvili[1], P. Le Quesne Stabej[1], C. Bacchelli[4], L. Jenkins[5], M. Bitner-Glindzicz[6], J. Hurst[2], N. Lench[3], M. Peters[7], P. Beales[1].* 1) GOSgene, Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom; 2) Clinical Genetics and Genomic Medicine, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom; 3) Present address Congenica Ltd, Bioinnovation Data Centre, Wellcome Genome Campus, Cambridge, United Kingdom; 4) Experimental and Personalised Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom; 5) NE Thames Regional Genetics Laboratory, Great Ormond Street Hospital, London, United Kingdom; 6) Genetics and Genomic Medicine, UCL Great Ormond Street Institute of Child Health, London, United Kingdom; 7) UCL Great Ormond Street Institute of Child Health and Great Ormond Street NHS Foundation Trust, London, United Kingdom.

For children with undiagnosed genetic diseases on paediatric intensive care units (PICU) there is a pressing need to identify the molecular cause of illness rapidly. This may inform clinical management and sometimes permit therapeutic intervention to minimise disease related morbidity. The use of whole genome sequencing (WGS) alleviates many of the deficiencies associated with targeted sequencing approaches but in turn introduces novel challenges that need to be overcome before WGS is suitable for routine diagnostic testing. Our Rapid Paediatric Sequencing (RaPS) study involves a co-ordinated multidisciplinary team, comprising the academic research group GOSgene based at the UCL Great Ormond Street Institute of Child Health and consultant clinicians at Great Ormond Street Hospital (GOSH) departments; PICU and clinical genetics and genomic medicine. We used clear patient selection criteria to prioritise use in individuals where an urgent genetic result may significantly influence their current management. Turnaround time is drastically reduced through use of advanced bioinformatics and phased variant reporting to clarify the variant interpretation process for referring clinicians. To assess the clinical utility of RaPS we processed 20 patient/parent trios from GOSH PICU. We identified a pathogenic mutation in 35% (7/20) of families (*POLE1, GLDC, TBCE, COL3A1, CHD7, WT1* and *NSD1*), a further 4 families (20%) harboured Variants of Uncertain Significance (VUS) (*SLC30A6, PIGT, IL2RG* and *G6PD*) and 9 families (45%) had no candidate variants. We highlight in detail 4 cases where management was clearly changed by the genetic result, including a patient with an atypical *NSD1* phenotype in whom an early diagnosis obviated the need for further diagnostic testing and allowed targeted therapy. For VUS we propose a data sharing strategy to improve the identification of novel genes. We also show it is possible to complete this procedure in as little as 5 days in a diagnostic environment. The use of RaPS requires a considerable investment in time and resources, however, we demonstrate empirically for the first time in a UK NHS setting the potential of RaPS to influence point of care decisions and provide a robust framework that can be followed by other laboratories. Specifically, we detail our patient selection criteria checklist, bioinformatics data analysis pipeline and variant calling strategy that led to the reduced turnaround time and high diagnostic yield.

2536W

Complete STK11 deletion and atypical symptoms in Peutz-Jeghers Syndrome. *Y.H. Hong.* Department of Pediatrics, Soonchunhyang University Bucheon Hospital, Bucheon, Gyeounggi, South Korea.

Peutz-Jeghers syndrome (PJS) is a rare autosomal dominant inherited disease characterized by gastrointestinal polyposis and mucocutaneous pigmentation. The gene involved in PJS is mapped to chromosome 19p13.3 and encodes a serine-threonine protein kinase (STK11) known as LKB1. Only a small number of large genomic deletions have been identified. We report a boy with PJS and multiple anomalies, neurologic manifestations. The boy was the second child of non-consanguineous and healthy parents. He was delivered at term, by vaginal delivery, after an uncomplicated pregnancy. The first child was an apparently normal boy. He had an atrial septal defect and underwent orchiopexy for cryptorchidism. He had global developmental delay and mental retardation. At about four years of age, black pigment spots appeared on his lips. He visited the emergency department complaining of acute hematochezia at 14 years old age. Colonoscopy showed a huge grape-shaped pedunculated polyp in the sigmoid colon. A few small sessile polyps were found in the terminal ileum, descending colon, and rectum. Polypectomy and multiple biopsies were obtained from the polyp. Histologically, the polyp fragments were hyperplastic. Initially, we performed PCR and direct sequencing using primers targeting the nine coding exons of the STK11 gene. No pathogenic or likely pathogenic variantswere detected by this direct sequencing analysis. Next, gene dosage analyses were performed by using a multiplex ligation-dependent probe amplification (MLPA) kit (SALSA MLPA P101-A2 STK11 probemix; MRC-Holland, Amsterdam, Netherlands). Deletion of all 10 exons of STK11 gene was detected. Although large-scale deletions from STK11 have been reported in PJS, including deletion of the entire STK11 gene, these appear to be rare. Whole-gene and exonic deletions and duplications cannot be detected by using conventional screening tests for point mutations and small deletions/insertions. Hence, the prevalence of STK11 mutations in PJS may have been underestimated. This case emphasizes the importance of genetic testing to identify the genotype-phenotype relationship in PJS. The MLPA assay is a valuable method for detecting STK11 mutations in PJS. More studies would be needed to search for large deletions in patients who present with PJS clinically and have atypical clinical features of PJS.

2537T

Whole *ROR2* gene deletion uncovering a pathogenic mutation in a patient with autosomal recessive Robinow syndrome. *B.M. Ferreira[1], A.R. Lima[1], T.M. Canó[2], M.A. Chiabai[3], R. Pogue[3], J.F. Mazzeu[1,4].* 1) University of Brasília, Brasília, Distrito Federal, Brazil; 2) Secretary of State for Health of the Distrito Federal, Distrito Federal, Brazil; 3) Catholic University of Brasilia, Distrito Federal, Brazil; 4) Robinow Syndrome Foudation, Anoka MN, USA.

Statement of purpose: Robinow Syndrome is a genetically heterogeneous disorder characterized by a triad of facial dysmorphis, including hypertelorism, short stature and genital hypoplasia with extreme clinical variability. More severe bone involvement and marked short stature are observed in the autosomal recessive form of the syndrome (RRS [MIM 602337]) caused by biallelic mutations in *ROR2* (receptor tyrosine kinase-like orphan receptor 2), whereas heterozygous mutations in *WNT5A* (DRS1 [MIM 180700]), *DVL1 (*DRS2 [MIM 616331]) and *DVL3* (DRS3 [MIM 616894]) have been identified in a subset of patients with autosomal dominant Robinow syndrome. These genes encode for components of the Wnt signaling complex which play a key role in the regulation of cell differentiation and patterning. Here we report a patient with recessive Robinow syndrome caused by a whole *ROR2* deletion and a hemizygous mutation in the other allele. **Methods used:** Patient is a 3-month old boy born to non-consanguineous parents with typical recessive Robinow syndrome phenotype, referred to our genetic service for molecular confirmation of the diagnosis. Mutation screening was performed by using next generation sequencing (NGS) using ION PGM™ Inherited Disease Panel as described by the manufacturer. Investigation of deletions and duplications on *ROR2* was performed by MLPA using kit P179. Chromosome Microarray analysis was performed for delimitation of the deleted segment using the Affymetrix 750K platform. **Summary of results:** NGS revealed a single pathogenic mutation in *ROR2* (c.1970G>A, p.Arg657His), allegedly in homozygosity. Patient was homozygous for other polymorphic variants in *ROR2*. Considering that the parents were not consanguineous we performed MLPA that revealed a deletion of all probes for *ROR2*. Chromosome microarray analysis delimited the deletion to a 470 Kb on 9q22 (arr[hg19] 9q22.31(94,381,136-94,851,388) x1). The deletion includes *ROR2* and *SPTLC1* genes. **Conclusion:** The patient here described presents a chromosomal microdeletion including *ROR2* gene uncovering a pathogenic mutation on the other allele of *ROR2*. These findings corroborate the diagnostic of RRS and sheds light to the importance of a criterious evaluation of NGS data for differentiation of hemizygous and homozygous mutations.

2538F

Expanding the spectrum of *TBL1XR* deletion: Report of a patient with brain and cardiac malformations. *S. Oliveira[1], A. Vaqueiro[2], C. Oliveira[2], M. Cordoba[3], B. Versiani[3], C. Carvalho[1], P. Rodrigues[4], J. Mazzeu[4], A. Pic-Taylor[1].* 1) Genetica e Morfologia, Universidade de Brasilia, Brasilia, Brasilia (Distrito Federal, Brazil; 2) Hospital de Apoio, Secretaria de Estado de Saúde do Distrito Federal, Brasília, Brazil; 3) Hospital Universitário, Universidade de Brasília, Brasília, Brazil; 4) Faculdade de Medicina, Universidade de Brasília, Brasília, Brazil.

The *TBL1XR1* gene (transducin beta-like 1 X-linked receptor 1-NM_024665 – OMIM 608628), also known as *hTBL1XR1, hTBLR1* and *IRA1*, is located on 3q26.32 and belongs to the *TBL1XR1* family. This gene encodes a nuclear protein of 515 amino acids, which is a component of SMRT/N-CoR co-repressor complexes and participates in the molecular switch of specific gene transcription. Deletions of the this gene have been described as associated to intellectual disability and dysmorphisms. Point mutations in this gene have been associated to autism and Pierpont syndrome. Here we describe a girl diagnosed with cardiac abnormalities - atrial and ventricular septum defects, a persistent truncus arteriosus and moderate mitral insufficiency - at three months of age. Language development was delayed and she displayed compulsive behaviour. Facial characteristics included a square face frontal bossing, sparse eyebrows, bilateral epicanthus, hooded eyes, midface hypoplasia, a bulbous nose, long philtrum, thin lips, crowded teeth, a narrow palate, and a large pointed chin. Hands were short with brachydactyly and bilateral clinodactyly of the fifth finger. Feet were small and padded, toes were small and splayed and she presented cutaneous syndactyly between the 2nd and 3rd toes. Her lower left limb was thicker than her right, with visibly protruding veins. Cranial tomography showed premature closing of the sagittal suture and anteroposterior growth of the cranium, although the scaphoid skull was not visible. In addition, we observed a Chiari I malformation, that has been asymptomatic to date. Chromosomal microarray analysis revealed a 1,3 Mb *de novo* microdeletion on chromosome 3q26.32: arr[hg19] 3q26.32(176,025,379-177,377,006) x1 encompassing one coding gene, *TBL1XR1*, and part of a long non-coding RNA *LINC00578*. Altogether, we can summarized the main clinical signs of the *TBLX1R1* gene deletion as: mild to moderate intellectual disability, delayed language development, a pointed chin, and, according to this study, cardiac and brain. The role of TBL1XR1 on Wnt signalling and the consequences of this pathway on both brain and cardiac development is in accordance with the phenotype of the patient described herein and supports *TBL1XR1* as a strong candidate to explain ID and cardiopathy associations. Therefore, the data provided here contributes in defining the phenotype linked with haploinsufficiency of the *TBL1XR1* gene and affords a better understanding of its function.

2539W

Study of genetic defects in patients with limb malformations. *A. Rai[1], A. Moirangthem[1], P. Srivastava[1], N. Gupta[2], RD. Puri[3], SJ. Patil[4], P. Ranganath[5], K. Gowrishankar[6], A. Dutta[7], JP. Soni[8], K. Mandal[1], SR. Phadke[1].* 1) Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Science, Lucknow, India; 2) Division of Genetics, Department of Pediatrics, All India Institute of Medical Sciences, New Delhi, India; 3) Institute of Medical Genetics & Genomics, Sir Ganga Ram Hospital, New Delhi, India; 4) Department of Medical Genetics, Narayana Multispecialty Hospital, Bangalore, India; 5) Department of Medical Genetics, Nizam's Institute of Medical Genetics, Hyderabad, Telangana, India; 6) Kanchi Kamakoti CHILDS Trust Hospital, Chennai, India; 7) North Bengal Medical College & Hospital Sushrutanagar, Siliguri, India; 8) Dr. S.N. Medical College, Jodhpur, India.

Congenital limb malformations occur in approximately 1 in 500 human live births which makes them one of the most frequent birth defect and includes both gross reduction defects and more subtle alterations in the number, length and anatomy of the digits. Increasing numbers of genes are being identified for limb abnormalities and identification of causative gene mutations is important for genetic counselling and also provide insight into mechanisms those control limb development. We have studied 62 cases with limb malformations for genetic etiologies using Sanger sequencing, cytogenetic microarray and exome sequencing. For monogenic disorder where causative genes were known, Sanger sequencing has been done. Cases of unknown etiology were subjected to chromosomal microarray and exome sequencing. For genes of monogenic disorders having more than 10 exons exome sequencing has been done. Sanger sequencing identified causative variants in 11 cases (7 were known and 4 were novel) out of 25 monogenic limb malformation cases. These include *ECEL1, FGFR2, ROR2, PITX1, TP63* and *SOST1* genes. Chromosomal microarray identified copy number variations in two cases out of 10 cases; 7q21.2-q22.1 deletion (7 MB) in a patient with split hand-split foot malformation and sensorineural hearing loss and 13q31.1 deletion (1.4 MB) in a patient with short 5th metacarpal and ptosis, genu valgum. By homozygosity mapping, two candidate genes were identified in a patient with Arthrogryposis Renal Dysfunction and Cholestasis which was further subjected to exome sequencing due to its large genes. Samples for exome sequencing were outsourced which were sequenced using Nextera Rapid Capture Exome from Illumina and raw data were analysed for disease causing variations. Exome sequencing identified causative variants in 19 out of total 27 cases. So in this study the total diagnostic yield of single gene sequencing, chromosomal microarray and exome sequencing is 51% which is showing the efficacy of these three techniques in evaluating the genetic cause of limb malformation patients. The mutation data was used for providing genetic counseling to the families. Exome sequencing data of 8 cases in whom the causative variants were not identified is being analysed for variants in novel genes.

2540T

AMELIE accelerates Mendelian patient diagnosis directly from the primary literature. *J. Birgmeier[1], J.A. Bernstein[2], G. Bejerano[1,2,3].* 1) Department of Computer Science, Stanford University, Stanford, California 94305, USA; 2) Department of Pediatrics, Stanford School of Medicine, Stanford, California 94305, USA; 3) Department of Developmental Biology, Stanford University, Stanford, California 94305, USA.

Rare diseases, in aggregate, affect up to 10% of the world's population. Patients with rare Mendelian diseases have 1-2 genetic mutations in a single gene primarily responsible for their phenotypes. The diagnosis of known Mendelian disorders requires matching a patient's phenotype and genetic variants against the published medical literature and can be highly labor intensive, with more than 40 hours of expert time needed for individual diagnoses. There are currently more than 5,000 diagnosable Mendelian diseases, caused by roughly 5,000 different genes, and manifesting in different subsets of over 10,000 documented phenotypes. Patients may present up to 300 candidate genes of 20,000 in our genome. To arrive at a differential diagnosis, the clinician must read about all of the patient's candidate genes to try and find one that may explain their patient. Our software system AMELIE (Automatic Mendelian Literature Evaluation) automates the process of literature curation for diagnosis of Mendelian diseases. AMELIE collects and parses hundreds of thousands of full-text articles from all of PubMed to find an underlying diagnosis that best explains the patient's phenotypes. Given a list of candidate genes and a patient's phenotypes, AMELIE uses its articles knowledgebase to rank the patient's genes by their likelihood of causing the patient's phenotypes, and offer the diagnostic paper(s) explaining its rationale. Diagnosis of singleton patients (where the patient, but not their relatives, is sequenced) is a common scenario that is both most challenging and most time-consuming. AMELIE's gene prioritization method was tested on 215 singleton patients with diagnosed Mendelian diseases. AMELIE ranked the causal gene in the top 2 in 63% of all cases and among the top 10 in 95% of all cases, out of a median of 124 candidate genes per patient. AMELIE significantly outperformed curation dependent methods. AMELIE's superior performance is explained by the fact that it holds over 3 times as many gene-phenotype relationships as are curated by HPO from OMIM and OrphaNet. We are launching a web portal where one can analyze patient cases based on the latest literature at AMELIE.stanford.edu. We will present on different patient cases from the Stanford Hospital whose diagnosis is greatly facilitated using AMELIE's automated literature-supported gene ranking.

2541F

Two children with copy number variants in the critical regions of both Wolf-Hirschhorn and Cri du Chat syndrome. *L. Andoni[1], L. Fuqua[1], K. Ho[1], J.C. Carey[2], C. Bupp[3].* 1) Lineagen Inc, Salt Lake City, UT; 2) Department of Pediatrics, University of Utah, Salt Lake City, UT; 3) Spectrum Health Medical Group, Grand Rapids, MI.

When multiple pathogenic variants co-occur in a patient, clinical diagnosis can be greatly aided by genetic testing. Here we describe two infants with copy number variants (CNVs) involving the critical regions for both Wolf-Hirschhorn (WHS) and Cri du Chat (CdCS) syndromes. One infant has a deletion in both the WHS and CdCS regions, while the other has a duplication of the WHS critical region, also known as 4p16.3 microduplication syndrome, and a deletion of the CdCS region. In this report, we delineate how features of these two children correlate and contrast with known features of these conditions. Intellectual disability, developmental delay, and multiple congenital anomalies are common to all three conditions, while each also has unique characteristics, such as distinct craniofacial features for WHS and a cat-like cry for CdCS. Patient 1 was diagnosed by chromosome microarray testing (CMA) in the neonatal period with a pathogenic 8.86 megabase (Mb) terminal deletion of 4p16.3-p16.1, 1.18 Mb deletion of 4q11-q12 of uncertain clinical significance, and pathogenic 14.05 Mb terminal deletion of 5p15.33-p15.2, all of which were *de novo*. Born at 34 weeks, the patient had microcephaly, congenital heart and kidney malformations, dysmorphic features, and brain abnormalities. Interestingly, WHS and CdCS were not initially suspected, although with age the patient now demonstrates more of a WHS phenotype. Patient 2 was diagnosed by CMA at 10 months of age with a pathogenic 10.96 Mb terminal duplication of 4p16.3-p16.1 and a pathogenic 29.77 Mb terminal deletion of 5p15.33-p13.3. Parental karyotype has not yet been performed. Born at 36 weeks, the patient had microcephaly, TE fistula, esophageal atresia, and congenital cardiac defects. 22q11.2 deletion syndrome was high on the initial differential diagnosis, but FISH testing for 22q11.2 deletion syndrome was negative. With age, the patient demonstrates more of a CdCS phenotype, but is lacking the characteristic cry. While careful clinical assessment and phenotyping are keys to diagnoses, these two unique cases also underscore the importance of CMA. Having multiple CNVs can lead to unique phenotypes, as evidenced by these patients that are not completely clinically consistent with either WHS or CdCS. Including the family in the decision-making process when ordering genetic testing is crucial and the psychosocial aspects of these decisions are underscored.

2542W

Recurrent unbalanced constitutional chromosomal translocation between chromosomes 8 and 12, der(8)t(8;12)(p23.1;p13.31), detected in three patients with similar phenotype. *D. Huang[1], L. Matyakhina[1], A. Wray[1], V. Nelson[1], R.L. Ladda[2], S. Sell[2], R. Davis[3], J. Richards[3], J. Condie[4], J. Meck[1].* 1) Cytogenomics, GeneDx, Inc., Gaithersburg, MD; 2) Department of Pediatrics, Penn State Hershey Children's Hospital, Hershey, PA; 3) Pediatric Neurology P.A., 7485 Sandlake Commons Blvd, Orlando, FL; 4) St. Luke's Children's Neurology, 100E Idaho St., Boise, ID.

   Recurrent constitutional reciprocal non-Robertsonian chromosome translocations are rare. To our knowledge, there are only three such translocations reported in the literature: the most common one is t(11;22) (q23;q11), while t(8;22) (q24.13;q11.21) and t(4;8)(p16;p23) are less frequently seen (PMID: 22260357 and 12058347). Recurrent unbalanced translocations, most notably der(4)t(4;11)(p16.2;p15.4), have been also reported (PMID: 21205869). Here we report an apparently recurrent unbalanced translocation between the short arms of chromosomes 8 and 12 found in 3 cases referred to our clinical diagnostic laboratory for chromosomal microarray. In all cases array detected a terminal 7 Mb deletion in 8p and a terminal 8.3 Mb duplication in 12p. Metaphase FISH using probes from within the deleted and duplicated regions performed on two probands showed that these aberrations were secondary to a derivative chromosome 8, der(8)t(8;12)(p23.1;p13.31). Parental metaphase FISH revealed normal results, consistent with a *de novo* occurrence of the unbalanced translocation. To date, three patients with apparently the same unbalanced translocation have been reported in the literature, one of which had clinical information (PMID: 21205869 and 22639464). Common clinical features include motor and language delay, hypotonia, childhood onset epilepsy, and autistic features. Additional features observed in more than one patient were facial dysmorphism, umbilical hernia, CNS abnormalities, and eye problems. The 8p23.1 and 12p13.31 chromosome regions are known to contain highly homologous LCR clusters: this genomic architecture likely underlies the recurrent occurrence of this translocation. In summary, the der(8)t(8;12) appears to be a recurrent *de novo* unbalanced translocation mediated by large LCRs and associated with the phenotype of developmental delay, hypotonia, childhood onset epilepsy, and autistic features.

2543T

The first Japanese patient of Mandibular hypoplasia, deafness, progeroid features, and lipodystrophy syndrome (MDPL) diagnosed via *POLD1* mutation detection. *I. Imoto[1], T. Kohmoto[1], T. Naruto[1], A. Goji[2], A. Okada[2], I. Yokota[2,3], Y. Kotani[2], A. Shimada[4], Y. Miyamoto[1], K. Masuda[1], S. Kagami[2].* 1) Dept Human Gnetics, Grad Sch Biomed Sci, Tokushima Univ, Tokushima, Tokushima, Japan; 2) Dept Paediatrics, Grad Sch Biomed Sci, Tokushima Univ, Tokushima, Tokushima, Japan; 3) Dept Pediatrics, Div Pediatric Endocrinol & Metabolism, Shikoku Medical Center for Children and Adults, Zentsuji, Kagawa, Japan; 4) Dept Otolaryngology, Grad Sch Biomed Sci, Tokushima Univ, Tokushima, Tokushima, Japan.

   Mandibular hypoplasia, deafness, progeroid features, and lipodystrophy syndrome (MDPL, MIM#615381) is a rare autosomal dominant systemic disorder resulting from heterozygous mutations in the *POLD1* gene (MIM#174761). MDPL is clinically characterized by prominent loss of subcutaneous fat, a characteristic facial appearance, metabolic abnormalities including insulin resistance and diabetes mellitus, and sensorineural deafness occurring late in the first or second decades of life. Until now, 18 patients including 15 Caucasian, 2 Hispanic, and 1 Indian cases affected by this syndrome have been described. Among analysed 13 cases, only two causative *POLD1* mutations, an in-frame deletion (Ser605del, 11 cases) and a missense mutation (R507C, 2 cases), have been identified, and most of them have occurred *de novo*. Although these *POLD1* mutations in hot spots causing MDPL may occur regardless of race/ethnicity, no East Asian cases including Japanese cases have been reported. We herein report the first East Asian/Japanese case of MDPL with one of known heterozygous *POLD1* mutations. A clinically undiagnosed 11-year-old, Japanese boy, who noted joint contractures at age 7 years, was affected with various other symptoms, such as short stature, truncal obesity, mandibular underdevelopment, pinched nose, tight facial skin around the eyes, prominent upper teeth, healing loss, hypogonadism. Targeted-exome sequencing (TES) using a TruSight One Sequencing Panel (Illumina) performed as a genome-first approach detected an in-frame heterozygous deletion in exon 15 of *POLD1*, NM_002691.3:c.1812_1814del, affecting the polymerase-active site, NM_002691.3:p.(Ser605del), which was confirmed by Sanger sequencing. This mutation has been shown to cause most cases of MDPL. Neither variations nor gross deletions in the coding regions of other progeroid-related genes were detected in TES. As a result of this molecular diagnosis and re-evaluation of the affected patient's clinical features, together with the clinical spectrum of patients harbouring *POLD1* mutations, the patient was diagnosed with the 19th MDPL and 14th *POLD1*–caused MDPL case reported worldwide caused by a known frame-shift deletion in POLD1. Notably, in addition, this patient is the first Japanese as well as East Asian case with MDPL, supporting the hypothesis that *POLD1* mutations causing MDPL, at lease this in-frame deletion mutation, commonly occur at hot spot irrespective of race/ethnicity.

2544F

Enhanced detection of uniparental disomy (UPD) and copy number variations (CNVs) with delineation of parental origin for clinical trio exome cases. *H. Dai[1], P. Liu[1,2], W. Bi[1,2], M. Wallkiewicz[1,2], M. Leduc[1,2], T. Santiago-Sim[1], I. Van den Veyver[1,3], S. Stover[4], J. George-Abraham[5], C. Qu[2], V. Patel[2], X. Wang[1,2], Y. Yang[1,2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Genetics, Houston,TX; 3) Obsterics and Gynecology, Baylor College of Medicine, Houston, TX; 4) Genetics, Texas Children's Hospital, Houston, TX; 5) Dell Children's Medical Center, Austin, TX.

   Introduction: Trio exome sequencing has been widely applied to provide molecular diagnoses for patients with a wide spectrum of genetic disorders. Yet the potential of trio exome data for identifying structural abnormalities and UPDs has not been fully utilized. **Objectives:** To develop a trio data analysis algorithm based on parental origins of alleles to identify UPDs and other structural abnormalities. **Methods:** SNP data from 1264 clinical trio exome cases were retrospectively analyzed using an algorithm based on the assignment of alleles by descendant. The patterns of uniparental disomy and other structural anomalies were flagged automatically based on the most likely parental origin of each allele or allele pair. For the confirmation of the results called by the computer program, the exome and/or cSNP array data were subsequently examined manually. **Results:** The findings include: 1) Seven (7/1264) UPD cases, including 5 for which UPD is the cause of the diseases in the patients. A heterodisomy UPD(14)mat, undetectable by microarray due to lack of absence of heterozysosity (AOH) on the chromosome, unveiled Temple syndrome presented in the first patient. An isodisomy UPD(15)mat was detected in the second patient confirming the clinical diagnosis of Prader-Willi syndrome. A complex UPD(16)mat with regions of isodisomy and heterodisomy was detected in the amniotic fluid sample of a fetus, indicating possible trisomy rescue during earlier fetus development. Two molecular diagnoses (Stargardt disease and Parkinson disease) were made in yet another patient with homozygous pathogenic variants in *ABCA4* and *PARK7* resulting from UPD(1)pat. Lastly, one case has mosaic UPD(1)mat exposing the pathogenic variant in *ZMPSTE24* associated with lethal restrictive dermopathy. 2) Nine cases with *de novo* del/dup events, of which the parental origins were also delineated, providing additional insights into disease etiology. Intriguingly a *de novo* loss on 6q27 and gain on 10q25.1q26.3 of paternal origin were detected in one patient diagnosed with 6p duplication by chromosome studies. Our results indicates that the cryptic 6q27 duplication based on karyotype actually resulted from translocation of the duplicated 10q25.1q26.3 bands. **Conclusion:** Expanding trio exome analysis to genome wide structural variations provides greater potentials to decipher the etiology of genetic diseases. This analysis has been incorporated into the routine trio exome analysis in our laboratory.

2545W

A rapid and reliable chromosome analysis method for products of conception using interphase nuclei. *R. Babu[1], D. Van Dyke[2], S. Bhattacharya[3], V. Dev[4], M. Liu[4], M. Kwon[4], G. Gu[4], P. Koduru[5], N. Rao[6], C. Williamson[2], E. Fuentes[1], S. Fuentes[1], S. Papa[1], S. Kopuri[1], V. Lal[3].* 1) InteGen LLC, Orlando, FL; 2) Mayo Clinic, Rochester, MN; 3) Dr. Lal Path Labs, New Delhi, INDIA; 4) Genetics Associates Inc., Nashville, TN; 5) UT Southwestern Medical Center, Dallas TX; 6) David Geffen UCLA School of Medicine, Los Angeles, CA.

   Background: Karyotype determination has a central role in the genetic workup of pregnancy loss, as aneuploidy and polyploidy are the main causes. Current methods of obtaining a karyotype using traditional cytogenetics, FISH with a limited number of probes, and CMA have limitations of culture failure, incomplete results, lower sensitivity, and longer turnaround time. **Methods:** A Standard Resolution Interphase Chromosome Profiling (ICP) probe set, a variation of a recently developed High Resolution probe set (Babu et al., in press), was developed that targets only the subtelomere and paracentromeric regions. Abnormalities commonly encountered in POC samples, such as trisomy and unbalanced translocations, can be easily detected with this design. To detect Robertsonian translocations, each acrocentric chromosome's paracentromeric area is targeted and a mixture of these five targets is used in a separate analysis. First, initial familiarization of signal patterns from the probe set was achieved, followed by validation of this method using 83 blind samples from miscarriages from three different laboratories. Finally, the clinical utility of the method was tested on 291 samples in two commercial reference laboratories on two different continents. A minimum of 20 interphase cells were analyzed for each chromosome to identify an abnormal clone. For detection of Robertsonian translocations, a normal cutoff was set at 20% for any two paracentromeric probes to co-localize by random chance or by satellite association. **R**esults: All 374 samples had results by the ICP method and the average turnaround was less than 48 hours. This new molecular approach not only detected all the chromosomal changes observed by current methods, but significantly improved the abnormality detection by characterizing derivative chromosomes and finding subtle subtelomeric deletions (Xq and 17q) and A balanced translocation involving 2q and 10q. All Robertsonian translocations were also detected. Abnormalities detected only by ICP included trisomy 4, 7, 8, 14, 15, 16, 19, 20 and 22. The abnormality rate was 54% on clinical samples from commercial lab 1 and 63% from lab 2. **Conclusion(s):** The attributes of Interphase Chromosome Profiling (ICP), namely 1) near 100% results, 2) greater sensitivity than conventional methods, and 3) no need for cell culture or DNA isolation, make this novel method an ideal choice for first-line testing in the genetic workup of miscarriages.

2546T

Clinical validation of whole-genome sequencing assay for constitutional disorders. *A. Abhyankar, W. Clarke, G. Narzisi, V. Felice, E. Dikoglu, L. Kozon, Y. Kasai, V. Jobanputra.* New York Genome Center, New York, NY.

Next-generation sequencing technology has come a long way in making high-throughput genetic screens accurate, affordable and accessible. Though whole-exome sequencing (WES) assays are being used in clinical settings, whole-genome sequencing (WGS) assays have not yet been widely adopted for clinical use. There are many advantages of using WGS over WES assay. This approach can be used to identify SV, and CNV events more accurately as compared to WES. Additionally, this approach allows assessment of non-coding regions of the genome in the clinical context. We performed analytical validation of whole-genome sequencing assay for the purpose of clinical testing of constitutional disorders. WGS was performed on 88 samples that included singleton, trios and quads with undiagnosed disease and healthy individuals. We performed the validation on DNA derived from whole blood, archived dried blood spot and saliva specimens. We also sequenced NA12878 cell line obtained from NIST. The results were compared with Genome In a Bottle high-confidence call set for NA12878 to define the accuracy and analytical specificity of our assay. To estimate the degree to which repeated sequence analyses give the same result - repeatability (inter-assay precision) and reproducibility (intra-assay precision), we sequenced six samples in triplicate for inter-assay evaluation and five samples for intra-assay evaluation. All genomes were sequenced on Illumina HiSeqX system at mean genome coverage of 30x. The resulting raw reads in FASTQ format were processed through analysis pipeline designed to call SNVs, Indels, CNVs, and variants in the mitochondrial genome. Considering the circular nature of mitochondrial genome we developed a specialized workflow to identify homoplasmic and heteroplasmic variations as well as to estimate the relative mitochondrial DNA copy number. We also developed workflow to analyze parent-child trios and quads. Quality control metrics were assessed at various stages of the workflow and cutoffs were determined to ensure high quality of the resulting data. Clinically relevant variant prioritization and reporting strategy was developed to provide accurate variant discovery, evaluation and reporting in the context of provided clinical information.

2547F

Analysis of 1374 comparative genomic hybridization (CGH) results: Indirect evaluation of carrier rate in a Colombian population. *M. Garcia-Acero[1], F. Suarez-Obando[1], A. Gomez[2].* 1) Genetics, Pontificia Universidad Javeriana, Bogota, Colombia; 2) Instituto de Referencia Andino.

Introduction: Comparative Genomic Hybridization (CGH) is a high-resolution technique that allows the detection of microdeletions and microduplications unrevealed in the conventional microscopic karyotype. These chromosomal abnormalities are a common cause of developmental delay, non-syndromic cognitive disorders, autism spectrum disorders and multiple congenital anomalies. Although a considerable number of these patients are suspected of having cytogenetic imbalances, only a few are positive for copy number abnormalities that explain the phenotype. On the contrary, unexpected findings have been documented that modify family counseling and medical care. **Methodology:** Descriptive analysis of results of 1374 CGH performed in a Reference Laboratory in Colombia during the period 2009 - 2015. **Results:** 495 chromosomal alterations were found through CGH analysis, 302 (61%) of which were classified as pathogenic (4 of these not associated with the clinical phenotype), 92 (18.5%) as non-pathogenic, 65 (13.1%) were defined as variants of uncertain significance and 36 (7.27%) as variants that constitute alterations of a single allele in autosomal recessive inheritance pathologies, contributing as carrier variants in the population. **Conclusions:** The detection of genomic imbalances will increase knowledge of the human genome through phenotype-genotype correlations. Additional findings in the test not associated with the phenotype, could be useful for genetic counseling. The analysis of additional findings of a bigger sample should allow the description of carrier states in the evaluated population.

2548W

10x Genomics® Chromium™ linked-read workflows fully optimized on PerkinElmer Sciclone® for high-throughput automation of exome and genome applications. *J. Garifallou[1], F. Mafra[1], R. Pellegrino[1], C. Siebert[2], B. Gerwe[2], M. Gonzalez[1], M. Benway[2], C. Kaminski[1], C. Kao[1], H. Hakonarson[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) PerkinElmer, Inc.

   Current massively parallel sequencing methods, while high-throughput and efficient in data generation, are limited due to the nature of the short read sequences that restrict certain applications such as the ability to phase/haplotype over long genomic distances, accurately map reads between highly homologous regions (e.g. genes vs. pseudogenes), and robustly detect particular types of structural variants. The Chromium instrument (10x Genomics) leverages advances in microfluidics technology and precision reagent delivery to generate a haplotype-level dilution of ~1 ng of high molecular weight DNA molecules in >1 million barcoded partitions, which creates a novel data type called 'Linked-Reads'. The in-partition biochemistry and subsequent processing generates Illumina®-compatible libraries in which all short reads that carry the same partition barcode are automatically associated with the same originating DNA molecule. Each short read retains up to hundreds of kilobases of positional context from that individual DNA molecule, thus enabling long-range dependent applications such as phasing, structural variant detection, and mapping reads of paralogous segments. Here we demonstrate how the PerkinElmer Sciclone® NGSx Workstation can successfully and fully automate the Chromium Genome workflow. We show the libraries have comparable quality to those of a manual preparation. Following Chromium partitioning, library generation, and pooling, samples were processed again on the Sciclone NGSx Workstation using an established automated workflow for exome/panel target capture using Agilent SureSelect™ baits. This end-to-end automated workflow was used to generate Linked-Read whole exome data on samples with unresolved structural rearrangements and difficult-to-map loci. Linked-Reads enable us to 1) fine-map structural rearrangements detected by karyotyping and 2) resolve variants in genes versus those in their homologous pseudogenes without invoking non-NGS methods such as MLPA or long-range PCR. The benefits of automation are critical to the scale-up of high-throughput projects in the clinical setting by removing manual variability and increasing efficiency/accuracy. Combining 10x Genomics technology with PerkinElmer automation yields a complete workflow solution for exome and panel-based Linked-Read sequencing, at scale, to expand the downstream application of regular short reads sequencing with minimal equipment and infrastructure changes.

2549T

Human ring chromosome atlas: A web-based registry and a comprehensive review of ring chromosome cases in the Chinese population. *Q. Hu[1,2], H. Chai[1], P. Li[1].* 1) Genetics, Yale School of Medicine, New Haven, CT; 2) Cell Biology and Genetics, Guangxi Medical University, Nanning, Guangxi, China.

   Background: Human constitutional ring chromosomes are a rare type of chromosomal structural abnormality. Ring chromosome syndrome featuring growth retardation and intellectual disability is likely attributed to the dynamic behavior of ring chromosome through cell cycles. Chromosomal specific phenotypes likely result from segmental losses and gains during the ring formation. Although recent applications of genomic analyses revealed various ring chromosome structures from an increasing number of case studies, there are no organized efforts for compilation and curation of ring chromosome findings. **Human Ring Chromosome Registry and Tested Cases**: We developed a web-based 'Human Ring Chromosome Registry' using Microsoft Access relational database with online implementation by active server page and internet information services. To test and validate this ring chromosome registry, we collected and input ring chromosome cases reported in the Chinese population thorough a search of Chinese medical literature and PubMed publications from 1979 to 2017. **Ring Chromosomes in Chinese Population:** From a total of 95 cases, ring chromosomes were detected in all chromosomes except chromosomes 7, 16, 17 and 19. The most frequently seen by a decreasing order of relative frequencies were ring chromosomes 13 (14%), X (11%), 15 (9%), 22 (9%), 14 (8%), 18 (7%), 6 (5%), and 9 (5%). Distal deletions ranging from 60 Kb to 9 Mb were noted in thirteen cases analyzed by microarray analysis and in two cases by genome-wide sequencing. Four cases were diagnosed prenatally, 68 cases were noted in perinatal or pediatric clinic assessment, and six cases were noted in infertility clinics. Shared phenotypes of developmental delay, dysmorphic facial features, microcephaly, hypotonia, speech difficulty, low birth weight, growth retardation and intellectual disability were noted in most autosomal rings. Chromosomal or regional specific phenotypes included Wolf-Hirschhorn syndrome in a ring chromosome 4, cri-du chat in a ring chromosome 5, epilepsy in ring chromosome 14, Tuner syndrome in ring chromosome X, and infertility in ring chromosomes 13, 21, 22 and Y. **Conclusions:** This newly-developed Human Ring Chromosome Registry summarized the relative frequencies, cytogenomic and clinical findings of reported constitutional ring chromosomes in the Chinese population. It represents an initial effort toward a systematic compilation of cytogenomic and phenotypic findings of ring chromosome cases.

2550F

Uniparental isodisomy X: Evidence for monosomy rescue. *K. Rudd[1], J. Schleede[1], S. Williams[2], K. Lee[3], J. Laffin[3], R. Pasion[1], J. Tepperberg[1], P. Papenhausen[1].* 1) Center for Molecular Biology and Pathology, Laboratory Corporation of America Holdings, Research Triangle Park, NC; 2) Savannah Perinatology Associates, Savannah, GA; 3) University of Wisconsin Madison, Wisconsin State Laboratory of Hygiene Clinical Genetics Laboratories, Madison, WI.

   10-15% of girls with Turner syndrome have mosaicism for 45,X and 46,XX cell lines. Though X chromosome mosaicism has important implications for clinical presentation and fertility, the meiotic and/or mitotic mechanisms for monosomy X mosaicism are not well understood. Here we report four cases of 45,X/46,XX mosaicism with uniparental isodisomy (UPiD) for the X chromosome. Subject 1 was first assessed prenatally due to IUGR and a positive non-invasive prenatal screen (NIPS) for monosomy X. Prenatal microarray analysis revealed normal copy number and homozygous SNPs across the entire X chromosome; microsatellite analysis confirmed paternal UPiD X. Though FISH and chromosome analysis of amniocentesis and peripheral blood samples revealed two X chromosomes in all cells, FISH of the placenta showed a single X chromosome in 67% of nuclei, consistent with mosaic monosomy X that may be confined to the placenta (CPM). The other females were tested after birth and had monosomic and disomic X cells in peripheral blood. Subject 2 was a newborn with a 45,X[9]/46,XX[11] karyotype who was tested due to a family history of intellectual disability (ID). Subject 3 was tested at 14 years old due to short stature and had a 45,X[25]/46,XX[5] karyotype. Unlike the other subjects, Subject 4 had a structural X chromosome abnormality related to her ID. At 5 years old, chromosome analysis revealed three cell lines: 46,X,+mar[10]/45,X[2]/46,XX[8]. The small marker chromosome was visible via microarray analysis as a 5.86-Mb copy number gain that flanked the X centromere and did not include the *XIST* gene. SNP microarray analysis revealed homozygosity across the entire X chromosome for all three postnatal cases. These results support a model where the conceptuses arose as 45,X females, likely due to a nullisomic gamete or loss of an abnormal sex chromosome, and then doubled the remaining X chromosome during development to generate 45,X/46,XX mosaicism with UPiD X. Rather than the more common trisomy rescue mechanism in autosomal UPD, we propose that UPiD X arises via monosomy rescue. These findings have important implications for monosomy X detected during NIPS, since a monosomy X cell line that is rescued early in embryogenesis could be CPM with minimal effects on the fetus. This could explain at least some of the monosomy X false positives from NIPS. Furthermore, cryptic monosomy X rescue likely contributes to the survival of some Turner syndrome fetuses detected prenatally.

2551W

Variant of Turner syndrome 45, X/46Xdel(X)(q21) mosaicism: A case report. *G. Giraldo[1], M. García-Acero[1,2], O. Moreno[1], F. Suarez-Obando[1,2], C. Cespedes[2], J. Perez[2], JC. Prieto[1], N. Fernandez[2], J. Aulí[2], A. Rojas[1].* 1) Instituto de Genètica Humana, Pontificia Universidad Javeriana, Bogotà, Bogotà D.C, Colombia; 2) Hospital Universitario San Ignacio, Bogota, Colombia.

   Introduction: Turner syndrome (TS) is the most frequent alteration in the sex chromosomes. It results from the complete or partial absence of the X chromosome, associated with another cell line, which might be 46/XX, 46/XY or other rearrangement. The characteristics of the syndrome are short stature, gonadal dysgenesis, cardiac and renal defects, webbed neck, low-set ears and skeletal anomalies. Patients with Xp deletion have low stature and different birth malformations. Those with Xq deletion often have gonadal dysgenesis. It has been suggested that the phenotype is closely linked to the associated cell line in case of mosaicism. **Case presentation**: A 16-year-old female presented to genetic consultation because of primary amenorrhea. She is a product of normal spontaneous vaginal delivery and had normal development milestones. Physical examination revealed a young female with few secondary sexual characteristics, wide neck and short stature. G-band karyotype showed 45,X[29]/46,X,del(X)(q21).nuc ish(DXZ1x1,SRY-,DYZ1-)[59/100] / (DXZ1x2,SRY-,DYZ1-) [41/100]. The interpretation of this analysis revealed Turner syndrome in mosaicism form with two different cell lines; one with full monosomy X and a second with distal deletion of the long arm of the chromosome X. Other complementary lab tests showed hypogonadotropic hypogonadism. A pelvic ultrasound showed infantile uterus without evidence of ovaries. **Discussion:** The absence of the short arm of the X-chromosome is associated with the classic Turner syndrome, while deletions of the long arm are associated with gonadal dysgenesis. Proximal deletions of Xq are related to primary amenorrhea, lack of breast development and complete ovarian failure, in that Xq13 is the key region for ovary maintenance, unlike the compromise of Xq21, they may menstruate. About deletions, it is crucial to detect not only the presence of a structural alteration but also the chromosomal level where it occurred and to define the deleted area. This precision will make it possible to establish the diagnosis and the prognosis since there are some patients with structural alterations of chromosome X who only present menstrual irregularities or premature menopause, and given the reports of cases with familial inheritance, the genetic counseling becomes more meaningful.

2552T

A unique case of male/female chimerism in buccal specimen due to a bone marrow transplant. *H. Risheg[1], R. Pasion[2], K. Barnes[3], M. Schwab[3], S. Schwartz[2].* 1) Laboratory Corporation of America® Holdings/Dynacare, Seattle, WA; 2) Laboratory Corporation of America®Holdings, Center for Molecular Biology and Pathology, Department of Cytogenetics, Research Triangle Park, NC; 3) Hackensack Meridian Health, Genetic Services, Hackensack, NJ.

Over the past decade microarray technology has allowed the detection of chromosomal aberrations not routine seen with standard analysis. However, it has also provided a better understanding of specimens from patients received in the laboratory. Here we describe a unique case of a 21 year old female with a clinical history of dysautonomia, cataracts, anxiety and short stature, and a history of acute lymphoblastic leukemia. Buccal swab was performed and sent in for SNP/CN microarray analysis which identified a sex chromosome mosaicism with 85% of cells with XX and 15% of cells with XY suggestive of chimerism. Based on the sex chromosome findings and the unusual pattern allele tracks and homozygosity seen, this was believed to be due to a bone marrow transplant from a male sibling, which was subsequently confirmed by the referring physician. Interestingly, the array also identified a mosaic 12 Mb interstitial deletion of 15q14-q21.1. However, this deletion was only present in 15% of the sample suggestive that the deletion is t in the male transplant donor rather than in the proband. By report this sibling has learning disabilities, which to date does not have an underlying etiology, but may be associated with the deletion. Overlapping deletions of this 15q deletion have been reported with clinical features that include craniosynostosis, congenital heart defects, limb abnormalities, urinary abnormalities and bone maturation. In this complicated array there were also multiple chromosomes identified with regions of allele homozygosity totaling 41.1 Mb, suggesting a distant parental relationship or common decent. This case highlights the clinical utility of the SNP microarray analysis in allowing the correct delineation of complex findings. In this patient it clearly indicated the true nature of the chimerism prior to clinical information being available and it was able to indicate that the deletion appears to be in the donor rather than proband.

2553F

Landscape of disorders of sex development with mosaicism. *A. Rojas[1], O. Moreno[1,2], M. García-Acero[1], AS. Clavijo[2], J. Perez[2], N. Fernandez[1], I. Zarante[1].* 1) Instituto de Genetic Humana, Pontificia Universidad Javeriana, Bogota, 1, Colombia; 2) Hospital Universitario San Ignacio, Bogota, Colombia.

Introduction: Disorders of sex development (DSD) are congenital conditions in which the development of chromosomal, gonadal or anatomical sex is atypical. The chromosomal sex based on the result from karyotype is currently considered to be the main finding for classifying the DSD. 45,X/46,XY mosaicism and variants of it are considered to be a phenotypically very heterogeneous condition, the clinical phenotype may range from female external genitalia with mild clitoromegaly, through ambiguous genitalia, to isolated hypospadias or normal male external genitalia. **Cases series:** We report 3 patients with a DSD diagnosis in whom chromosomal mosaicism was found in a sample of peripheral blood or gonadal tissue. The chromosomal analysis included high-resolution G and FISH with a specific probe to evaluate the presence of the SRY gene in blood or gonadal tissue when it was possible. 2 patients were assigned male gender and 1 patient was assigned female gender. One of the boy clinically with hypospadias and bilateral cryptorchidism showed a karyotype 46,XX [100] with intraabdominal finding of Wolff duct derivatives in left side and Mullerian duct derivatives in right side with histology of fibrous bands of ovarian stroma. Gonadal karyotype showed RG 1: mos 92,XXXX[8]/46,XX[42] LG 2: mos 92,XXXX[12]/46,XX[38].The other boy clinically with hypospadias and bilateral cryptorchidism showed a karyotype 45,X[4]/46,XY[96] with histological finding in gonadal biopsy of gonadoblastoma and gonadal karyotype showed mos 45,X[97]/46,XY[3]. The girl had hypertrophic clitoris, aorta coartaction and short stature in karyotype was evident mos 45,X[19]/46,XY[23]/47,XYY[8].ish Yp11.31(SRY-)/idic(Y)(p11.31)(SRY++)/idic(Y)(p11.31)(DYZ1-,SRY++)x2. **Discussion:** Mosaicism is defined as the presence of two or more chromosomal lineages in the same individual. The different distributions of the 45,X and 46,XY chromosomal cell lines among the tissues in individuals with this mosaicism presumably reflect the wide variety of phenotypes observed. In our study, the individuals with 45,X/46,XY mosaicism ranged from phenotypically men with hypospadias and to Turner syndrome phenotype. This explain one of the most important and difficult points in treating patients with 45,X/46,XY mosaicism, the decision about the sex of rearing.

2554W

HLA-B*1502 genotyping for the prevention of carbamazepine induced severe cutaneous adverse drug reactions (SCARs) in a children's hospital. *H. Law[1], L. Ching[1], W. Low[1], Y. Tan[1], A. Lai[1], D. Chan[2].* 1) Genetics Service, KK Women's & Children's Hospital, Singapore, Singapore; 2) Neurology Service, KK Women's & Children's Hospital, Singapore, Singapore.

Carbamazepine (CBZ) is a common medication prescribed to control recurrent seizures. Clinical studies have demonstrated a strong association between HLA-B*1502 allele and risk of CBZ - induced Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) which are known to be associated with significant mortality and long term morbidity, with an odds ratio of 1357 for Han Chinese. Since 2013 genotyping for the HLA-B*1502 allele has been made mandatory prior to the initiation of CBZ therapy in new patients of Asian ancestry. Our hospital started offering the test in 2014. Genotyping was carried out using the One Lambda-Micro SSP HLA Class I B Locus Specific Kit (SSP1B). Most of the results were returned within 24 hours. In 3 years (2014 to 2016), 309 patients aged 25 days to 20 year old (median age 7) were tested. HLA-B*1502 was found to be positive in 34 (11.00%). Breaking down into 3 main ethnicities, the Chinese, Malay, and Indian, which make up 76.2%, 15.0% and 7.4% of the population respectively, the Malay was found to have the highest positive rate (9 out of 59, or 15.25%), followed by the Chinese (21 out of 195, or 10.77%). None of the 26 Indian patients tested carried the HLA-B*1502 allele. However, a relatively high positive rate (4 out of 29, or 13.79%) were found in the "other" ethnic group. This "other" ethnic group, made up of migrants from Indonesia, northern Malayan peninsula and other mixed ethnicities, accounts for about 1.4% of the population. Since the implementation of HLA-B*1502 genotyping there were no new cases of SJS and TEN. In contrast there were 5 cases of carbamazepine-associated SJS /TEN in this hospital in the 7 years prior to implementation of testing. HLA-B*1502 testing has been effective in reducing morbidity due to administration of CBZ.

2555T

Clinical whole genome sequencing in a pediatric hospital. *C. Saunders[1,2,4], E. Farrow[1,3,4], N. Miller[1], M. Gibson[1], L. Zellmer[1,2,4], D. Faller[1,2], S. Soden[1,3,4], I. Thiffault[1,2,4].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO 64108; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospital, Kansas City, MO 64108; 3) Department of Pediatrics, Children's Mercy Hospital, Kansas City, MO 64108; 4) University of Missouri-Kansas City School of Medicine, Kansas City, MO 64108.

Clinical exome sequencing is a widely used and powerful tool for the diagnosis of rare genetic diseases, affording coverage of the coding region at a fraction of the cost of running clinical whole genome sequencing (cWGS). However, there are pros and cons to both methods, with exomes offering higher sequencing depth, albeit more biased coverage. Paradoxically, the laboratory time to process a genome is shorter; but the costs of sequencing and storage are higher. Although WGS includes noncoding regions, they are largely unanalyzed in the clinical setting. For all these reasons, the use of WGS is generally limited to the research setting. Here we report for the first time, the use of cWGS in a pediatric setting, including our process, ordering metrics, diagnostic yield, and reimbursement. Parental samples are requested for concurrent WES, however only a single report, including inheritance, is issued for the proband. All consents are done by genetic counselors. Secondary findings of known and/or expected pathogenic variants, are reported for the recommended ACMG gene list, if requested. In an 18th month period, 74 cWGS were ordered with 73% performed with WES on both biological parents, ~5% on proband only. Insurance preauthorization was obtained for 73% patients and 26% were inpatient (no preauthorization required). Average insurance reimbursement was 28%, with yield for commercial payers significantly higher, at 66%. Moreover, 22% received definitive diagnoses, including 1 patient with 2 diagnoses. An additional 3 patients have likely positive results, pending functional studies. Of the 25% definitive or potential diagnoses, 47% were *de novo*. Incidental findings were reported in 4% of families who requested them; with 8% of families opting out. In two cases, cWGS revealed the breakpoints of large deletions and absence of heterozygosity in recessive genes harboring a missense variant on the other allele. Having this more complete information lead to a more direct and potentially faster diagnosis that may have been otherwise overlooked, however, in the majority of cases exome sequencing would have been comparable. Overall diagnostic yields for cWGS are similar to those reported for WES in this small but unbiased study, suggesting WES is currently more cost effective. .

2556F

Single exon resolution copy number analysis significantly increases clinical sensitivity of NGS. *R.R. Kelly, N.A. Rouse, W.A. Langley, P.L. Nagy.* MNG Laboratories, 5424 Glenridge Drive NE, Atlanta, GA 30342.

The detection of copy number variants (CNVs) is essential to providing a comprehensive approach to reaching a differential diagnosis. Typically, a large-scale method such as array-based copy number analysis is used to detect larger variations across the genome. While these methods offer a more efficient approach to covering a wide genomic area, they often fail to identify smaller CNVs that may be causative of a patient's phenotype. In our custom next generation sequencing (NGS) panels, we seek to provide a more comprehensive diagnosis through the detection of CNVs down to a single exon resolution. Using a custom designed capture reagent and the EXCAVATOR copy number analysis software, we are able to detect CNVs of greater than ten exons across all genes included in our panels. Additionally, we are able to detect CNVs at single exon resolution through the use of supplementary intronic probes for all genes with known pathogenic CNVs of ten exons or fewer. Since the inclusion of copy number analysis with our custom NGS panels in September 2016, we have seen a 1.6% increase in the total number of positive reports issued. Here we present multiple case studies in which single exon resolution copy number analysis was crucial in reaching a diagnosis across a variety of phenotypes.

2557W

A comparative study of the CYTAG® CGH and CYTAG® SuperCGH DNA labeling kits to detect CNVs with small amounts of DNA. *A.L. Mosca-Boidron, N. Marle, M. Payet, C. Ragon, J. Szczepaniak, D. Callegarin, M. Poulleau, P. Callier.* Laboratoire de Génétique Chromosomique et Moléculaire, DIJON, France.

Array comparative genomic hybridization (aCGH) is a powerful clinical diagnostic tool that can be used to evaluate copy number variation (CNV) changes in the genome. The detection of CNVs (gains and deletions) depends on the quality and quantity of DNA, especially for prenatal or oncology samples. Until now, aCGH in diagnosis has been done using more than 1μg of DNA, as recommended by several manufacturers. The principle of aCGH is based on the quantitative comparison between a test DNA and a reference DNA, which depends on the labeling of DNA. Here, we present a new aCGH labeling kit called CYTAG® SuperCGH labeling kit that accurately detects CNVs with less than 100ng of DNA and compare this new kit with the one previously used in our laboratory, namely CYTAG® CGH Labeling kit. Validation of the labeling procedure with this new kit included a comparison between the two commercial kits on 100ng of seven different samples: three Glioma cases (96ng, 32ng and 8ng), one postnatal case with three DNA quantities (90ng, 50ng and 30 ng), and one prenatal case with 90ng. CNVs in tumor samples (Glioma), postnatal and prenatal samples were compared. QC metrics were compared for samples labeled separately using each kit. For Glioma, DLRS were 0.43 vs 0.34 and 0.26 vs 0.21 with the traditional and new labeling kit, respectively. The Glioma sample with only 8ng of DNA gave a DLRS of 0.19 with the new kit but no comparison was possible as there was not enough DNA for the other kit. For postnatal cases, the DLRS were 0.14 vs 0.12; 0.17 vs 0.13; 0.21 vs 0.14. Finally, DRLS were 0.21 vs 0.16 for the prenatal case. The two kits allowed the detection of CNVs with better DLRS spread values and higher signal-to-noise ratios with the new CYTAG® SuperCGH labeling kit. Requirements for validation were achieved successfully and the comparative study showed that the labeling efficiency of the CYTAG® SuperCGH kit was higher than that achieved with the CYTAG® CGH kit. This study showed that the CYTAG® SuperCGH labeling kit was especially useful for CGH experiments using prenatal, postnatal and tumor samples. Moreover, we showed that Glioma-CNV could be detected in as little as 8ng of DNA. This technology is particularly useful in diagnosis where small amounts of DNA are available, and offers new possibilities for aCGH analysis in prenatal, postnatal or oncology, where the samples are poor and precious.

2558T

Gene-specific criteria for PTEN variant curation. *J. Mester[1], R. Ghosh[2], T. Pesaran[3], R. Karam[3], R. Huether[4], K. Hruska[1], H. Costa[5], K. Lachlan[6], J. Ngeow[7], J. Barnholtz-Sloan[8], L. Zhang[9], K. Sesock[10], S. Plon[2], M. Hegde[11], C. Eng[8,10].* 1) GeneDx, Gaithersburg, MD; 2) Baylor College of Medicine, Houston, TX; 3) Ambry Genetics, Aliso Viejo, CA; 4) Tempus, Chicago, IL; 5) Stanford University School of Medicine, Stanford, CA; 6) Princess Anne Hospital, Southampton, UK; 7) National Cancer Centre Singapore, Singapore; 8) Case Comprehensive Cancer Center, Cleveland, OH; 9) Memorial Sloan Kettering Cancer Center, New York, NY; 10) Cleveland Clinic Foundation, Cleveland, OH; 11) Emory University, Atlanta, GA.

PTEN was the first gene within the ClinGen Hereditary Cancer domain for which an expert variant curation group was assembled. The group was tasked with optimizing the ACMG/AMP Variant Interpretation Guidelines to specifically curate *PTEN* variants, with the hope of resolving variants with a ClinVar status of "conflicting interpretations of pathogenicity" or "uncertain significance". We present our finalized benign and draft pathogenic *PTEN*-specific variant curation criteria. For benign criteria, modifications to current guidelines include setting allele frequency thresholds for stand-alone and strong criteria; creating criteria for homozygous observations, observations in trans with known pathogenic variants, and lack of segregation; and adding a supporting-level criterion for functional evidence. These criteria achieved a benign or likely benign classification for 13/15 (86.7%) consensus ClinVar benign/likely benign variants used as a test set. Optimizations to ACMG/AMP pathogenic criteria include setting boundaries for functional domains and use of very strong vs. moderate-level evidence for truncating variants; defining functional assays for strong or an added supporting-level evidence; defining phenotype required for supporting-level evidence; establishing numbers of meioses and families for segregation criteria; and allowing use of a strong criterion for more than one *de novo* occurrence without maternity/paternity confirmation. Given the paucity of known benign missense variants for validation of *in silico* predictors, the group did not adopt the existing pathogenic or benign supporting-level computational criteria for missense variants. The group also did not adopt the existing supporting-level evidence for a "reputable source" reporting a variant as pathogenic or benign given the requirement that classifications made by this group be highly evidence-based. Following testing of the pathogenic criteria on a set of 15 consensus ClinVar pathogenic/likely pathogenic variants, all criteria will be used to curate a final test set of ClinVar variants with conflicting interpretations or of uncertain significance. The final criteria will be incorporated along with the curation process into the expert panel application. These criteria will be compared with gene-specific optimizations developed by other ClinGen expert panels to identify common and unique modifications to the ACMG/AMP guidelines.

2559F

Covering all the bases: Case vignettes and diagnostic pipelines at the Stanford Center for Undiagnosed Diseases. *A.M. Dries[1,2], J. Kohler[1,2], D.B. Zastrow[1,2], D. Bonner[1,2], J. Davison[1,2], L. Fernandez[1,2], C. McCormack[1,2], M. Majcherska[1,2], S.B. Montgomery[3,6], D. Waggot[1,2], S. Marwaha[1,2], C. Prybol[1,2], N. Friedman[1,2], U.D.N. Members[9], J. Sampson[1,8], J. Hom[1,7], P. Fisher[1,4,8], E.A. Ashley[1,2,6], J. Bernstein[1,4,5], M. Wheeler[1,2].* 1) Stanford Center for Undiagnosed Diseases, Stanford University, Stanford, California 94305-5406; 2) Division of Cardiovascular Medicine, Stanford University, Stanford, California 94305; 3) Department of Pathology, Stanford School of Medicine, Stanford, California 94305-5324; 4) Department of Pediatrics, Stanford School of Medicine, Stanford, California 94305-5208; 5) Lucille Packard Children's Hospital Stanford, Palo Alto, California 94304; 6) Department of Genetics, Stanford School of Medicine, Stanford, California 94305-5208; 7) Department of Medicine, Stanford School of Medicine, Stanford, California 94305; 8) Department of Neurology, Stanford School of Medicine, Stanford, California 94305; 9) NIH Undiagnosed Diseases Network, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892.

Since the launch of the Undiagnosed Diseases Network (UDN), the Stanford Center for Undiagnosed Diseases (CUD) has evaluated more than 60 cases of very rare or unique disease. Of those cases at the CUD with initial genetic and clinical evaluation completed, over 50% of these cases have either a clinical or genetic diagnosis while other cases remain unsolved. Initial evaluation at the Center includes: precise phenotyping, exome or genome sequencing, metabolomic studies, and network-wide clinical case matching. For the unsolved cases with completed initial evaluations, unique strategies are being employed to improve the diagnostic yield. The Stanford Center for Undiagnosed Diseases uses several multi-omics pipelines to investigate these initially unsolved cases. These clinical case driven pipelines include analyzing research genetic variants at the UDN Model Organism Screening Core, multi-tissue RNA-seq, immune profiling, microbiome analysis, long read genome sequencing, metabolomics, deep sequencing, and mosaicism analyses. With these additional tools, the Stanford Center for Undiagnosed Diseases can further evaluate these undiagnosed cases. Six unsolved cases are presented here and highlight these comprehensive diagnostic pipelines.

2560W

VarSome, the Human Genomic Variant Search Engine. *A. Massouras.* Saphetor, Lausanne, Vaud, Switzerland.

There is now a wealth of information available about human genomic variants. However, this is usually distributed across multiple, independent databases, which greatly complicates the collection of available information on a given variant. Here, we present VArSome, a novel knowledge base and aggregator for human genomic variants. It provides an intuitive web-based interface, offering access to the more than 30 billion items of variant and gene annotation we have accumulated in less than a second per query. Varsome provides functional annotation, calculated in real time using the RefSeq and Ensembl transcript databases; it correctly matches (and displays) equivalent indels to annotations; it displays disease and phenotype associations, drug interactions, relevant clinical trials, population frequencies, and prediction scores from a multitude of sources. The large number of external sources is regularly updated to the most recent release. Variants can be searched in VarSome in many different ways including dbSNP rsIDs, explicit descriptions using HGVS notation or genomic coordinates (e.g. BRAF:V600E, chr6-161127501-A-G) or even by pasting a line from a VCF file. Varsome is not limited to known variants: the user may obtain functional annotation for any variant. Users can mark favorite variants, post comments or link articles to specific variants which will be visible to anyone else who visits that variant. This provides an easy and centralized way of sharing information about known or novel variants. Varsome also provides stable, unique links to any variant, known or not, facilitating the sharing of variants between collaborators. VarSome is freely available to the community on varsome.com.

2561T

Going beyond the ACMG recommendations for reporting secondary findings: From decision-making to follow-up. *N.T. Strande[1], J. Booker[1], A.K. Foreman[1], G.T. Haskell[2], K. Lee[1], J.M. O'Daniel[1], B. Powell[1], M. Roche[1], B.A. Seifert[1], J.P. Evans[1], K.E. Weck[1], J.S. Berg[1].* 1) UNC Hospitals, Chapel Hill, NC; 2) Duke University, Durham, NC.

As clinical laboratories implement genome-scale sequencing for diagnostic purposes, they must consider secondary findings (SF) and decide which to report. The American College of Medical Genetics and Genomics (ACMG) recommends reporting known or expected pathogenic variants in 59 genes deemed highly actionable, however only a small number of disease-associated genes were systematically reviewed for inclusion on the list. Furthermore, the full phenotypic spectrum and penetrance associated with suspected pathogenic variants in these genes has not been established. The North Carolina Clinical Genomic Evaluation by Next-generation Exome Sequencing (NCGENES) project analyzed participant exomes for pathogenic and likely pathogenic variants in a list of 165 medically actionable genes partially overlapping the ACMG recommended list and some additional conditions. The overall rate of SF was 2.99% (20/669) consistent with prior studies and did not vary by age, sex, or race. Variants reported as SF were found in 13 genes, including 4 on the ACMG list (*BRCA2, MSH6, KCNQ1,* and *SCN5A*) and 9 others (*BRIP1, PALB2, FH, BCHE, SERPIN1C, F8, HBB, PROC,* and *BTD*). Most SF were in *SCN5A* (3 cases) and *F8* (3 cases). Clinical follow-up (analyte testing and symptom review) to confirm a diagnosis consistent with the SF occurred in 12/20 cases: 6 negative, 3 positive, and 3 unclear. Negative follow-up evaluation could be due to variant over-interpretation (false positive) or an overdiagnosis (due to variable expressivity and/or incomplete penetrance). SF suggestive of long QT (LQT) syndrome comprised the majority of results with negative follow-up evaluation (4/6). The remaining 2 negative follow-up evaluations were in patients with *F8* SF and normal factor VIII levels. A third patient with a family history of hemophilia had a hypomorphic F8 allele, but subclinical levels of factor VIII complicating clinical management. Although many of our SF were negative upon follow-up, it is unclear whether these patients will remain asymptomatic or if these results represent true false positives. Our results suggest a need to better define the spectrum of pathogenic variants found in medically actionable genes and their associated phenotypes, and highlight the need to better define the penetrance and expressivity of SF in asymptomatic populations. It will be critical to set thresholds for return of results that strike a balance between case finding versus false positive/overdiagnosis results.

2562F

The discussion of whether in vitro fertilization (IVF) or preimplantation genetic diagnosis (PGD) in inv(9) carriers. *H. Chen[1], W.E.I. Wu[1,2].* 1) Nanjing Medical University, Nanjing, China; 2) Jiangsu Province Hospital.

Objective:;Research on whether and which type of INV(;9);need PGD. Design:;453 Couples having INV(;9);were analyzed in this retrospective case-control study,and some data come from periodical literature.Blastomeres or blastocysts from these couples were analyzed by next-generation sequencing technique. Data was tested by the chi-square test. Results:;For various types of inv(9),35.1% were inv(9)(p12;q21),21.7% were inv(9)(p12;q13). The number of all fresh embryos and frozen embryo transplantation period is 723, the clinical pregnancy rate was 42.5%, biochemical pregnancy rate was 50.3%, the abortion rate was 5.4%. There were 63 couples not pregnant with two or more IVF, with the exception of the uncommon subtype ,(p11; q13) is the highest rate subtype with not pregnant, followed by (p12; q21).The age of Most of them are over 35 years old. The inverted segment located at the end of the chromosome such as p24, and q34 had a lower clinical pregnancy rate. In 41 couples with PGD,;the number of transplantation period is 42, the clinical pregnancy rate was 66.7%, biochemical pregnancy rate was 71.4%, the abortion rate was 11.9%. The clinical pregnancy rate of PGD is higher than the rate of IVF, chi-square is equal to 9.45, the difference has statistical significance,and the miscarriage rate of PGD is higher than IVF abortion rate, chi-square is equal to 3.1, the difference do not has statistical significance. Conclusions:;In this study, the distribution of the inv(;9);subtype is in accordance with previous report that the common breaking point of INV(;9);located in 9p12 or 9q13-21.1. Regular IVF can solve reproductive problems of most of the patients,but the pregnancy rate of only one transplantion is low. Compared IVF with PGD, PGD can increase the clinical pregnancy rate, but the impact on the abortion rate is not clear.In the case of recurrent miscarriage, advanced age, and the large inversion fragment were advanced to undergoing PGD. Besides on the basis of this study and the previously report that inv (9) (p24.3; q34.1) has had live births with adverse clinical phenotype, for segments of inversion is located at the end of chromosomes also recommend PGD. In addition depend on the report that the segment q21 - q22 may be pathogenic, the patients with q21-q22 inversion may be recommended to PGD to reducing the risk of neonatal have disease due to chromosomal problems.

2563W

Improved molecular tracking of individual genomes for clinical whole-genome sequencing. *S. Batalov, Y. Ding, S.A. Nahas, S. Chowdhury, J. Cakici, S. Caylor, S. Kingsmore, D.P. Dimmock, N. Veeraraghavan.* Rady Children's Institute for Genomic Medicine, San Diego CA.

The speed, accuracy and lowering costs of whole-genome sequencing (WGS) can change the acute care of infants in the intensive care units (ICU). In a clinical setting, it is imperative to establish a robust and scalable system for chain-of-custody of every patient's genome through all steps of the high volume rapid sequencing lab and *in silico* workflows. Beyond routine workflow solutions, such as barcoding of tubes and plates, most clinical laboratories employ some method of molecular fingerprinting of genomic samples. Many of these solutions are costly or not fast enough to be deployed in a clinical laboratory performing large numbers of rapid WGS tests. Short-tandem repeat (STR) variations have been useful in assigning identity in wide-ranging areas of human genetics. We present a scalable and cost-effective framework based on established STR markers, specifically tuned for ultra-rapid turn-around-time, and that is well suited to an Illumina-based clinical whole genome workflow. Upon parental consent, for each entering DNA sample, duplicate aliquots are stored – one for the Rapid WGS workflow, and the other for later confirmatory PCR-guided Sanger sequencing for the confirmation of pathogenic variants prior to the return of patient results, under IRB protocol 160468. PCR-guided Sanger sequencing and identity testing involves using a very small portion (0.5-1 ng) of extracted gDNA used in the conventional and inexpensive GlobalFiler STR allelotyping workflow using a standard ABI 3500 or 3100 sequencer. The identity match of the two aliquots is ensured by using short-read WGS sequencing data after Illumina 2500/4000 sequencing by in silico independent determination of the STR allelotyping using a modified lobSTR (Gymrek M. et al. (2012)) computational workflow. To reduce the computational time, the BAM file can be restricted to the vicinity of the 18 STR regions (as well as two non-STR X/Y-linked markers). Because all accurate genomic mapping are readily available after the computational workflow, the overhead for the STR allelotyping implementation is minimal (<5 CPU-minutes per sample; the reagent cost is ~ \$25 / sample). The output alleles for all STR markers establish a unique fingerprint that ensures specificity of at least $1/10^{15}$. Recent improvements to the workflow (including additional non-STR marker (Y-indel) and a comprehensive cross-match testing in typical and in consanguineous families) will be presented.

2564T

Workshop in genomic medicine for paediatric specialists. *A.D. Gilbert[1], E. Bedoukian[7], L. Kyriakopoulou[6], C.R. Marshall[1,5,6], M.S. Meyn[1,2,3,4], S.M. Jamal[2,6], N. Monfared[1], S. Bowdin[1,2,3].* 1) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Departments of Paediatrics, University of Toronto, Toronto, ON, Canada; 4) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 7) Children's Hospital of Philadelphia - Roberts Individualized Medical Genetics Center, Philadelphia, PA, USA.

   Background: The clinical application of exome and genome sequencing is revolutionizing the diagnosis of genetic disorders due to the high diagnostic yield of these genetic tests. Not only does this impact the services available to a family, it can provide an explanation, recurrence risk, anticipatory guidance and, in many cases, management improvement. Since there are not enough genetic health care professionals to facilitate the coordination and testing of all children suspected of a genetic disorder, more responsibility will rest with the paediatric specialist. The literature shows that the medical genetic knowledge base of the practicing non-genetics clinician is insufficient. Hence, our own ASHG: Genomic Medicine Education Consortium strives to improve access and quality of genomic educational resources for the practicing clinician. **Objective:** To establish the SickKids paediatric faculty's comfort level with counseling and offering genomic testing, a survey was electronically distributed. Based on the results of this e-survey, a half-day workshop was developed. **Method:** The Centre for Genetic Medicine hosted a 2.5-hour, small group, didactic workshop focusing on the key components of performing clinical genomic sequencing and returning results in the pediatric setting. The target audience was pediatric specialists at SickKids. **Results:** Twenty-five individuals registered for the workshop, and 14 attended. Data was obtained from question sets distributed at the beginning and end of the workshop, in addition to a feedback evaluation. **Conclusions:** Based on the feedback, this workshop was a needed addition to the genomic education of SickKids providers. 91% of the respondents said they would recommend the workshop, 67% thought the 2.5 hour time-frame was "just right" in length, the majority gave all the lecturers positive feedback, and 55% agreed strongly they were "more prepared to effectively integrate genome testing into my practice." Despite the apparent success from these parameters, the 44% no-show rate suggests another mode of information delivery should be considered. Thus, we will be offering 1-hour in-service educational sessions for paediatric specialists with delivery of a comprehensive slide deck designed so that any genetic provider can deliver with confidence.

2565F

Developing frameworks to evaluate diagnostic genomic testing strategies for rare disease and cancer. *R. Scott[1], Z.C. Deans[2], E.R.A Thomas[1], A. Hamblin[1,3], E.V. Davison[2], E. Graham[2], R. Snow-Miller[2], S. Wordsworth[4], C. Turnbull[1], N. Murugaesu[1], S. Henderson[1,2], T. Fowler[1], M.J Caulfield[1], S.L. Hill[2].* 1) Genomics England, London, United Kingdom; 2) NHS England, London, United Kingdom; 3) Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom; 4) University of Oxford, Oxford, United Kingdom.

   Advances in genomic technologies, including the feasibility of whole genome sequencing (WGS) have been demonstrated by the UK 100,000 Genomes Project and other published studies, to present a significant opportunity to improve genomic testing in rare disease and cancer patients. In response the National Health Service (NHS) in England is planning a substantial reconfiguration of genomic testing covering the complete testing spectrum from single gene through to WGS. This approach will deliver mainstreaming of cutting edge genomic technologies into the NHS. To inform the testing strategy in this reconfigured service, a working group was formed consisting of experts in clinical and laboratory genetics, cancer specialists, health economists, healthcare policy and commissioning experts, and patient representatives. The group sought to develop objective tools to evaluate the optimal genetic testing strategy for each clinical indication, recognising the need to be flexible and responsive to the emerging evidence base. Separate approaches were developed in rare disease and cancer. The tools took into account technical considerations such as number of known genes, pace of new discoveries, variant types relevant to indication, and the suitability of different technologies to detect these. Health economic assessment determined the value of the genomic testing strategy, not only of variant detection but also the overall patient care pathway. The tools were optimised with disease area experts before being applied across the spectrum of rare inherited disease and cancer. In rare disease, the resulting recommendations for each clinical indication included guidance on optimal test technology, family structure (trio or proband-only sequencing), and technical requirements. In future, annual review of the recommendations will allow incorporation of new genetic knowledge, and adaptation as sequencing technology and costs evolve. In cancer the systematic evaluation of genomic tests for a range of targets, in different tumour types, with clinically relevant testing times, will enable access to genomic medicine for all patients using the most effective and efficient approach. These tools will be presented together with an overview of their outputs and implications for the architecture of a comprehensive genomic testing strategy for the NHS.

2566W

Benchmarking the quality of diagnostic next generation sequencing. *S. Deans[1], M.W. Gutowska-Ding[2], F. Khawaja[1], K. Brugger[3], J. Wook Ahn[4], S. Abbs[3], J. Coxhead[5], P. Westwood[6], K. Thompson[7], C. Boustred[8], G. Taylor[4], E. Souche[9], S. Sander[10], B. Ylstra[11], S. Patton[2].* 1) UK NEQAS, Edinburgh, United Kingdom; 2) EMQN, Manchester, United Kingdom; 3) Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom; 4) Viapath Genetics Laboratories, London, United Kingdom; 5) Institute of Genetic Medicine, Newcastle University, Newcastle, United Kingdom; 6) NHS Greater Glasgow & Clyde, Glasgow, United Kingdom; 7) The Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom; 8) Genomics England, London, United Kingdom; 9) Center for Human Genetics, Leuven, Belgium; 10) MUMC Clinical Genetics, Maastricht, Netherlands; 11) VU University Medical Center, Amsterdam, Netherlands.

Next Generation Sequencing (NGS) is being introduced into clinical genetics laboratories worldwide. It is not possible for the large amount of data generated by NGS to be duplicated by alternative methods to enable internal validation of all results, therefore external assessment of data quality and consistency is required. The European Molecular Genetics Quality Network (EMQN) and the UK National External Quality Assessment Service (UK NEQAS) for Molecular Genetics have developed a joint External Quality Assessment (EQA) scheme for NGS, with the aims to assess and improve quality, enable laboratories to benchmark their NGS service against others, work towards consistency of reporting clinical results generated by NGS and contribute towards defining best practice. EMQN and UK NEQAS offer many disease-specific EQA schemes, and the objectives for developing an NGS EQA were to make it generic (independent of genes tested, diseases, and platforms used), applicable to as many users as possible and to provide them with a broad range of quality indicators on their NGS data. So far, four pilot schemes have been provided. Laboratories were sent a genomic DNA or FFPE sample for sequencing using their usual approach, submit technical details, and genotypes of reportable variants detected. The sequence of the DNA provided was validated in three clinical laboratories using different NGS platforms. The two latest EQA runs (2015, 2016) were divided into Germline and Somatic NGS, in order to address the different challenges of NGS in these settings. Participating laboratories has increased from 30 in 2013, to 303 in 2016. We have established a pipeline to produce a consensus genome based on the EQA detected variants, reducing the need to rely on time-consuming and expensive prior validation. This has required the development of sophisticated tools for the integration and benchmarking of NGS data. These NGS germline and somatic EQAs are technology and testing context independent and seamlessly fit with wet laboratory and bioinformatics processes. Disease-specific EQA has improved the quality of results and consistency in clinical reports. The NGS EQA results enable clinical diagnostic laboratories to address the quality of their NGS testing. These technical NGS EQAs will play an important role in enabling laboratories to benchmark this new technology, assess the accuracy of data and facilitate high quality reporting for patient benefit.

2567T

Correctly building, evaluating and using clinical grade pathogenicity classifiers for variant of unknown significance. *G. Bejerano[1,2,3], K.A Jagadeesh[2], J.M. Paggi[2], J.S. Ye[4], J.A. Bernstein[3].* 1) Developmental Biology, Stanford University, Stanford, CA; 2) Computer Science, Stanford University, Stanford, CA; 3) Pediatrics, Stanford University, Stanford, CA; 4) Biology, Stanford University, Stanford, CA.

The diagnosis of known monogenic Mendelian disorders can be highly labor intensive, with more than 40 hours of expert time needed for individual diagnoses, when each patient may present up to 300 candidate genes. One way to alleviate clinician load in patient data analysis (and reanalysis) is to predict the likelihood that any observed rare variant in the patient genome may be pathogenic. The goal, per patient, is to predict as many variants of unknown significance (VUS) as "likely benign" while never misclassifying the 1-2 actual pathogenic variants in the patient as such. Many current pathogenicity classifiers either do not focus on Mendelian diseases (e.g., in their choice of training data), or test their predictions on data other than actual patient data. We have recently published (1) a novel clinical-grade Mendelian disease non-synonymous VUS pathogenicity classifier named M-CAP (Mendelian Clinically Applicable Pathogenicity). M-CAP can be explored via a dedicated web portal (2) and has also been kindly integrated into Annovar and dbNSFP. In our paper we show that very popular variant pathogenicity classifiers misclassify up to 38% of known pathogenic Mendelian mutations. We show M-CAP to outperform these methods at all thresholds and correctly dismisses 60% of rare, missense VUS in a typical genome at 95% sensitivity. Here we will extend on our published work in two ways: First, we will discuss the recent flourish of pathogenicity prediction tools. We will discuss some inadvertent errors that are possible in building such tools, or in comparing these tools to each other. This includes conflating rare and common variants, inflating performance by testing on data that was used by some classifiers, or the features they use, for training, and poorly balancing positive and negative data points during training, testing or comparison. We will show how omitting any of these rules results in misleading conclusions. We will discuss ways of circumventing these pitfalls, improving any classifier performance, and performing a fair assessment of multiple classifiers performance. We will also discuss, and give examples of how to best use these classifiers in the clinic. Finally, we will present our most recent unpublished work to improve and extend VUS pathogenicity prediction from the coding into the non protein coding genome. (1) Jagadeesh et al, Nat Genet 2016 https://doi.org/10.1038/ng.3703 (2) http://bejerano.stanford.edu/mcap.

2568F

Paternal inheritance of X chromosome CNVs aid in the interpretation of pathogenicity. *R. Burnside[1], M.K. Rudd[1], I. Gadi[1], K. Phillips[1], A. Penton[1], V.R. Potluri[2], H. Risheg[3], J. Schleede[1], S. Schwartz[1], J.H. Tepperberg[1], P. Papenhausen[1].* 1) Cytogenetics, Laboratory Corporation of America Holdings, Research Triangle Park, NC; 2) Cytogenetics, LabCorp, Houston, TX; 3) Cytogenetics, LabCorp, Seattle, WA.

   Clinical microarray analyses are performed for an ever-expanding number of indications including infertility, intellectual disability, congenital anomalies, and parental follow up studies. CNVs found on the X chromosome in female patients are particularly difficult to interpret due to protection offered from random X inactivation. Parental studies of X-linked CNVs are important since paternal inheritance from a clinically normal father supports a benign interpretation. We report on a large cohort of X chromosome CNVs identified in female probands by microarray analysis which were paternally inherited. From August 2008-November 2015, 1784 X chromosome CNVs were reported in males and females, including those resulting from unbalanced structural rearrangements of the X chromosome. Of these, 1073 (60%) were found in females and 428 (24% of total 1784) were followed up with at least one parent. Of the 428 cases with follow up, 217 were female patients and 63 CNVs were paternally inherited. Two fathers were followed up after abnormal microarray results for their daughters performed at outside labs, both of whom were found to be clinically normal carriers. Of the 63 total confirmed paternal CNVs, eleven were deletions and 52 were copy number gains. Four deletions occurred in the PAR1 region, and three deletions included the STS gene region. Two females had paternally inherited exonic DMD deletions, and two other females with exonic DMD deletions were shown not to be maternally inherited (no paternal studies were done). The CNVs ranged from 53 kb to 2.28 Mb. While clinical information on most fathers was lacking, three paternal carriers were reported to be similarly affected as their daughters, suggesting that the duplicated regions may be associated with X inactivation skewing or escape inactivation and affect both males and females. These data will aid in the interpretation of pathogenicity for paternally inherited X chromosome CNVs.

2569W

Considering other mechanisms of gene regulation in disorders of sex development. *M. Molina[1], M. García-Acero[1,2], O. Moreno[1], F. Suarez[1,2], C. Cespedes[2], J. Perez[2], N. Fernandez[2], J. Prieto[1], J. Auli[2], A. Rojas[1].* 1) Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, 11, Colombia; 2) Hospital Universitario San Ignacio, Bogota, Colombia.

   Introduction: Disorders of Sex Development (DSD) are a heterogeneous group of entities related to sex assignment. The DSD are classified according to the categories of the International Consensus of Chicago (2006) as 46, XX, 46, XY or variants of the sex chromosomes like 45,X or 47,XYY among others. Sexual differentiation is a complex, dynamic and genetically controlled process that depends on a delicate balance between sexual and autosomes genes. These genes are responsible for the tissue differentiation and the correct production of steroidogenic enzymes. Nowadays with the use of available techniques of molecular diagnosis, only 30 to 40% of individuals with DSD achieve diagnostic etiology. Therefore, in search of further causes of DDS, epigenetic modifications, must be considered as a relevant component of gene expression regulation in sexual differentiation. Methodology: transdisciplinary group of DDS of the Hospital Universitario San Ignacio performed a clinical and genetic characterization of 24 patients. The chromosomal analysis included high-resolution G and R-banding and FISH with a specific probe to evaluate the presence of the SRY gene. For assessing alterations in the number of copies of the NR0B1, SOX9, SRY, WNT4 and NR5A1 genes MLPA (Multiplex Ligation-dependent Probe Amplification) analyzes were performed using the SALSA MLPA P185-C1 Intersex MRC-Holland kit. Results: Twenty-four patients with DSD were analyzed. The clinical characteristics of these patients included hypospadias (29%), genital ambiguity (37%), male infertility (4%), primary amenorrhea (21%) and gonadal tumor (8%). 14 patients (58%) were assigned male gender, and 10 patients (42%) were assigned female gender. 19 individuals (79%) were 46, XY, 3 individuals (12%) were 46, XX, two patients (8%) presented mosaicism. Analysis of SRY-FISH was positive in all individuals XY. MLPA analysis was performed in all patients. Discussion: Most of our patients had no cytogenetic correlation with the assigned phenotype, making it difficult to assign diagnostic but justifying further studies. However, these results highlight the importance of carrying out the karyotype as an approach to etiological diagnosis, although it has been a questioned technique since it does not solve the controversy among gender assignment. It is worthwhile considering the hypothesis that some other genes are involved in the gonadal development and are susceptible of expression regulation by epigenetic mechanisms.

2570T

Identifying single fetal trophoblastic cells in the maternal circulation: A modified NGS genotyping method. *X. Zhuo, Q. Wang, L. Vossaert, A. Kim, S. Qdaisat, R. Salman, A. Breman, A. Beaudet.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

  Background: One of the challenges for cell based non-invasive prenatal testing (CB-NIPT) is to distinguish individual presumptive fetal cells from maternal cells in female pregnancies. We have sought a rapid, robust, versatile and low cost method to facilitate this assessment.   **Methods:** For samples from 12-16 weeks gestation pregnancies, fetal trophoblastic cells were picked from maternal blood using previously described methods (Breman et al. PMID:27616633). The isolated cells were processed using various types of whole genomic amplification (WGA) prior to genotyping. Multiple highly polymorphic genomic regions (including HLA-A and HLA-B) with 10-20 very informative SNPs around 200 bps interval were amplified with a modified method based on another published method (Debeljak et al. PMID:25132481). To enhance the power of identification, about 40 human identification (HID) amplicons were also utilized. Additional genotyping assays were established for common cystic fibrosis, Tay Sachs, and sickle cell mutations. For these disease mutations, testing was carried out using unfixed and fixed heterozygous single lymphoblasts and using normal fetal cells and results were compared for three different WGA protocols. NGS was performed on both Ion Torrent and Illumina platforms.   **Results:** The genotyping method allowed reliable differentiation of fetal and maternal cells. Using multiple highly polymorphic amplicons each including >10 highly informative SNPs with >1000 reads for haplotyping, maternal and fetal cells could be distinguished in >70% of the samples. SNP typing of additional HID amplicons with >200 reads was used to resolve uninformative cases, so all together >90% cases were identified. A paternal DNA sample is not required for use of this method to distinguish maternal and fetal cells in informative cases. The assay also successfully detected mutations of interest in several spike-in diseases cells.   **Conclusion:** We found this method to be superior to SNP arrays, multiplex genotyping methods, and STR analysis considering cost (<$10 prep + sequencing), throughput (2-6 hrs prep for 1-96 cells), and reliable distinction of fetal from maternal cells. This method derives its power by combining a highly informative haplotyping within individual reads and SNP genotyping. Comparison of different WGA protocols suggested modest bias during amplification. This method will facilitate genotyping to confirm fetal cells within an overall platform for performing CB-NIPT.

2571F

Comprehensive analysis of CYP2D6 variants and copy numbers using reverse-hybridization and real-time PCR based assays. *C. Oberkanins, A. Berndt, K. Gesson, H. Puehringer.* ViennaLab Diagnostics, Vienna, Austria.

  Introduction: The cytochrome P450 2D6 (CYP2D6) is an important liver enzyme involved in the metabolism of up to 25% of clinically used drugs. The CYP2D6 gene is highly polymorphic, with numerous (sub)variants described in the Human Cytochrome P450 Database (www.cypalleles.ki.se). While the most frequent allelic variations are caused by single nucleotide polymorphisms and small insertions/deletions, highly homologous regions in the CYP2D6 gene locus facilitate unequal cross-over leading to large deletions, duplications and gene conversions. Material and Methods: We developed a reverse-hybridization assay (PGX-CYP2D6 XL StripAssay) for the simultaneous detection of 19 sequence variations in the CYP2D6 gene, which define the most prevalent alleles impacting enzyme activity in Caucasians. For the detection of copy number variations a real-time PCR based assay (CYP2D6 RealFast CNV Assay) was established. The StripAssay and real-time PCR assay were validated on 118 and 98 samples, respectively. Results: The PGX-CYP2D6 XL StripAssay correctly identifies allelic variants with normal (*1, *2, *35, *39), reduced (*9, *10, *17, *29, *41) and no (*3 to *8, *11, *12, *14, *15, *40, *58) enzyme activity, and hence allows the classification of individuals into extensive (EM), intermediate (IM) and poor (PM) metabolizers. In addition, ultrarapid (UM) metabolizers and CYP2D6*5 carriers can be identified by quantifiying their abnormal copy number status using the CYP2D6 RealFast CNV Assay. Both assays demonstrate a test accuracy of >0.99. Conclusions: The metabolizer phenotype of patients treated with CYP2D6 substrates can be accurately determined by the combined use of both assays.

2572W

A highly specific, cost-effective solution utilizing a unique 2-enzyme system for SNP genotyping in pharmacogenetic studies. *D. Tsang[1], J. Graham[2], K. Datta[1], K. Beltz[1], S. Rose[1], Y. Wang[1], M. Fiddler[2], Y. Bao[1].* 1) Integrated DNA Technologies, Redwood City, CA; 2) Insight Medical Genetics, Chicago, IL.

   Introduction: Drug-gene interactions have been widely studied to help predict the efficacy and potential toxicity of drug therapies. The cytochrome P450 enzyme family is primarily responsible for the metabolism of pharmaceutical drugs. In particular, CYP2D6 is important as it is responsible for a wide variety of phenotypic outcomes and has a highly homologous pseudogene that complicates genotype calling. IDT rhAmp™ SNP Genotyping Assays combine a unique 2-enzyme system with DNA-RNA hybrid primers to detect genetic variations, including single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels). Allele-specific primers (ASPs) contain a 5' universal tail, a single RNA base targeting the SNP, and a 3' terminal blocking group. The thermostable RNase H2 only cleaves the primer at the RNA base, when the primer is hybridized to its perfectly matched target. Primer cleavage releases the 3' blocking group, which activates the primer for extension by a thermostable DNA polymerase engineered for enhanced specificity. **Methods**: rhAmp ADME assays, which target SNPs in genes known to be involved in drug absorption, distribution, metabolism, and excretion [ADME; published by the Clinical Pharmacogenetics Implementation Consortium (CPIC)], were generated using an optimized assay design algorithm that ensures high target specificity. The assays were tested with rhAmp Genotyping Master Mix using either gBlocks® Gene Fragments as artificial templates for positive controls or genomic DNA extracted from human whole blood, saliva, and buccal swab samples. rhAmp SNP reactions were run on a BioMark™ Real-Time PCR system (Fluidigm) utilizing the 96.96 Dynamic Array™ IFC for Genotyping (Fluidigm). **Result**: We successfully demonstrated >99% call rate for the SNP targets, and genotyping results were concordant with established orthogonal methods. CYP2D6 gene variants were accurately genotyped using genomic DNA without need for preamplification of the gene. **Conclusion**: rhAmp SNP Genotyping products provide a cost-effective and highly specific genotyping solution for pharmacogenetic studies. The assays are compatible with different sample types, and the results are concordant with those from an established technology.

2573T

Long read capture sequencing for clinical applications. *K.C. Worley[1,2], Y. Liu[2], Q. Meng[2], Y. Han[2], G.M. Weissenberger[2], J. Hu[2], S. Richards[1,2], D.M. Muzny[2], R.A. Gibbs[1,2].* 1) Dept Molecular & Human Gen, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

   We use long read sequencing technologies including Pacific Biosciences RSII and Sequel and Oxford Nanopore for genomic applications such as *de novo* genome assembly and structural variant (SV) investigation. We report here the use of target enrichment capture with these long reads for assessment of repeats and phasing of heterozygous variants.   Clinical exome sequencing with short read Illumina technology does not adequately address all known Mendelian diseases. Distinguishing between normal, pre-mutation, and full mutation triplet repeat expansion alleles is difficult or impossible with short reads. Phasing compound heterozygous mutations in a gene to determine if the mutations are in cis (on the same allele) or trans is also difficult or impossible with short reads. The goal of this project is to build a cost-effective sequencing platform for variant discovery in complex and difficult clinically important genomic regions.   The initial capture design includes 94 genes with a total targeted capture size of 13.4 Mb. Nine of the genes are repeat expansion containing genes. The remaining genes were selected for their importance in clinical genetic diagnosis with both small genetic changes and SVs. Intact genomic DNA from 6 breast cancer patient samples was sheared to 2kb~6kb and amplified by ligation-mediated PCR. The pre-capture library was hybridized with NimbleGen SeqCap EZ platform probes and the capture products sequenced using RSII or Sequel platforms.   In the preliminary analysis, filtered subreads were mapped using the long read aligner BLASR; the resulting BAM files were used to generate capture metrics and processed by PBHoney to generate SV calls. Robust capture efficiency was demonstrated with 94% of the targets covered at 20x with 1~2 cells. More than half of the repeat expansion disease genes were captured successfully. A HapMap DNA yielded contiguous coverage at 100x in high GC regions including the FMR1 5' UTR enabling determination of the copy number of this CGG repeat.   We will report the initial results using long read sequencing technologies for targeted sequencing to assess genetic variation including SNPs, small indels and large SVs. We will also report laboratory experiments to refine and improve the regional capture and computational experiments to improve the sequence data assessment. Multiplex capability and customizable target selection provide a foundation for the clinical implementation of cost-effective long read sequencing.

2574F

Clinical reassessment of post-laboratory variant call format (VCF) files.
L.F. Al Subaie[1], S. Alturki[2], A. Alothaim[2], M. Alfadhel[1], A. Alfares[2]. 1) Division of Genetics, Department of Pediatrics, King Abdulaziz Medical City, Riyadh, Saudi Arabia; 2) Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City, Riyadh, Saudi Arabia.

    Next generation sequencing (NGS) has been leading the genetic study of human disease for the past ten years. And since then we have been dealing with a huge number of sequence variant data generated from both whole genome (WGS) and whole exome sequencing (WES) that is stored in Variant Call Format (VCF) files. To our knowledge there are few discussions about the utility of VCF files reanalysis. Here we retrospectively analyzed twenty samples. The samples were divided into WGS with 14 cases (70%) and WES with 6 cases (30%). All cases of WGS had a previous negative results from outside lab except one case of WGS had a previous result of variant of uncertain significance. and one WES case had a previous result of variant of uncertain significance. We reported seven cases (n=20) differently from the outside laboratory, this account for almost ~35% and mainly due to the ability to gather more information about the patient's phenotype. One WGS case changed to negative case despite the previous report of variant of uncertain significance. Also, we identified variants related to the patient's phenotype in six cases; Two of them were WGS with and four of them were WES with previous negative results. Comprehensive phenotyping of the individuals is crucial step in identifying candidate phenotype-related variants. We outline the benefit from access to the patient's medical records and communication with referring physicians. This study findings match up with previous researches that highlights particularly the importance of patient's clinical data for variant classification.

2575W

A single assay system for CNV, AOH, and Seq Var genetic testing. *S. Shams[1], A. Venier[1], S. Verma[1], A. O'Hara[1], I. Simonic[2].* 1) BioDiscovery, Inc., El Segundo, CA; 2) Addenbrooke's Hospital, Cambridge, UK.

    Traditionally, large structural genomic aberrations, such as Copy Number Variations (CNV) and regions with Absence of Heterozygosity (AOH), have been the domain of Cytogenetics, using methods such as karyotyping, FISH, and Chromosomal MicroArrays (CMAs). Similarly, smaller structural aberrations, Sequence Variants (Seq Var), have been the domain of Molecular Genetics using tools such as PCR, Sanger sequencing, and more recently Next-Generation Sequencing (NGS). In this paper, we present an approach that utilizes NGS technology to simultaneously detect variants regardless of size and nature, e.g. CNV, AOH, and Seq Var. The benefits of integrated review include facilitating identification of heterozygous events, as well as providing rapid results when urgent diagnosis is required. The core component of the integrated system is a software platform that effectively extracts CNV and AOH data using read depth coverage, and using an integrated filtering and visualization framework, presents a single view to perform clinical case review in individual as well as trio samples. We will describe this system and workflow in detail and demonstrate the utility of the approach on a set of anonymized clinical samples referred for various constitutional conditions. These samples have been processed for molecular testing using the Illumina TrueSight One panel with targeted coverage across more than 4800 genes associated with disease. In the present study, we will evaluate the utility of obtaining CNV and AOH data from the same NGS data using the BAM MultiScale Reference algorithm and integrating this data along with Seq Var data into a single analysis pipeline. The results of this analysis will be presented.

2576T

Copy number variant discrepancy resolution using the ClinGen dosage sensitivity map. *E.R. Riggs[1], T. Nelson[1], E. Andersen[2,3], S. Aradhya[4], O. Canto[5], B.A. Hilton[3], V. Jobanputra[6,7], S. Kantarci[8], W.A. Khan[9], H. Kearney[10], C.P. Lorentz[10], P. Paulraj[3], D. Pineda-Alvarez[11], R.A. Rowsey[10], D. Ritter[12], Y. Shen[13], M.D. Speevak[14,15], E. Thorland[10], C.L. Martin[1].* 1) Autism & Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA; 2) ARUP Laboratories, Salt Lake City, UT; 3) Department of Pathology, University of Utah, Salt Lake City, UT; 4) Invitae, San Francisco, CA; 5) GeneDx, Gaithersburg, MD; 6) New York Genome Center, New York, NY; 7) Columbia University Medical Center, New York, NY; 8) Quest Diagnostics Nichols Institute, San Juan Capistrano, CA; 9) Department of Genetics and Genomic Sciences, Icahn School of Medicine, Mount Sinai, New York, NY; 10) Department of Pathology and Laboratory Medicine, Mayo Clinic, Rochester, MN; 11) Courtagen Life Sciences, Inc., Woburn, MA; 12) Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, TX; 13) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 14) Trillium Health Partners, Credit Valley Site, Mississauga, ON; 15) Department of Laboratory Medicine and Pathobiology, University of Toronto, ON.

Resolution of inter-laboratory conflicts in clinical variant interpretation is a critical step towards standardizing interpretations and improving patient care. Evaluating interpretation discrepancies in copy number variants (CNVs) is complicated by the fact that variants are often unique; with the exception of recurrent segmental duplication-mediated events, there is rarely another CNV with the same genomic coordinates available for comparison. As such, CNVs are typically evaluated by their overlapping genomic content, with a particular focus on genes in the region that may be subject to dosage sensitivity. The Clinical Genome Resource (ClinGen) is an NIH-funded effort dedicated to identifying clinically relevant genes and variants. As part of ongoing curation efforts, ClinGen evaluates genes and genomic regions for evidence of dosage sensitivity; to date, over 1000 genes/regions have been evaluated, 318 of which meet criteria to be considered known haploinsufficient (HI) or triplosensitive (TS) genes/regions. In general, we recommend that deletions containing known HI genes and duplications containing known TS genes be interpreted as "likely pathogenic" (LP) or "pathogenic" (P), and that CNVs involving the same known dosage sensitive genes receive the same clinical interpretation regardless of other factors, such as size. Currently, ClinVar contains 19,000 CNVs; of these, there are approximately 3594 discrepancies (19%) amongst CNVs of the same copy number with at least 50% overlap. It is possible that some of these conflicts, particularly older ones, could be resolved by comparing against the current ClinGen dosage map; evidence for dosage sensitivity may have emerged since the CNVs were last evaluated. Within the ClinVar CNV dataset, we identified 896 deletions overlapping known HI genes and 37 duplications overlapping known TS genes with interpretations other than LP or P. ClinGen will work with the original submitters of these data to determine whether changing these classifications is appropriate within the context of the particular clinical case(s). Updating these calls to reflect the current understanding of dosage sensitivity for these genes/regions could result in the resolution of up to 25% of existing conflicts in ClinVar CNV data. Efforts such as this one to address "easily resolvable" discrepancies will put more consistent interpretation data into the public domain and allow laboratories to focus attention on more complex discrepancies.

2577F

Defining quality standards for clinical whole exome sequencing: A national collaborative study of the Dutch Society for Clinical Genetic Laboratory Diagnostics (VKGL). *M.G. Elferink[1], I.J. Nijman[1], M.A. Haagmans[2], A.H.B. Bollen[3], R.W.W. Brouwer[4], B. de Koning[5], L.F. Johansson[6], D. van Beek[7], O.R. Mook[2], M. van Slegtenhorst[8], S. Stegmann[5], B. Sikkema-Raddatz[6], M. Nelen[9], Q. Waisfisz[7], M. Weiss[7], J.D.H. Jongbloed[6], N. van der Stoep[3], K.L.I. van Gassen[1].* 1) Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; 2) Department of Clinical Genetics, Academic Medical Center, Amsterdam, the Netherlands; 3) Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands; 4) Center for Biomics, Erasmus Medical Center, Rotterdam, the Netherlands; 5) Department of Clinical Genetics, University Medical Center Maastricht, Maastricht, the Netherlands; 6) Department of Genetics, University Medical Center Groningen, Groningen, the Netherlands; 7) Department of Clinical Genetics, VU University Medical Center, Amsterdam, the Netherlands; 8) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 9) Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands.

Clinical whole exome sequencing (WES) has proven to be very effective for diagnosing heterogeneous genetic diseases, and has therefore been adopted in standard Dutch genetic diagnostic services. Quality standards have been described for gene panel based next generation sequencing, but for clinical WES these have not been determined yet. In this study, we present a nationwide quality assessment scheme within all eight university medical centers in the Netherlands in order to improve and harmonize quality and to formulate a quality standard for clinical WES. The exome of Genome-in-a-Bottle (GIAB) sample NA12878 was sequenced and processed by the Dutch genetic diagnostic centers according to their own standard diagnostic protocols. Anonymized VCF and BAM files were collected and used for comparisons. VCF files were used to calculate variant detection sensitivity and precision using the GIAB high confidence call set (v2.19). BAM files were used to calculate informative exome coverage statistics based upon a standardized target defined by all coding exons of UCSC and Ensembl ±20bp intron flanks. Results showed that single nucleotide variant (SNV) detection sensitivity varied between 96.3% and 99.1% (average 98.1%) and that INDEL detection sensitivity varied between 86.1% and 96.1% (average 90.4%). Precision of SNV detection was between 98.0% and 99.8% (average 98.9%), and precision for INDEL detection was between 74.8% and 97.0% (average 91.0%). The mean coverage of the exome varied between 55X and 152X. Completeness of the exome (defined as the percentage of bases with ≥15x informative coverage) varied between 90.0% and 96.4% (average 93.5%). Downsampling analysis indicated that completeness is a good predictor for quality and can be used to determine the required sequencing output for a desired quality threshold. The nationwide quality assessment scheme shows an acceptable concordance between the Dutch genetic diagnostic centers, although harmonisation of lab protocols and bioinformatics may lead to improved concordance and quality. We propose the following quality standards for a clinical WES based upon the GIAB high confidence call set (v2.19): a minimum sensitivity of 97% for SNVs and 85% for INDELs, and a minimum precision of 98% for SNVs and 85% for INDELs. These quality assurance standards will be deposited as field standards at the Dutch accreditation council for use in ISO15189 laboratory accreditation.

2578W

Characterization of incidental findings identified by targeted testing for gene deletions and duplications. *A.M. Janze, T. Brandt, A. Ondov, L. Butler, G. Richard, K. Hruska, J. Meck. GeneDx, Gaithersburg, MD.*

A clinical diagnostic laboratory sometimes encounters an unexpected test result that is unrelated to the indication for testing. These events are considered incidental findings (IFs). While clinicians are familiar with unexpected results in the context of whole genome array or whole exome sequencing, IFs may also be detected on targeted multi-gene panels that have array CGH testing as a component of the test or as a method of confirmation. Microarray designs have probe placement in targeted genes of interest and, in addition, in surrounding genes and in intergenic stretches of DNA to create a backbone for the array. While bioinformatic techniques are used to mask areas outside of targeted regions, viewing the whole of the data is a commonly employed quality control measure that may incidentally reveal constitutional or acquired large copy number variants (CNVs) involving genes not included on the testing panel. Our laboratory has developed strategies for evaluating and reporting unexpected CNV findings. Incidental findings are deemed reportable if they contain a targeted gene, or if they lack a targeted gene but are pathogenic, are medically actionable, and are not already known. In this retrospective study we reviewed reported incidental findings from exon array cases that were unrelated to the indication for testing. Of the most recent 22,489 cases, there were only 26 (0.1%) with unrelated reportable IFs. Of these, 16 (62%) contained a targeted gene and 10 (38%) did not. Five (19%) of the patients had an acquired abnormality likely related to a hematologic malignancy, 21 (81%) had a constitutional aneuploidy, and 1 had a large genomic duplication. The patients with an identified acquired abnormality were either being tested for a cancer-related indication (and may have been undergoing chemotherapy) or were elderly. Of the reported constitutional aneuploidies, 18 (90%) were sex chromosome aneuploidies and 2 (10%) were trisomy 21. Additionally, the referring providers were previously aware of at least 3 (14%) of the observed constitutional aberrations. These data show that detection of IFs by exon array testing at our laboratory is rare and typically involves either a relatively benign sex chromosome aneuploidy or a finding already known to the provider. We propose a standard and consistent policy for reporting incidental findings to be considered by all genetic testing laboratories.

2579T

Diagnostic digital PCR copy number assay for NKX2-1 related disorders. *K.M. Robbins[1], KMB. Vinette[1], SM. Kirwin[1], KW. Gripp[1,2], VL. Funanage[1]. 1) Molecular Diagnostics Laboratory, Nemours/A. I. duPont Hospital for Children, Wilmington, DE; 2) Division of Medical Genetics, Nemours/A. I. duPont Hospital for Children, Wilmington, DE.*

The *NKX2-1* gene encodes a transcription factor that is critical for embryonic development of the brain, thyroid, and lung. Haploinsufficiency of *NKX2-1* is responsible for NKX2-1 related disorders including benign hereditary chorea (BHC; OMIM #118700) and choreoathetosis, congenital hypothyroidism, with or without pulmonary dysfunction (CAHTP; OMIM 610978). These rare autosomal dominant disorders occur due to either mutations or copy number variations in *NKX2-1*. *NKX2-1* related symptoms typically appear before age five years and characteristics may include early onset chorea, delayed motor development, and dysarthria; however, the disease is not associated with cognitive deterioration. This disorder typically improves or does not progress after the second decade of life. Choreoathetosis, congenital hypothyroidism, with or without pulmonary dysfunction, also known as brain- thyroid- lung-syndrome, has a more severe phenotype. The characteristics of this disorder include neonatal respiratory distress, involuntary movements, congenital hypothyroidism with elevated levels of thyroid stimulating hormone, developmental delays, hypotonia, and ataxia. Our molecular diagnostic laboratory previously developed a multiplex PCR with fluorescently labeled primers to detect copy number/gene dosage to determine if one or more exons of *NKX2-1* are deleted. The goal of this work was to determine if a copy number/dosage assay utilizing droplet digital PCR was comparable to the current assay. Digital polymerase chain reaction (dPCR) enables determination of absolute copy number analysis without using known reference controls and positive controls. Additionally, this assay will improve our turnaround. We used the QuantStudio 3D digital PCR system to develop a Taqman copy number assay for each of the three exons of *NKX2-1*. Optimization of the *NKX2-1* copy number assays was necessary throughout assay development. Through troubleshooting of these assays, we determined that predigestion of the genomic DNA was required in order to improve signal separation and amplification. After optimization, each *NKX2-1* dPCR assay was validated with several calibrator controls and deletion positive controls. We will use the dPCR *NKX2-1* copy number/dosage assay as proof of principle; in the future, we plan to develop additional dPCR copy number/dosage assays for diagnostic testing of other rare pediatric disorders.

2580F

An open-source quality control monitoring system for clinical NGS.
N.R. Tawari, J.J.W. Seow, S. Ang, P. Dharuman, P. Ng. Computational and Systems Biology, Genome Institute of Singapore (GIS), Singapore, Singapore, Singapore.

In the clinic, regular quality control (QC) meetings are conducted where laboratory tests' performances are reviewed and compared with historical data. We present an open-source, interactive, record-keeping QC system for clinical implementation of NGS. ChronQC tracks and displays NGS QC data such as total reads, depth of coverage, % duplicates, and Ti/Tv ratio. Chron-QC captures QC data from raw NGS files (e.g. bam and VCF files) and stores the metrics in a database. ChronQC then automatically generates interactive time-series plots for various metrics, allowing comparison of the current run to historical runs. In QC meetings, users can record their notes and corrective actions directly onto the graphs for long-term record-keeping. Outliers are flagged based on 1) hospital-defined thresholds derived from clinical validation experiments and 2) Westgard rules (e.g. ± 2 standard deviations). ChronQC can identify bias, excessive scatter, shift, or trends in the clinical results, so that corrective and preventive actions can be taken to ensure patient test results remain clinically valid. The source code is available at https://github.com/nilesh-tawari/ChronQC.

2581W

Application of comprehensive actionable medical panel and whole exome sequencing in critical care of pediatric patients. *H. Wang[1], B. Wu[1,2], Y. Lu[1], H. Sun[3], P. Fang[3], X. Fan[4], W. Zhou[1,2].* 1) Molecular Diagnosis Laboratory of Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defects, The Translational Medicine Center of Children Development and Disease of Fudan University, Shanghai 201102, China; 2) Division of Neonatology, Children's Hospital of Fudan University, Key Laboratory of Neonatal Diseases, Ministry of Health, Shanghai 201102,China; 3) Wuxi NEXTCODE, 288 Fute Zhong Road, Waigaoqiao Free Trade Zone Shanghai 200131, China; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX, USA.

Background Comprehensive evaluation of critically ill patients to identify or rule out genetic causes is a key in precision medicine. NGS technology based whole exome sequencing (WES) is desirable in this aspect, which cost is considerably high for low-income families. We developed a sequencing test as the comprehensive actionable medical panel (CAMP), including ~3,000 genes thoroughly evaluated by medical expertise. CAMP is aim for low-cost and rapid turnaround time (TAT). **Methods** We continuously evaluated about 4,800 pediatric patients referred to the Molecular Diagnosis Laboratory of Children's Hospital of Fudan University. Both CAMP and WES options are offered to the families. 2/3 of them have chosen CAMP test. Data analysis and clinical genetic reports were performed in the Molecular Diagnosis Laboratory. For WES, average target sequencing depth of 100X is required; and mean depth of 200X is achieved for CAMP samples. The TATs are 3-4 weeks and 6-12 weeks for CAMP and WES, respectively. **Results**. There are 586 of WES cases and 635of CAMP cases received diagnosis. In WES, the largest component of patients is of neonate in NICU and neuromuscular disorders. There are 47.68% patients are neonate. In the diagnosed patients, 51.30% (298/586) of them are neonate onset patients. Copy number variations analyzed using deep sequencing is also accounted. 79 patients are diagnosed as chromosome deletion/duplication. In the 233 autosomal dominant gene mutations, 89 (38.20%) of them are confirmed as de novo. For CAMP cases, 50.87% (323/635) of the diagnosed patients are neonate onset. There are 220 autosomal dominant gene mutations, 215 autosomal recessive, and 121 are diagnosed as CNV. For cases received positive findings, the parents did further consultation of the possible treatment improvement, the future of the patients, or consulted family planning on prenatal and preimplantation options. For negative results, families are offered research opportunities including enrollment in Undiagnosed Disease Program center (UDP) of our Hospital. **Conclusion** This is the largest cohort utility of genetic diagnosis in Chinese pediatric clinics. We present an all-in-one pipeline from clinic patient samples to laboratory genetic data analysis and molecular diagnosis in a pediatric specialize hospital. The precise genetic diagnosis can change clinical treatment, change the life quality of patients. Family planning can greatly benefit from genetic diagnosis results.

2582T

Structural variant detection with optical mapping and microfluidic partitioning: A t(9;13) case report. *D. Baldridge[1], D. Wegner[1], J. Wambach[1], T. Graves-Lindsay[2], K. Meltz Steinberg[2], R. Fulton[2], F.S. Cole[1].* 1) Department of Pediatrics, Washington University in St. Louis, St. Louis, MO; 2) The McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO.

Detection and characterization of structural variants (SVs) including copy number variants, translocations, inversions, and complex re-arrangements are important for increasing diagnostic yield, for understanding genotype-phenotype correlation, and for complementing whole exome and whole genome sequencing. As a proof-of-principle approach to evaluate recently developed technologies in a single sample with a known, complex chromosomal translocation (46,XX,der(9)t(9;13)(p23;q13)), we performed 1) next-generation optical mapping in collaboration with Bionano Genomics, and 2) microfluidic-based partitioning in collaboration with 10X Genomics. The subject's phenotypic features include global developmental delay, hearing loss, autism spectrum disorder, dysmorphic facial features, polydactyly, vertebral anomalies, scoliosis, coarctation of the aorta, bicuspid aortic valve, and superior mesenteric artery syndrome. The SV detection methods that we used successfully identified the deletion and translocation previously identified by karyotype. Characterization of specific genomic locations of the chromosome 9 breakpoint and the translocated portion of chromosome 13 by optical mapping or microfluidic-based partitioning are reported, with resolution capability of 2-3kb and single base pair, respectively. This single case suggests that detection and characterization of SVs will benefit from new sequencing and mapping technologies.

2583F

Confirmation of copy number variations from massively parallel sequencing using a chromosomal microarray with single exon level detection. *J.M. Devaney, B.W. Meltzer, M. Tun, S.L. Everhart, M.B. Seprish, K.P. Cusmano-Ozog.* Laboratory Medicine, Children's National Health System, Washington, DC.

Statement of Purpose: The adoption of copy number variation (CNV) analysis in the clinical diagnostic laboratory will have an effect on the rate of positive analysis for the discovery of pathogenic variants. This will have a major consequence on the field of medical genetics. It will help refine genotype-phenotype relationships in known disorders and will lead to the discovery of new syndromes. We present a process for CNV analysis with MPS as a discovery method coupled with the confirmatory use of a microarray with capabilities to discover single exon deletion/duplication events. Methods: We examined 42 samples with MPS and a microarray capable of analyzing single exon CNV events. This group included 14 Coriell Institute for Medical Research cell lines with described CNVs and 32 samples with CNVs discovered using MPS. For the MPS, genomic DNA was prepared for targeted sequencing using the TruSight One panel. The samples were analyzed on an Illumina NextSeq 500. We performed alignment to the hg19 genome with BWA (v 0.7.7) and GATK (v1.6) to generate BAM files. The BAM files were analyzed with the NextGENe software (v2.4.0.1; SoftGenetics) for exon level CNV detection. For the exon level array, we used a whole genome microarray (Applied Biosystems, Thermo Fisher) with probes targeting single nucleotide variants and CNVs for single exon resolution focusing data analysis in 5,376 genes. Results: Thirty-nine samples were evaluated with 50 CNVs and three negative samples without deletions/duplications. In this sample set, there were 29 deletions from single exons to 25.9 MB and duplications from two exons to entire chromosome duplications. In addition, there were three samples with negative deletion/duplication results for single genes. A partial deletion spanning two exons in the *PLP1* gene was not seen on MPS but was present on array CNV analysis. The sensitivity, positive predicative value, and the negative predictive value were all 100%. Conclusions: Massively parallel sequencing is a powerful analysis tool for detecting small genomic losses and gains that are relevant to human genetic disease. The MPS technique for CNV discovery can be fraught with false positives and the use of second method for confirmation of those results will add to its clinical viability. The combination of MPS and arrays will add to the detection of CNVs in clinical samples and present a reliable method allowing for reporting of these results with confidence.

2584W

New systematic rubric for clinical interpretation of copy number variants (CNVs) improves interpretation consistency across laboratories. *D. Pineda Alvarez[1], E. Riggs[2], E. Andersen[3], A. Cherry[4], S. Kantarci[11], H. Kearney[5], A. Patel[6], G. Raca[8], D. Ritter[7], S. South[9], E. Thorland[5], S. Aradhya[10], C. Martin[2] On behalf of the ACMG/ClinGen Copy Number Variant Interpretation Guidelines Working Group.* 1) Courtagen Life Sciences, Woburn, MA; 2) Geisinger Health System, Lewisburg, PA; 3) ARUP Laboratories, Salt Lake City, UT; 4) Stanford University, Palo Alto, CA; 5) Mayo Clinic, Rochester, MN; 6) Columbia University Medical Center, New York, NY; 7) Baylor College of Medicine, Houston, TX; 8) Children's Hospital of Los Angeles, Los Angeles, CA; 9) Ancestry.com, San Francisco, CA; 10) Invitae, San Francisco, CA; 11) Quest Diagnostics, San Juan Capistrano, CA.

Analysis of CNVs by chromosomal microarray analysis (CMA) is the first-tier genetic test in patients with neurodevelopmental disorders and/or multiple congenital anomalies. In addition, due to advancements in microarray and sequencing technologies, CNVs are now being analyzed at higher resolutions extending down to single-exon CNVs. Despite the existence of CNV interpretation standards from ACMG and mounting experience from laboratories analyzing CNVs, inconsistencies in clinical interpretation persist, a large portion of which can be attributed to differences selecting and weighing evidence used to classify the variants. The current ACMG CNV interpretation guidelines provide a high-level review of evidence that may be considered to determine pathogenicity of CNVs in a diagnostic setting. However, to improve consistency of CNV interpretation across laboratories, the ACMG and the Clinical Genome Resource (ClinGen) have established a collaboration to update these guidelines with a more standardized clinical classification framework. We have developed a framework to interpret deletions, duplications, CNVs involving non-coding regulatory loci and intragenic CNVs, which uses a point-based, hierarchical scoring system to determine the pathogenicity of CNVs by assigning and weighing evidence from several categories. These categories include overlap with CNVs reported in affected or unaffected individuals, presence of dosage-sensitive genes, case-control studies, segregation data, and number of protein coding genes. The metric was initially tested using a set of 33 deletions with defined clinical significance from previous clinical laboratory testing. Each deletion was evaluated independently by two geneticists using the new schematic, and 14 of these were also analyzed using the existing guidelines. Concordance between evaluators and both rubrics were compared. The new rubric determined the correct interpretation in more than 80% of the CNVs, and improved the concordance of clinical interpretation between evaluators from 40% to 80%. A similar analysis is now in progress for duplications. The framework will be further tested by a broader group of clinical laboratory geneticists to identify nuances and refine its guidance. Given that the updated guidelines are being developed to be applied to any method capable of defining discrete CNV events, this rubric is expected to have broad impact by providing a robust system to support the consistent interpretation of CNVs.

2585T

No consistent phenotype in patients with copy number variants of the SHOX downstream regulatory domain. *A. Wray, J. Meck, L. Matyakhina.* GeneDx, Gaithersburg, MD.

Copy number variants (CNVs) within the SHOX downstream regulator domain (DRD) have been reported in cohorts of individuals with Leri-Weill dyschondrosteosis (LWD), a skeletal dysplasia characterized by short stature and/or Madelung deformity, and isolated idiopathic short stature (ISS). Concern for ascertainment bias in studies performed on individuals with a clinical diagnosis of LWD or ISS led us to retrospectively review from our diagnostic laboratory 18,646 postnatal chromosomal microarray (CMA) cases for deletions (dels) or duplications (dups) involving the SHOX DRD. In total, 107 patients had CNVs within the SHOX DRD (0.57%): 85 with a recurrent 47.5 kb del [no other deletions within this region were detected] (22 F, 63 M) and 22 had duplications (11 F, 11 M). Nineteen patients (17 with del, 2 with dup) were excluded from analysis either due to presence of a known pathogenic CNV fully explaining their phenotype (n=14) or lack of clinical information (n=5). Twelve of the remaining 88 cases (9%) had an additional CNV of uncertain significance (6 dels, 2 dups, 1 female patient had LWD) or large regions of homozygosity (3 dels, 1 dup). There was no consistent phenotype in individuals with CNVs in the SHOX DRD. Reported test indications primarily included the standard neurodevelopmental phenotypes which generally trigger CMA testing: autism, learning disability/intellectual disability/developmental delay, dysmorphic features, congenital heart defect, and seizures. A minority of cases (n=8; 9%) reportedly had a spectrum of phenotypic features including short stature (n=6, 3 F and 3 M) or possible LWD (n=2, both female), but not ISS. For 20/88 probands parental testing was performed, and showed that all CNVs involving SHOX DRD were inherited. Two mothers (10%) had a phenotype suggestive of LWD: 1 with short stature and bowing of the legs, 1 with possible abnormality of the elbow without reduced movement. Our findings suggest that CNVs within the SHOX DRD are relatively common, are typically inherited, may be seen in association with other CNVs, and either are not associated with a consistent phenotype or show greatly reduced penetrance. Further large cohort studies would be necessary to determine if CNVs in SHOX DRD might act as a modifying factor for developing LWD or ISS.

2586F

Clinical variant reclassification and scaling support for the return of updated genetic results. *H.L. Rehm[1,2,3], S. Harrison[1], S. Baxter[1,3,4], M. Oates[1,5], L. Babb[1,4], M. Varugheese[4], E. Clark[4], S.J. Aronson[1,5], M. Lebo[1,2].* 1) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA; 2) Department of Pathology, Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 3) The Broad Institute of MIT and Harvard, Cambridge, MA; 4) GeneInsight, Sunquest Information Systems, Boston, MA; 5) Research Information Sciences and Computing, Partners HealthCare, Cambridge, MA.

   Variant interpretations are constantly changing as 1) new evidence is identified, 2) variants are reassessed with new guidelines, 3) labs resolve interlaboratory differences in interpretation enabled by data sharing and 4) variants are interpreted by experts. This creates challenges for labs, physicians and patients to manage updates and adjust care accordingly. Given that irreversible decisions can be initiated in symptomatic patients and at risk individuals, it is critical that knowledge updates be communicated as rapidly as possible. Yet patients change health care providers frequently and laboratories lack reimbursement to support report updates, both contributing to challenges in supporting this critical aspect of genomic medicine. We analyzed our data from the Partners Laboratory for Molecular Medicine over twelve years to understand variant reclassification rates. In that time period, 2,519 variant reclassifications occurred across 2,192 variants out of 18,766 total variants (12%). Of these reclassifications, 1,390 (55%) were changes in confidence (e.g. Likely Pathogenic (LP) vs Pathogenic (P) or Likely Benign (LB) vs Benign (B)) unlikely to impact clinical care; however, 1,129 (45%) were more substantive changes, more likely have impact. Interestingly, while 83% of variants moved to a category of more certainty (VUS to LP, LP to P, VUS to LB, LB to B), 17% moved to a less certain (16%) or opposing (between P/LP and B/LB; 1%) category. These later changes are the most critical for communication as they are unanticipated, and physicians and patients are less likely to check in with the lab for an update on non-VUS variants. This creates enormous challenges for labs wishing to provide updates, yet have difficulties re-contacting the original ordering provider and lack the resources required for amendments in the absence of reimbursement. To address the challenge of variant reclassification, we support three main mechanisms to deliver updates back to patients: 1) amended reports (upon request or for the most impactful and unanticipated changes), 2) automated updates delivered through a web-based portal (GeneInsight) with electronic health record integration where available, and 3) publically sharing current variant interpretations through ClinVar. This talk will present the details and the benefits and drawbacks of each approach, our experience supporting each method and considerations of legal liabilities and professional best practices.

2587W

Linked-read sequencing for molecular cytogenetics. *S. Garcia, S. Williams, J. Herschleb, P. Marks, A.W. Xu, M. Schnall-Levin, D. Stafford, D.M. Church.* 10x Genomics, Pleasanton, CA., USA.

   Structural variants (SVs), including deletions, duplications, insertions, inversions, and translocations, are important contributors to disease. However, they remain one of the most difficult types of variation to accurately ascertain, in part because they tend to cluster in duplicated and repetitive regions. Clinically, aCGH and karyotype remain the most commonly used assays for genome-wide SV detection, though there is clear potential benefit to an NGS-based assay that accurately detects both SVs and SNVs. Recently, laboratories have demonstrated that low coverage long-insert WGS (liWGS) can identify both CNVs and balanced rearrangements at a cost and speed similar to aCGH/ karyotype. This introduces the possibility of a hybrid NGS assay (liWGS+genome/exome) but liWGS protocols are difficult and time-consuming to establish. Linked-Read sequencing is a relatively simple, fast, and cost-effective alternative that is applicable to both genome and targeted assays. Linked-Reads are generated by performing haplotype-level dilution of long input DNA molecules into >1 million barcoded partitions, generating barcoded short reads within those partitions, and then performing short read sequencing in bulk. We performed 30x Linked-Read genome sequencing on 24 GetRm CNVPanel samples with known balanced or unbalanced SVs. Downsampling was performed on a subset to determine the lowest sequencing depth required to detect variation. 22/24 variants were identified, and another known event was called as a candidate. The remaining sample had a balanced structural variant involving 16q11.2, a region known to be difficult to assess due to gaps in the reference assembly. With current algorithms, CNVs can be called with as little as 1-2x sequencing depth (5-10Gb) while balanced events require on the order of 10x coverage. We demonstrate that there is specific signal in the data for balanced events down to ~2x coverage. Similar to recent observations, we find complexity at breakpoints not previously detected by conventional technologies.

2588T

High throughput linked-read sequencing for improved variant detection. *A. Fehr, S. Garcia, M. Prout, C. O'Keeffe, A. Price, A.W. Xu, S. Williams, C. Catalanotti, P. Marks, D.M. Church.* 10x Genomics, Pleasanton, CA., USA.

Causal single nucleotide variants (SNVs) and small indels tend to impact genes across the genome, including in regions with known paralogs that are difficult to assess with short read (SR) next generation sequencing (NGS) techniques. To address variant detection shortcomings, we developed a technology that retains long-range information while maintaining the power, accuracy, and scalability of short read sequencing. The 10x Genomics Chromium™ Genome solution utilizes haplotype-level dilution of high molecular weight DNA molecules into >1 million barcoded partitions to create a novel data type referred to as 'Linked-Reads'. These Linked-Reads (LR) enable high-resolution genome analysis with minimal DNA input (~1 ng). Here, we describe a high throughput workflow that interfaces an Agilent Bravo Liquid Handling platform with the Chromium system to automate the LR library preparation solution. Using this reliable and reproducible workflow, we generated both genome-level Linked-Read NGS data in addition to using Agilent SureSelect Human V6 exome baits to capture a subset enriched for exonic targets, on multiple reference samples including NA12878. Of particular interest are the exome data wherein for ~60x median coverage on NA12878 samples, LR sequencing yielded 73.4% of bases with >10x coverage at mapq>30 compared to 71.4% for SR. For clinically-relevant genes in loci with known paralogs, this metric ranged between 73.1 and 84.4% for LR but only 53.6-75.4% for SR. The fraction of exons with at least 99% of bases with >10x coverage at mapq>30 in difficult loci ranged from 71.3-77.6% for LR and 48.6-64.8% for SR. With this additional coverage, LR detected 1,597 novel SNPs in degenerate regions with 78.6% in the coding sequence of 414 total genes, versus only 527 SNPs for SR. Improvements were also seen for whole genome LR data including 74,667 novel SNPs in degenerate regions versus only 32,295 for SR. Single exon deletions and duplications are not routinely assessed in exome sequencing because coverage is highly variable but haplotype reconstruction in LR data improves the signal:noise ratio for duplications and reveals deletions as a complete loss of coverage on one haplotype. For NA12878 50 exon-level deletions are called with LR with a heterozygous sensitivity of 72.7% (PPV= 72%) and a homozygous sensitivity of 83.3% (PPV= 100%). Duplications are robustly detected when phasing is present over the affected region.

2589F

A clinical molecular genetics laboratory experience with whole exome sequencing. *J. Machado, D. Allingham-Hawkins, B. Dorshorst, G. Londre, E. McPherson, R. Steiner, G. Sun, J. Weber.* PreventionGenetics, Marshfield, WI.

Whole exome sequencing (WES) is a test that is intended for health care providers to look for a genetic diagnosis when the clinical phenotype is unclear and/or previous test results have been uninformative. We performed WES on patients and family members' (where available) DNA in a clinical diagnostic laboratory utilizing Agilent Clinical Research Exome v1 hybridization probes on Illumina's NextSeq 500 using 150 by 100 bp paired end reads. Quality metrics included >97% of target bases at >20x, and mean coverage of target bases at >120x. Sequencing data was analyzed using our internal Titanium-Exome software, where output data was converted to FASTQ using Illumina Bcl2Fastq 1.8.4, and mapped by BWA. Variant calls were made by the GATK Haplotype caller and annotated using in-house software and SnpEff. Variants were filtered and annotated using VarSeq (www.goldenhelix.com). Variant interpretations and reports were made by PreventionGenetics' PhD molecular geneticists. The top clinical indications for whole exome sequencing included 45% for neurological disorders, 31% for multiple congenital abnormalities, and 14% for neuromuscular diseases. We observed positive molecular diagnoses in 32% of cases, 52% of cases were indeterminate, and 16% of cases were negative. We observed a similar positive rate for singleton and trio (i.e. proband and parents) exomes of 35% vs. 31%, respectively. About 50% of positive cases were due to an autosomal dominant disorder, and from these 56% of variants were classified as likely pathogenic due to *de novo* variant status, emphasizing the importance of trio testing. In fifty-nine genes recommended by the American College of Medical Genetics and Genomics (ACMG) that have potential medical value that are separate from the primary clinical indication (e.g. secondary findings), we found approximately 1.7% of exome cases had a pathogenic/likely pathogenic variant in one of these genes. We also gave the option of selecting additional Mendelian disorders genes, which may result in disease, and found that 35% of individuals were positive for a potentially pathogenic disease. We also recently gave the option of obtaining carrier status for pathogenic/likely pathogenic variants for recessive diseases. Nearly all individuals (94%) carried at least one positive finding. In conclusion, whole exome sequencing is a powerful tool for determining a genetic diagnosis in a clinical laboratory setting.

2590W

Diagnostic exome sequencing identifies a homozygous whole-gene deletion of *DPY19L2* that was not detected by a high-density single nucleotide polymorphism (SNP) array. *S. Sajan[1], S. Saliganan[1], KL. Helbig[1], B. Barrows[2], R. Martinez[2], Z. Powis[1], R. Burns[3], L. Rohena[3], JA. Massie[4], E. Spiteri[1], WA. Alcaraz[1].* 1) Department of Clinical Genomics, Ambry Genetics, Aliso Viejo, CA; 2) Clinical Laboratory Department, Ambry Genetics, Aliso Viejo, CA; 3) Departments of Pediatrics and Obstetrics and Gynecology, San Antonio Military Medical Center, San Antonio, TX; 4) Department of Obstetrics and Gynecology, Madigan Army Medical Center, Tacoma, WA.

While both exome sequencing and high-density single nucleotide polymorphism (SNP) arrays are capable of detecting copy number variants (CNVs), neither methodology is comprehensive and each can miss certain regions of the genome. In a clinical setting, SNP array testing usually precedes exome sequencing primarily due to its lower cost and ability to reliably detect large CNVs that are at least 100 kilobases (kb) in length. It is therefore important for SNP arrays to be able to detect, at the very least, known clinically relevant CNVs that meet the length cutoff so as to minimize additional unnecessary investigations towards a genetic diagnosis. We present a case example of an adult male with globozoospermia, a spermatogenic defect that can cause male infertility, who pursued diagnostic exome sequencing at our clinical genetic testing laboratory after having a negative CytoscanHD Array (Affymetrix) result. His exome data showed complete or almost-complete absence of coverage in all coding exons of *DPY19L2*, a gene known to cause globozoospermia when homozygously deleted. Homozygous deletions of this gene, which are approximately 200kb, are a common cause of this condition due to nearly-identical low copy repeats on either side of this gene that facilitate non-allelic homologous recombination. We confirmed the deletion in the patient by polymerase chain reaction-based amplification of three of the 22 coding exons (3, 11, and 22) of this gene, which showed no amplification. In contrast, two of six coding exons of the nearest gene *TMEM5* located 111kb away amplified successfully. Due to the presence of four pseudogenes in the genome, probe coverage of *DPY19L2* on the CytoscanHD array is poor, with only three intragenic SNPs in the entire 110kb length of the gene. However, homozygous deletions of *DPY19L2* have previously been detected by array comparative genomic hybridization (aCGH). This case illustrates the utility of exome sequence data in revealing CNVs previously missed by a high-density SNP array and suggests that a clinical high-density aCGH may be a favorable alternative for cases where discovering pathogenic CNVs is the primary intent before pursuing other genetic diagnostic tests.

2591T

Towards automated variant pathogenicity assessment: A Bayesian classification framework. *L.G. Biesecker[1], M.S. Greenblatt[2], S.M. Harrison[3], R.L. Nussbaum[4], S. Prabhu[5], K.M. Boucher[6], S.V. Tavtigian[7], ClinGen Sequence Variant Interpretation Working Group.* 1) NHGRI/NIH, Bethesda, MD; 2) University of Vermont, Burlington, VT; 3) Partners Healthcare Laboratory, Boston, MA; 4) Invitae Corp, San Francisco, CA; 5) Dept of Genetics, Stanford University, Palo Alto, CA; 6) Dept of Medicine, Univ Utah, Salt Lake City; 7) Dept of Oncologic Sciences, Huntsman Cancer Inst, Salt Lake City, UT.

In 2015, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) published recommendations for interpretation of Mendelian disease gene variants (PMID: 25741868). The committee recognized the guideline as a starting framework that would evolve. One goal of the Clinical Genomics Resource (ClinGen) Sequence Variant Interpretation (SVI) Group is to refine and evolve those guidelines. We have initiated this by asking whether the criteria could be reframed as a Bayesian heuristic using conditional probabilities scaled to the categorical descriptors. The recommendations were evaluated to determine whether the rules are internally consistent and whether the qualitative, categorical "rules for combining criteria" to evaluate variants were quantitatively consistent. To do this, the criteria for and against pathogenicity were translated into a naïve Bayesian classifier, assuming four levels of evidence and exponentially scaled odds of pathogenicity. Results of this analysis show that the ACMG/AMP were in general valid and compatible with a quantitative framework and that 16 of 18 of ACMG/AMP pathogenicity combining criteria were internally consistent. However, two out of the 18 combining criteria were inconsistent (Likely Pathogenic combination i and Pathogenic combination ii). We will recommend changes for these two combining rules. The Bayesian classifier was extended to model additional combinations that currently lead to a VUS pathogenicity level (by including evidence for and against pathogenicity), for which we show that certain combinations can lead to posterior probabilities in the range of likely pathogenic or even pathogenic, and can be readily calculated. Finally, this formal Bayesian framework will facilitate the transition from categorical evidence (e.g., allele frequency above or below a set threshold) to continuous variable evidence types, which extract maximal information from quantitative datasets (e.g., ExAC or functional data). By transforming the ACMG/AMP recommendations into a quantitative, Bayesian framework, we provide a mathematical foundation to what was a qualitative, categorical heuristic and validates the fundamental approach adopted by the ACMG/AMP. More importantly, it advances the science of variant interpretation by providing a mathematical foundation to further refine evidence and combining rules, and supports efforts to automate components of variant pathogenicity assessments.

2592F

Protocols to keep NGS gene panels and annotation content current. *N.A. Rouse, M.J. Snyder, T.J. McMahan, W.A. Langely, P.L. Nagy.* MNG Laboratories, 5424 Glenridge Drive NE, Atlanta, GA 30342 USA.

With decreasing sequencing costs and growing numbers of out of the box bioinformatics solutions, the landscape of clinically relevant genes and variants is changing at an unprecedented rate. Staying at the forefront of clinical relevancy requires frequently scheduled curation and updating of both gene lists and annotation databases. A two-fold approach is taken to keep data interpretation up to date. Gene list classifications are updated quarterly and annotation databases are updated monthly, coinciding with ClinVar's monthly release. Gene classification is determined by the combined use of OMIM and ClinVar and review of the scientific literature. Newly pathogenic or reclassified genes are individually examined to confirm correct placement for each panel based on multiple lines of independent scientific evidence. This gene curation results in a quarterly custom capture redesign and NGS panel updates. To ensure that annotations are precisely defined and up to date, our bioinformatics solution processes variant and gene annotation from external databases monthly through a normalization pipeline. Frequently updating custom capture reagent in conjunction with updating annotation content on a monthly basis increases diagnostic sensitivity in NGS panels.

2593W

Elective whole genome testing in clinical practice. *D. Bick[1,2], K. East[1,2], W. Kelley[1,2], C. Aiken[2], A. Crouse[1], H. Jacob[1], A. Hott[1], N. Lamb[1], M. Schroeder[1], C. Birch[1], D. Brown[1], L. Handley[1], S. Newberry[1], E. Worthey[1], M. Cochran[1,2].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) Smith Family Clinic for Genomic Medicine, Huntsville, AL.

Genomic testing (whole genome and whole exome sequencing) has entered routine clinical practice worldwide for the diagnosis of genetic disorders. Recently, individuals who do not have a standard indication for sequencing have begun to request genomic testing. These tests are referred to as elective genomic tests. As the availability of testing has spread and the cost has diminished, more individuals are choosing to have elective testing. Elective genomic testing generally includes one or more of following categories of results for adults interested in their personal or family risk: (1) Variants that are associated with diseases already present in the individual. (2) Secondary or incidental variants that result in a Mendelian disorder where the individual has not manifested the disorder. (3) Carrier status for recessive disorders. (4) Pharmacogenomic variants. (5) Multi-SNP genetic risk profiles for actionable common adult-onset diseases. In July 2016, we initiated Insight Genome, an elective whole genome sequencing and pharmacogenetic test for adults. Insight Genome patients provide medical history, medical records through their physician, have a physical examination, pre-test genetic counseling and post-test results discussion. Clinical whole genome sequencing and pharmacogenetic results are returned to the patient and referring provider. To date 30 individuals have participated in the program. Among the first 24 to complete testing and reporting, 7 (29%) were found to have a variant associated with their phenotype (a primary finding). One was found to have a variant not associated with their phenotype (secondary finding). 17 (71%) were found to be carriers for 1 to 5 autosomal recessive disorders. 12 (50%) were found to have variant(s) where guidelines recommend selection of an alternative drug affecting 3 or more medications. Patients received findings that may explain part of their medical history. Few of these individuals had a standard indication for diagnostic testing. Many patients were carriers for disorders not covered by available pan-ethnic carrier screening and were found to harbor important pharmacogenomics variants. Both patients and their providers have been enthusiastic about and receptive to the results and have indicated that they plan to use this information for future care. Elective genomic testing provides important medical information that individual and their physicians can use today. .

2594T

Clinically significant small indels detected by whole genome sequencing: A proof of concept case series. *C.M. Brown[1], A. Gross[1], K. Bluske[1], N.J. Burns[1], A. Chawla[1], A.J. Coffey[1], R.T. Hagelstrom[1], M.C. Jones[2], A. Malhotra[1], D. Masser-Frye[2], D.L. Perry[1], A. Scocchia[1], E. Thorpe[1], J.L. Williams[3], M.S. Williams[3], R.J. Taft[1].* 1) Illumina Inc., 5200 Illumina Way, San Diego, CA 92122; 2) Rady's Children's Hospital San Diego, Division of Dysmorphology/Genetics, 7920 Frost Street, Suite 200, San Diego, CA, 92123; 3) Geisinger Health System, Genomic Medicine Institute, Weis Center for Research, 100 North Academy Ave., Danville PA 17822.

Insertions and deletions (indels) between ~10 bp and 10 kb represent a diagnostic challenge and often are not captured by current genetic testing platforms including chromosomal microarray (CMA), whole exome sequencing (WES) and many next generation sequencing panels. Small indels can be detected by targeted techniques including multiplex ligation probe amplification or custom CMA, but these methods require prior identification of candidate genes or loci. Consequently, routine clinical testing for small indels is restricted to single gene or panel analyses. Although the prevalence of disorders caused by small indels is unknown, it is likely that individuals with genetic disorders caused by these variants go undiagnosed, particularly when non-specific features or atypical presentations make the differential diagnosis challenging. Whole genome sequencing (WGS) is a diagnostic platform which has the capability of simultaneously detecting small nucleotide variants, like WES, and indels greater than 10 kb, like CMA, across the genome. There is also emerging evidence that WGS has the capability to detect small indels on a genome-wide scale. Three cases with clinically significant indels less than 10 kb detected by WGS (TruGenome Undiagnosed Disease Test, Illumina Inc.) are presented; a heterozygous 8.4 kb loss at Xq13 (NM_018486.2:c.1112-7725_*615del) causing a partial deletion of *HDAC8* in a child with X-linked Cornelia de Lange syndrome, a 1.7 kb deletion in *CHRNB1* (NM_000747.2: c.821-990_1045-311del) in trans with a missense variant, p.Pro271Leu (NM_000747.2:c.812C>T), in a child with congenital myasthenic syndrome and a 46 bp deletion in *ECEL1* (NM_004826.2:c.110_155del) in trans with a nonsense variant, p.Gln265Ter (NM_004826.2: c.793C>T) in a child with distal arthrogryposis type 5D. Orthogonal testing performed to validate these findings is also presented. This case series illustrates that in addition to capturing variant types typically detected by WES and CMA, WGS has the potential to detect indels less than 10 kb, addressing a known diagnostic gap in clinical genomics.

2595F

Safely washing and reusing pipette tips for Next Generation Sequencing (NGS) qPCR. *A. Graham.* Grenova, LLC, 483 Southlake Blvd., Richmond VA 23236.

Statement of Purpose: Disposable pipette tips provide an easy and carryover free process for liquid transfers. Once discarded, these polypropylene plastics enter the waste stream and are either incinerated as biohazard waste or moved to a landfill. Because of potential carryover, washing pipette tips for reuse is not generally accepted as safe for many scientific assays. Using recently introduced technology, Grenova tip washer devices demonstrate that pipette tips can be safely and effectively washed for NGS qPCR reuse. **Methods**: Validation test methods were performed with 384 Hamilton CO-RE nested tip racks. Carryover studies were performed by soaking control and experimental wash pipette tip racks in a DNA solution. Control group pipette tips were subsequently used in qPCR preparation. Experimental pipette tips were washed then used in qPCR preparation. **Results**: Amplification resulted in a five orders of magnitude difference between control and washed pipette tips. Control tips qPCR resulted in expected Ct values indicating normal amplification. Washed pipette tips qPCR amplification resulted no amplification or Ct values between 25.4 – 30. **Conclusion**: This study indicates that pipette tips washed and dried in a Grenova tip washer performed as expected. Washing and reusing over 15,000 pipette tips daily rather than discarding, resulted in over $500,000 cost savings in addition to waste reduction for the NGS labs.

2596W

A randomized controlled trial of rapid whole genome sequencing for neonatal genetic diagnosis. *S. Kingsmore[1], J.A. Cakici[1], M.M. Clark[1], L.K. Willig[2,3,4], N.M. Sweeney[1,6], E.G. Farrow[2,3,4], C.J. Saunders[2,4,5], I. Thiffault[2,4,5], N.A. Miller[2], L. Zellmer[2], S.M. Herd[2], A.M. Holmes[3], S. Batalov[1], N. Veeraraghavan[1], L.D. Smith[2,4], D.P. Dimmock[1], S. Leeder[3,4], J.E. Petrikin[2,3,4].* 1) Rady Children's Institute for Genomic Medicine, San Diego, CA; 2) Center for Pediatric Genomic Medicine, Children's Mercy - Kansas City, Kansas City, MO; 3) Department of Pediatrics, Children's Mercy - Kansas City, Kansas City, MO; 4) School of Medicine, University of Missouri-Kansas City, Kansas City, MO; 5) Department of Pathology, Children's Mercy - Kansas City, Kansas City, MO; 6) Department of Pediatrics, University of California – San Diego, Rady Children's Hospital, University of California – San Diego, CA.

There is increasing evidence that genome sequencing has diagnostic utility in the acute care of infants with genetic diseases in neonatal and pediatric intensive care units (NICU, PICU). Since genetic diseases are a leading cause of mortality in infants and disease progression can be very rapid, genetic diagnoses must be made quickly to permit consideration of precision medicine interventions in time to decrease morbidity and mortality. Though there is evidence that standard whole genome and whole exome sequencing (WGS, WES) have higher diagnostic rates than other genetic tests in children with likely genetic diseases, diagnoses generally occur too late to guide NICU and PICU management. In a retrospective study, 57% of infants received genetic diagnoses by rapid whole genome sequencing (rWGS) in a median of 23 days (day of life [DOL] 49). We report results from the first randomized controlled trial comparing trio rWGS plus standard tests (cases) with standard tests alone (controls) in a regional NICU and PICU from 2014 - 2016. The primary end-point was rate of genetic diagnosis within 28 days of enrollment.   65 families with infants <4 months old and suspected of a genetic disease completed the study. The study was terminated early due to loss of equipoise: 63% (21) of 33 controls received standard WGS, WES, or next-generation sequencing panels. Nevertheless, an intention-to-treat analysis showed the rate of confirmed genetic diagnosis within 28 days of enrollment to be higher in cases (31%, 10 of 32) than controls (3%, 1 of 33; p=0.003). Among infants enrolled in the first 25 days of life, the rate of diagnosis by DOL 28 was higher in cases (32%, 7 of 22) than controls (0%, 0 of 23; p=0.004). Overall time to diagnosis and age at diagnosis also differed significantly between cases and controls after accounting for non-proportional rates of diagnosis (p=0.040 and p=0.043, respectively). We suggest that rWGS should be considered as a first-tier genetic test in NICU and PICU infants with likely genetic diseases. Since genetic diseases are among the leading cause of death in infants, these results may have broad implications for pediatric precision medicine.

2597T

Challenges and solutions for FFPE DNA quantitation. *K. Plasman, T. Boonefaes, T. Martens, T. Duthoit, T. Montoye.* Trinean, Gentbrugge, Belgium.

As large collections of FFPE samples have been gathered over many years, FFPE tissue derived DNA or RNA are popular subjects in genetic studies and cancer research. The success of downstream tests with DNA isolated from FFPE tissue blocks however cannot be guaranteed as samples can be fragmented, chemically altered and/or heavily contaminated with carry-over molecules from the source material and extraction buffer constituents. Proper quantification and purity analysis after isolation are therefore desirable. In this poster we present experimental data to demonstrate the advantages and shortcomings of the most used FFPE DNA quantification technologies, being A260 absorbance, Qubit fluorescence and DropSense/cDrop. Finally, we propose ways to standardize FFPE DNA QC in order to achieve higher success-rates of downstream assays.

2598F

CLIP-Cap: Combined Long-Insert Paired-End and Capture sequencing, a novel method for the analysis of complex genomic aberrations. *C. Purmann[1], X. Zhu[1], D. Palejev[2], J. Bernstein[1], J.F. Hallmayer[1], A.E. Urban[1].* 1) Stanford University, Palo Alto, CA, United States; 2) Bulgarian Academy of Sciences, Sofia, Bulgaria.

CLIP-Cap is a method that allows to completely resolve complex chromosome rearrangements in the human genome using standard laboratory equipment and an efficient and multiplexable workflow. CLIP-Cap is a <u>c</u>ombination of <u>l</u>ong-<u>i</u>nsert <u>p</u>aired-end and <u>cap</u>ture sequencing. It is based on creating a paired-end sequencing library where the inserts of genomic DNA are of variable length in the kilobase-range (i.e. 2-8 kbp as opposed to 400-800 bp in standard paired-end sequencing), also known as mate-pair or jumping library. From this paired-end library a semi-targeted capture is carried out using capture oligomers representing one of the chromosomes involved in the structural genome aberration at hand. This is then followed by limited (i.e. benchtop or multiplexed) 'next-generation' DNA sequencing and data analysis with standard software tools (for mate pair mapping, split-read analysis and read-depth analysis, respectively). We resolved the three main Philadelphia Chromosome aberrations, in three different cancer cell lines (Kasumi-4, which contains a three-way translocation, SUP-B15 and K562). Furthermore, we resolved the complex rearrangements in four cases with developmental disorders that all involved chromosome 22, but also other chromosomes (9, 10 or 11) or multiple different aberrations (e.g. deletions and dicentric inversions) present in a mosaic manner. Currently there is no single approach to resolve typical complex chromosomal rearrangements that are a common hallmark of cancer and also of developmental disorders. We have tested CLIP-Cap on seven different cases of genomic rearrangements to demonstrate that it is sufficient to resolve a wide range of scenarios, the only requirement at the outset being knowledge of one of the chromosomes involved in the aberration. .

2599W

Frequencies of *BCHE* variants in a large cohort of US individuals. *G. Zhu, H. Fang, F.G. Espin, E.S. Dawson, T.A. Moreno, A.L. Del Tredici.* Millennium Health, LLC, San Diego, CA.

The *BCHE* gene encodes butyrylcholinesterase (also known as plasma cholinesterase or pseudocholinesterase), an enzyme responsible for the hydrolysis and deactivation of the neuromuscular blocking agent succinylcholine. BChE deficiency increases the risk of prolonged neuromuscular blockade and respiratory depression in response to succinylcholine. The common BCHE variants atypical (A, rs1799807) and Kalow (K, rs1803274), and rarer variants fluoride-1 (flu-1, rs28933389), fluoride-2 (flu-2, rs28933390) and silent-1 (sil-1, rs398124632) are associated with various degrees of reduction in the activity of the BChE enzyme. A real-time PCR-based method was developed to test for these variants in a large (n > 13,000), multi-ethnic cohort in the United States. Test subjects received pharmacogenetic testing and were not selected based on planned surgeries or history of any perioperative complication. The allele frequencies of A, K, flu-1, flu-2 and sil-1 were 0.016, 0.199, 0.0008, 0.0047, and 0.0004, respectively, in this cohort. Approximately 0.06% of individuals were predicted to have severe BChE deficiency because they were homozygotes or compound heterozygotes of A, flu-1 and/or flu-2. Approximately 8% of individuals were predicted to have moderate BChE deficiency because they had two copies of K or one copy of A, flu-1, flu-2 or sil-1. Compared to other ethnic groups, Caucasians were predicted to have the highest frequencies of both severe and moderate BChE deficiencies. This is the first screening of *BCHE* genetic variants in a large cohort and, especially, this is the first time frequencies of *BCHE* genetic variants are estimated in African-American, Asian and Hispanic ethnic groups in the US.

2600T

WGS is an imperfect but valuable tool for predicting the risk of genetic disease in children. *M.S. Meyn[1,2,3,4,5], N. Monfared[1,2], C.R. Marshall[6,7], D.J. Stavropoulos[7,8], R. Basran[7], R.H. Hayeems[2,9,10], D. Merico[4,6], M. Szego[6,11,12], R. Zlotnik-Shaul[5,11,15], C. Shuman[1,2,3,4], T. Nalpathamkalam[4,6], G. Pellecchia[4,6], M. Girdea[4,13], M. Brudno[4,13], R.D. Cohn[1,2,3,4,5], S.W. Scherer[2,3,4,6,14], P.N. Ray[2,3,4,6,7], S.C. Bowdin[1,2,5].* 1) Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 4) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 8) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada; 9) Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON, Canada; 10) The Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada; 11) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 12) Centre for Clinical Ethics, St. Joseph's Health Centre, Toronto, ON, Canada; 13) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 14) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 15) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada.

To pilot paediatric genomic medicine we developed the SickKids Genome Clinic, a research project centered on diagnostic whole genome sequencing (WGS) of children undergoing genetic evaluation. With parents' consent, we also search for predictive secondary variants (PSVs) associated with occult or future disease.   For our first 100 patients, analysis of WGS data for candidate PSVs revealed 4 variants associated with the ACMG list of reportable disorders. Expanding the search to 3000+ ClinGen genes identified returnable PSVs in more than 25% of children. Four children had 2 PSVs and one child had 3. The majority were rare variants in genes that cause dominant Mendelian disorders. The rest included variants linked to attenuated Mendelian phenotypes, common disease risk variants, an X-linked variant and a protective variant.   To date, reverse phenotyping of the proband and 1st degree relatives found evidence of the phenotype predicted by the PSV in 44% of cases. On multiple occasions, evidence for the predicted phenotype was hidden "in plain sight," as the patient's medical record contained clinical or laboratory findings consistent with the predicted phenotype that had not been acted upon (*e.g.,* long QT syndrome).   As all predictive secondary variants in our dataset were inherited, we compared the predictive performance of WGS to a) routine family histories obtained by the referring clinical geneticists/genetic counselors, b) comprehensive family histories obtained by the research team *prior* to disclosure of PSVs, and c) family histories obtained by the research team *after* disclosure of PSVs. Importantly, family histories obtained prior to disclosure of PSVs by either the referring clinician or the research team did not predict the phenotypes associated with the PSVs. However, in almost 50% of cases in which the PSV was found to be penetrant, parents revised their family histories in response to being informed of the phenotypes predicted by the PSV and reported previously undisclosed PSV-associated phenotypes in the proband and/or family members.   Our results suggest that, routine family histories substantially underestimate genetic risk, as many parents edit family history and clinicians frequently target their family history inquiries. Predictive analysis of WGS data currently outperforms family history in identifying the risk of occult or future genetic disease in children, but this greater sensitivity is accompanied by a significant false positive rate. .

2601F

Undiagnosed Diseases Network (UDN) successes in precision medicine. *J. Phillips[1], A. Robertson[1], A. Bican[1], J. Cogan[1], L. Rives[1], M. Koziura[1], L. Germano[1], V. Hannig[1], L. Fairbrother[1], J. Pfotenhauer[1], E. Brokamp[1], P. Clark Pollock[1], B. Small[1], J. Sergent[2], J. Duis[1], J. Newman[2], R. Hamid[1].* 1) Depart of Pediatrics, Vanderbilt University Medical Center, Nashville, TN; 2) Depart of Medicine, Vanderbilt University Medical Center, Nashville, TN.

Aims: Evaluate the clinical impact of individual diagnoses on treatment & management in a personalized approach to UDN patients. **Methods:** Review of our first 15 UDN diagnoses, based on: A) Medical record reviews, B) OMIM/SimulConsult searches, C) Clinical & research consultations & D) NGS using PhenoDB & Omicia Opal pipelines, BioVU phenotype searches, structural biology modeling & co-segregation studies. **Results:** Age(yrs), gender (F/M)-symptoms; diagnoses & therapy were: 1) 4F-fragile skin; Severe Progeria due to *ZMPSTE24* fs; statin & bisphosphonate based on mouse model, 2) 13F-vision loss & hyperlysinemia; NADK2 Deficiency due to *NADK2* variants; NADH substrate trial, 3) 18F-renal disease & pulmonary hypertension (PAH); Systemic Lupus based on renal biopsy & ANA & PAH due to *ENG* variant; hydroxychloroquine & methotrexate, 4) 32F-recurring angioedema & fever; PFAPA Syndrome due to *NLRP3* & *IL17A* variants; IL1 antagonist, 5) 37F-deafness & neuropathy; Riboflavin (B2) Transporter Deficiency based on phenotype & B2 challenge test; high dose B2, 6) 48M-head & neck nodules, & cerebral artery block; IgG4 Related Disease based on phenotype, biopsy, imaging & IgG4 levels; Rituximab, 7) 3M-delay & brain volume loss; GTP Cyclohydrolase 1 Deficiency due to *GCH1* variants; L-Dopa/Carbidopa, 8) 6F-delay & CHD; Stormorken Syn & asplenia due to a *STIM1* variant; antibiotics prn, 9) 4M-delay & inner ear anomalies; phenotypic expansion of CGD2j due to a heterozygous *COG4* variant, confirmed by model organism; NIH study (NIHS) & clinical trial (CT), 10) 31M-blindness & renal failure; Senior Loken 5 due to *IQCB1* variants, 11) 58F-Familial Paragangliomas due to a *SDHD* exon 4 deletion; NIHS & CT, 12) 2M-hypertelorism, ASD & skin pigmentation; Diploidy/Triploidy Mosaic, XLMR & Van Maldergem Syn from skin biopsy karyotypes, *THOC2 & FAT4* variants, 13) 3F-seizures & brain volume loss; CGD 1K due to *ALG1* variants; NIHS & CT, 14) 11F-muscle weakness; Charcot Marie Tooth 2S due to *IGHMPB2* variants; CT & 15) 3M-delay & nystagmus; Coffin Siris due to an *ARID1B* term variant. NGS was needed in 87% of cases (not 5-6) for diagnosis, & multiple diagnoses were made in 13% of cases. **Conclusions:** 1) We found that for 12/15 (80%) of our UDN diagnoses, the diagnosis led to precise & directed therapy &/or a clinical trial & 2) UDN cases offer unique opportunities to facilitate research, apply principles of precision medicine & offer specific, individualized treatment.

2602W

Integration of sequence data from 150,000 individuals provides new insights for variants involved in cardiomyopathy. *E.J. Mazaika[1,2], M. Ahmad[1,2], R. Govind[1,2], N. Whiffin[1,2,3], E. Edwards[1,2], B. Weisburd[4,5], M. Solomonson[4,5], A. Harper[2], R. Walsh[1,2], X. Zhang[1,2], A.E. Wilk[1,2], A. O'Donnell-Luria[4,5], H. Najgebauer[1,2], P.J.R. Barton[1,2], S.A. Cook[1,2,6], D. MacArthur[4,5], J.S. Ware[1,2,3,5].* 1) National Heart and Lung Institute, Imperial College London, UK; 2) Royal Brompton and Harefield NHS Foundation Trust, London, UK; 3) MRC London Institute of Medical Sciences, Imperial College London, UK; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 6) National Heart Centre Singapore, Singapore.

The gnomAD browser brings together genome and exome sequence data of 138,632 unrelated individuals from diverse populations, providing an unprecedented reference resource, particularly for variant analyses that leverage allele frequency information. By contrast, current repositories of genetic variation for Mendelian diseases do not record allele frequency in cases, critically limiting the application of contemporary quantitative approaches in clinical variant interpretation. We present an aggregated case series comprising genetic data from more than 10,000 cardiomyopathy patients, together with a web-browser interface (VariantFX) that allows for seamless integrative analysis with the gnomAD resource. This is the largest genetically characterized cardiomyopathy series to date, stratified by cardiomyopathy subtype and ethnicity, and includes representation of ethnic groups in whom the architecture of cardiomyopathy has been previously uncharacterized. We leverage this data to derive quantitative parameters for the interpretation of individual variants observed in cardiomyopathy, including estimates of the probability that a variant is causal in the disease, and variant penetrance. Integrative analysis of regional clustering of rare variation in cases with regional constraint in gnomAD yields insights into the molecular mechanisms of pathogenicity for missense variants in cardiomyopathy genes, and allows for extrapolation of quantitative measures for interpretation of novel and very low frequency variants. Population analyses identify distinct disease architectures in different ethnic groups. In addition to the new dataset and resulting biological insights, we have developed a data integration and visualization platform explicitly designed to be readily expanded to additional cardiovascular diseases and genes, and for modular redeployment in other disease areas. A containerized architecture allows for distributed VariantFX instances hosting new disease-specific case series, with native integration of the gnomAD resource implemented via API. Our resource is available via cardiac.variantfx.org, and the variantfx framework is available open source at https://github.com/ImperialCardioGenetics/variantfx.

2603T

Analysis of 181 selected genes associated with dilated cardiomyopathy by next-generation sequencing. *K. Al Harbi[1], Z. A. Bhuiyan[2], AM. Abdallah[3], S. Justin Carlus[1,4].* 1) Pediatrics Department, College of Medicine, Taibah University, Al-Madinah, Saudi Arabia; 2) Molecular Genetics Laboratory, Lausanne, Switzerland; 3) West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust , Birmingham , UK; 4) Global Gene Corp Pvt. Ltd, India.

Inherited cardiomyopathy is the major cause of sudden cardiac death (SCD) and heart failure (HF). Genetic defects could be elucidated in 50-70% of patients with cardiomyopathy. Mutations have been described in about 70 genes; this number is on the rise. Currently, there is no information available on genetic pathology of various cardiomyopathies in Saudi Arabia. We undertook the pilot initiative for the first time to systematically screen the sporadic Dilated cardiomyopathy (DCM) patients in Saudi Arabia, to understand the prevalence of this disease entity, prominent genes involved into the pathology of cardiomyopathies and to identify several founder mutations in our Kingdom. **Methods:** A cohort of 45 samples - 15 sporadic DCM samples and 30 unaffected parents' samples as controls were collected. Their parents displayed no phenotypic evidence of DCM on screening by echocardiography. Initially next generation sequencing was performed using NimbleGen custom design Cardiomyopathy Enrichment Panel (181 genes) in conjunction with Illumina HiSeq 2500® instrument. Our sequencing panel covered coding exons, exon-intron boundaries and known variants in intronic regions of target genes. **Results:** NGS identified 12 putative pathogenic variants in various cardiac expressed genes (LAMA4, MYH7, ACADVL, SCN5A, RBM20, TCAP, DPP6, PKP2) in 7 patients (46.6%). No genetic variants were identified in 8 cases (53.3%). **Conclusion:** Our data support the potential role of the detected variants in pathogenesis of DCM. Mutations in LAMA4, MYH7, SCN5A, RBM20, ACADVL and TCAP have been reported to be associated with DCM and HCM. Variants in PKP2 gene was associated with Arrhythmogenic right ventricular dysplasia 9. This initial analysis suggests that these mutations are very likely pathogenic and confirmed by the Sanger sequencing. .

2604F

Loss of ADAMTS3 activity causes Hennekam lymphangiectasia-lymphedema syndrome 3. *P. Brouillard[1], L. Dupont[2], R. Helaers[1], R. Coulie[1], G. Tiller[3], J. Peeden[4], A. Colige[2], M. Vikkula[1].* 1) de Duve Institute, Human Molecular Genetics, Universite catholique de Louvain, Brussels, Belgium; 2) Université de Liege, Laboratory of Connective Tissues Biology, Liege, Belgium; 3) Vanderbilt University, Medical Center, Nashville, TN, USA; 4) University of Tennessee Medical Center, East Tennessee Children's Hospital, Knoxville, TN, USA.

Primary lymphedema is due to developmental and/or functional defects in the lymphatic system. It may affect any part of the body, with predominance for the lower extremities. Twenty-seven genes have already been linked to primary lymphedema, either isolated, or as part of a syndrome. The proteins that they encode are involved in VEGFR3 receptor signaling. They account for about one third of all primary lymphedema cases, underscoring the existence of additional genetic factors. We used whole-exome sequencing to investigate the underlying cause in a non-consanguineous family with two children affected by lymphedema, lymphangiectasia and distinct facial features. We discovered bi-allelic missense mutations in *ADAMTS3*. Both were predicted to be highly damaging. These amino acid substitutions affect well-conserved residues in the prodomain and in the peptidase domain of ADAMTS3. *In vitro*, the mutant proteins were abnormally processed and sequestered within cells, which abolished proteolytic activation of pro-VEGFC. VEGFC processing is also affected by CCBE1 mutations that cause the Hennekam lymphangiectasia-lymphedema syndrome syndrome type1. Our data identifies *ADAMTS3* as a novel gene that can be mutated in individuals affected by the Hennekam syndrome. These patients have distinctive facial features similar to those with mutations in CCBE1. Our results corroborate the recent *in vitro* and murine data that suggest a close functional interaction between ADAMTS3 and CCBE1 in triggering VEGFR3 signaling, a cornerstone for the differentiation and function of lymphatic endothelial cells.

2605W

Titin rare genetic variants in arrhythmogenic cardiomyopathy. *R. Celeghin, E. Lazzarini, I. Rigato, K. Ludwig, M. Cason, G. Thiene, C. Basso, B. Bauce, K. Pilichou.* Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, Italy.

Introduction. Arrhythmogenic cardiomyopathy (AC) is a heart muscle disease characterized by life-threatening ventricular arrhythmias and progressive dystrophy of the ventricular myocardium with fibro-fatty replacement. Due to an estimated prevalence of 1:2000-1:5000, AC is define as rare disease with autosomal-dominant transmission and incomplete penetrance. It is considered a disease of intercalated discs, however about 1% of AC patients harbor a mutation in non-junctional proteins such as RYR2, TTN, LMNA, DES, PLN. Non-junctional proteins mutation frequency in AC patients is probably underestimated due to gene size-related sequencing and interpretation difficulties. In fact 60,000 *TTN* missense variants have been identified with the advent of massive parallel sequencing. The purpose of this study was to assess the frequency of *TTN* variants in AC patients. **Methods.** Eighty-nine consecutive AC index cases, fulfilling 2010 Task Force Criteria, underwent DNA sequencing on the MiSeq platform (Illumina) using the Trusight Cardio panel (174genes). Variant selection was based on current ASHG guidelines and a threshold <0,0002 minor allele frequency in the general population. Cascade genetic screening in probands' families was performed to study the segregation of variants and their impact on the clinical phenotype. **Results.** Variant filtering for aminoacid conservation and pathogenicity based on at least two *in silico* prediction algorithms (i.e Polyphen-2, SIFT, Mutation Taster) reduced the number of initially detected rare *TTN* variants from 49 (in 41 AC patients) to 29 (in 23 AC patients 26%). Only, three out of the 23 AC index cases carried an additional putative pathogenic variant in AC related genes (*DSP*, *DSC2*, *PKP2*). **Conclusions.** Preliminary data demonstrates a high frequency of rare *TTN* variants in AC patients. Additional clinical studies as to the possible phenotypical differences between *TTN*-only carriers and those carrying additional variants in other AC-related genes will help to clarify the causative or modulatory role of *TTN* in the etiopathogenesis of AC.

2606T

Novel heterozygous mutations of *KCNQ1* gene in a Jervell and Lange-Nielsen syndrome patient with gastric neuroendocrine tumor. *K. Choi, J. Kim, Y. Jee.* Department of Laboratory Medicine, Dankook University Hospital, Cheonan-si, Chungnam, South Korea.

The Jervell and Lange-Nielsen syndrome (JLNS) is a rare autosomal recessive syndrome characterized by congenital hearing loss, prolongation of the QT interval, ventricular arrhythmias, and high risk of sudden death. JLNS has been primarily caused by homozygous variants in *KCNQ1* or *KCNE1* gene. There have been very few clinical case reports on JLNS in Korea. Recently, we identified novel heterozygous variant of *KCNQ1* gene using massively parallel sequencing method in a Jervell and Lange-Nielsen syndrome patient, who underwent surgery due to gastric neuroendocrine tumor. She complained epigastric pain for 6months and her laboratory results showed iron deficiency anemia. Her electrocardiogram revealed a markedly prolonged QT interval. Targeted next generation sequencing analysis were performed and it revealed the presence of novel heterozygous variant (c.826T>C, p.Ser276Pro). The same variant was found in the parents, a sister, and two children, and all of them appeared to have prolonged QT interval. This is the first report of gastric neuroendocrine tumor in a Jervell and Lange-Nielsen syndrome patient although iron-deficiency anemia and gastric hyperplasia have been reported previously.

2607F

MCTP2 gene change detected by whole exome sequencing in an infant with endocardial fibroelastosis syndrome who underwent heart transplantation. *D. Ercelen[1], B. Monteleone[2], G. Wallis[3].* 1) Carolinas Health System, Charlotte, NC; 2) Carolinas Health System, Charlotte, NC; 3) Carolinas Health System, Charlotte,NC.

This is the first report of a patient with Endocardial Fibroelastosis (EF) Syndrome. Whole exom analysis demonstrated a variant of unknown change in MCTP2 gene: c.2156A>G (p.Y719C) that was inherited from the father. This variant was not detected in the NHLBI Exome Sequencing Project, indicating it is not a common benign variant. In silico analysis predicts this variant is probably damaging to protein structure and function. 5 days old girl was born at 39 weeks with a large left ventricle cardiac mass detected by ECHO. Birth weight: 3.74 kg, length: 49.5 cm,head circumference: 35.5 cm. Physical exam was normal. MRI of the brain and kidneys were normal. Her clinical findings were not sufficient for a diagnosis of Tuberosclerosis with normal TSC1 and TSC2 genetic. Patient did not respond to several courses of sirolimus. Cardiac MRI showed several masses in the left ventricle, causing severe obstruction and diastolic dysfunction. She subsequently developed signs of atrial hypertension with concomitant pulmonary hypertension. and received heart transplant at 9 months of age. Endocardial fibroelastosis (EF)syndrome is a pronounced, diffuse thickening of the ventricular endocardium that presents as an unexplained heart failure in infants and children. The disease can be primary or secondary to various congenital heart diseases. Primary EF is not associated with any significant cardiac structural abnormalities, and one of the causes of unexplained heart failure. 25% of children who were transplanted for dilated cardiomyopathy demonstrated significant endocardial fibroelastosis on histopathology of the explanted hearts. MCTP2 gene are calcium-binding domains works in proteins involved in signal transduction or membrane trafficking. In animal models, MCTP2 gene has been shown to be associated with cardiogenesis. More, disruption of MCTP2 has been reported to contribute to congenital left heart obstructive cardiac defects in humans. Previously, 2 patients with 15q26.2 deletion involving MCTP2 gene was reported with EF syndrome caused by pre and postnatal left ventricular outflow tract (LVOT) obstruction.Their mother had similar heterozygous deletion with normal ECHO. The penetrance for the cardiac phenotype is not complete with Mctp2 knockdown. In summary, we have preliminary evidence that variants in MCTP2 gene are associated with primary and secondary EF syndrome. More patients are warranted to investigate these associations further.

2608W

Use of the ClinGen clinical validity framework to evaluate the strength of evidence for genes implicated in hypertrophic cardiomyopathy. *J. Goldstein[1], J. Ingles[2], C. Caleshu[3], E. Corty[1], S. Crowley[1], K. Dougherty[4], J. McGlaughon[1], L. Milko[1], A. Morales[5], B. Seifert[1], C. Semsarian[2], N. Strande[1], C. Thaxton[1], K. Thomson[6], P. van Tintelen[7], K. Wallace[1], R. Walsh[8], J. Ware[8], Q. Wells[9], N. Whiffin[8], L. Wikowski[10], R. Hershberger[5], B. Funke[10], ClinGen Cardiovascular Clinical Domain Working Group.* 1) UNC-Chapel Hill, Chapel Hill, NC; 2) Sydney Medical School, University of Sydney, Australia; 3) Stanford University, CA, USA; 4) Eastern Virginia Medical School, VA, USA; 5) Ohio State University, OH, USA; 6) Oxford Medical Genetics Laboratory, Oxford, UK; 7) Academic Medical Center, Amsterdam, Netherlands; 8) Imperial College London, London, UK; 9) Vanderbilt University Medical Center, TN, USA; 10) Harvard Medical School/MGH, MA, USA.

Hypertrophic cardiomyopathy (HCM) is defined as unexplained left ventricular hypertrophy (LVH) in the absence of other causes. The clinical presentation ranges from asymptomatic LVH to progressive heart failure, arrhythmia, and sudden cardiac death. In the majority of cases, HCM is caused by a pathogenic variant in one of 8 genes encoding components of the cardiac muscle sarcomere. Z-disc and calcium-handling genes have also been implicated. Commercial laboratory gene testing panels can include 50 or more genes with variable levels of evidence to support their association with HCM. Using the Clinical Genome Resource (ClinGen) gene-disease clinical validity framework, we curated publically available data to classify the clinical validity for 55 genes and HCM. The gene list was based on (1) Presence on >30% of HCM gene panels in the NCBI Genetic Testing Registry, (2) Inclusion in the TruSight HCM gene panel (Pua et al, 2016, PMID: 26888179), and (3) Expert opinion. Curations were performed by biocurators, overseen by a cardiac expert, and discussed on regular conference calls to reach a final gene-disease validity classification. Initially, we targeted genes that appeared to be associated with isolated HCM. Of 37 genes curated to date, 8 had a definitive association, 2 strong, 4 moderate, 16 limited, and 7 had no reported evidence. As we began to curate genes associated with HCM as part of a wider syndrome (e.g. GLA, FHL1) we recognized the need to fully evaluate the range of disorders caused by all of the genes on our list, and whether it was appropriate to "split" the curation to include only individuals with isolated HCM or "lump" cases into a wider syndrome. Following the recommendations of the newly-formed ClinGen Lumping and Splitting working group, we binned all 55 genes on our list into one of three groups: genes causing isolated HCM, genes associated with multiple cardiac phenotypes, and syndromic genes that include HCM as part of the phenotypic spectrum. We will present the rationale for the binning process and our curation results to date. We conclude that the ClinGen clinical validity classifications will be beneficial to the development and interpretation of genetic testing in patients with HCM.

2609T

Cardiologic findings in 101 patients with Williams Syndrome. *R.S. Honjo[1], J.R.M. Ceroni[1], P.H.S. Castro[1], J.G. Salomão[1], S.P.O. Saldaña[1], D.C.S. Bermeo[1], E.A. Furusawa[2], G.N. Leal[3], J. Wagenführ[4], E.A. Zanardo[5], L.D. Kulikowski[5], D.R. Bertola[1], C.A. Kim[1].* 1) Genetics Unit, Instituto da Criança, São Paulo, São Paulo, Brazil; 2) Nephrology Unit, Instituto da Criança - HC-FMUSP, São Paulo - SP, Brazil; 3) Echocardiography Service, Instituto da Criança - HC-FMUSP, São Paulo - SP, Brazil; 4) Pediatric Cardiology Unit, Instituto da Criança - HC-FMUSP, São Paulo - SP, Brazil; 5) Medical Investigation Laboratory (LIM 03) - FMUSP, São Paulo - SP, Brazil.

Objective: To identify and describe the cardiologic findings of 101 patients with Williams Syndrome (WS) confirmed by molecular studies. **Methods:** Retrospective and observational study of the cardiologic findings in 101 patients with molecularly confirmed WS, evaluated from 2008 through 2016. **Results:** The total cohort consisted of 101 patients, 62.4% males and 37.6% females. A total of 77 (76.2%) WS children had a congenital structural heart defect. Overall 44/77 (57.1%) patients with heart anomalies had a single defect, while the remaining 33/77 (42.9%) had combined heart defects. The most common abnormality was supravalvular aortic stenosis (48%), found in 37 patients, of whom 15 (40.5%) had another heart defect. Other defects included: Aortic stenosis (7.8%), pulmonary stenosis (26%), stenosis of the branch pulmonary arteries (14.3%), mitral valve prolapse (19.5%), mitral insufficiency (11.7%), aortic coarctation (7.8%), aortic insufficiency (3.9%), ventricular septal defect (3.9%), atrial septal defect (3.9%), bicuspid aortic valve (2.6%). Dilated aortic root, tricuspid insufficiency and hypertrophic cardiomyopathy were all found each in one patient (1.3%). A total of 28/80 (35%) had normal ECG, while left ventricle overload, the most common ECG abnormality, was found in 5 patients (6.3%). Other findings: abnormal ventricular depolarization (2.5%), right atrial overload (1.2%), conduction delay (1.25%), and ectopic atrial rhythm (1.2%). High blood pressure was diagnosed in 42 of 101 (41.6%) patients. Heart transplantation was carried out in a 19-year-old female patient who presented supravalvular aortic stenosis and refractory congestive heart failure. This patient had been previously submitted to aortic stenosis correction twice. Unfortunately, the patient died due to CMV infection on the 47th day after transplant. **Conclusions:** A high incidence of potentially hazardous congenital heart diseases has been reported in patients with WS. There are few descriptions of heart transplantation procedures in patients with genetic syndromes and, up to date, only one in a WS patient. Cardiovascular diseases are responsible for the greatest morbidity and mortality in WS, but are preventable and treatable with the appropriate care. It is recommended that all patients have periodic cardiology and nephrology follow-up, in order to prevent, manage, and treat possible complications and improve both quality of life and survival rates. (Support: CNPQ / FAPESP).

2610F

Toward genetics-driven early intervention in dilated cardiomyopathy: The DCM Precision Medicine Study. *D.D. Kinnamon[1], A. Morales[1], D.J. Bowen[2], W. Burke[2], R.E. Hershberger[1] on behalf of the DCM Consortium.* 1) Division of Human Genetics, Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, OH; 2) Department of Bioethics and Humanities, University of Washington, Seattle, WA.

The etiology of idiopathic dilated cardiomyopathy (DCM) is unknown by definition, but its familial subtype is considered to have a genetic component. We hypothesize that most idiopathic DCM, whether familial or non-familial, has a genetic basis, in which case a genetics-driven approach to identifying at-risk family members for clinical screening and early intervention could reduce morbidity and mortality. Based on this hypothesis, we have launched the NHLBI- and NHGRI-funded DCM Precision Medicine Study, which aims to enroll 1,300 individuals (600 non-Hispanic African ancestry, 600 non-Hispanic European ancestry, and 100 Hispanic) who meet rigorous clinical criteria for idiopathic DCM along with 2,600 of their relatives. Enrolled relatives undergo clinical cardiovascular screening to identify asymptomatic disease, and all individuals with idiopathic DCM undergo exome sequencing to identify relevant variants in genes previously implicated in DCM. Genetic test results are returned by genetic counselors 12-14 months after enrollment. The data obtained will be used to describe the prevalence of familial DCM among idiopathic DCM cases and the genetic architecture of idiopathic DCM in multiple ethnicity-ancestry groups. We are also conducting a randomized controlled trial (NCT #: NCT03037632, https://www.clinicaltrials.gov) to test the effectiveness of *Family Heart Talk*, an intervention to aid family communication, for improving uptake of preventive screening and surveillance in at-risk first-degree relatives. In this presentation, we describe the design and analysis plan for the DCM Precision Medicine Study while summarizing current progress and lessons learned since accrual began in June 2016. At its conclusion, we anticipate that the study will demonstrate that idiopathic DCM has a genetic basis and guide best practices for a genetics-driven approach to early intervention in at-risk relatives.

2611W

Familial TAPVR with 15q11.2 (BP1-BP2) microdeletion. *Y. Kuroda[1,3], I. Ohashi[1], T. Naruto[1], K. Ida[1], Y. Enomoto[1], T. Saito[2], J. Nagai[2], K. Kurosawa[1].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Clinical Laboratory, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Department of Pediatrics, University of Tokyo, Tokyo, Japan.

Congenital heart diseases (CHDs) were found in 20-30 % of the 15q11.2 (BP1-BP2) deletion patients. We reported familial TAPVR patients with microdeletion of 15q11.2. Patient 1 was the first child of nonconsanguineous and healthy parents. He was delivered at a gestational age of 37 weeks and 3 days after an uneventful pregnancy. His birth weight was 2486 g. ;Echocardiography showed total anomalous pulmonary venous return (TAPVR) type 1a and surgical repair was performed at the next day after birthday. He started controlling head at 4 months, rolling over at 5 months, sitting at 7 months, ad walking at 17 months. He spoke the first word at 1 year, and two-word sentences at 3 years of the age. He had learning disability and attended a special class. Patient 2 was the second child of parents of Patient 1. She was delivered at a gestational age of 38 weeks. Her birth weight was 3066 g. At a gestational age of 34 weeks, the diagnosis of TAPVR type 3 was confirmed by fetal echocardiography. After birth, surgical repair was performed. She died of renal failure due to chylothorax at 8 months-old of the age. Patient 3 was the fourth child and younger sister of dizygotic twins. She was delivered at a gestational age of 36 weeks and 3 days by emergency caesarean section because of threatened premature delivery. Her birth weight was 2823g. At a gestational age of 31 weeks, the diagnosis of TAPVR was confirmed and surgical repair was performed at the birthday. At the age of 4 years and 6 months, she had speech delay. In Patient 1, karyotyping revealed normal karyotype. No deleterious mutation was identified by exome sequencing. Array CGH revealed an approximately 222 kb deletion at 15q11.2, spanning from position 22,860,857 to 23,082,821 bp (GRCh37/hg19). FISH analysis confirmed the deletion in Patient 1, 3 and their father. Healthy sibling (the third child) and their mother did not have the deletion. More than 27 patients of CHD have been reported associated with the 15q11.2 deletion, and no familial case has been reported. The penetrance of CNV was estimated as 10.4%. In this family, higher penetrance was observed, since TAPVR were shown in all three siblings harboring the CNV.

2612T

Genetic testing in arrhythmogenic cardiomyopathy: New insights into the disease determinants and new doubts. *E. Lazzarini, R. Celeghin, K. Ludwig, M. Cason, I. Rigato, D. Corrado, C. Basso, G. Thiene, B. Bauce, K. Pilichou.* Dept of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, Italy.

   Background: Arrhythmogenic Cardiomyopathy (AC) is an inherited cardiac disease (MIM#107970) characterized by fibro-fatty replacement of myocardium and increased risk of sudden cardiac death. AC is mainly linked to variants in desmosomal genes, detected in about 50% of cases. **Purpose**: AC familial cascade screening can be life-saving, thus the aim of the study is to define the genetic spectrum of AC in a large patient cohort. **Methods**: 189 consecutive AC index cases (73F, 116M, mean age 35±12y) fulfilling revised 2010 Task Force diagnostic criteria, were enrolled at the Referral Clinical-Genetic Centre of Padua University during years 2013-2017. Molecular genetic testing was performed on a Miseq platform using a cardiovascular gene panel of 174 (TruSightCardio, Illumina). Alignment and variant calling were realized by Miseq Reporter and VariantStudio softwares (Illumina). Variants filtering and prioritization were carried out considering a population frequency <0.02%, protein conservation (10 species), clinical data, literature and co-segregation analyses. Potentially pathogenic variants and regions with insufficient coverage were Sanger sequenced. **Results**: Almost 50% (94/189) of AC index cases carried a rare nucleotide variant in AC-related genes, of whom 44.5% (84/189) were desmosomal variant carriers and 5.3% (10/189) displayed rare variants in non-desmosomal AC-related genes (*CTNNA3, DES, SCN5A, FLNC, TMEM43*). Specifically, 15 were single-variant *DSP* carriers, 7 *DSG2*, 29 *PKP2*, 7 *DSC2*, 2 *JUP,* and 24 patients were compound or digenic heterozygous carriers in desmosomal genes and 2 *CTNNA3*, 3 *DES,* 2 *SCN5A*, 1 *FLNC,* and 2 *TMEM43*. Moreover, all 94 index cases carried additional rare variants (according to the minor allele frequency ≤ 0.5%, ASHG guidelines) in AC-unrelated genes, mostly encoding for sarcomeric proteins (*OBSCN, MYH7, MYBPC3*) or ion channels subunits (*CACNA1B, CACNA1C, SCN10A, KCNQ1*); the significance of these additional variants remains to be elucidated. **Conclusions**: Analysis of a large AC cohort broadens our knowledge regarding the genetic complexity of the disease. AC patients carried a combination of pathogenic variants and/or multiple rare variants in AC-unrelated genes. Our data highlights the importance of AC family cascade genetic screening in order to determine the clinical significance of rare genetic variants.

2613F

A report of a patient with hypertrophic cardiomyopathy without myopathy associated with FLNC heterozygous pathogenic variant and review of literature. *B. Monteleone[1], C. Strachan[2].* 1) Carolinas Health System,UNC Charlotte, Charlotte, NC; 2) UNC Charlotte.

   This is a 7-month-old male diagnosed with cardiomyopathy status post heart transplant. Dilated and hypertrophic cardiomyopathy panel to Gene Dx were all normal. Chromosome microarray showed normal male, 46, XY. Urine Hex4 and GAA enzyme levels were normal. Metabolic screening was normal. His physical exam was otherwise normal. Whole exome analysis identified a heterozygous likely pathogenic variant in FLNC gene: IVS21-1G>C (c.3791-1G>C) that was inherited from his father. His family history is significant for father's sudden cardiac death, and resuscitated and ICD was placed. FLNC gene is essential for sarcomere attachment to the plasmatic membrane. Patient and his father have no sign of myopathy or any other neurological issues. It is known to be associated with myofibrillar myopathy 5 as well as cardiomyopathy, familial restrictive 5 (RCM5), and Cardiomyopathy, familial hypertrophic 26 (CMH26) that leads to cardiomyopathy, arrhythmias as well as sudden death. Patients with myopathy has been reported to have risk to develop cardiomyopathy as well. However, there is little evidence that patients with cardiomyopathy and harbor FLNC mutations do not develop myopathy. More studies are needed to understand patients with FLNC gene mutations who present with cardiomyopathy initially and their risk to develop subsequent myopathy.

2614W

Comparison of genetic architecture of isolated left ventricular non-compaction cardiomyopathy and familial dilated cardiomyopathy as assessed by whole exome sequencing. *L. Piherová[1], H. Hartmannova[1], V. Stranecky[1], A. Pristoupilova[1], K. Hodanova[1], V. Baresova[1], M. Kubanek[2], A. Krebsova[2], R. Kockova[2], T. Palecek[3], F. Malek[4], V. Melenovsky[2], S. Kmoch[1].* 1) Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles Univesity, Prague, Czech Republic; 2) Department of Cardiology, Institute of Clinical and Experimental Medicine Prague, Czech Republic; 3) Second Internal Clinic, First Faculty of Medicine, Charles Univesity, Prague, Czech Republic; 4) Department of Cardiology, Na Homolce Hospital, Prague, Czech Republic.

Isolated left ventricular noncompaction cardiomyopathy (LVNC) is an unclassified cardiomyopathy which may overlap with genetic forms of dilated cardiomyopathy. We aimed to compare genetic architecture of LVNC and familial DCM (DCM) to elucidate whether LVNC represents a distinct disease entity. A representative cohort of 42 patients with LVNC and 245 probands with familial DCM from three major Czech hospitals was assessed by whole exome sequencing (WES; SeqCap EZ MedExome, Roche, USA; Illumina, USA). Detected variants were confirmed by Sanger sequencing and their segregation was assessed in available families. Diagnosis of LVNC was made by two specialists based on echocardiographic criteria and/or cardiac magnetic resonance imaging. Three probands with LVNC had familial disease with either LVNC (2 cases) or hypertrophic cardiomyopathy (1 case) in a first degree relative. In LVNC patients, WES revealed one pathogenic mutation in 55% of cases, ≥2 pathogenic mutations in 20% and an inconclusive result in 25% cases. In two cases we have found large deletion of 2 sarcomeric genes. The corresponding results of WES in familial DCM were as follows: one pathogenic mutation in 30%, ≥2 pathogenic mutations in 43% and an inconclusive result in 27% cases. Our data indicate that LVNC is predominantly caused by mutations affecting sarcomeric genes. On the other hand, truncating mutations of *TTN* were the most common etiology of familial DCM followed by mutations of sarcomeric and desmosomal genes.

2615T

High incidence of *SCN5A* c.5350 G>A (p.Glu1784Lys) heterozygotes and prolonged QT intervals in a large biobank cohort. *M.L.B. Schwartz[1], J.Z. Luo[1], A.C. Sturm[1], M.F. Murray[1], R. Pulk[1], M.S. Williams[1], D. Hartzel[1], F.E. Dewey[2], U.L. Mirshahi[1], K. Manickam[1], T. Mirshahi[1].* 1) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 2) Regeneron Genomics Center, Tarrytown, NY.

Background The MyCode Community Health Initiative at Geisinger Health System has over 140,000 participants who have agreed to sample collection, genomic sequencing, use of their Electronic Health Record (EHR) data and the return of actionable genomic results. One variant of clinical interest is *SCN5A* c.5350 G>A (p.Glu1784Lys), a well-described cause of both Long QT syndrome type 3 (LQT3) and Brugada syndrome. **Methods** A bioinformatics analysis found *SCN5A* variants, including the c.5350 G>A (p.Glu1784Lys) variant, among the first 60,873 participants sequenced. Personal and family histories of cardiac phenotypes in variant carriers were ascertained by an EHR data pull and chart review. A relatedness analysis was conducted using the Pedigree Reconstruction and Identification of a Maximum Unrelated Set (PRIMUS) package. Comparison of QTc intervals was done using a one-way ANOVA with Tukey's post-hoc test. **Results** Eighteen people had the *SCN5A* c.5350 G>A (p.Glu1784Lys) variant. Fifteen (86.7% female, mean age 55.7y) had ≥1 EKG available for review. The mean QTc interval in carriers was 557.3 ms in males and 483 ms in females, which was significantly different from the Geisinger EHR average (male 429.8 ms, p<0.0001 female 437.6 ms, p<0.0001). Chart review was allowed for 14 patients and revealed that 6 (42.9%) had an EKG flagging a prolonged QT interval; however, only 1 had prolonged QT on the EHR problem list. One 84-year-old male with multiple cardiac comorbidities had a history of sudden cardiac arrest and an ICD and a 27-year-old female had a history of unexplained syncope. All 14 patients had a family history of a "heart disorder" including 8 with a first-degree relative (FDR) with a "heart attack" (7/8 male), including 2 male FDRs with an event in their 20s. Only two variant carriers were FDRs. No patient had a clinical diagnosis of LQT3, Brugada syndrome or genetic testing independent of the study. **Conclusions** While 18 carriers of the previously reported *SCN5A* c.5350 G>A (p.Glu1784Lys) variant were found in this cohort, the variant is absent in other large genomic databases such as ExAC and gnomAD. It is unclear if a regional founder effect influenced variant frequency in this cohort. Some variant carriers had prolonged QT intervals but were not recognized to have LQT3 by their clinical teams. One hypothesis for the sex ratio skew in variant carriers is higher disease-related mortality among male carriers.

2616F

Germline loss-of-function mutations in *EPHB4* cause a second form of capillary malformation–arteriovenous malformation (CM-AVM2) deregulating RAS-MAPK signaling. *M. Vikkula[1], M. Amyere[1], N. Revencu[2], R. Helaers[1], E. Pairet[2], J.B. Mulliken[3], L.M. Boon[1,4].* 1) de Duve Institute, Universite catholique de Louvain, Brussels, Belgium; 2) Center for Human Genetics, Cliniques universitaires Saint-Luc, Universite catholique de Louvain, Brussels, Belgium; 3) Vascular Anomalies Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; 4) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques universitaires Saint-Luc, Universite catholique de Louvain, Brussels, Belgium.

BACKGROUND Most AVMs are localized and occur sporadically; however they also can be multifocal in autosomal dominant disorders, such as Hereditary Hemorrhagic Telangiectasia (HHT) and Capillary Malformation–Arteriovenous Malformation (CM-AVM). Previously, we identified RASA1 mutations in 50% of patients with CM-AVM. Herein we studied non-RASA1 patients to further elucidate the pathogenicity of CMs and AVMs. METHODS We conducted a genome-wide linkage study on a CM-AVM family. Whole exome sequencing was also performed on 9 unrelated CM-AVM families. We identified a candidate-gene and screened it in a large series of patients. The influence of several missense variants on protein function was also studied in vitro. RESULTS We found evidence for linkage in two loci. Whole-exome sequencing data unraveled four distinct damaging variants in *EPHB4* in five families that co-segregated with CM-AVM. Overall, screening of *EPHB4* detected 47 distinct mutations in 54 index patients: 27 lead to a premature stop codon or splice-site alteration, suggesting loss of function. The other 20 are non-synonymous variants that result in amino-acid substitutions. In vitro expression of several mutations confirmed loss of function of EPHB4. The clinical features included multifocal CMs, telangiectasias, and AVMs. CONCLUSIONS We found *EPHB4* mutations in patients with multifocal CMs associated with AVMs. The phenotype, CM-AVM2, mimics *RASA1*-related CM-AVM1 and also HHT. *RASA1* encoded p120RASGAP is a direct effector of EPHB4. Our data highlights the pathogenetic importance of this interaction and indicts EPHB4-RAS-ERK signaling pathway as a major cause for arterio-venous malformations.

2617W

Predisposition genetic screening for actionable cardiovascular conditions in patients undergoing heritable cancer syndrome testing: Prevalence of pathogenic variants in 10,812 individuals. *S. Yang[1], E. Esplin[1], E. Haverfield[1], S. Aradhya[1], R. Nussbaum[1,2].* 1) Invitae, San Francisco, CA; 2) UCSF, San Francisco, CA.

Background The ACMG published recommendations for reporting medically actionable pathogenic variants in a set of 59 curated genes as secondary findings in patients undergoing whole exome or genome sequencing. We report an initial estimate of the prevalence of secondary findings in genes implicated in dominant cardiovascular disorders in 10,812 individuals with personal or family history of cancer, and stratify the prevalence by ethnicity. **Methods** Per an IRB approved protocol, we analyzed data from 10,812 de-identified patients, originally referred for cancer genetic testing, for 68 genes associated with Mendelian cardiovascular conditions considered medically actionable by in-house experts; which encompass all the 30 cardiovascular disease genes recommended by ACMG for secondary findings. We identified known pathogenic (P) or likely pathogenic (LP) variants and predicted LOF variants and stratified the patients according to reported ethnicity. **Results** In 10,812 de-identified individuals, we found 250 patients (2.31%) with P/LP variants in 68 cardiovascular disorder-related genes. The prevalence is lower if only consider the ACMG 30 is considered. We also observed that in this cohort, 37 patients (0.34%) have positive findings in both cancer and cardiovascular genes. **Conclusions** Overall, 2.31% of unaffected individuals have positive findings in 68 genes associated with Mendelian cardiovascular disorders; the prevalence is lower (1.64%) for 30 cardiovascular ACMG genes. People with Caucasian, black and Ashkenazi Jewish ethnicity have a higher prevalence than Hispanic individuals while people of Asian descent have the lowest rate. We also observed 0.34% of patients referred for heritable cancer syndrome testing carried P/LP variants in both cancer and cardiovascular genes. This information could be potentially important for guiding the choice of chemotherapy with cardiotoxic effects. The overall prevalence estimate is likely underestimated. Nonetheless, the findings observed in this study contribute to our knowledge concerning genetic screening for heritable cardiovascular disorders in individuals with a personal or family history of cancer.

	All (n=10812)	Caucasian (n=7320)	Hispanic (n=706)	Black (n=667)	Ashkenazi Jewish (n=497)	Asian (n=241)
Positive findings in-house panel	250 (2.31%)	172 (2.35%)	11 (1.56%)	15 (2.25%)	13 (2.26%)	2 (0.83%)
Positive findings in ACMG panel	120 (1.64%)	120 (1.64%)	10 (1.42%)	12 (1.80%)	8 (1.61%)	2 (0.83%)

2618T

Novel *CACNA1C* variant associated with Long QT syndrome in a multi-generational family. *V. Novelli[1,4], F.D Tiziano[2], G. Pelargonio[1], G. LaRosa[2], P. Zeppilli[3,4], F. Crea[2], M. Genuardi[1].* 1) Institute for Genomic Medicine, Fondazione Policlinico Gemelli, ,Catholic University of the Sacred Heart, Rome, Italy; 2) Department of Cardiovascular Sciences,Catholic University of the Sacred Heart, Fondazione Policlinico Gemelli, Rome, Italy; 3) Sports Medicine Unit, Catholic University of Sacred Heart, Rome, Italy; 4) Center, Benito Stirpe, for the prevention of the Sudden Cardiac Death in the young athletes. Fondazione Policlinico Gemelli,Roma Italy.

   Background Long QT syndrome (LQTS) is a potentially lethal cardiac disease characterized by a prolongation of the QT interval and by a propensity for life-threatening arrhythmias. Genetic alterations in the *KCNQ1*, *KCNH2* and *SCN5A* genes, encoding for ion cardiac channels, account for the 75-80% of the genotype-positive patients. Additional minor genes have been involved in this conditions. Recently, genetic variants in the *CACNA1C* gene, encoding for the a-subunit of the cardiac L-type calcium channel (LTCC), have been identified in 2 families: one with LQTS and one with a complex phenotype of LQTS, hypertrophic cardiomyopathy, sudden death and congenital heart defects. Here we show the results from a LQTS multigenerational pedigree with negative screening in the main LQT genes. **Methods** A 43 years old woman was referred to our center for evaluation of long QT (QTc>490 ms) and aspecific T-wave abnormalities on ECG. Two of her three sons (age 8y and 14y) also had a diagnosis of LQTS (QTc=510ms and 500ms) with a syncopal episode reported in the younger brother. To identify a causative genetic variant associated with the phenotype in this family, we performed genetic testing, using a custom targeted panel, containing 12 genes, associated with the disease. **Results and Conclusions** A novel variant in *CACNA1C*, c.4879 T>C, p.(Tyr1627His) was identified in the proband and in her two affected sons. Cascade screening in the family members confirmed the presence of the variant also in the mother and in the brother of the proband, who both had borderline QT interval (QTc>460ms). The variant was not presented in her sister and in the third son who had a normal QT interval. The variant is not reported in GnomAD (gnomad.broadinstitute.org) and in literature. **Conclusion** Co-segregation of this novel *CACNA1C* variant with the clinical phenotype of LQTS in multiple family members supports the hypothesis that *CACNA1C* is a novel player in the spectrum of LQTS. This work demonstrates the importance of using targeted sequencing panels in LQTS, updated with the new disease-associated genes, in order to identify causative variants.

2619F

Mutation spectrum of Long QT Syndrome in Singapore. *R.Y.Y. Yong[1], B.Y Tan[2], L.SH Gan[1], P.S Wasan[1], M.AL Teo[1], C.K Ching[2], D.TT Chong[2], K.L Ho[2], L. Wang[4], W. Chow[4], J.TL Choo[5], S. Moochhala[1], E.PH Yap[6], M. Uttamchandani[1], S.C Seow[3], W.S Teo[2].* 1) Defence Med & Environmental, DSO National Labs, Sinagpore, Singapore; 2) National Heart Centre, Singapore General Hospital, Singapore; 3) Cardiac Department, National University Hospital, Singapore; 4) Headquarters Medical Corps, Singapore Armed Forces, Singapore; 5) KK Women's and Children's Hospital, Singapore; 6) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.

   Long QT syndrome (LQTS) is an inherited arrhythmia syndrome with an increased risk of sudden death. It is characterized by a prolonged QT interval on 12-lead ECG that can progress to a polymorphic ventricular tachycardia known as torsades de points in the absence of structural abnormal heart. Clinically, torsades de points can produce syncope, ventricular fibrillation and even sudden cardiac death. The prevalence of LQTS is reported to be 1 in 2000 in Caucasians. LQTS is known to be genetically heterogenous with at least 16 genes been implicated. However, mutations in 3 genes (*KCNQ1, KCNH2, SCN5A*) can account for 70% to 75% of diagnosed patients in Caucasians. Genetic testing plays an important role in the diagnosis of LQTS according to expert consensus statement.   This study investigated the mutation spectrum of the 3 genes (*KCNQ1, KCNH2, SCN5A*) in 26 unrelated LQTS patients in Singapore. LQTS cases are diagnosed with a Schwartz score >3 in the absence of a secondary cause for QT prolongation. Bi-directional Sanger sequencing was carried out on all the exons cum flanking sequence of the 3 genes, including 5' and 3' UTRs. The minor allele frequency of all potential mutations in normal controls was determined by genotyping 105 healthy Singapore Chinese controls, and with reference to ExAC and 1K Genome databases. Putative mutations were identified in 13 probands (Combined diagnostic yield 50%), with one proband harbouring 3 mutations in *SCN5A*. A total of 15 mutations were identified, including 4 for *KCNQ1*, 5 for *KCNH2*, and 6 for *SCN5A*. These comprised of 12 missenses, 1 nonsense and 2 indel/frameshift. Many of these mutations have been reported in recent literature, except for 2 novels for *KCNH2*. Among this patient cohort, family samples were available for 7 probands. Five families have mutation identified demonstrating Mendelian segregation. The study is on-going. To our knowledge, this is the largest study of LQTS mutation spectrum in Singapore to date.

2620W

Clinical evolution and recommendations for management of the smooth muscle dysfunction syndrome due to mutations of the ACTA2 arginine 179. *E. Regalado[1], J. De Backer[2], A. Braverman[3], S. Benedict[4], T. Bradley[5], M.E. Brickner[6], K. Chatfield[7], A. Child[8], K. Holmes[9], G. Iannucci[10], T. Morisaki[11], H. Morisaki[11], S. Morris[12], A.L. Mitchell[13], J. Ostergaard[14], J. Richer[15], D. Sallee[10], L. Mellor-Crummey[1], S. Shalhub[16], A. Estrera[1], P. Musolino[17], A. Yetman[18], R. Pyeritz[19], D. Milewicz[1], Montalcino Aortic Consortium.* 1) University of Texas Health Science Center Houston, Houston, TX, USA; 2) University Hospital Ghent, Ghent, Belgium; 3) Washington University School of Medicine, St. Louis, MO, USA; 4) The University of Utah School of Medicine, Salt Lake City, UT, USA; 5) University of Saskatchewan, Saskatoon, SK, Canada; 6) University of Texas Southwestern Medical Center, Dallas, TX, USA; 7) University of Colorado, SOM, Aurora, CO, USA; 8) St George's University of London, London, UK; 9) Oregon Health and Science University, Portland, OR, USA; 10) Emory University School of Medicine, Atlanta, GA, USA; 11) Tokyo University of Technology School of Health Sciences, Tokyo, Japan; 12) Texas Children's Hospital, Houston, TX USA; 13) University Hospitals, Cleveland, OH, USA; 14) Aarhus University Hospital, Aarhus, Denmark; 15) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 16) University of Washington, Seattle, WA, USA; 17) Harvard Medical School, Boston, MA, USA; 18) Children's Hospital & Medical Center, University of Nebraska, Omaha, NE, USA; 19) Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.

Smooth muscle dysfunction syndrome (SMDS) due to heterozygous ACTA2 arginine 179 mutations is characterized by complex congenital heart defects, pulmonary arterial hypertension, cerebrovascular disease, thoracic aortic aneurysms and dissections, arterial aneurysms, and chronic conditions affecting organs dependent on smooth muscle actin. We examined the clinical history and outcomes of 33 SMDS cases identified worldwide and developed recommendations for evaluation, surveillance, and management of this condition. All 33 cases had congenital mydriasis and presented with a patent ductus arteriosus or aorto-pulmonary window and required surgical repair during infancy (median age at repair at 61 days). Cerebrovascular disease characterized by Moyamoya-like distal internal carotid artery stenosis was present in 53% of cases and small vessel disease as evidenced by T2 weighted- hyperintense periventricular white matter lesions on magnetic resonance imaging in 95% of cases. Overt ischemic strokes occurred in 9 cases (27%), with 7 of the cases presenting before the age of 10 years, and seizures occurred in 6 cases (19%) between the ages of 1 and 19 years. Thoracic aortic disease requiring ascending aortic aneurysm repair or presenting as dissections was fully penetrant by the age of 25 years. Severe cerebrovascular disease delayed aortic aneurysm repair in 3 cases, leading to deaths due to aortic rupture. Three cases had subclavian artery aneurysms resulting in thromboembolic episodes and multiple surgical operations to repair the aneurysms. Pulmonary complications, including pulmonary arterial hypertension or chronic bronchopulmonary disease, were prevalent in 73% of cases and across all age groups. Gastrointestinal complications including intestinal malrotation, immotility or gall bladder sludge or stone were observed in 58% of cases. Hypotonic bladder and urinary tract complications were noted in 64% of cases. Cryptorchidism was present in 40% of males. Premature deaths occurred in 8 cases between the ages of 0.5 and 32 years due to pulmonary disease, aortic dissection or rupture, or stroke. These data demonstrate the clinical evolution of SMDS and emphasize the importance of early detection, multi-disciplinary and coordinated care, and timely management of the SMDS. These findings define the SMDS complications so that therapies can be initiated and tested to reduce the high mortality and morbidity in this patient population.

2621T

HIVEP2 as a cause hypoplastic left heart syndrome: A report of a rare condition in an 18 month old boy. *H. Holway[1], B. Monteleone[2].* 1) UNC Chapel Hill, Chapel Hill, NC; 2) Carolinas Health System, UNC Charlotte.

Human immunodeficiency virus type 1 enhancer binding protein (HIVEP2) is part of a family of zinc-finger-containing transcriptional proteins that known to be involved in brain growth and development as well as cellular immunity, bone remodeling, and adipogenesis. HIVEP2 gene defects have been previously associated with intellectual disability, developmental delay, behavioral problems, and dysmorphic feature. This is the 10th case of pathogenic HIVEP2 mutation reported in the literature to our knowledge. Previously described changes in patients include nonsense, missense, and frameshift mutations. Here we describe an 18 month old patient with hypoplastic left heart syndrome (HLHS), hydrocephalus, and global developmental delays. He is the second child born to a G3P1012 mother at 39 weeks and 2 days of gestation. He was born via unremarkable vaginal delivery. Apgars were 8 and 9 at 1 and 5 minutes. Birthweight was 4.43 kg (98 %) with length 49 cm (28%) and head circumference 34 cm (32%). He later developed hydrocephalus requiring shunt placement and feeding issues requiring G-tube placement; he also has a tracheostomy. Other medical comorbidities include t-cell lymphopenia, hypgammablobulinemia, hyperopia, and astigmatism. On physical exam, he has distinctive features including broad, depressed nasal bridge, upturned nasal tip, wide set eyes, full cheeks, high forehead, and macrocephaly. He is social and interactive. This patient was initially evaluated by genetics at 2 days of life for HLHS. A routine chromosome analysis and microarray showed normal, 46, XY. At seven months of age, the Invitae Isolated and Syndromic Congenital Heart Defect was drawn and also returned negative. Ultimately, whole exome analysis (WES) was performed, showing likely pathogenic change to HIVEP2 (MRD 43; OMIM #616977). The change to HIVEP2: c.1799C>A (p.Ser600*) is anticipated to cause premature protein termination. This patient had no other pathogenic changes on WES, indicating that mutations in HIVEP2 most likely contributed to his significant cardiac anomalies. Further research is necessary to understand the role of HIVEP2 as a potential cause of HLHS.

2622F

***EIF2AK4* mutations are not likely a common genetic modifier of disease in BMPR2 mutation positive pulmonary arterial hypertension patients.** *K. Sumner[1], Y. Ha[1], J. Sanders[1], A. Morris[2], E. Paul[2], D.H. Best[1,3,4].* 1) R&D Molecular Genetics, ARUP Laboratories, Salt Lake City, UT; 2) Genomics, ARUP Laboratories, Salt Lake City, UT; 3) Dept. of Pathology, University of Utah School of Medicine, Salt Lake City, UT; 4) Dept. of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.

   Background: Pulmonary arterial hypertension (PAH) is a rare, but lethal disorder most commonly caused by autosomal dominant (with reduced penetrance) mutations in the *BMPR2* gene. PAH can also result from mutations in the *ACVRL1, ENG, CAV1, KCNK3,* and *EIF2AK4* genes. A second-hit model has been theorized to account for the reduced penetrance seen with *BMPR2* mutation positive patients. A recent study by Eichstaedt, et al. reported a mutation in *BMPR2* that showed reduced penetrance in a family of ten. Interestingly, a second mutation in the *EIF2AK4* gene was also identified and was found to segregate with disease. The authors of this study suggested that *EIF2AK4* mutations were likely to be a genetic modifier in *BMPR2* mutation positive patients. Based on this study we decided to sequence the *EIF2AK4* gene in a cohort of *BMPR2* mutation positive patients. **Method:** PAH patients with a known *BMPR2* mutation (n=14) were evaluated by Sanger sequencing for mutations in the coding regions and intron/exon boundaries of the *EIF2AK4* gene. **Results:** None of the 14 PAH patients with known *BMPR2* mutations were found to have mutations in the *EIF2AK4* gene. **Conclusion:** The results of our study suggest that *EIF2AK4* mutations are not likely to be a common genetic modifier in patients with *BMPR2*-related PAH.

2623W

Loss-of-function variant in *FNDC3B* is associated with dominant pulmonary arterial hypertension in a pedigree. *M. Cousin[1], N. Boczek[2], P. Byers[3], D. Riegert-Johnson[4], E. Klee[1,2,5].* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA; 3) Department of Pathology, Collagen Diagnostic Laboratory, University of Washington, Seattle, WA 98195-7655, USA; Department of Medicine (Medical Genetics), University of Washington, Seattle, WA 98195, USA; 4) The Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL 32224, USA; Center for Individualized Medicine, Mayo Clinic, Jacksonville, FL 32224, USA; 5) Department of Clinic Genomics, Mayo Clinic, Rochester, MN 55905, USA.

   A 47-year-old woman presented with pulmonary arterial hypertension (PAH), dyspnea, lower extremity edema, arthralgia with gait disturbance, severe osteoarthritic changes in the spine and severe scoliosis, Raynaud's phenomenon, joint laxity, and periodontal disease requiring removal of all but two teeth by age 28. She has a strong family history of PAH, scoliosis, periodontal disease, and joint laxity. *C1S* and *C1R* alterations were ruled out by gene sequencing. Clinical WES on the proband and her daughter, sister, and maternal aunt revealed a heterozygous loss-of-function (LOF) variant in *FNDC3B* (c.1154delC, p.P385Lfs*11) in all four affected individuals. The daughter had not manifested PAH and retained teeth at 26 years old. *FNDC3B*, the Fibronectin Type III Domain Containing 3B gene, also known as factor for adipocyte differentiation 104 (FAD104), encodes a protein with a single transmembrane domain and nine fibronectin type-III domains. It has been seen deleted in two children with mild facial dysmorphisms and one of the two reported children had respiratory distress syndrome. *FNDC3*B has also been associated with central corneal thickness and intraocular pressure phenotypes in GWAS of glaucoma, but has not yet been associated with a single gene disorder. *FNDC3B* is expressed broadly, but highest in adipose, arterial, and lung tissues and is intolerant to LOF variation (pLI = 1.00). *FNDC3B* has reported roles regulating adipogenesis, osteogenesis, and lung development with *fdnc3b[-/-]* mice unviable after birth due to cyanosis-associated lung dysplasia with absent gas exchange. Surfactant-releasing cells in the lungs of these mice were immature with decreased surfactant-associated protein expression. FNDC3B has been shown to bind to and down regulate SMAD1/5/8 phosphorylation and is a negative regulator of BMP/SMAD signaling in a study of calvarial bone formation and osteoblast differentiation. Pathogenic alterations in *BMPR2*, which signals through SMAD1/5/8, are a known cause of FPAH. Consistently, FNDC3B has been shown to activate the TGF-β pathway in cancer studies. Thus, the LOF *FNDC3B* variant segregating with disease in this family lies in a gene shown to be critical for normal lung development and implicated in pathways associated with PAH and connective tissue dysfunction. *FNDC3B* is a strong candidate gene that may explain the dominant PAH in this family.

2624T

Mutation spectrum of the *KCNQ1*, *KCNH2*, and *SCN5* genes for the long QT syndrome in Korea. *M. Kim[1], S. Kim[2], E. Bae[3], J. Noh[3], S. Cho[1], M. Seong[1], S. Park[1].* 1) Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of Laboratory Medicine, National Medical Center, Seoul, Korea; 3) Department of Pediatrics, Seoul National University Hospital, Seoul, Korea.

   Long QT syndrome (LQTS) is an inherited cardiac disease with multiple causative genes and characterized by QT interval prolongation, arrhythmia, and sudden cardiac death. Among Koreans, both the prevalence of symptomatic patients and apparently healthy but with abnormal electrocardiogram subjects are at least as high as in other ethnicities; however, their genetic mutation spectrum has not yet been elucidated comprehensively. We aimed to characterize the mutation spectrum and characteristics in Korean LQTS patients through analyzing three causative ion-channel-encoding genes. For 57 unrelated LQTS probands and their 67 family members, mutations in the genes *KCNQ1*, *KCNH2*, and *SCN5A* were identified by Sanger sequencing and gene dosage analyses using multiplex ligation-dependent probe amplification. This study revealed that 45.6% of probands (26 patients) harbored causative mutations in one of the LQTS genes: *KCNQ1* (53.8%, 14 probands), *KCNH2* (23.1%, 6 probands), *SCN5A* (23.1%, 6 probands). The *KCNQ1* exon deletions were frequent (28.6% of *KCNQ1* mutations, 4 probands). The *KCNH2* mutations were clustered in exon 7 (66.7%). The *SCN5A* mutations were much frequent than in previous reports. De novo mutation rate was 11.5%. This study indicated that the *KCNQ1* is the most important causative gene among Koreans, and mutations were heterogeneous in three genes. Mutation was not identified in approximately half of the probands, therefore other multi-gene panel tests may be helpful.

2625F

Whole-exome sequencing identified a *de novo PDE3A* mutation causing autosomal dominant hypertension with brachydactyly. *D. Wang[1,2], Z. Li[3,4], C. Li[3,4], J. Liu[1,2], D-W. Wang[3,4].* 1) The Center for Clinical Reproductive Medicine, Jiangsu Province Hospital, Nanjing, Jiangsu, China; 2) State Key Laboratory of Reproductive Medicine, Jiangsu Province Hospital, Nanjing, Jiangsu, China; 3) Division of Cardiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China; 4) Genetic Diagnosis Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.

   Purpose: Autosomal dominant hypertension with brachydactyly type-E syndrome caused by mutations in *PDE3A* gene was firstly reported in 2015. To date, there are only two reports of this kind of hypertension. Other patients are still lack of genetic diagnosis. **Methods:** Whole-exome sequencing was performed in an 18-year-old female with clinical diagnosis of Hypertension with brachydactyly syndrome (HTNB). After bioinformatics analysis and healthy control database filtering, potential disease-causing variants in genes associated with HTNB were validated using Sanger sequencing in the proband and her healthy parents. To make sure the mutation is not a rare single nucleotide polymorphism, we also re-sequenced it in 800 ethnically and geographically matched healthy controls by Sanger sequencing. **Results:** We identified a novel *de novo* heterozygous missense mutation c.1346G>A in exon 4 of the *PDE3A* gene. This mutation was carried by the patient and was absent in her healthy parents. The mutation, resulting in p.Gly449Asp, was located in a highly conserved cluster region of reported *PDE3A* mutations. Importantly, this mutation was predicted to affect the protein function by both SIFT (score=0) and Polyphen-2 (score=1). **Conclusion:** Our study presents a third report linking a novel *de novo PDE3A* mutation to autosomal dominant hypertension with brachydactyly type-E syndrome.

2626W

A novel workflow for analysis of whole genome sequencing in cardiac disease using tissue-specific biological datasets. *S.M. Hosseini[1], R. Manshaei[1,2], E. Liston[1,2], R.R. Chaturvedi[3,4], R.H. Kim[1,5,6], S. Bowdin[1,2,4].* 1) Ted Rogers Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, ON, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 3) Division of Cardiology, The Hospital for Sick Children, Toronto, Canada; 4) Department of Paediatrics, University of Toronto, Toronto, Canada; 5) Fred A Litwin Family Centre in Genetic Medicine, University Health Network, Toronto, Canada; 6) Department of Medicine, University of Toronto, Toronto, Canada.

Next-generation sequencing (NGS) technology has greatly enhanced our ability to investigate the genetic basis of mono- and oligo- -genic diseases. Yet, interpreting a multitude of variants is a major challenge. Available guidelines rely heavily on variant level evidence to screen variants and using gene level biological data to boost diagnostic yield has not been fully examined. We developed a computational framework, based on the MUGQIC (McGill University and Génome Québec Innovation Centre) NGS analysis pipeline, by integrating cardiac specific annotations. This set of annotations takes advantage of functional, cross-species and tissue evidence to prioritize genes before genome interpretation. We also used Gpsy, a cross-species gene prioritization bioinformatics tool (http://gpsy.genouest.org/). Ten datasets encompassing molecular interaction, animal model, expression, proteome and disease-specific data were selected for annotation through extensive literature review. These annotations fall into five different categories: functional, cross-species, tissue, disease expression signatures and predictive. For each gene, gaining support from two evidence categories was considered a positive prediction. To investigate the performance of our strategy, we compared 26 established genes for hypertrophic and dilated cardiomyopathy (positive set) against 34 retinitis pigmentosa genes (negative set). Our strategy correctly predicted 22 cardiomyopathy genes and ruled out 29 of the negative set (sensitivity = 0.85, p=5.3E-4; specificity = 0.85, p=3.9E-5) showing significant advantage over an organ blind framework (diagnostic OR = 31.9). Functional and tissue datasets were the categories with highest information content (TPR 0.65 & 0.77 vs. FPR 0.12 & 0.09 respectively) with the least contribution from disease expression signatures (TPR 0.077 vs FPR 0.029). Gpsy had a mediocre performance when using the top 100 ranking genes for heart development (TPR 0.35 vs FPR 0). These values remained robust to other negative gene sets. Our approach for prioritizing variants leverages the wealth of biological data to facilitate downstream genome interpretation. Efforts are underway to further quantify our framework and to apply our strategy to other diseases predominantly affecting a specific organ system.

2627T

Novel PIEZO1 mutations in patients with autosomal recessive hereditary lymphedema. *L. Lai[1,2], M.H. Witte[3], M. Bernas[3], M.F. Hammer[4], R.P. Erickson[5].* 1) Laboratory Medicine, UCSF, San Francisco, CA; 2) Chemistry and Biochemistry, U of Arizona, Tucson, AZ; 3) Surgery, U of Arizona, Tucson, AZ; 4) Neuroscience and Anthropology, U of Arizona, Tucson, AZ; 5) Pediatrics, U of Arizona, Tucson, AZ.

Mutations in PIEZO1, a mechanically activated ion channel, cause an autosomal dominant dehydrated hereditary stomatocytosis and autosomal recessive hereditary lymphedema. We report the finding of novel PIEZO1 mutations in a multi-generation consanguineous family with many affected subjects in each generation. Peripheral blood DNA samples from male subject A and two of his affected first cousins (B, C) were subjected to whole exome sequencing using the Agilent SureSelect Exon V6 target enrichment kit followed by paired-end sequencing on Illumina HiSeq 2500. Only variants shared by all three cousins were further investigated. We identified compound heterozygous variants, a missense variant c.7316G>A;p.Gly2439Asp and a nonsense variant c.3373C>T, in PIEZO1 gene in subjects A and B, and homozygous variant c.7316G>A in subject C. Neither variants have been listed in the publicly available human genomics databases 1000 Genomes, Exome Sequencing Project or the Exome Aggregation Consortium. Family segregation analyses using Sanger sequencing confirmed that subject A inherited variant c.7316G>A from his affected mother and variant c.3373C>T from unaffected father, and that neither of the identified variants were present in his unaffected sister. Furthermore, the computational programs SIFT, PolyPhen2, MutationTaster, MutationAssessor, and FATHMM MKL Coding, all predict variant 7316G>A to be pathogenic. Further study of the variant c.7316G>A found that it is the last nucleotide of the exon 50 (total 51 exons), and predicted to cause loss of the canonical splice donor site and creation of alternative splice sites, resulting in loss of function, either from gain of stop codon or frame shift. RNA sequencing to verify the result of this putative alternation in splice sites is ongoing. Additionally, subject A, but not B or C, also carries a missense variant c.1331A>G;p.Gln444Arg in FOXC2, previously reported in a child with lymphedema, varicose veins, and aortic valve stenosis. This variant was found in subject A's affected mother, but not in unaffected father, brother, sister, or affected brother. Whether the mother's lymphedema is associated with this FOXC2 variant requires further investigation. In summary, the identification of two novel pathogenic variants in PIEZO1 responsible for autosomal recessive hereditary lymphedema further confirms the role of channelopathy in the pathogenesis of lymphedema.

2628F

Exome sequencing in children with pulmonary arterial hypertension demonstrates a different genetic architecture of disease compared to adults. *C.L. Welch[1], N. Zhu[1,5], C. Gonzaga-Jauregui[2], L. Ma[1], H. Qi[4,5], A.K. King[2], U. Kishnan[1], E.B. Rosenzweig[1,6], D. Ivy[8], E.D. Austin[9], W.C. Nichols[10], A. Sawle[3], J.G. Reid[2], J.D. Overton[2], Y. Shen[3,5,6], A.R. Shuldiner[2], A. Baras[2], F. Dewey[2], W.K. Chung[1,3,7].* 1) Department of Pediatrics, Columbia University, New York, NY, USA; 2) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, New York, USA; 3) Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, New York, USA; 4) Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, USA; 5) Departments of Systems Biology, Columbia University, New York, New York, USA; 6) Departments of Biomedical Informatics, Columbia University, New York, New York, USA; 7) Department of Medicine, Columbia University Medical Center New York, New York, USA; 8) Children's Hospital Colorado, Department of Pediatric Cardiology, Denver, Colorado, USA; 9) 9Division of Allergy, Immunology and Pulmonary Medicine Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; 10) Division of Human Genetics, Cincinnati Children's Hospital Medical Center and the Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA.

Pulmonary arterial hypertension (PAH) is a rare disease characterized by distinctive changes in pulmonary arterioles that lead to progressive elevation of pulmonary arterial pressure, right-sided heart failure, and a high mortality rate.   To study the genetics of PAH and compare the differences between children and adults, we selected 134 pediatric-onset and 289 adult-onset PAH patients. Among pediatric-onset patients, 28 had familial PAH (FPAH) and 106 had sporadic, idiopathic PAH (IPAH). Among adult-onset patients, 85 had FPAH and 204 had IPAH. FPAH cases were initially evaluated by Sanger sequencing of *BMPR2* and *ACVRL1*, and *BMPR2/ACVRL1* mutation-negative FPAH cases and all IPAH cases were evaluated by exome sequencing.   We found a similar frequency of *BMPR2* mutations in pediatric and adult PAH patients: 50% in FPAH and 8-11% in IPAH patients. However, pediatric-onset patients were significantly more likely to carry mutations in *TBX4* than adult-onset IPAH patients (11/106, 10% versus 2/204, 1%; odds ratio=11.6; p-value=0.0002). About 5-14% of pediatric and 2-12% of adult PAH patients had mutations in other known risk genes (*ACVLR1*, *BMPR1A*, *BMPR1B*, *CAV1*, *EIF2AK4, ENG, KCNK3, SMAD4 and SMAD9*). In 34 pediatric IPAH patients without mutations in known risk genes, we observed a marginally significant enrichment of *de novo* likely gene damaging (LGD) and deleterious missense (predicted by MetaSVM) variants in genes that are highly expressed in the developing lung and heart (enrichment=4.2x, p-value=1.6E-03; based on background mutation rate). In summary, mutations in all known PAH genes accounted for 64-70% of FPAH (both pediatric-onset and adult-onset), 26% of pediatric-onset IPAH, and 11% of adult-onset IPAH. Notably, in pediatric patients without mutations in known risk genes, there was evidence for the existence of novel IPAH genes due to *de novo* LGD variants, suggesting the need for additional studies with expanded pediatric IPAH cohorts to enable identification of novel genetic causes and to provide insights into the pathophysiology and potential new targets for treatment for PAH.

2629W

CRISPR-Cas9 mediated knockout of SEL1L and proteasomal inhibition reveal divergent degradation pathways for corresponding LDLR and VLDLR disease-causing mutants. *B.R. Ali, P. Kizhakkedath, A. John, L. Al-Gazali.* College of Medicine and Health Sciences, United Arab Emirates University, Al-Ain, Al-Ain, United Arab Emirates.

Terminally misfolded proteins in the early secretory pathway are mainly degraded by the ubiquitin-proteasome systems known as ER associated degradation (ERAD) which has been implicated in the pathogenesis of numerous congenital disorders. Therefore, ERAD has been a promising target for therapy of such diseases. In this study we compare and contrast the degradation behavior of equivalent disease-causing mutations in LDLR and VLDLR, two LDLR family receptors. HRD1-SEL1L knockout (KO) HEK293 cell lines have been generated using CRISPR/Cas9. The missense mutants were generated by QuikChange site-directed mutagenesis. The expressed proteins in HeLa and HEK293 cells were analyzed biochemically and by microscopy. We found that the disequilibrium syndrome-causing VLDLR missense mutants (p.D487Y, p.D521H and p.C706F) are retained in the ER whereas only p.D482H and p.C667F of the LDLR corresponding mutants are retained. Two of the ER-retained VLDLR mutants were found to be aggregation-prone leading to ER stress as measured by spliced-XBP mRNA levels. However, there was no differences between LDLR wild type and mutants with regard to their aggregation status or ER stress. Treatment with inhibitors of autophagy seemed to stabilize the degradation of wild type receptors. In addition, p.C706F VLDLR mutant stabilized during early phase of both proteasomal and autophagy inhibition. Cycloheximide chase analysis in SEL1L-KO cells revealed that the HRD1-SEL1L complex could be involved in the clearance of ER-retained VLDLR mutants. The downstream degradation and ER stress induction behavior of ER-retained corresponding LDLR and VLDLR mutants are divergent suggesting different clearance mechanisms despite their structural similarities.

2630T

Thoracic aortic disease outcomes in 987 cases with pathogenic variants in *ACTA2*, *PRKG1, TGFBR1, TGFBR2* and *SMAD3* ascertained by the Montalcino Aortic Consortium. *D. Milewicz[1], J. De Backer[2], C. Boileau[3,10,11], S. LeMaire[4], R. Jeremy[5], R. Pyeritz[6], S. Morris[7], A. Braverman[8], A. Evangelista[9], G. Jondeau[3,10], E. Regalado[1], Montalcino Aortic Consortium.* 1) University of Texas Health Science Center at Houston, Houston, TX, USA; 2) University Hospital Ghent, Ghent, Belgium; 3) LVTS, INSERM U1148, Université Paris Diderot, Hôpital Bichat, Paris, France; 4) Baylor College of Medicine, Houston, TX, USA; 5) Royal Prince Alfred Hospital, New South Wales, Australia; 6) Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA; 7) Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA; 8) Washington University School of Medicine, St. Louis, MO, USA; 9) Hospital Universitari Vall d´Hebron, Barcelona, Spain; 10) Centre de Référence Maladies Rares Syndrome de Marfan et pathologies apparentées, Service de Cardiologie, Assistance Publique-Hôpitaux de Paris, Hôpital Bichat, Paris, France; 11) Département de Génétique, Assistance Publique-Hôpitaux de Paris, Hôpital Bichat, Paris, France.

The Montalcino Aortic Consortium was established to characterize the phenotype associated with novel heritable thoracic aortic disease (HTAD) genes. Thoracic aortic disease presentations and outcomes in the initial 987 cases from North America, Europe, Asia, and Australia with pathogenic variants in *ACTA2* (n=319), *TGFBR2* (n=265), *SMAD3* (n=190), *TGFBR1* (n=176) and *PRKG1* (n=37) were analyzed. Average age at first aortic event, defined as aortic aneurysm repair or aortic dissection, was significantly older among *SMAD3* mutation carriers (mean 46 years, standard deviation (SD) 11.5) when compared to *ACTA2* (mean 34, SD 13.9; p<0.0001), *PRKG1* (mean 34, SD 12.4; p<0.0001), *TGFBR1* (mean 32, SD 14.3; p<0.0001), and *TGFBR2* (mean 30, SD 16.9; p<0.0001) mutation carriers. For *TGFBR1* and *TGFBR2* mutation carriers, these aortic events were more frequently elective repair of aortic aneurysms, whereas for *SMAD3*, *ACTA2* and *PRKG1*, the vast majority of presenting aortic events were aortic dissections. Overall cumulative risk of an aortic event was lowest among *SMAD3* mutation carriers compared to *ACTA2* (p=0.003), *PRKG1* (p<0.0001), *TGFBR1* (p=0.0007), and *TGFBR2* (p<0.0001). Cases with the only variant in *PRKG1* that causes aortic disease (p.R177Q) have the highest cumulative risk of aortic events, but childhood onset of aortic events (<13 years) only occurred in cases with *TGFBR2* (n=18), *TGFBR1* (n=5) and *ACTA2* (n=6) mutations, including recurrent mutations disrupting the ACTA2 arginine 179, TGFBR2 aspartic acid 446 and arginine 528 residues. Gender influenced the risk of aortic event only among cases with *TGFBR1* (p<0.0001) and *ACTA2* (p=0.02) mutations with higher risk of aortic event in men than women. Additionally, risk of aortic events was found to be higher among cases recruited in North America compared to cases recruited in Europe (p=0.002). These findings indicate thoracic aortic disease phenotype varies with the underlying HTAD gene and often the specific mutation in the gene, gender and site of recruitment. Thus, these data support gene- and mutation-specific management of thoracic aortic disease of HTAD cases and argue against grouping these patients together under one diagnosis, as well as provide baseline data to determine if the proposed management of these patients improves outcomes.

2631F

Exome sequencing of 103 Williams syndrome cases rules out variation in the remaining elastin allele as a major contributor to variance in blood pressure and arterial stenosis. *P.C.R. Parrish[1], M. Lugo[1], K. Shields[2], Y.P. Fu[1], R.H. Knutsen[1], B.A. Kozel[1].* 1) National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD; 2) Department of Pediatrics, Washington University School of Medicine, St. Louis, MO.

Elastin (*ELN*) is an extracellular matrix protein that allows tissues to stretch and recoil. Diseases of elastin insufficiency include supravalvular aortic stenosis (SVAS), which is caused by mutations in the *ELN* gene, and Williams syndrome (WS), which is caused by the deletion of 26-28 genes including *ELN* at the 7q11.23 locus. Patients with these conditions have stiff and narrow arteries, in addition to focal stenosis and hypertension. As in many diseases of haploinsufficiency, phenotype severity varies widely among affected individuals. A possible source of this variation is changes in the remaining *ELN* allele which impact the residual quantity and quality of elastin deposited. To assess the effect of quantitative differences in elastin on vascular phenotype, 80 C57Bl/6 *Eln*[+/-] × DBA/2 WT F2 intercross mice underwent vascular phenotyping. Aortic desmosine, an elastin-specific crosslink, was then measured and correlated with blood pressure, aortic diameter and ascending aortic length. For human genetic studies, vascular phenotypes were obtained from a review of medical records from 103 WS subjects. Whole-exome sequencing was performed and common and rare *ELN* variants were evaluated for association with SVAS and hypertension. Within *Eln* genotypes, no correlation was identified between desmosine concentration and mouse aortic diameter, aortic length and systolic blood pressure (all R^2 ≤0.06). Similarly, in human studies, Fisher's exact tests revealed no association between *ELN* variants and SVAS or hypertension. Burden testing on rare exonic variants (n=11 for MAF <10%, n=5 for MAF <5%) revealed no correlation with SVAS (MAF <10%, p=0.5 and MAF <5%, p=0.4) or hypertension (MAF <10%, p=0.5 and MAF <5%, p=0.6). Common *ELN* variants (MAF≥10%) were analyzed by single-variant testing for all intronic (n=7) and exonic (n=1) variants. Neither type of variation was significantly associated with SVAS (intronic alleles, p>0.1 for all and exonic, p=0.3) or hypertension (intronic, p>0.7 and exonic, p=0.4). With more than 100 rare disease patients studied, this is the largest WS modifier study to date. The results show that genetic variation in the remaining elastin allele is unlikely to play a large role in modifying vascular disease severity, although more modest effects may be detected by increasing sample size. Future modifier work will focus genes and pathways outside the WS locus.

2632W

Treatment of RIT1-associated cardiomyopathy with trametinib: Initial results in two patients. *G. Andelfinger[1], C. Marquis[2], M.J. Raboisson[2], A. Karalis[3], S. Waldmüller[4], G. Wiegand[5], B.D. Gelb[6], M. Zenker[7], M. Hofbeck[5], M.A. Delrue[3].* 1) Cardiovascular Genetics, CHU Sainte Justine, Department of Pediatrics, Université de Montréal, Montréal, QC, Canada; 2) Service of Cardiology, CHU Sainte Justine, Department of Pediatrics, Université de Montréal, Montréal, QC, Canada; 3) Service of Genetics, CHU Sainte Justine, Department of Pediatrics, Université de Montréal, Montréal, QC, Canada; 4) Institute of Medical Genetics, University Hospital of Tübingen, Germany; 5) Department of Pediatric Cardiology, Pulmonology and Paediatric Intensive Care Medicine, University Children's Hospital Tübingen, Germany; 6) Mindich Child Health and Development Institute and Departments of Pediatrics and Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Institute of Human Genetics, University Hospital Magdeburg, Germany.

RASopathies are associated with substantial cardiovascular morbidity, mainly due to the occurrence of hypertrophic cardiomyopathy (CMP). CMP with biventricular obstruction carries a particularly severe prognosis, and therapeutic options are limited. In animal models, MEK/ERK inhibition alleviates RASopathy-associated phenotypes. Here, we report the treatment of two infants with severe obstructive CMP caused by *RIT1* mutations with trametinib. The first patient (P1), diagnosed prenatally with semilunar valve thickening, developed rapidly progressive CMP after birth despite high-dose propranolol treatment (8 mg/kg/d). The second patient (P2) was diagnosed in the neonatal period with pulmonary stenosis and hypertrophic CMP. Distinct de novo RIT1 mutations were identified in each patient: S35T and F82L, respectively. CMP progressed rapidly, resulting in severe bilateral outflow obstruction with consecutive pulmonary edema despite propranolol treatment (9 mg/kg/d). P2 was resuscitated and required mechanical ventilation at the age of 12 weeks. Having exhausted conventional treatment for both patients, we used pharmacological MEK inhibition as last resort after approval of local ethics boards for off-label use and obtaining informed consent. We started trametinib at 14 and 13 weeks of age respectively, at a dose of 0.025 mg/kg/d, and monitored side effects. After three months of treatment, we observed dramatic improvement of cardiac status. In P1, outflow tract obstruction decreased (RVOT, peak 70 mmHg to 42 mmHg; LVOT, peak/mean 71/23 to 37/15 mmHg), and progression of hypertrophy ceased. In P2, hypertrophy and pulmonary edema regressed, allowing extubation at 18 weeks of age. Side effects were limited to accentuation of physiological alopecia in P1 and a persistent acneiform eruption in P2, which was treated topically. In summary, trametinib proved to be an effective, well-tolerated treatment for *RIT1*-associated progressive CMP in two selected patients over a short time period. This preliminary experience raises important questions for the treatment of similar cases, in particular with respect to long-term effects, long-term side effects and dosing regimens. MEK inhibition may prove most effective during a fixed time window before the onset of irreversible cardiac remodeling in RASopathies, including those caused by genes other than *RIT1*. In addition, future studies will also need to address the effect of MEK inhibition on other organ systems.

2633T

Genetic architecture of pulmonary arterial hypertension with congenital heart disease. *N. Zhu[1,4], C. Gonzaga-Jauregui[2], C. Welch[1], L. Ma[1], E.B. Rosenzweig[1,5], U. Kishnan[1], A.K. King[2], J.G. Reid[2], J.D. Overton[2], A.R. Shuldiner[2], A.B. Baras[2], F.D. Dewey[2], Y. Shen[3,4,5], W.K. Chung[1,3,6].* 1) Department of Pediatrics , Columbia University , New York, NY; 2) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 3) Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY; 4) Departments of Systems Biology, Columbia University, New York, NY; 5) Departments of Biomedical Informatics, Columbia University, New York, NY; 6) Department of Medicine, Columbia University Medical Center, New York, NY.

Among group I pulmonary arterial hypertension (PAH), approximately 40% of children and 10% of adults have congenital heart disease (CHD). Left-to-right (systemic-to-pulmonary) shunts increase pulmonary blood flow and the risk of PAH, but not all PAH-CHD is associated with prolonged increased pulmonary blood flow, and the PAH may be due to another cause. Missense variants in *BMPR2*, a major gene contributing to familial and idiopathic PAH, have been identified in 5-7% of PAH-CHD patients, suggesting that the majority of cases are not due to *BMPR2* mutations.   To identify novel genetic causes of PAH-CHD, we performed reflexive *BMPR2* sequencing and MLPA, followed by exome sequencing in a total of 270 PAH-CHD patients, including both parents in 52 patients without mutations in established PAH genes. Twenty-eight patients (10.3%) carried likely pathogenic variants in established PAH-risk genes, including *TBX4* (11) and *BMPR2* (11). We estimated the contribution of *de novo* variants in novel candidate genes. Based upon the background mutation rate, we identified a significant enrichment of missense *de novo* variants in PAH-CHD cases in genes highly expressed in the developing lung (p-value=0.02, enrichment =1.71), explaining ~14% of cases.   Additionally, we assessed the extremeness of deleteriousness of rare variants in each patient using population sampling probability (PSAP). Using the PSAP-score based association test, we identified a new candidate risk gene *SOX17*, a transcription regulator involved in Wnt signaling and required for normal looping of the embryonic heart tube. *SOX17* is intolerant of deleterious mutations (ExAC pLI = 0.87, misZ = 3.25). Eight PAH-CHD patients (p-value<1e-6) with either a patent ductus arteriosus or secundum atrial septal defect carry 5 predicted loss of function (LoF) and 3 novel missense that are predicted to be more deleterious than the most deleterious variants in the gene in 99.9% population. Interestingly, 3 patients carried 2 novel missense and 1 LoF variants in a separated cohort of 426 PAH patients without CHD, supporting *SOX17*'s role in PAH.   In summary, we characterized the genetic architecture of PAH-CHD and found there is a lower contribution from known PAH risk genes than in patients without CHD, with 5% associated with *TBX4* mutations and a potentially substantial contribution from *de novo* missense variants. Finally, we identified a novel candidate PAH-CHD risk gene *SOX17* harboring rare, predicted deleterious variants.

2634F

Probing for modifiers of X-inactivation in a cohort of Amish families with hemophilia B. *A. Ozel[1], K. Desch[2], D. Siemeniak[3], M. Heiman[4], A. Shapiro[4], D. Ginsburg[1,3], J. Li[1].* 1) Dept. of Human Genetics, University of Michigan-Ann Arbor, Ann Arbor, MI; 2) Dept. of Pediatrics, University of Michigan-Ann Arbor, Ann Arbor, MI; 3) HHMI, University of Michigan - Ann Arbor, Ann Arbor, MI; 4) Indiana Hemophilia and Thrombosis Center, Inc., Indianapolis, IN.

To balance gene dosage in mammals between males and females, one of the X chromosomes is randomly inactivated early in female embryonic development in all cells except oocytes, with gene expression primarily restricted to genes on the active X. On average, each X-chromosome will be inactivated in ~1/2 of somatic cells in a given female. The presence of preferential or non-random inactivation of the X chromosome, a.k.a. X-inactivation skewing, in a female carrier for an X-linked recessive disease, can lead to deviation from the 50/50 distribution of cells expressing the normal and mutant chromosomes. This skewing could be by chance or directed by genetic factors. Coagulation Factor IX (Xq27.1-q27.2) is produced by liver hepatocytes. Therefore, plasma Factor IX levels should be an excellent proxy for skewing of X-inactivation in carriers of severe hemophilia B, an X-linked recessive disorder. We measured plasma Factor IX activity levels in 166 female carriers from three large Amish pedigrees with severe Hemophilia B. Factor IX activity in carriers ranged from 14-142%. We used deviation from 50% Factor IX activity levels as a quantitative trait in linkage and association testing. Genotyping was done using the Human Omni Express-24 BeadChip and Factor IX activity were measured using a chromogenic assay. Heritability (h^2) of the deviation in activity levels from 50% was estimated as 6.6% (using *SOLAR*), similar to the results from *Merlin-Regress* (14.6%). Genome-wide association analysis didn't identify a major locus for X-inactivation skewing, though a suggestive signal was detected at the *DBC1* locus on Chromosome 9 (rs1013282(G), MAF=0.42, p=4.3E-7). This top SNP didn't exhibit a significant cis-eQTL in any of the tissues in the *GTeX* database, but was found to alter 7 regulatory motifs as reported by *HaploReg*. The lowest p-value observed on Chromosome X was at the *FGF13* locus (rs7052771(A), MAF=0.081, p=1.6E-6) with no significant cis-eQTLs in any of the tissues in the *GTeX* database and evidence for alteration of 2 regulatory motifs by *HaploReg*. Linkage analysis on the female carriers in these 3 pedigrees using *Merlin-Regress* and *SOLAR* also failed to detect any significant linkage or association with skewed X inactivation suggesting the absence of a single, specific X-inactivation modifier in these families. These data are consistent with the relatively low heritability estimates and suggest that the majority of skewed X-inactivation is due to non-genetic factors.

2635W

Novel genetic associations for blood pressure identified via gene-alcohol consumption interaction in about 570K individuals. *M.F. Feitosa[1], A.T. Kraja[1], M.A. Province[1], D. Levy[2].* 1) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, Saint Louis, MO; 2) Framingham Heart Study, Framingham, MA and the Population Sciences Branch, National Heart, Lung, and Blood Institute, Bethesda, MD; on behalf of the Gene-Lifestyle Interactions Working Group.

Blood pressure (BP) levels are influenced by alcohol consumption even after adjusting for environmental confounders. Genome-wide association (GWA) of gene-by-alcohol consumption status for BP may identify additional loci and contribute to the understanding of alcohol-BP metabolism. We conducted GWA analyses with joint (2df) test of main genetic (G), alcohol exposure (E) and their interaction (G×E) effects, using 1000G imputed markers on systolic BP, diastolic BP, mean arterial pressure, and pulse pressure. In Stage-1, meta-analyses with double genomic corrections were applied in up to 130,828 individuals, and in each ancestry: African (AA: N=21,417), European (EA: N=91,102), Asian (ASA: N=9,838), Hispanic (HA: N= 4,056) and Brazilian (N= 4,415). 3,713 SNVs representing 126 loci (+/-500kb SNV-distance) that reached suggestive associations (*P*≤ 1E-06) were analyzed in Stage-2 (Replication). Stage-2 included 440,824 individuals from AA (N=5,041), ASA (N=141,026), EA (N=281,380), and HA (N=13,377). We identified 5 novel loci (in EA and in multi-ancestry) and 44 BP-known loci (in multi-ancestry and 37 in EA). Further, we meta-analyzed Stage-1 and Stage-2 combined using 3,713 SNVs (126 loci) and found 21 potential novel loci (p< 5E-08), where 18 of them are in AA, 2 in EA, and 1 in multi-ancestry. In addition, Stages 1 and 2 yielded 20 BP-known loci (18 in EA, 12 in ASA and 4 in AA). Many of our genes belonging to the above loci (e.g., *PINX1, GATA4, FTO, BLK, GABBR2, CHAT,* and *BAG3*) have been linked to alcohol dependence. Some genes belonging to these loci (*KCNMB2, BAG3, SGK1,* and *CDH17*) have also been reported as suggestive in associations with BP traits in humans or animal models. Our novel variants interacting with alcohol consumption for all four BP traits on 8p23.1 region (*e.g., PINX1, BLK* and *MIR4286*) show evidence of cis- and trans- regulatory functions. Our study extends evidence from previously reported BP-findings in European ancestry cohorts to other ancestries (*e.g., LSP1-TNNT3* in Asians). Our novel and known BP loci are enriched for several biological terms and pathways that capture important information for BP and cardiovascular disease risk factors, alcohol dependence, personality, neurodegenerative and neuropsychiatric disorders. These findings may provide insights into mechanisms by which alcohol consumption impacts BP and hypertension risk. *This work was supported in part by the NHLBI grant HL118305.*.

2636T

Genome-wide association study of cardiotoxicity and cardiac gene expression in NCCTG N9831 (Alliance) Adjuvant Trastuzumab Trial. *D.J. Serie[1], J.E. Crook[1], B.M. Necela[2], T.J. Dockter[3], G. Colon-Otero[4], E.A. Perez[4], E.A. Thompson[2], N. Norton[2].* 1) Health Sciences Research, Mayo Clinic, Jacksonville, FL; 2) Department of Cancer Biology, Mayo Clinic, Jacksonville, FL; 3) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 4) Department of Hematology/Oncology, Mayo Clinic, Jacksonville, FL.

The major clinical side-effect of the *ERBB2* targeted breast cancer therapy, trastuzumab, is decline in left ventricular ejection fraction (LVEF). Improved methods are required to better identify patients susceptible to cardiotoxicity. NCCTG N9831 trial compared adjuvant doxorubicin and cyclophosphamide followed by either weekly paclitaxel (Arm A); paclitaxel then trastuzumab (Arm B); or concurrent paclitaxel and trastuzumab (Arm C) in patients with HER2-positive breast cancer. GWAS was performed across 618,863 SNPs and 1,191 patients. We used linear regression to identify SNPs associated with decline in LVEF, adjusting for age, baseline LVEF, and anti-hypertensive medications. We used Predixcan to perform association testing of decline in LVEF with the estimated heritable gene expression levels in left ventricular tissue of patients in N9831, using genotype and expression data from 159 individuals in the GTEX database as a training set. SNPs at six loci were associated with decline in LVEF, ($p=7.73 \times 10^{-6}$ to 8.93×10^{-8}): *LDB2*, *BRINP1*, chr6 intergenic, *RAB22A*, *TRPC6* and *LINC01060*, in patients who received combination chemotherapy and trastuzumab (Arms BC, N=800). We replicated association, p<0.05, with SNPs previously associated with anthracycline cardiotoxicity at *CBR3* and *ABCB1*. Predixcan estimated heritable levels of cardiac gene expression at 4,853 genes, of which, 228 were associated with decline in LVEF, p<0.05. Estimated *TRPC6* and *CBR3* expression levels were associated with decline in LVEF, p=0.005 and 0.08 respectively. Our study identified six putative novel cardiotoxicity loci in patients treated with combination chemotherapy and trastuzumab and replicated known associations of anthracycline induced cardiotoxicity. In the case of *TRPC6* and *CBR3*, cardiotoxicity may be related to genetic variants that alter cardiomyocyte gene expression.

2637F

The association of TMPO and RYR1 genes with cardiovascular diseases in a Turkish Cypriot Family. *M.C. Ergoren[1], R. Kalkan[2], B. Turkgenc[3], S.G. Temel[4].* 1) Medical Biology, School of Medicine, Near East University, Nicosia, Cyprus; 2) Faculty of Medicine, Department of Medical Genetics, University of Near East, Nicosia, Cyprus; 3) Genetic Diagnostic Center,University of Acibadem, Istanbul,Turkey; 4) Faculty of Medicine, Department of Histology and Embryology, University of Uludag, Bursa, Turkey.

The protein encoded by Thymopoietin (TMPO) gene may play a role in the assembly of the nuclear lamina, and thus help maintain the structural organisation of the nuclear envelope. Furthermore, Ryanodine receptor is encoded by Ryanodine Receptor 1 (RYR1) gene and functions as a calcium release channel in the sarcoplasmic reticulum but also serves to connect the sarcoplasmic reticulum and transverse tubule. Mutations in TMPO and RYR1 genes are associated with dilated cardiomyopathy and mini core myopathy with external ophthalmoplegia, respectively. The 50 years old male proposiatus was referred to our clinic after having artery blockage of 90% and 99% that was fixed by stent operation. This was followed by by-pass cardiac surgery. He is the middle son of his parents and the family history was significant for sudden death of his father (<40 yrs), and his mother and his younger brother has similar phenotype with him and had by-pass heart surgery in their early ages. TMPO and RYR1 exon sequencing of the probing revealed c.289A>C (p.Lys97Gln) heterozygote missense variation (rs778356103) in exon 2 of TMPO gene and c.1354G>A (p.Glu452Lys) heterozygote missense variation within exon 13, c.6310C>T (p.Arg2104Cys) (rs541757529) heterozygote missense variation in exon 30 and a heterozygote c.9743G>A (p.Arg3248Gln) (rs553187437) missense variation within exon 66 of RYR1 gene. Heterozygote missense variation c.1354G>A (p.Glu452Lys) in RYR1 gene is not identified before, so it is de novo. None of these variants have been reported before in the mutational databases and additionally, no phenotype has been reported with these variants. Moreover, population studies showed that minor allele frequencies of those variants are much more less than 1% and they are located within conserved region between species. To conclude, we found a direct association with three variants that in exon regions of both TMPO and RYR1 genes with cardiovascular diseases in a Turkish Cypriot family.

2638W

Association of a polymorphism in *ITGB3* with resistance to clopidogrel in early acute coronary syndrome in an admixed population from Colombia. *A.V. Valencia-Duarte[1], N. Gallego[1], L. Franco[1,2], C.A. Velarde[1], D.M. Valencia[1], K.J. Betancur[3], I.C. Ortiz[1], H. Cuervo[1,3], N. González[3], C. Rúa[4], C. Saldarriaga[1,3], C.A. Martínez[1,3], N. López[3], L.M. Parra[1], P. Parra[1], E.A. Jiménez[1], S. Carvalho[1], C.I. Vergara[5], J. García[6], G. Bedoya[7].* 1) Faculty of Medicine, Universidad Pontificia Bolivariana, Medellín, Antioquia, Colombia; 2) National Faculty of Public Health, University of Antioquia, Medellín, Antioquia, Colombia; 3) CardioVID Clinic, Medellín, Antioquia, Colombia; 4) Hematology Department, Hospital San Vicente Fundación,, Medellín, Antioquia, Colombia; 5) School of Medicine, Johns Hopkins University, Baltimore, MD.,USA; 6) Faculty of Medicine, University of Antioquia, Medellín, Antioquia, Colombia; 7) Institute of Biology, University of Antioquia, Medellín, Antioquia, Colombia.

   Dual antiplatelet therapy with clopidogrel and aspirin is the standard treatment to prevent ischemic events in patients undergoing percutaneous coronary intervention (PCI), which reduces the rate of recurrent infarctions and thrombotic events. However, the antiaggregant response of clopidogrel varies individually and at population levels due to different factors including genetics. We aimed to evaluate the association between genetic variants in genes related to absorption, metabolic activation and biologic activity of clopidogrel with the response to clopidogrel and recurrent coronary events in an admixed population from Colombia. The study included 212 patients with diagnosis of early acute coronary syndrome, undergoing PCI and with stenosis ≥ 50%. Platelet aggregation was measured using light transmittance aggregometry with adenosine diphosphate as agonist. Outcome variables were clopidogrel resistance, defined as platelet aggregation ≥70% after 600 mg loading dose of clopidogrel (38 resistance/153 non resistance) and risk of major adverse cardiac events (MACE), such as infarction, stent thrombosis or death during 1 year of follow-up (9 events/203 no events). Polymorphisms in the genes *ABCB1* (rs1045642), *CYP2C19* (rs4244285), *CYP3A5* (rs776746), *CYP1A1* (rs1048943) *ITGB3* (rs5918) and *P2RY12* (rs6785930 and rs6809699) were genotyped using RT-PCR. Association analyses were performed using logistic regression, including age, sex, socioeconomic status, cholesterol levels and treatment as covariates in the model. rs5918 SNP in *ITGB3* was associated with clopidogrel resistance (CC/CT adjusted OR: 0.261; 0.071 - 0.961, P-value: 0.043) suggesting that patients carrying the one or two copies of the C allele in rs5918 have lower risk to present clopidogrel resistance. Genetic variants located in *P2RY12* were associated with clopidogrel resistance in the bivariate analysis (AA/AC OR: 2.418; 1.058 - 5.524, P-value: 0.033) but the results were not significant after adjusting by covariates. There was no association between any analyzed polymorphism and MACE. In summary, clopidogrel resistance has a high frequency among the evaluated population and it might be explained by a genetic variant in the *ITGB3* gene. Further analysis are needed to understand the biological mechanisms underlying the observed association.

2639T

Identifying and characterizing causal genes in GWAS-identified loci for heart rate variability using high-throughput, image-based screens in zebrafish larvae. *B. von der Heyde[1], A. Emmanouilidou[1], T. Klingström[1], S. Vicenzi[1], H. Brooke[2], A. Allalou[3], M. den Hoed[1].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Institute of Environmental Medicine, Karolinska Universitet, Stockholm, Sweden; 3) Department of Information Technology, Uppsala Universitet, Uppsala, Sweden.

   Introduction: Heart rate variability (HRV) is associated with an increased risk for cardiovascular disease (CVD), the major cause of death worldwide. A recent meta-analysis of genome-wide association studies (GWAS) in data from 53,172 individuals has identified variants in eight loci that show robust associations with HRV. The causal genes and mechanisms through which these loci influence heart rate variability are still unknown. We aimed to characterize positional candidate genes for HRV using a zebrafish model system. **Methods:** We prioritized 18 candidate genes for functional follow-up based on results from five bioinformatic tools. Twelve of these genes had a total of 16 orthologous genes in zebrafish. We targeted these 16 orthologues in three pools of up to eight targeted genes using a multiplex CRISPR-Cas9 approach in fertilized eggs from fish that transgenically express green fluorescent protein on smooth muscle cells, which enables visualization of the beating heart. After three months, founder mutants were crossed, and heart rate and rhythm were visualized and quantified *in vivo* in the offspring by recording 30 sec videos of the atrium at 2 and 5 days post-fertilisation (dpf, repeated measures). Heart rate and two different measures for HRV were quantified in batch mode using a custom-written MatLab script. CRISPR-Cas9 induced mutations were quantified using next-generation sequencing and data were analyzed using mutually adjusted multiple linear/logistic regression models (additive model). **Results:** We have so far characterized one of three multiplex mutant lines and show that: 1) each additional disrupted allele in rgs6 increases the odds of cardiac arrests at 2dpf by 2.52 (95% CI 1.13-5.61, n=372); 2) each additional disrupted allele in gngt1 is associated with a 0.21±0.09 higher RMSSD (n=303), a 3.49 higher odds of a reduced contractility (95% CI 1.29-9.43, n=326) and an 8.88 higher odds of an abnormal morphology of the heart (95% CI 2.16-36.48, n=297) at 5dpf; and 3) each additional affected allele in kiaa1755 results in a 0.14±0.06 (SD) higher heart rate (n=289). **Conclusion:** Zebrafish larvae facilitate large-scale, objective, image-based genetic screens of heart rate and rhythm, and allowed us to identify previously unanticipated genes that influence heart rate variability and early cardiac development (gngt1); the risk of cardiac arrests (rgs6); and heart rate (kiaa1755).

2640F

Combining Mendelian genetics and genetic epidemiology identifies *APOL3* as a new gene for abdominal aortic aneurysm. *J.M. Albuisson[1,2,3], M.J. Bown[4], R.S. Kiando[2], L. Lecordier[5], S. Uzereau[5], P. Julia[6], N.J Samani[4], N. Bouatia-Naji[2], E. Pays[5], X. Jeunemaitre[1,2,3].* 1) AP-HP, Hôpital Européen Georges Pompidou, Département de Génétique. Service de Médecine Vasculaire et Centre de Référence des Maladies Vasculaires Rares, Paris, France; 2) INSERM, U970, Paris Centre de Recherche Cardiovasculaire-PARCC, Paris, France; 3) Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine, Paris, France; 4) Department of Cardiovascular Sciences, University of Leicester, Leicester LE2 7LX, UK; 5) Laboratoire de Parasitologie Moléculaire, IBMM, Université Libre de Bruxelles, Gosselies, Belgium; 6) Unité de chirurgie vasculaire, Hôpital Européen Georges Pompidou, Paris, France.

Abdominal aortic aneurysm (AAA) is a common disorder with a strong hereditary component and rare fully penetrant, dominant inheritance. However, no gene of Mendelian AAA was identified to date. The power of recently developed strategies comes from combining exome or genome sequencing with new genetic association analysis tools, allowing for the identification of rare variants with stronger effects. We used this strategy and first performed exome sequencing in a large 3-generations pedigree showing dominant inheritance. Five affected cases had exome analysis, completed by segregation analysis in 5 affected and unaffected relatives. Using classical filtering tools and segregation, we identified only one candidate variant, R377W, in *APOL3*, a gene of unknown function. A Z-score of 4,14 (p=2.10^{-5}) at *APOL3* locus confirmed linkage with the condition. Gene-centered association analysis was performed, focused on *APOL3* rare and very rare variants. A large cohort of 2500 sporadic and 500 familial AAA cases was compared to 3000 age and sex-matched controls by sequencing and genotyping. 12 rare variants (MAF<1%) and 7 very rare variants (MAF<1‰) were identified. A significant excess of very rare variants was found in the AAA cases compared to controls (Burden test; p=0.03). Western-blot and immunohistochemistry demonstrated the expression of APOL3 protein in the smooth muscle cells and endothelium of one affected case, and in control samples. Inflammatory and viral stimuli like TNFα and TLR3 strongly induce APOL3 expression in endothelial cells and trigger cell death. APOL3 function is still unknown, but it belongs to a family of 6 proteins regulating apoptosis. We show that APOL3 regulates the cell cytoskeleton in various cell types. The mutant *APOL3* R377W is suggested to induce a dominant–negative effect rather than a loss of function, as a heterozygous nonsense variant is present at very high frequency in the African-American population (EXAC database, rs11089781: MAF=0.22). Preliminary observations after transfection of smooth muscle cells by the R377W *APOL3* mutant shows alteration of the cellular phenotype in terms of survival, shape, and cytosqueleton dynamics. Identification of an *APOL3* mutation indicates that AAA can result from underlying genetic variants that significantly alter the phenotype of vascular smooth muscle cells, through a mechanism distinct from atherosclerosis.

2641W

Is HDL-C causally associated with risk of cardiovascular diseases in Han Chinese? A Mendelian randomization study with 10,000 subjects. *M.I. Biradar[1], K.M. Chiang[2], H.C. Yang[3], W.H. Pan[1,2].* 1) Taiwan International Graduate Program in Molecular Medicine, National Yang-Ming University and Academia Sinica, Taipei, Taiwan; 2) Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan; 3) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

Plasma high density lipoprotein-cholesterol (HDL-C) level is known as a risk factor for cardiovascular disease and is influenced by various genetic & environmental factors. Plasma HDL-C level has been consistently associated with reduced risk of cardiovascular diseases, but whether this association is causal has not been carefully examined in Han Chinese. Considering the fact that genotypes are randomly distributed during meiosis, independent of non-genetic confoundings, and not modified by the disease process; Mendelian randomization can be applied to test whether the HDL-C is causally associated with CVD. Recent GWAS studies have identified several loci associated with HDL levels, but these loci have been primarily reported in European and Caucasian populations. In order to identify the loci associated with HDL levels in the Chinese population, we have conducted a two-stage GWAS study using 5000 Han Chinese data from Taiwan Biobank. Replication was performed in another 5,000 subjects of Han Chinese ethnicity. Further, we performed Mendelian randomization analysis using weighted genetic risk score (wGRS) as the instrumental variable. A total of 77 SNPs was significantly associated with HDL-C levels. These SNPs are located in *APOA5, BUD13, CETP, DOCK6, APOA5-BUD13* cluster*, LIPC, LIPG, and LPL & ZPR1*. Due to high linkage disequilibrium among the SNPs in each region, only one most significant SNP was chosen for each gene to compute the wGRS. Analyzing data from 10,000 Han Chinese, we found that although HDL-C level was significantly associated with the status of CVD; HDL-C wGRS, an instrumental variable representing the lifelong genetic exposure dosage, was not with the development of CVD event with Cox Proportional Hazard Model. In conclusion, we have identified and confirmed the 10 SNPs associated with the HDL-C level, but Mendelian randomization study does not support its causal relationship with CVD risk in Taiwan Han Chinese.

2642T

Genome-wide association study of congenital heart disease in the UK Biobank. *A. Córdova-Palomera[1], S. Gustafsson[2], E. Ingelsson[3], J.R. Priest[1,4].* 1) Department of Pediatrics, Stanford University, Stanford, CA; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 4) Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA.

Affecting nearly 1 per 100 live births, congenital heart disease (CHD) is the most common congenital anomaly in newborns. While previous findings indicate an important etiological role for inherited factors such as structural variation, additional genetic variants influencing disease risk remain to be discovered.   Using publicly available resources from the first release of data from the UK Biobank, a genome-wide association study of 1,387 CHD cases and 114,509 controls was performed with PLINK; putative associations of rare genetic variants were validated with Firth's penalized regression. Additionally, genetic correlation analysis of the most frequent CHD subtypes (aortic valve stenosis, atrial septal defect, left ventricular outflow tract obstruction, mitral valve abnormalities and other septal defects) was implemented using LD score regression (LDSC).   Markers at twenty-six loci displayed statistically significant associations with CHD, including rare coding variants in *DSG2* (rs536617217, Firth's penalized regression OR: 8.02, *p* = 1.8e-9) and *ITGB2*, loci previously associated with CHD-related phenotypes. In addition, genetic correlations were detected between the main diagnostic subtypes.   The results suggest novel candidate loci as determinants of genetic risk for CHD in the general population, and indicate a shared genetic architecture between different types of CHD. Follow-up analyses in additional datasets (Pediatric Cardiac Genomics Consortium and second UK Biobank data release) will be performed to confirm the findings. Though use of adult populations to study childhood disease is prone to selection and survivorship biases, large-scale public datasets offer an unprecedented opportunity to study the genetic architecture of childhood diseases, such as CHD.

2643F

Genetic correlation of human lipidomic endophenotypes and cardiometabolic phenotypes in the Busselton Family Heart Study. *G. Cadby[1], PE. Melton[1], J. Hung[2], J. Beilby[3,4], MP. Dubé[5], F. van Bockxmeer[6], GF. Watts[2,7], J. Blangero[8], PJ. Meikle[9], EK. Moses[1].* 1) Centre for Genetic Origins of Health and Disease, The University of Western Australia and Curtin University, Western Australia; 2) School of Medicine, The University of Western Australia, Western Australia; 3) Busselton Population Medical Research Institute Inc., Western Australia; 4) PathWest Laboratory Medicine WA, Western Australia; 5) Université de Montréal Beaulieu-Saucier Pharmacogenomics Centre, Montreal Heart Institute, Quebec; 6) School of Surgery, The University of Western Australia, Western Australia; 7) Lipid Disorders Clinic, Department of Cardiology, Royal Perth Hospital, Western Australia; 8) South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley School of Medicine, Texas; 9) Metabolomics Laboratory, Baker Heart and Diabetes Institute, Victoria.

Cardiometabolic disease (CMD) is the leading cause of death globally. Genetic investigations into the human lipidome (>1000 lipid species) may provide insight into CMD risk. Due to the biologically simpler nature of lipid species, their determinants are likely to be closer to the causal action of genes than more complex integrated lipid measures. The aim of this study was to estimate the heritability of lipidomic endophenotypes and their genetic correlation with CMD.   Targeted lipidomic profiling was performed on 4671 individuals from the Busselton Health Study (BHS) using electrospray ionization-tandem mass spectrometry to quantify >300 lipid species from >20 lipid classes. The BHS is a community-based study that included both related and unrelated individuals. To exploit both known and unknown relatedness, we estimated empirical kinship probabilities using LDAK. The general linear mixed model incorporated in SOLAR was used to estimate narrow-sense heritabilities of lipid species and classes, and genetic correlations between lipid classes and CMD traits: HDL-C, LDL-C, triglycerides (TG), systolic blood pressure (SBP), diastolic blood pressure (DBP) and HOMA-IR. Analyses were adjusted by age, sex and use of lipid-lowering and antihypertensive medications.   Over 98% of the lipid species were significantly heritable (h^2: 0.11-0.52) and all lipid classes were significantly heritable (h^2: 0.15-0.50). The most heritable lipid classes were total monohexosylceramide (MHC) and total acylcarnitine (both h^2=0.50, $P<1\times10^{-40}$). Total MHC was genetically correlated with TG (r_g=-0.26, $P=6\times10^{-5}$) and LDL-C (r_g=0.25, $P=1\times10^{-5}$). LDL-C was genetically correlated with 22 of the 24 lipid classes (h^2: 0.25-0.78). The largest genetic correlations were between TG and total diacylglycerol (r_g=-0.88, $P=1\times10^{-24}$) and total triacylglycerol (r_g=0.81, $P=3\times10^{-23}$). DBP was not genetically correlated with any lipid class and SBP was only correlated with lysophosphatidylcholine (r_g=0.38, $P=1\times10^{-4}$). HOMA-IR was only genetically correlated with total sulfatide (r_g=-0.32, $P=1\times10^{-4}$); HDL-C showed the strongest genetic correlation with total sulfatide (r_g=0.47, $P=5\times10^{-14}$). We have demonstrated, using empirically-derived kinship estimates, that lipidomic endophenotypes are heritable and genetically correlated with CMD. Future work includes identifying lipid species that represent the best potential CMD-related endophenotypes, and whole genome sequencing to identify causal genetic variants.

2644W

Genome-wide association analysis identifies multiple loci associated with coronary artery calcification in Koreans. *S. Choi[1,2], H. Lee[1,2], H. Park[1,2], S. Choi[1,2], E. Choe[1,3], S. Oh[1,4], C. Lee[1,2], E. Shin[5].* 1) Internal medicine, Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, South Korea; 2) Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea; 3) Department of General Surgery, Seoul National University College of Medicine, Seoul, Korea; 4) Department of Family Medicine, Seoul National University College of Medicine, Seoul, Korea; 5) DNA Link, Inc., Seoul, Korea.

Background/Aims: Coronary artery calcium score (CACS) represents a measure of overall coronary atheromatous plaque burden and is a strong predictor of coronary heart disease. The development of coronary atherosclerosis is associated with multiple genetic and environmental factors, and life style. We report an analysis of genome-wide association studies (GWAS) for coronary artery calcification in Koreans. **Methods:** We performed a GWAS to identify genetic factors related to coronary artery calcification in large Korean population based samples of 1,729 subjects using the Affymetrix Axiom® Korean Biobank Array 1.0, customized biobank genotyping arrays. We replicated the data in another sample of 460 subjects. Quantitative CACS were calculated according to the Agatston method. We investigated SNPs that are associated with severe coronary calcification, defined as above 90th percentile of CACS by each 5-year age group and gender category in the Korea Initiatives on Coronary Artery calcification (KOICA) registry (85,945 asymptomatic Koreans, 69.3% in male). The genetic analysis was used with additive model by SAS version 9.1 (SAS Institute Inc., Cary, NC, USA.) and PLINK version 1.9 (https://www.cog-genomics.org/plink2). **Results:** Rs2227593 located in the SERPINC1 gene in chromosome 1 was significant associated with sever coronary artery calcification; $p = 3.5 \times 10^{-4}$, OR 0.65 (95% CI 0.51-0.82) in the discovery set and $p = 4.6 \times 10^{-2}$, OR 0.65 (95% CI 0.42-0.99) in replication set. Rs16981779 and rs76229963 located in the ANKRD60, C20orf85, PPP4R1L gene in chromosome 20 were significant associated with sever coronary artery calcification; $p = 6.0 \times 10^{-4}$, OR 2.2 (95% CI 1.4-3.3) in the discovery set and $p = 2.2 \times 10^{-2}$, OR 2.6 (95% CI 1.3-5.3) in replication set and $p = 5.0 \times 10^{-4}$, OR 2.3 (95% CI 1.5-3.5) in the discovery set and $p = 1.6 \times 10^{-2}$, OR 2.7 (95% CI 1.3-5.5) in replication set, respectively. Rs16981779 and rs76229963 were in strong linkage disequilibrium (R-square = 0.78 and D' = 0.95). **Conclusions:** We identified associations among these loci with coronary artery calcification. These findings provide new insights into the genetic susceptibility to coronary atherosclerosis in Koreans.

2645T

The search for coronary heart disease biomarkers: A large scale reanalysis of gene expression data. *B. Cunha, F. Vargas, L. Cordeiro.* UNIRIO, Rio de Janeiro, Brazil.

Cardiovascular disease remains the leading global cause of death and, more specifically, coronary heart disease (CHD) still represents a serious public health problem throughout developed countries. In the United States approximately 16 million individuals over 20 years of age have CHD and its prevalence is projected to increase by more than 18% in 2030. Direct and indirect costs of heart disease are expected to be over $200 billion in 2017 and it is estimated that in every 40 seconds one North American will die of cardiovascular disease and another will suffer a myocardial infaction (MI). In spite of its clinical relevance, precise risk stratification of CHD patients is still lacking, posing one of the biggest challenges in modern cardiology. The objective of this study is to identify specific gene expression signatures associated with different stages of CHD. We undertook a reanalysis of public gene expression data related to CHD and constructed a unique dataset of over 440 microarray samples collected from peripheral blood mononuclear cells of healthy patients and patients with angina and myocardial infarction (MI). We identified different gene expression patterns between healthy controls and the MI group – 192 differentially expressed genes (DEGs) – as well as between the angina and MI groups – 134 DEGs. Interestingly, there was no DEGs between the angina group and healthy controls. Top DEGs between healthy controls and MI patients are: *CXCL2, IL1B, IL1R2, NR4A2, CCL20, LILRB2, CSTA, RPS4Y1, TREM1, NFIL3, MGAM, C5AR1, IRAK3, MGP*, and *PTX3*. The top most DEGs between the MI group and the angina group are: *S100A12, AQP9, TREM1, NR4A2, SERPINA1, CHI3L1, IER3, MGP* and *PLBD1*. All DEGs are upregulated in both comparisons, *i.e.*, there was not a single downregulated gene in the MI group relative to both the angina group and healthy controls. Pathway analysis shows that several signaling pathways are deregulated in the MI group: cell cycle regulation, cell differentiation, TNF signaling, TLR signaling, NF-κB signaling and hematopoietic stem cell activation; all previously associated with CHD. We also found that the IL-17 signaling pathway is suppressed in the MI group, lending support to the hypothesis of its protective role. Moreover, the rheumatoid arthritis signaling pathway is activated in the MI group as well. Taken together, these results corroborate the significant role of inflammatory signaling pathways in the etiopathogenesis of coronary heart disease.

2646F

Identification of rare variation influencing CVD risk in Mexican Americans. *J.E. Curran[1], N.B. Blackburn[1], J.M. Peralta[1,2], M.A. Almeida[1], H.H. Goring[1], L. Almasy[3,4], M.C. Mahaney[1], D.C. Glahn[5,6], R. Duggirala[1], J. Blangero[1].* 1) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX, USA; 2) Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, AUS; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 5) Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 6) Olin Neuropsychiatry Research Center, Institute of Living, Hartford Hospital, Hartford, CT, USA.

Cardiovascular disease (CVD) is the leading cause of mortality worldwide and encompasses conditions of the heart and vasculature. CVD risk is attributable to several factors including blood pressure, smoking, increasing age, sex, and genetics. Although CVD risk is heritable and there have been a number of successes localizing QTLs that influence disease risk, the genetic basis is still relatively unknown and likely differs substantially between families. Here we show the utility of whole genome sequencing (WGS) for the discovery of rare CVD risk variants. We have prospectively followed 1,222 Mexican American individuals from large pedigrees in the San Antonio Family Heart Study and have their histories of heart attack, heart surgery for arterial blockages, and incidences of cardiac-related death. The large extended pedigrees provide opportunity to potentially capture sufficient numbers of very rare functional variants to permit variant-specific association testing. The age-standardized prevalence of CVD in this cohort is 16% and we have calculated the heritability of CVD using the normal threshold model to be 25.9% (P=0.01). Using WGS, we searched for rare (MAF ≤ 0.01), non-synonymous, highly deleterious (CADD ≥ 15) variants influencing the risk of CVD. We identified multiple rare variants significantly (after taking into account the multiple testing scheme) associated with CVD in our large pedigrees including an apparently Hispanic-specific *ALKBH3* variant (*G95R*) and an *ATG4C* variant (*K326Q*) that show a greater than 6-fold increase in risk overall and are suggestive of early onset CVD. Other interesting variants include a pair of variants in compete linkage disequilibrium on chromosome 1 in the *PIGC* and *RABGAP1L* genes whose heterozygotes exhibit more than 2 times the risk for disease as their common homozygotes. Our results suggest that extended pedigrees are useful for the capture of rare functional variants of relevance to CVD risk. Given the rarity of such variants, it is unlikely that traditional statistical replication will be possible and we are actively pursuing biological experimental validation. Given the large effect sizes of these variants, our results suggest the likelihood of widely varying genetic causes of individual risk and family specificity of predominant causal factors in CVD.

2647W

Pharmacogenetic study on clopidogrel response among Filipinos. *E.C. Cutiongco de la Paz[1,2], R.G. Sy[1,2,3], R.P. Tiongco II[3], J.B. Nevado, Jr[1,2], E.B. Llanes[3], J.A. Magno[3], D.D. Ona[3], F.R. Punzalan[3], P.M. Reganit[3], L.G. Santos[3], R.L. Santos-Cortes[4], J.A. Aherrera[3], L.L. Abrahan IV[3].* 1) Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila, Manila, Philippines; 2) Philippine Genome Center, University of the Philippines; 3) Section of Adult Cardiology, Department of Medicine, Philippine General Hospital, University of the Philippines Manila, Philippines; 4) Department of Otolaryngology, School of Medicine, University of Colorado, USA.

Dual antiplatelet therapy with clopidogrel and aspirin is the standard of care for patients with acute coronary syndromes, and is often recommended to prevent acute thrombotic events. Clopidogrel selectively and irreversibly inhibits the P2Y12 ADP receptor, which is responsible for platelet activation and aggregation. However, response to this antiplatelet has been significantly variable, and nonresponse may lead to fatal outcomes. The use of genetic markers may provide guided therapy to prevent such outcomes. An observational study involving Filipino participants (80 non-responders and 165 responders; 150 males and 95 females) taking clopidogrel was conducted. DNA from blood samples were extracted and were genotyped using customized Illumina Goldengate microarray chips for 384 gene variants. From these variants, two were found to be of statistical significance after adjusting for multiple testing. These are the SNPs affecting apolipoprotein A-V (APOA5) and matrix metalloproteinase 3 (MMP3), which are associated in regulating plasma triglyceride levels and progression of atherosclerosis, respectively. Both variants were observed to be more frequent among non-responders: APOA5 (allelic trend, OR 2.77, p-value 0.00000049) and MMP3 (allelic trend, OR 4.50, p-value 0.0000016), and both are not previously associated with clopidogrel nonresponse. Further validation studies may need to be done before concluding a definite association between these markers and clopidogrel response.

2648T

Arrhythmia and night vision blindness: Chicken and egg? Or could be chicken or egg? *A. Faucon, X. Zhang, P. Evans, A. Konkashbaev, P. Straub, N. Cox.* Human Genetics, Vanderbilt University, Nashville, TN.

   PrediXVU, the gene by medical phenome catalog created by applying PrediXcan to BioVU, the biobank at Vanderbilt University, currently serves results on about 10,000 European descent subjects. The top gene associated with atrial fibrillation is *LRIT3*, Leucine-rich repeat, immunoglobulin-like and transmembrane domain 3, which is a Mendelian gene in which recessive mutations cause night vision blindness. This is an intriguing pattern because drugs used to treat arrhythmias can have as a side effect night vision blindness. *LRIT3* is thought to direct the localization of *TRPM1* to dendritic tips of ON-bipolar cells in the eye, as well as altering cone synapse formation. We identified an additional 22 genes with associations to both cardiac arrhythmia phenotypes and visual disturbances including night vision blindness: *BBOX1, BCAN, BPIFC, C20orf197, CLCNKB, ELP2, F2R, FAM115C, GALNT14, GOLT1B, HGS, IL36RN, MAF1, NR2F6, PIK3C2G, PPIP5K2, PTGFRN, RAD9A, RCBTB1, SCOC, SLC25A4,* and *STRADA*. A number of these are channel genes, while others have been previously reported to be involved in drug response phenotypes.

2649F

Associations of circulating protein levels with lipid fractions in the general population. *S.M. Figarska[1], S. Gustafsson[2], J. Sundström[3,4], J. Ärnlöv[5,6], L. Lind[3], E. Ingelsson[1,2].* 1) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA, USA; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden; 4) Uppsala Clinical Research Center; 5) Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Sweden; 6) School of Health and Social Sciences, Dalarna University, Falun, Sweden.

   Background and aims: Revealing patterns of associations between circulating protein and lipid levels could improve biological understanding of cardiovascular disease, the main cause of death worldwide. The aim of this study was to investigate associations between protein biomarkers shown to be related to cardiovascular disease, and triglyceride (TG), total cholesterol (TC), LDL and HDL cholesterol (LDL-C and HDL-C) levels in three population-based cohorts. **Methods:** We measured lipid fractions and plasma protein levels using the Olink ProSeek CVD I or II+III arrays in fasting individuals without lipid-lowering treatment from the EpiHealth (n=2086, 52% women, median age 61 years), PIVUS (n=833, 51% women, all aged 70 years) and ULSAM (n=598, all men aged 77 years) studies. We included 57 proteins that overlapped between the different arrays and that passed QC in all three cohorts. A discovery analysis was performed in EpiHealth in a regression framework, and associations with FDR<0.05 were further tested in a meta-analysis of PIVUS and ULSAM, where p-value of 0.05 was considered a successful replication (validation FDR of 0.1%). All analyses were adjusted for sex, age, BMI, smoking, glucose levels, systolic blood pressure, blood pressure medication, diabetes medication and CVD history. **Results:** Out of 57 tested proteins, 42 demonstrated a replicated association with at least one lipid fraction; 35 were associated with TG, 15 with TC, 9 with LDL-C, and 24 with HDL-C. Among these associations, we found kidney injury molecule (KIM1), tumor necrosis factor receptor 1 and 2 (TNF-R1 and TNF-R2) to be associated with all four lipid fractions. Further, 17 proteins were related to both TG and HDL in a consistent and biologically expected manner, i.e. higher TG and lower HDL or *vice versa*. For instance, higher stem cell factor (SCF) and growth hormone (GH) were associated with lower TG and higher HDL; thus, a pattern generally associated with lower insulin resistance and CVD risk. Another common pattern of associations was concomitantly higher TG, TC and LDL, which often is associated with higher CVD risk. This was observed for example for higher C-X-C motif chemokine 16 (CXCL16), Gal3 (Galectin-3) and monocyte chemotactic protein 1 (MCP-1). **Conclusions:** Our comprehensive analysis of plasma proteins and lipid fractions of >3,500 individuals from the general population provide new information about lipid biology and inform about potential links to cardiovascular disease.

2650W

Association of common variants in arrhythmogenic cardiomyopathy desmosomal genes with ECG traits in the general population. *L. Foco[1], C. Fuchsberger[1], M. Gögele[1], F. Murgia[1], G. Schmidt[2], A. Rossini[1], P.P. Pramstaller[1,3], C. Pattaro[1].* 1) Institute for Biomedicine, Eurac research, Bolzano, Bolzano, Italy; 2) Medizinische Klinik und Poliklinik, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; 3) Department of Neurology, General Central Hospital, Bolzano, Italy.

   Purpose Arrhythmogenic cardiomyopathy (AC) is a rare condition, with familial inheritance, featuring life-threatening arrhythmias and cardiomyocyte replacement with fibro-fatty cells. Such alterations are reflected in both depolarization and repolarization abnormalities at the electrocardiogram (ECG) level. Mutations in genes encoding cardiac desmosomal proteins are an established cause of AC. Given that common genetic architecture is often observed between Mendelian diseases and related complex traits, we hypothesised that common variants in desmosomal genes, involved in the electrical coupling between cardiomyocytes, could influence the cardiac conduction traits in the general population. **Methods** Through linkage-disequilibrium analysis of SNPs from the HumanOmniExpressExome Bead Chip imputed using 1000 Genome panel Phase 3, we selected 2429 SNPs in the desmosomal genes Desmoplakin (*DSP*), Plakophilin 2 (*PKP2*), Desmoglein 2 (*DGS2*), Desmocollin 2 (*DSC2*), and Junction Plakoglobin (*JUP*). SNPs were tested for association with P wave, PR, QRS, and QT intervals, derived from 10 sec ECGs, in 4979 participants to the Cooperative Health Research in South Tyrol (CHRIS) study. We fitted age- and sex- adjusted linear mixed models accounting for population relatedness and assuming additive genetic effects. Multiple testing corrected significance was set at 1.9×10^{-4} based on the percentage of genome covered by the tested regions. **Results** Two independent SNPs, rs115171396 (MAF=0.02) and rs72835665 (MAF=0.49) in *JUP*, were associated with P wave length, with effects of 4.87 (SE=1.08, P-value=6.6×10^{-6}) and -1.10 (SE=0.27, P-value=4.5×10^{-5}) milliseconds (ms) per copy of the minor allele, respectively. Variant rs2744389 in *DSP* was associated with QRS duration: effect= -1.10 ms (SE=0.24, P-value=3.5×10^{-6}). We did not find evidence of association with PR interval and QT duration. **Conclusions** While awaiting independent replication, the results suggest that variants in desmosomal genes could impact the cardiac electric function in the general population. Ongoing whole exome sequencing analyses will allow finer mapping of the investigated genes.

2651T

Genome wide association study identifies nine novel loci for subclinical atherosclerosis traits and highlights genetic correlation with clinical cardiovascular disease. *N. Franceschini[1], C. Giambartolomei[2], J. Liang[3], J.P. Casas[4], C.J. O'Donnell[5].* 1) Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Human Genetics, University of California Los Angeles, Los Angeles, California, USA; 3) Department of Epidemiology & Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 4) Farr Institute of Health Informatics Research, UCL Institute of Health Informatics, University College London, London, UK; 5) Cardiology Section and Center for Population Genomics, Boston Veteran's Administration (VA) Healthcare, Boston, MA, USA.

   Carotid intimal medial thickness (cIMT) and carotid plaque are measures of subclinical atherosclerosis associated with clinical cardiovascular disease events - myocardial infarction, stroke and death. Genetic studies may help to identify biologic pathways contributing to subclinical and clinical manifestations of atherosclerosis. To search for genetic loci for subclinical atherosclerosis, we conducted meta-analyses of genome-wide association study (GWAS) in up to 71,129 participants for cIMT and up to 41,145 participants for carotid plaque. cIMT and plaque measures were evaluated using high-resolution B-mode ultrasonography. Each study imputed up to the Phase 1 integrated reference panel from the 1000 Genomes Project data and followed standardized protocols for quality control and association analyses. Fixed effect meta-analyses identified nine novel loci for cIMT (1q32.2, *ARP6AP1IL, AIG1, PIK3CG, MCPH1, VTI1A, SGK223, CBFA2T3*) or plaque (*TRIM44*) at *P* value < 5.0 x 10^{-8}, and replicated five previously described loci for cIMT (*ZHX2, PINX1, APOE*) and plaque *(EDNRA, PIK3CG).* To better understand the shared genetic influences between subclinical atherosclerosis and atherosclerotic clinical disease, we tested the genetic correlation between cIMT or carotid plaque with coronary heart disease or stroke using genome-wide summary statistics from our meta-analyses and the CARDIoGRAMPlusC4D and the METASTROKE consortia using the LD score regression approach. Both cIMT and carotid plaque showed significant genetic correlations with coronary heart disease and ischemic stroke (all *P*<0.05), though the magnitude of correlation was almost twice stronger for carotid plaque and coronary heart disease (0.54) or ischemic stroke (0.60) compared to cIMT (0.20 and 0.30 for coronary heart disease and ischemic stroke, respectively). When looking at stroke sub-types (cardio-embolic, large vascular disease strokes) against cIMT and carotid plaque, the magnitude of the correlation was similar though with larger uncertainty (all *P*>0.05). In conclusion, we provide here the largest genomic study of cIMT and carotid plaque using high-density imputed genotypes. Our meta-analyses of cIMT and carotid plaque identified nine novel loci. The correlations with coronary heart disease and stroke highlight novel biological pathways for further evaluation.

2652F

66 novel loci detected in a trans-ethnic genome-wide association study of blood pressure in over 440,000 participants from the Million Veteran Program and UK Biobank. *A. Giri[1], A. Hung[2], J.N. Hellwege[1], E. Torstenson[1], O. Wilson[2], Y.V. Sun[3], H. Warren[4], M. Caulfield[4], P. Elliott[5], P.W.F. Wilson[6], P.S. Tsao[7], C.P. Kovesdy[8], K.A. Birdwell[9], D.R. Velez Edwards[10], C. O'Donnell[11], T.L. Edwards[1], the UK Biobank, and on behalf of the VA Million Veteran Program.* 1) Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN; 2) VA TVHS Nashville, Division of Nephrology & Hypertension, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Epidemiology, Emory University Rollins School of Public Health; Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA; 4) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; 5) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, United Kingdom; 6) Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA; 7) VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 8) Nephrology Section, Memphis VA Medical Center, Memphis TN; 9) Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 10) Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center; Tennessee Valley Healthcare System VA, Nashville, TN; 11) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Blood pressure (BP) is a heritable trait and high BP is associated with significant global morbidity and mortality. We conducted a genome-wide association study (GWAS) of systolic (SBP) and diastolic blood pressure (DBP) using electronic heath record (EHR) data from the Million Veteran Program (MVP) cohort (n > 301,000) and the UK Biobank (UKB; n ~ 145,000). DBP and SBP were regressed onto additively encoded genotypes imputed to the 1000 Genomes reference panel, adjusting for age, age-squared, sex, body mass index, and top 10 principal components. Primary analyses were conducted by self-reported race/ethnicity status for MVP (White non-Hispanic, Black non-Hispanic and Hispanic) and UKB separately, followed by inverse-variance weighted fixed-effects meta-analysis across strata using METAL. Meta-analyses identified 233 associated loci with p-value < 5×10^{-8}; 4,982 SNPs at 111 loci were associated with DBP explaining 1.1% of the phenotypic variance, and 7,873 SNPs at 145 loci were associated with SBP explaining 1.6% of the variance. Among these, we detected 66 novel signals, 12 of which were significant in both SBP and DBP, with 33 and 21 unique previously unreported signals for SBP and DBP, respectively. Novel signals explained 0.4% and 0.3% of the variance in SBP and DBP respectively, increasing the total variance explained by 25% for each trait. Furthermore, we provide confirmatory evidence for 167 previously identified BP loci at genome-wide significance. The strongest evidence for association with SBP was rs12705390 near *CCDC71L – PIK3CG* (p-value = 1.44×10^{-45}), while the top association with DBP was rs12509595 near *PRMD8 – FGF5* (p-value = 6.14×10^{-35}). The most significant novel locus was near *RXFP2* at rs1869800 (p-value = 4.27×10^{-26}) for SBP and rs7983337 (p-value = 1.78×10^{-13}) for DBP (r^2=0.99). In DBP alone, the most significant novel signal was near *TRIM39 – RPP21* (rs3094704; p-value = 6.13×10^{-11}), while rs10265221 (p-value = 5.82×10^{-14}) in *PRKAG2* was the strongest novel signal with SBP. Trans-ethnic fine-mapping analyses, which leverage the differences in linkage disequilibrium patterns between populations, are likely to refine associated loci into credible sets of potentially causal SNPs. These results represent the largest reported discovery-stage GWAS of BP traits to date and can offer new insights into vascular biology.

2653W

Prevalence of variants of *FVL*, *PTH*, *PAI-1*, *MTHFR* and *EPCR* among Cardiovascular patients. *M. Hosseini moghadam, N. Mahdieh.* Cardiogenetic Research Laboratory , Rajaie Cardiovascular Medical and Research Center, Tehran, Tehran, Iran.

Introduction Venous thromboembolism (VTE) is a multifactorial condition with genetic and acquired risk factors. Thrombosis could be associated with cardiovascular disorders. Here, we study five common variants G1691A (Factor V Leiden, *FVL*), G20210A (Prothrombin, *PTH, FII*), 4G/5G (Plasminogen activator inhibitor-1, *PAI-1*), A4600G (Endothelial protein C receptor, *EPCR*) and C677T (Methylene-tetrahydrofolate reductase, *MTHFR*) referring Rajai Center. **Material and Method** We studied a case series of 50 CVD patients (24 women and 26 men) withVTE condition that referred to our center. After, DNA extraction, PCR-RFLP was applied to determine the variants. **Results** The allele frequencies of *FVL* (G1691A), *PTH* (G20210A), Plasminogen activator inhibitor-1, *PAI-1* (4G/5G), Endothelial protein C receptor, *EPCR* (A4600G) and Methylene-tetrahydrofolate reductase, *MTHFR* (C677T) were 0.02, 0.03, 0.67, 0.11 and 0.3 respectively. **Conclusion** *PAI-1* (4G/5G) and *MTHFR* (C677T) are the common variants among VTE patients in our population. In addition, *FVL* (G1691A) and *PTH* (G20210A) variants are not common in our population. **Keywords:** Thrombophilia; venous thromboembolism (VTE); cardiovascular disease (CVD).

2654T

Elucidating the molecular causes of severe hypercholesterolemia in Finland. *N. Junna, P. Ripatti, I. Surakka, S. Ripatti, E. Widén.* Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland.

Severe hypercholesterolemia is a major risk factor for coronary artery disease (CAD). Patients suffering from its familial inherited forms have more than 10-fold increased risk for CAD. Although CAD may be prevented by lowering cholesterol levels, less than 20% of hypercholesterolemia patients are identified or treated in Finland and only half of the patients carry hypercholesterolemia-associated mutations. Our study aims at unraveling the genetic architecture underlying severe hypercholesterolemia in Finland by studying individuals with high LDL cholesterol (>=5 mmol/l), who are identified in our ongoing population study GeneRISK, encompassing 10,000 subjects aged 45-64 in Southern Finland (www.generisk.fi). Hitherto 6,817 study participants have been recruited. 4% of them had severe hypercholesterolemia, of which only 5% received lipid-lowering therapy. 200 hypercholesterolemia cases have been screened for the 7 most common known familial hypercholesterolemia-associated *LDLR*-mutations enriched in Finland. Surprisingly, all subjects were mutation negative. Whole exome sequencing of 141 subjects has not revealed other previously known hypercholesterolemia-causing mutations in the *LDLR*, *APOB* or *PCSK9* genes either. Nonetheless, 4 subjects carried rare or novel variants in *APOB* with probably damaging *in silico* predictions (P877L, R1689H, I2305V and N3257S). Polygenic modeling (genomic risk score (GRS) encompassing 142 SNPs explaining 17% of the LDL cholesterol variation in the population) suggested slight clustering of polygenic risk in 34 (24%) of the exome sequenced patients (GRS within the top 10% of population distribution). Lifestyle factors, such as consumption of saturated fats, did not correlate with LDL cholesterol levels. In conclusion, treatment of hypercholesterolemia appears gravely suboptimal and the contribution of known hypercholesterolemia-associated mutations may be smaller than previously anticipated in Southern Finland. At present, the exome sequencing analyses of up to 200 hypercholesterolemia cases are still ongoing. Our intention is to evaluate identified putative novel pathogenic variants by seeking replication in a larger Finnish population cohort, and to assess function both by *in silico* and *in vitro* methods.

2655F

Gender specific modification of heart failure with preserved ejection fraction risk by mitochondrial haplogroups. *R.T. Levinson[1], E. Farber-Eger[1], J.D. Mosley[1], D.C. Samuels[2], Q.S. Wells[1].* 1) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN.

Heart failure (HF) is a complex syndrome characterized by dyspnea, fatigue, reduced exercise tolerance, and fluid retention, that may result from a variety of structural or functional cardiac impairments. HF is frequently partitioned into subgroups using cutoffs of left ventricular ejection fraction (EF), with patients having an EF ≤40% termed HF with reduced EF (HFrEF) and those with EF ≥50% having HF with preserved EF (HFpEF). HFpEF disproportionately affects women, and no current therapies exist. A consistent feature of HF is altered mitochondrial (mt) function and cardiac bioenergetics, including oxygen availability and mt ATP production. However, the role of mt genetic variation in HF and its subtypes is incompletely understood. We proposed to evaluate the role of mt genetics in HF and its subtypes, including the gender specific influence of mt genetics. We curated electronic health records (EHRs) from BioVU, the Vanderbilt DNA biobank, and assigned HF status in 6,512 individuals (55% male, median age 66) of European ancestry with genotyping available on the MEGA chip. EFs extracted from echocardiography data were used to assign HF subtype (HFpEF or HFrEF) when possible. Our analysis dataset consisted of 3,259 HF cases and 3,253 controls. In total, 2,170 (67%) of HF cases were assigned to a subtype (HFpEF=964 and HFrEF=1,206). We extracted 894 mtSNPs and group assignment was done with Haplogrep. MT haplogroups (HG) were tested for association with HF, HFpEF, and HFrEF. We also conducted gender stratified analyses and analyses accounting for the interaction between gender and HG. Analyses were adjusted for age, BMI, myocardial infarction, hypertension, diabetes, and estimated glomerular filtration rate (eGFR). HG Uk was significantly associated with risk of HFpEF in males (OR[95%CI]= 1.46 [1.11, 1.92], p=0.006) but not in females. HG T was protective for HFpEF in males (OR[95%CI]= 0.61[0.41, 0.92], p=0.02), but had a risk effect in females (OR[95%CI]= 1.54 [1.08, 2.19] , p=0.01). Terms for the interaction of male gender with both Uk (p=0.03) and T (p=0.001) were significant when added to their respective models. No mt HGs were significantly associated with HF status or HFrEF in any analysis. These results establish a possible role for mt genetics in risk of HFpEF, and the gender specific effects of mt variation.

2656W

Family study of noncompaction cardiomyopathy shows variability of cardiac phenotype within and between families. *D. Majoor-Krakauer[1], J.I. van Waning[2], M. Michels[2], R.M.W. Hofstra[1], K. Caliskan[2], M.A. van Slegtenhorst[1].* 1) Clinical Genetics, Erasmus Medical Center, Rotterdam, Netherlands; 2) Cardiology, Erasmus Medical Center, Rotterdam, Netherlands.

The objective of this study was to explore which cardiomyopathies occur in families of NCCM patients. In total 152 families of NCCM patients were screened for familial cardiomyopathy at the Erasmus Medical Center. The families had genetic counseling and DNA testing and were enrolled retrospectively from 2005 till 2017. In 61 (41%) of the 152 families 108 relatives were diagnosed with a cardiomyopathy. The diagnoses of the affected relatives were: 78 NCCM, 13 HCM and 17 DCM. In 37 (24%) families all affected relatives had NCCM; 24 / 37 of the families had a (L)PV ((likely) pathogenic variant) in one of the 52 tested cardiomyopathy genes. Sixteen (of 24 (likely)pathogenic variants involved the *MYH7* gene. The (L)PV in these 16 families all occurred in the head region of this gene. In contrast *MYH7* (L)PV associated with DCM or with familial NCCM and DCM were located in other parts of the *MYH7* gene. In 16 (13 with a (L)PV) families the combination of NCCM and DCM was observed in the family. In these NCCM/DCM families *MYH7* (5, 31%) and *TTN* (4, 25%) was involved. In seven (5 with (L)PV) families HCM was diagnosed in relatives. These were mostly families of patients with complex *MYBPC3* genotypes, in which relatives with HCM had a single *MYBPC3* (L)PV. In conclusion, familial NCCM is likely to be caused by defects in the head of the *MYH7* gene. DCM and HCM may occur in familial NCCM. The cause of the phenotypic variability in families remains to be elucidated.

2657T

Novel compound homozygous mutations in genes involved in mitochondrial function associated with sudden death with cardiac fibrosis in infancy. *K. Mittal[1], T.T. Koopmann[1], D. Chatterjee[1], J. Cameron[2], T. Paton[3], A. Rafiq[1], S. Aafaqui[1], D. Chiasson[2], C. Ackerley[2], M. Fatah[1], C. Marshall[2], R.H. Hamilton[1].* 1) Translational Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Genetics and Genome biology, The Hospital for Sick Children, Toronto, Ontario, Canada.

Sudden death is a rare event between age 1 and 2 years, with most sudden infant death syndrome (SIDS) victim usually being healthy infants less than one year of age. Although the causes of SIDS remain largely unknown, several recent reports have suggested the involvement of inherited cardiac diseases as a causal factor underlying SIDS. The genetic analysis of the SIDS cases has revealed genes encoding cardiac ion channels and in cardiomyopathy-associated genes. A role for abnormal/ defective energy production in SIDS infants is warranted because infants with SIDS tend to be sleepier, less active, and have lower activity scores. In the present study, we identified a consanguineous family originally from Pakistan who had lost three of five children to sudden death between the ages of 8 and 18 months. We performed homozygosity mapping and whole exome sequencing (WES) followed by segregation analysis and phenotype genotype correlation to identify the genetic mechanism responsible for the sudden deaths in this family. High-density homozygosity mapping revealed a contiguous block of homozygous SNVs at the loci on chromosomes 4, 8 and 15, shared between the two affected individuals, but not the parents. WES analysis revealed the presence of two homozygous pathogenic mutations. One variant was found in the gene encoding for Isocitrate Dehydrogenase 3A (IDH3A), a human mitochondrial NAD+ -specific isocitrate dehydrogenase alpha subunit that is part of the citric acid cycle, and the second was found in mitochondrial inorganic pyrophosphatase (PPA2). We have identified a case of compound homozygous mutations in association with a consanguineous family with sudden infant death syndrome. The functional assays showing no enzyme activity in either of the genes, in the patient, combined with the bioinformatics support a causal link. Together, this data support a deficient energy production process related to perturbations in the IDH3A and PPA2 functions. Our findings highlight the importance of genes involved in mitochondrial function in SIDS and warrant further replication studies. Further work is required to more precisely delineate the pathogenic mechanisms underlying mutations in mitochondrial proteins involvement in SIDS.

2658F

Identifying new therapeutical targets for congestive heart failure. *A. Moreira, F. Vargas, L. Cordeiro.* UNIRIO, Rio de Janeiro, Brazil.

Though gene expression analysis has allowed greater understanding of the pathophysiological development of congestive heart failure (CHF), clinical and tissue heterogeneity in microarray samples, in addition to the small number of cases in any individual study, makes it difficult to correctly interpret the results of such studies. Through reanalysis of public microarray gene expression data we were able to construct a dataset of about 200 left ventricular samples. Classifying patients in three groups: healthy, non-ischemic CHF, and ischemic CHF, we found 100 differentially expressed genes (DEGs) between healthy subjects and non-ischemic HF patients, and 96 DEGs between controls and ischemic HF samples. Analysing DEGs with the greater expression log fold changes allowed us to pinpoint a few therapeutic targets. It is worth mentioning the genes *FOS*, *SERCA3* and *OGN,* which are already subject of clinical investigation. Pathway analysis uncovered several altered biomolecular pathways, specially the ones directly involved with the reorganization of the extracellular matrix and cell adhesion, specially the BMP signalling, integrin and heparin binding, all of them presenting clear therapeutic opportunities. Other pathways related to inflammatory response, cellular proliferation, proteolysis, blood pressure regulation and oxidative stress may also be potential objects of future clinical research in CHF.

2659W

Polygenic hyperlipidemias and coronary artery disease risk. *P. Ripatti[1], J.T. Rämö[1], S. Söderlund[2], I. Surakka[1], N. Matikainen[2,3], P. Pajukanta[4], A.S. Havulinna[1,5], C. Ehnholm[5], V. Salomaa[5], R.K. Wilson[6], N.B. Ferimer[7], A. Palotie[1,8,9,10,11,12], M-R. Taskinen[2], S. Ripatti[1,13,14].* 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Research Programs Unit, Diabetes & Obesity, University of Helsinki, and Heart and Lung Centre, Helsinki University Hospital, Helsinki, Finland; 3) Endocrinology, Abdominal Center, Helsinki University Hospital, Helsinki, Finland; 4) Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California Los Angeles (UCLA), Los Angeles, California, USA; 5) Department of Health Solutions, National Institute for Health and Welfare, Helsinki, Finland; 6) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA; 7) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA; 8) Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 9) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA; 10) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; 11) The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 12) Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA; 13) Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 14) Wellcome Trust Sanger Institute, Cambridge, UK.

Hyperlipidemia, particularly increased LDL cholesterol (LDL-C) and triglycerides (TG), are established and treatable risk factors for coronary artery disease (CAD). In addition to well-known familial hypercholesterolemia (FH) mutations, hyperlipidemias have also been shown to be caused by polygenic burden. Whereas monogenic FH increases CAD risk considerably likely due to high lifelong exposure to LDL-C, it is unclear whether high polygenic load of LDL-C or TG increasing variants causes similarly increased CAD risk in a population or familial setting. We constructed clinical CAD endpoints from the FINRISK population cohort ($n = 20\ 499$) and a collection of 86 Finnish dyslipidemia families ($n = 1316$) with high total cholesterol (TC) and TG traits (identified from probands with premature CAD) linked with hospital discharge, causes-of-death and prescription drug registries. We tested if high polygenic scores for LDL-C or TG (> 90th percentile) were associated with increased CAD risk in hypercholesterolemic/-triglyceridemic (LDL-C or TG > 90th percentile) population samples ($n = 2249$ and 1709, respectively) or family members ($n = 230$ and 166, respectively). High polygenic score for LDL-C was associated with a 0.8 (0.7-0.8) mmol/l increase in LDL-C values and a 21 % increase in CAD risk in all population samples (HR 1.21 [0.99-1.48]). In hypercholesterolemic population samples, the polygenic LDL-C increase was 0.2 (0.1-0.3) mmol/l and did not confer additionally increased CAD risk (HR 1.06 [0.75-1.52]). High polygenic score for TG was associated with a 0.6 (0.5-0.7) mmol/l increase in TG values and a 26 % increase in CAD risk in all population samples (HR 1.26 [1.04-1.54]). Further, in hypertriglyceridemic population samples, high polygenic score for TG increased TG values by 0.5 (0.3-0.8) mmol/l and CAD risk by 63 % (HR 1.63 [1.07-2.48]). Only 4.6 % of the hypertriglyceridemic individuals were treated with fibrates during follow-up. In family-based samples, similar effects on lipid values were observed. High TG due to polygenic burden increases CAD risk more than high TG due to other causes. Measuring polygenic scores for TG may refine risk estimation for CAD in young and/or hypertriglyceridemic individuals.

2660T

Evaluating the role of genetic variants on blood cell count variability in the Jackson Heart Study. *J.R. Shaw[1], R.W. Corty[2], L.M. Raffield[2], A.P. Reiner[3], J.G. Wilson[4], L.A. Lange[1], W. Valdar[2], E.M. Lange[1].* 1) University of Colorado Anschutz Medical Campus, Aurora, CO; 2) University of North Carolina, Chapel Hill, NC; 3) University of Washington, Seattle, WA; 4) University of Mississippi, Jackson, MS.

The realization of precision medicine will require understanding how genetic variation can impact both the mean and variability of patient characteristics. While genome-wide association studies (GWAS) have identified thousands of genetic variants associated with differences in mean values of human quantitative traits, variants associated with trait variability around the mean (vQTL) remain relatively uninvestigated in human populations. Variants associated with trait variability may represent direct genetic influence on phenotypic variability, or indirect influence due to genetic interaction or imperfect linkage disequilibrium with a causal variant. Identifying vQTLs can therefore improve power to detect genetic interactions and aid in finemapping quantitative trait loci. Building on the R package DGLM, we used the double generalized linear model to perform a genome scan on 2257 unrelated African American participants from the Jackson Heart Study (JHS), simultaneously modeling the effects of each variant on the mean and variance of cardiovascular disease-related blood cell count measures including platelet and red and white (WBCC) blood cell counts. Covariates included age, sex, and ten principal components for ancestry. Our models showed good control of type I error for all traits. Our top finding for a vQTL was at chromosome 7 variant rs12535735 (p=1.7x107) for WBCC, an intronic variant in DPP6 that with no evidence of a mean effect (p=0.79). The top mean effect, also for WBCC, was observed at chromosome 1 variant rs2518564 (p=1.3x1072) in the previously established DARC region. Interestingly, this variant also showed nominal evidence for a variance effect (p=7.3x103), suggesting either a genetic interaction (epistasis or environmental) or moderate linkage disequilibrium with the causal variant (likely rs2814778, a previously reported null variant in DARC that is monomorphic in non-African descent populations). Our future plans include using JHS whole genome sequence data and targeted genetic interaction tests to further interrogate this region. Finally, we will increase our analysis sample size using additional cohorts and expand the phenotypes tested to both identify variants that directly influence the variance of cardiovascular disease-related quantitative traits, and to identify subsets of mean-effect variants for follow-up of possible interaction or haplotypic effects.

2661F

Elevated genetic risk for coronary artery disease increases hospitalization burden and mortality. *M. Sjögren, P. Almgren, O. Melander.* Clinical Sciences, Malmö, Lund University, Malmö, Sweden.

Background: Coronary artery disease (CAD) is one of the leading causes of mortality and a heavy burden on health care worldwide. Several genome wide association studies (GWAS) have identified common variants associated with CAD. Combining these SNPs into a genetic risk score (GRS) can estimate an individual's genetic burden and individuals in the highest quintile of GRs have nearly doubled risk of CAD. **Aim:** The aim of our study was to investigate whether GRS for CAD could predict future hospitalization and mortality in individuals from Malmö Diet and Cancer study (MDC). **Materials & Methods:** MDC is a populations based prospective study from the city of Malmö in southern Sweden. 23 594 individuals without CAD at baseline and with complete data for all covariates were investigated. The weighted GRS was constructed from 50 SNPs previously associated with CAD at GWAS significant level. The GRS was divided into quintiles and three groups were compared; low genetic risk (quintile 1, reference group), intermediate genetic risk (quintiles 2-4) and high genetic risk (quintile 5). Negative binomial regression was used to investigate hospitalizations and Cox proportional Hazards model was used for mortality analysis, all analyses were adjusted for age, sex, follow-up time, hypertension, diabetes, smoking, ApoA1 and ApoB. Data on hospitalizations and mortality were retrieved from Swedish National Inpatient registry and Swedish National Cause of Death Registry. **Results:** Individuals with higher genetic risk of CAD were hospitalized more often compared to individuals with low genetic risk (Intermediate genetic risk (IGR): IRR 1.07 [95% CI 1.02-1.12], p=0.005; High genetic risk (HGR): 1.10 [1.04-1.16], p=0.001) and this was mainly due to cardiovascular causes (IGR: 1.17 [1.09-1.26], p=0.00003; HGR: 1.31 [1.20-1.43], p=5.17 x 10[-10]). No association was found to hospitalization due to non CVD causes. Individuals with high and intermediate genetic risk were approximately 16% and 9% more likely to die of any cause during the follow-up as compared to individuals with low genetic risk. These individuals had highly increased risk of CVD mortality (IGR: HR 1.23 [1.08-1.39], p=0.001; HGR: 1.44 [1.25-1.66], p=6.56 x 10[-7]) but not the risk of mortality due to other causes. **Conclusion:** Our results suggest that genetic predisposition for CAD can predict hospitalization burden and premature death, especially due to cardiovascular causes, independently of traditional risk factors.

2662W

Uncovering the genetic determinants of variation in arterial stiffness through joint location and scale association testing. *D. Soave, H.A. Edgington, P. Awadalla.* Informatics and Bio-computing, Ontario Institute for Cancer Research, Toronto, Ontario, Canada.

   Arterial stiffness (AS) plays a major role in hypertension in aging adults and is associated with cardiovascular events such as stroke and heart attack. As individuals age, the structure and function of the arterial wall changes leading to increased stiffening. An improved understanding of the molecular mechanisms underlying AS will help to identify high-risk individuals and better inform therapeutic targets, however, our current knowledge of the genetic architecture underlying AS is limited. GWAS results from the Framingham Heart Study discovered several SNPs significantly associated with AS; however, regulatory variation associated with AS has not been investigated. With the goal of gaining better insight into the genetic determinants of AS we conducted a study integrating genomic and transcriptomic data with measures of the central augmentation index (cAI), a proxy for AS, in a sample of 700 individuals from Quebec. Specifically, we first modeled transcript expression levels as a function of cAI, controlling for age at measurement. We also conducted genome-wide expression quantitative traits loci (eQTL) mapping using main effects and interactions between genetic variant and cAI. Throughout our study, we incorporated joint location and scale testing to increase the power to identify genetic factors involved in complex genetic architecture, such as gene-gene interaction, not being directly modeled. Our findings show that gene-transcript relationships depend on AS variation. Additionally, we uncovered gene-transcript signatures associated with AS using our joint modeling approach that would otherwise be left undetected through traditional location-only testing.

2663T

Genetics and outcome of noncompaction cardiomyopathy: A Dutch multicenter study. *J. van Waning[1], Y.M. Hoedemaekers[2], K. Caliskan[5], M.A. van Slegtenhorst[1], J.D.H. Jongbloed[2], K.Y. van Spaendonck[4], A.F. Baas[3], R.H. Lekanne Dit Deprez[4], D. Dooijes[3], R.M.W. Hofstra[1], A. Ijpma[1], D. Majoor-Krakauer[1], Dutch NCCM consortium.* 1) Clinical Genetics, Erasmus Medical Center, Rotterdam, the Netherlands; 2) Clinical Genetics, University Medical Center Groningen, the Netherlands; 3) Clinical Genetics, University Medical Center Utrecht , The Netherlands; 4) Clinical Genetics, Amsterdam Medical Center Department, Amsterdam, the Netherlands; 5) Cardiology, Erasmus Medical Center, Rotterdam, the Netherlands.

   This is a large multicenter cohort study with 327 (52 children and 275 adult) Noncompaction cardiomypathy (NCCM) patients from four major cardiogenetic centers in the Netherlands. This study outlined the differences and similarities in genetic causes between children and adults patients. According to the results of DNA testing of cardiomyopathy gene panels and family histories the 327 NCCM index patients were classified into three groups; 1) genetic; with a mutation, ie (likley) pathogenic variant, 2) probably genetic; familial cardiomyopathy without a (likely)pathogenic variant, or 3) sporadic; no family history, no (likley)pathogeneic variant. In this way we presented novel insight in long-time follow up separately for children and adults from these different groups of patients. Genetic NCCM occurred in 32% (81 adults; 23 children), probably genetic in 16% (45 adults; 8 children) and 52% of the patients were sporadic (149 adults; 21 children). *MYH7*, *MYBPC3* and *TTN* explained 71% of genetic NCCM. Older NCCM patients were more likely sporadic (OR 0.987 per year; CI 95% 0.975-0.99; p=0.029). Genetic NCCM, with multiple mutations was more frequent in children (p=0.04). Left ventricular (LV) dysfunction was more pronounced in genetic patients (p=0.02), in particular with *TTN* or multiple mutations (p=0.02). Risk for major adverse cardiac events (MACE) was related to LV dysfunction (HR 2; p=0.01), age at diagnose before one year (HR 3; p=0,05), and multiple mutation in *MYBPC3* (HR 5; p=0,01). *MYH7* related NCCM had low risk of MACE. In conclusion, almost half the patients had (probably) genetic NCCM. The most important causes for genetic NCCM were mutations in *MYH7*, *MYBPC3* and *TTN* gene. Young children were more likely to have severe disease caused by complex genotypes. In the large proportion of cases that we classified as sporadic, we observed a better prognosis. Major cardiac events in the sporadic patients were less related to NCCM than in the patients with a mutation. This suggests a clinically distinct form of NCCM in sporadic, mostly adult cases, and has implications for revising the management of these cases and their families. From a clinical perspective age and genetics are important predictors of outcome of NCCM, highlighting the role of genetics testing and a detailed family history in management of and clinical surveillance of NCCM.

2664F

Evaluating the burden of pathogenic variants for the inherited arrhythmia syndromes. *Y.P. Fu[1], J. Kaltman[2].* 1) Office of Biostatistics Research, Division of Cardiovascular Sciences, NHLBI, NIH, Bethesda, MD; 2) Heart Development and Structural Diseases Branch, Division of Cardiovascular Sciences, NHLBI, NIH, Bethesda, MD.

Background: Inherited arrhythmia syndromes (IAS), including long QT syndrome (LQTS), Brugada syndrome (BrS), catecholaminergic polymorphic ventricular tachycardia (CPVT), short QT syndrome (SQTS), idiopathic ventricular fibrillation (IVF), and progressive cardiac conduction system disease (PCCD), are important causes of sudden cardiac death. Family-based genetic studies have identified many pathogenic variants underlying these diseases but the burden of these variants in the general population was rarely investigated. **Objective:** To evaluate the distribution of pathogenic variants associated with IAS among the 8 populations in the Genome Aggregation Database (gnomAD), and examine the functional impact of variants in different protein domains in the 3 most commonly mutated genes, *KCNQ1*, *KCNH2*, and *SCN5A*. **Methods:** We extracted disease-causing (DM) and likely disease-causing (DM?) variants associated with IAS from the Human Gene Mutation Database (HGMD) and merged them with the gnomAD dataset (n=123,136) to assess the overall and population-specific allele frequencies (AF). The function of individual variants was predicted by MetaSVM. We also evaluated the intolerance of genetic sub-regions using the Residual Variation Intolerance Score (sub-RVIS) in *KCNQ1*, *KCNH2*, and *SCN5A*. **Results:** A total of 2129 IAS variants were identified across 58 genes from the HGMD, where 583 (27.4%) variants were also presented in the gnomAD. Among these 583 disease-causing variants, 135 (23.2%) had an AF ≥ 0.01% and 27 (4.6%) had an AF ≥ 0.1% in the overall gnomAD subjects. The AF distribution across the 8 sub-populations in gnomAD was significantly different (p<0.0001), where the proportion of AF ≥ 0.01% ranged from 7.5% in the Finns population to 25.6% in those without assigned ethnicity. The prediction algorithm MetaSVM estimated that 50.8% of the gnomAD variants with AF ≥ 0.01% were functionally tolerated, compared to 22.6% of gnomAD variants with AF < 0.01% (p<0.0001). In *KCNQ1*, *KCNH2* and *SCN5A*, the variant distribution was significantly different by functional domains: most pathogenic variants located in the ion channel domains were not observed in the gnomAD, and many gnomAD variants located in regions without specific function (all p<0.0001). **Conclusion:** A significant number of disease-causing IAS variants in HGMD have a relatively common allele frequency and insufficient evidence for pathogenicity. The AF distribution also varies across different populations.

2665W

The importance of epistatic interactions and fitness costs in congenital heart disease. *E.O. Akhirome[1], S.D. Regmi[1], R. Magnan[1], Y. Qin[1], J.M. Cheverud[2], P.Y. Jay[1].* 1) Pediatrics & Genetics, Washington University School of Medicine, St. Louis, MO; 2) Biology, Loyola University Chicago, Chicago, IL.

Although congenital heart disease (CHD) is the most common type of birth defect, its genetic basis is still poorly understood. Efforts to identify a causal mutation fail in ~70% of cases. Even when one is found, the same mutation can cause different heart defects or sometimes none. This complexity suggests that CHD risk depends on epistasis – interactions between multiple genetic elements. Here, we demonstrate the importance of epistasis in the pathogenesis of CHD. As in humans, *Nkx2-5* haploinsufficiency elicits a range of incompletely penetrant heart defects in our mouse model. We identified genetic loci that epistatically interact with *Nkx2-5*[+/-] (G×G$_{Nkx}$) and each other (G×G×G$_{Nkx}$) to modify risk for distinct heart defects. We found a positive correlation between a defect's fitness cost and the magnitude of its associated epistatic effects, meaning that epistasis influences the risk for severe defects more than mild defects. For instance, one G×G×G$_{Nkx}$ interaction explained 50% of individuals with severe atrioventricular septal defects. Remarkably, G×G×G$_{Nkx}$ effects were 3 times more likely to suppress heart defect risk when the alleles at both modifier loci were inherited from the same inbred strain rather than different strains. Our findings unveil the oft-overlooked contributions of epistasis to the genetic architecture of CHD and other complex traits. Moreover, these results underscore the importance of epistasis for phenotype prediction in personalized medicine.

2666T

African ancestry genome- and transcriptome-wide association study of blood pressure detects nine novel loci in a large cohort from the Million Veteran Program. *J.N. Hellwege[1], A. Giri[1], J. Liang[2], E.S. Torstenson[1], O. Wilson[3], Y.V. Sun[4], P.W.F. Wilson[5], P.S. Tsao[6], C.P. Kovesdy[7], K.A. Birdwell[8], N. Franceschini[9], X. Zhu[2], D.R. Velez Edwards[10], C. O'Donnell[11], A. Hung[3], T.L. Edwards[1], the COGENT BP Consortium; on behalf of the VA Million Veteran Program.* 1) Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN; 2) Department of Population Quantitative Health Science, School of Medicine, Case Western Reserve University, Cleveland, OH; 3) VA TVHS Nashville, Division of Nephrology & Hypertension, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Epidemiology, Emory University Rollins School of Public Health; Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, GA; 5) Atlanta VAMC and Emory Clinical Cardiovascular Research Institute, Atlanta, GA; 6) VA Palo Alto Health Care System; Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 7) Nephrology Section, Memphis VA Medical Center, Memphis, TN; 8) Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 9) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 10) Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center; Tennessee Valley Health Systems VA, Nashville, TN; 11) VA Boston Healthcare, Section of Cardiology and Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

African Americans (AA) experience a high burden of hypertension and related complications, but have been underrepresented in genetic studies of blood pressure traits. To identify novel loci for blood pressure (BP) and hypertension in AAs, we performed a genome-wide association study (GWAS) of systolic (SBP) and diastolic blood pressure (DBP) in the largest group of AAs to date in a single cohort (N ~ 59,900) from the Million Veteran Program (MVP). We evaluated BP and GWAS data imputed to the 1000 Genomes reference panel in common variants (frequency > 1%) with additive genetic models adjusted for age, age-squared, sex, body mass index, and the top ten principal components. We then performed *in silico* replication and meta-analysis with previously calculated summary statistics from the COGENT BP Consortium study of AA BP genetics (N ~ 31,000). GWAS in MVP detected 99 SNPs representing 10 loci associated with SBP and 155 SNPs from 10 loci associated with DBP at p-value $\leq 5\times10^{-8}$. Seven loci for SBP and four for DBP were previously identified. Summary statistics from 396 lead SNPs (p-value < 1×10^{-5}) were obtained from COGENT and after meta-analysis all loci identified in MVP were more significant. The most significant SNP for DBP after meta-analysis was at *CACNA1D* (p-value = 1.26×10^{-12}), while the most significant SNP for SBP had p-value = 7.02×10^{-13} ($r^2 = 0.27$). The top novel results that have not been previously reported in studies of BP in African, Asian, or European populations were at *KCNN3* (rs1984285; p-value = 1.10×10^{-8}) for SBP and *HTR4* (rs4554204; p-value = 3.84×10^{-11}) for DBP. Additional novel loci include *LRRC38* and *RXFP2* for SBP and *OXSR1, CCR8, EIF1B, CCDC12*, and *RBMS2* for DBP. We further evaluated the relationship between genetically predicted gene expression and BP traits in 44 tissues based on the genotype-tissue expression (GTEx) project data using summary statistics from the MVP GWAS with the MetaXcan method. Increased predicted expression of *CCDC71* in pancreas was associated with DBP (p-value = 1.97×10^{-8}), while predicted reduced expression of *TMEM110* (near the *CACNA1D* GWAS signal) in the nucleus accumbens-basal ganglia of the brain was associated with SBP (p-value = 1.12×10^{-8}), suggesting that *TMEM110* may be the relevant gene in that region. Further studies are underway with additional independent AA samples and meta-analysis of all available variants from COGENT and MVP.

2667F

Exome sequencing identifies multiple variants associated with glycA, a novel biomarker of cardiovascular events. *L.C. Kwee[1], S. Giamberardino[1], J. Overton[2], J. Reid[2], S.G. Gregory[1], R. McGarrah[1], A. Baras[2], C.B. Newgard[1,3], W.E. Kraus[1], S.H. Shah[1].* 1) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 2) Regeneron Pharmaceuticals, Inc., Tarrytown, NY; 3) Sarah W. Stedman Nutrition and Metabolism Center, Duke University, Durham, NC.

GlycA is a novel NMR-based inflammatory biomarker associated with incident cardiovascular (CV) events and all-cause mortality. The glycA signal reports on N-acetyl methyl groups of several glycoproteins in the acute-phase response, and exhibits less intra-individual variability than other available inflammatory biomarkers. We hypothesized that genetic variation may explain some inter-individual variability in glycA levels, and that associated variants may help clarify the biological underpinnings of the relationship between glycA and poor CV outcomes. We conducted an association study using whole-exome sequencing and glycA levels from subjects in the CATHGEN biorepository of cardiac catheterization patients. After QC, data were available from 5882 subjects of European ancestry (EUR) and 1671 subjects of African ancestry (AFR). Using ancestry-stratified linear models, we tested 328,608 variants (MAF>1%) for association with glycA levels under an additive genetic model. Our initial model adjusted for sex, age, batch, and three ancestry-based principal components. Variants associated with glycA (p<5×10^{-8}) were then tested in a clinical model including diabetes, hypertension, hyperlipidemia, smoking, BMI, coronary artery disease, and ejection fraction. In the initial model, nine variants were associated with glycA levels in AFR individuals and 19 were associated in EUR individuals. All of the AFR variants and 14 of the EUR variants lie in a 200kb region on chr16 spanning several genes (*PKD1L3, HP, HPR, DHODH, TXNL4B, PMFBP1*), including the top two overall hits: rs5471 (p=2×10^{-21}, AFR) and rs11648353 (p=8×10^{-13}, EUR). Three of the chr16 variants show opposite directions of effect in the two subgroups. The remaining EUR variants are in *PMS1* (min. p=6×10^{-9}) and *TNC* (min. p=2×10^{-8}). Variants remained associated with glycA levels after adjustment for clinical covariates. Conditional analyses suggest five independent chr16 loci associated with glycA in EUR samples and two in AFR samples. Using exome sequencing in a large CV cohort, we have identified genetic loci associated with inter-individual variation in glycA, a novel inflammatory biomarker. These loci highlight potential genetic markers of high CV risk individuals and may help elucidate the genetic architecture of glycA. The lack of overlap with genetic loci for other inflammatory biomarkers (e.g. CRP) suggests that glycA potentially mediates CV risk through different inflammatory pathways.

2668W

Genetic determinants in the *LILR* gene family predicting statin intolerance. *M.K. Siddiqui[1], D. Carr[2], M. Pirmohamed[2], P. Ridker[3], D. Chasman[3], C.N.A. Palmer[1].* 1) Clinical and Molecular Medicine, University of Dundee, Dundee, United Kingdom; 2) University of Liverpool, United Kingdom; 3) Brigham and Women's Hospital, Harvard Medical School, USA.

The *LILR* are a gene family located in the leukocyte receptor complex on chromosome 19; encoding inhibitory and stimulatory receptors. There is increasing interest in their role in immune system function, as certain regions are known to interact with the HLA class I molecules, and might have immunomodulatory functions. In the sub-family *LILRB5* (Asp247Gly), is associated with muscle breakdown enzymes: serum creatine kinase (CK) and lactate dehydrogenase levels. Evidence suggests the variant is also associated with statin intolerance and myalgia. Due to their potentially interconnected nature, we decided to further explore the gene family and their role in statin intolerance. An association study of this region with statin intolerance was performed in the GoDARTS population (University of Dundee) and the results were meta-analyzed with cases of statin-induced myopathy from the CPRD-STAGE cohort (University of Liverpool). The meta-analysis showed a strong signal from SNPs in *LILRA3* and *LILRB2*. The strongest candidate SNP in *LILRB2*, (His20Arg) was selected, OR 1.48 (95% CI: 1.19, 1.85), *P-value*=4×10^{-4}. This variant was in strong linkage disequilibrium (LD) with neighbouring SNPs showing congruent effects. His20Arg was not in LD with the previously characterized myalgia variant- *LILRB5* Asp247Gly (r^2 = 0.04, D' = 0.54). Conditional analyses in GoDARTS showed carriers of both the *LILRB2* Arg37 and *LILRB5* Asp247 variants had higher odds of developing statin intolerance OR 1.44 (95% CI: 1.002, 2.06), *P-value*=0.04. The joint effect of these variants was replicated in the JUPITER trial, when examining the changes in CK levels between the start and end of the trial, as well as for the development of myalgia. We present a novel region of interest in the *LILR* family for statin intolerance and myalgia. The findings show the independent roles that individual LILR genes might play in statin intolerance and predisposition to statin-independent myalgia. These findings lend more evidence to the hypothesis of the immune system's involvement in statin intolerance, muscle repair and maintenance. The findings warrant further investigation into the immune-physiology of statin-induced muscle damage.

2669T

Mendelian randomization to identify causal risk factors for atrial fibrillation. *L. Weng[1,2], S.A. Lubitz[1,2,3], AFGen Consortium.* 1) Cardiovascular Research Center, Mass General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA; 3) Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA.

Background: Atrial fibrillation (AF) affects nearly 34 million individuals worldwide and is associated with substantial risks for stroke, dementia, heart failure, and death. Multiple clinical risk factors for AF have been identified. Understanding which risk factors are causally related to AF may help direct efforts focused on AF prevention. **Methods:** We tested whether a comprehensive set of established clinical AF risk factors are causally associated with AF using a two-sample Mendelian randomization approach. From published genome-wide association studies of each risk factor, we identified genetic variants associated with the risk factors (systolic blood pressure, body mass index, estimated glomerular filtration rate, smoking, type 2 diabetes, and coronary artery disease). Variants with p-values exceeding study-specific significance thresholds were included in our analyses. Variant effect sizes for AF were obtained from a prior genome-wide association study of AF. We used Wald-type estimation to calculate associations between risk factors and AF for each variant. We then summarized the effects across variants to evaluate the unconfounded estimates for each risk factor using both fixed and random-effects meta-analyses. Additionally, we evaluated associations after excluding potentially pleiotropic SNPs to minimize bias. **Results:** Genetically driven blood pressure and body mass index were both positively associated with AF risk. Each 10 mm Hg higher systolic blood pressure was associated with a 1.21-fold increased odds for AF (95%CI 1.05-1.39; n=25 variants). Each one kg/m^2 higher body mass index was associated with a 1.08-fold increased odds for AF (95% CI 1.04-1.13; n=32 variants). The results were similar after removing potentially pleiotropic variants. We did not observe significant associations between AF and genetically-determined estimated glomerular filtration rate, smoking, type 2 diabetes, or coronary artery disease. **Conclusions**: We found that increased blood pressure and body mass index are likely causally related to AF. Public health efforts focused on improving blood pressure and reducing body mass index may lower the incidence of AF and morbidity attributable to the arrhythmia.

2670F

Localization and dissection of a major quantitative trait locus influencing vascular endothelial growth factor plasma levels. *M. Almeida[1], M.P. Johnson[1], J.M. Peralta[1], J.E. Curran[1], E.K. Moses[2], J. Blangero[1]. 1) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley. One West University Blvd. STDOI Modular Building #100 Brownsville, TX 78520; 2) Centre for Genetic Origins of Health and Disease (GOHaD) The University of Western Australia (M409) 35 Stirling Highway CRAWLEY WA 6009 Australia.*

The use of large human pedigrees to investigate the association between genetic variants and phenotypes of interest has gained a new impulse by the development of cost-effective sequencing platforms. They are able to detect millions of rare genetic variants exclusive to a single family or to a population under study. The San Antonio Family Heart Study (SAFHS) is a long running (~26 years) genomics project with considerable contributions for the identification of genetic *loci* associated with human complex diseases. The SAFHS now has approximately 1,600 complete genomes from 50 large extended families of Mexican American individuals living in the South Texas. We identified over 29 million SNVs (Single Nucleotide Variants) and among them 18.2 million were rare variants (MAF<= 0.01). Rare genetic variants are generally missed in conventional genome wide association studies or are captured in too few numbers to permit variant-specific analysis in sequencing studies of unrelated individuals. Vascular endothelial growth factor (VEGF) plays a central role in angiogenesis and SNVs within the gene have been associated with arteriosclerosis, diabetes, preeclampsia and cancer. We isolated plasma derived VEGF and quantified total protein levels in a set of 1,210 SAFHS participants. VEGF concentrations were highly heritable (h^2=0.7) in our sample. We searched for QTLS influencing VEGF levels using empirical IBD matrices in SOLAR and localized a major QTL (LOD = 10.84) on chromosome 6 near the *VEGFA* structural locus. We tested SNVs in this QTL region for association with VEGF concentration using a linear mixed model and significant genetic association were detected. The leading association was the common SNV rs6921438 (p-value=$2.07×10^{-47}$) located in a highly conserved sequence region 200 kb downstream of the gene *VEGFA*. Conditional multi-variant analysis in the gene region revealed 5 independent putative functional variants accounting for 23.4% of the total phenotypic variation including a rare variant, rs73736911, with a very strong biological effect (+0.7 SD). These SNVs completely accounted for the original linkage signal (residual LOD= 0.07). These potential functional variants will require additional independent biological validation. Our results highlight the unique power of whole genome sequencing for QTL dissection when applied to quantitative risk factors in large human pedigrees.

2671W

A genome-wide gene by cigarette smoking interaction study on elevated blood pressure. *M. Kang, J. Sung. Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, South Korea.*

Introduction: Epidemiologic studies showed hypertension and cigarette smoking might result in a synergistic effect on the atherosclerotic cardiovascular disease risk, but the susceptible genes for this interaction are not well investigated. Although dozens of genetic markers were associated with hypertension, fewer studies were performed about genes that interact with smoking behavior to increase blood pressure. A gene-environment wide interaction study(GEWIS) is an effective approach particularly designed to identify these interactive genes. **Methods:** 18,025 individuals from two Korean cohorts, the subcohorts of the Korean Genome and Epidemiology Study(KoGES), were involved. Systolic and diastolic blood pressure were measured following the American Heart Association(AHA) guideline, and smoking status(current and ever smoking) were assessed using valid questionnaires. We conducted an exhaustive GEWIS on elevated blood pressure(EBP) for 4.1 million SNPs imputed using the 1000 genomes GRCh37 reference panel: the risk of EBP was adjusted for age, age^2, sex, and BMI in each logistic regression model. We applied various exhaustive scans(case-control, case-only, and empirical Bayesian approaches) and two-step methods(Hybrid, Cocktail, and EDGxE) in parallel. For exhaustive scans, the genome-wide significance level of $p<5×10^{-8}$ was applied to screen GxE interactive genes; step-wide penalties according to the marginal *p*-value were applied for two-step methods. **Results:** We identified five loci which modified the risk of EBP by current smoking: three loci near known hypertensive genes(*ABCG8*, *ATP2B1*), and two loci near *CWC27*, *PRSS45*. About 4.1% of Korean population had risk alleles of the locus near *ABCG8* which synergistically increase the risk of EBP with current smoking. The heritability of EBP explained by *ABCG8* was 0.9% by a conventional genome-wide association study(GWAS), but was increased to 5.7% when smoking habits were considered in a GEWIS. All of the identified GxE interactive SNPs from a discovery set were replicated with the other study population: *p*<0.01 was applied for the replication *p*-value. **Conclusion:** The identified SNPs in this GEWIS might be used as genetic markers to give a personalized guideline for those who are particularly susceptible to smoking consequences. Considering substantial differences between the conventional marginal GWAS and GEWIS, some amount of missing heritability should be explained by discovering more GxE interactive loci.

2672T

Large-scale validation of zebrafish larvae as a model system for genetic screens in dyslipidaemia, atherosclerosis and coronary artery disease.

M. Bandaru[1,2], A. Emmanouilidou[1,2], P. Ranefall[2,3], B. von der Heyde[1,2], T. Klingström[1,2,4], J. Ledin[2,4], A. Larsson[5], C. Wählby[2,3], E. Ingelsson[6], M. den Hoed[1,2]. 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Information Technology, Uppsala University, Uppsala, Swedeny; 4) Department of Organismal Biology, Uppsala University,, Uppsala, Sweden; 5) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 6) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford.

Background: Genome-wide association studies have identified 77 loci that are robustly associated with coronary artery disease (CAD). In all but a few of these loci the causal genes and mechanisms remain unknown. Results from small-scale studies suggest that zebrafish larvae represent a promising model system for genetic screens in dyslipidemia, early-stage atherosclerosis and CAD. We aim to confirm or refute these results in a large-scale study, expand the phenotypic pipeline, and increase the throughput. **Methods:** At the core of our setup is an automated positioning and imaging system that allows visualization and quantification of atherogenic traits in ~100 zebrafish larvae per day at 10 days post-fertilization, by making use of fluorescent transgenes and dyes. We used a three-tiered approach to validate the zebrafish model system: 1) a dietary intervention to examine the effect of overfeeding and cholesterol supplementation (N=2193); 2) a treatment regime with atorvastatin and ezetimibe (N=956); and 3) a genetic screen for zebrafish orthologues of *LDLR*, *PCSK9*, *APOB* and *APOE* using a multiplex CRISPR-Cas9 approach (N=2x384). After imaging, whole-body lipid and glucose levels were assessed using enzymatic assays, and CRISPR-Cas9 target sites were sequenced on a MiSeq. **Results:** Overfeeding and cholesterol supplementation have independent pro-atherogenic effects, including elevated total cholesterol and triglyceride levels, more vascular deposition of lipids and oxidized LDLc, and more co-localization of lipids with macrophages and neutrophils. Treatment with atorvastatin and ezetimibe results in lower whole-body total cholesterol, LDLc and triglyceride levels, as well as in less vascular lipid deposition and less co-localization of lipids and macrophages. Finally, mutations in *APOE* orthologues result in higher whole-body LDLc levels and more co-localization of lipids with macrophages or neutrophils compared with wildtypes. Mutations in *APOB* orthologues tend to result in higher LDLc levels, more vascular lipid deposition, and more co-localizing lipids and neutrophils. Treatment with lipid lowering drugs and mutations in pcsk9 both result in higher whole-body glucose levels. Data from all larvae combined show that atherosclerosis in 10-day-old zebrafish larvae is mainly driven by higher triglyceride but not LDLc levels. **Conclusion:** Zebrafish larvae can be used to systematically identify and characterize causal genes for CAD.

2673F

Sequence data processing and analysis of 70,000 human genomes in the NHLBI TOPMed sequencing program.

T. Blackwell[1], G. Abecasis[1], W. Gan[2], S. Gogarten[3], C. Jaquish[2], H.M. Kang[1], C.C. Laurie[3], G. Papanicolaou[2], J.G. Wilson[4], Q. Wong[3], Study Investigators of the NHLBI TOPMed Sequencing Program. 1) TOPMed Informatics Research Center, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 2) National Heart, Lung and Blood Institute, NIH, Bethesda Md; 3) TOPMed Data Coordinating Center, Department of Biostatistics, University of Washington, Seattle, Wa; 4) VA Medicl Center and the University of Mississippi Medical Center, Jackson Mississippi.

The National Heart Lung and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) initiative aims to stimulate discovery of fundamental mechanisms that underlie heart, lung, blood and sleep disorders by adding whole genome sequence data to existing high quality phenotypes and clinical diagnoses. The first two years of the project have generated deep whole genome sequence data for > 70,000 diverse samples, of which 46% have non-European ancestry. These populations are traditionally under-represented in genomic research. This project represents a collaboration between diverse groups of scientists, 7 genome sequencing centers, a project data coordinating center, an informatics research center and staff at the National Institutes of Health and the National Center for Biotechnology Information. Here, we describe the process of integrating sequence data across participating studies and sequencing centers being carried out at the project's Informatics Research Center (IRC). Currently, > 70,000 samples from 31 studies have been sequenced to an average sequencing depth of 38.3x (after duplicate removal), resulting in 98.9% of the genome covered to depth 10 or greater. The current data freeze with 18,526 sequenced individuals yields 199.5 million SNPs and 9.9 million short indels, of which 43.8% and 44.9% are singletons, respectively. Each individual carries a non-reference allele at between 3.5 and 4.5 million sites, depending on ancestry. As of abstract submission, variant filtering is ongoing for a new data set which will cover 64,960 individuals. Deep sequencing, PCR-free sample preparation protocols and harmonized data processing allow us to achieve < 5×10^{-5} genotype discordance when duplicate samples are sequenced at different sites. A primary goal of the project is to allow investigators employing many different study designs and phenotypes to benefit from access to high-quality sequence data and analysis. There are also many opportunities for combined analysis of project data, including a diverse catalog of genetic variation, a potential resource of deeply sequenced controls for genetic association studies, and an imputation reference panel. We discuss some of these opportunities and associated logistical and technical challenges, as well as opportunities for any scientist to access project results.

2674W

Genome-wide association study using whole-genome sequencing recapitulates both rare and common risk alleles for Brugada syndrome. *R. Redon[1], M. Karakachoff[1], P. Lindenbaum[1], J. Barc[1], S. Le Scouarnec[1], A. Boland[2], D. Bacq[2], V. Meyer[2], J. Giemza[1], J.B Gourraud[1], A. Leenhard[3], P. Guicheney[4], E. Genin[5], V. Probst[1], C. Dina[1], J.F. Deleuze[2], J.J. Schott[1], the RHYTHMOGEN consortium.* 1) L'institut du thorax, INSERM, CNRS, UNIV Nantes, CHU Nantes, Nantes, France; 2) Centre National de Recherche en Génomique Humaine, CEA, Evry, France; 3) Department de Cardiology, Bichat Hospital, AP-HP, Paris, France; Université Paris Diderot, Sorbonne Paris Cité, Paris, France; 4) INSERM UMR 1166, Faculty of Medicine, UPMC, Paris, France; 5) INSERM UMR 1078, CHRU Brest, UNIV Brest, Brest, France.

   The Brugada syndrome (BrS) is an inherited cardiac disorder associated with ST-segment elevation in the right precordial leads and a high risk for sudden cardiac death. Although rare variants in *SCN5A* (encoding the major cardiac Na-channel isoform) are found in ~20% of cases, the genetic basis of the disorder remains largely unresolved. In a previous genome-wide association study on 312 patients with BrS, we had identified 3 common haplotypes predisposing to the disorder, including two at the *SCN5A-SCN10A* locus. Additionally, by testing the burden of rare coding variation in 45 known or suspected arrhythmia-susceptibility genes among 167 patients with BrS, we have recently reported a significant enrichment in rare coding variation (with a minor allele frequency below 0.1%) only for *SCN5A*. Here we aimed to assess whether association testing based on whole genome sequencing would allow for simultaneous detection of both rare and common alleles associated to BrS. After sequencing the whole genomes of 328 French index cases with BrS and 404 ancestry-matched reference individuals to depth greater than 30X on Illumina HiSeq X station, we applied classical association scanning and gene burden testing on genotypes produced by multi-sample calling in line with the Broad/GATK best practices. By this strategy, we could reproduce both the previously reported association signals at the *SCN5A-SCN10A* locus and the significant enrichment in rare coding variation in the *SCN5A* gene. These results demonstrate the relevance of whole-genome sequencing to interrogate the full spectrum of genetic variation, from rare variants to common haplotypes, associated with susceptibility to complex disorders. We are currently investigating whether our whole-genome sequencing approach enables identifying additional low-frequency as well as structural variants conferring higher susceptibility to BrS, which could hardly be detected using former strategies.

2675T

Genome-wide association study of susceptibility to rheumatic heart disease in South Asians: Preliminary results. *K. Auckland[1], A. Mentzer[1], S. Kumar[2], M. Bhatt[3], N. Garg[3], J. Kado[4], M.L. Perman[4], A. Steer[5], A. Hill[1], B. Mittal[2], T. Parks[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Babasaheb Bhimrao Ambedkar University, Lucknow, Uttar Pradesh, India; 3) Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India; 4) Fiji National University, Suva, Fiji; 5) Murdoch Children's Research Institute, Melbourne, Victoria, Australia.

   Background: Rheumatic heart disease (RHD), a chronic consequence of an abnormal immune response to *Streptococcus pyogenes*, affects 16 million individuals worldwide, causing an estimated 300,000 premature deaths each year. Previously a major public health concern in Europe and the United States, the disease remains a leading cause of morbidity and premature mortality in developing countries. Recently, large-scale genetic studies have begun to provide much needed insight into the underlying pathogenesis. We therefore sought to identify common genetic variants affecting susceptibility to RHD in South Asians. **Methods:** We undertook a genome-wide association study of RHD susceptibility in 1163 individuals of Indian ancestry recruited in Fiji (n=309) and Northern India (n=854). Patients with incident or prevalent RHD were recruited as cases, while members of the general population were recruited as controls. We genotyped all individuals at ~250,000 variants using the Illumina HumanCore platform before estimating the genotypes of an additional ~5,000,000 variants by imputation. We used linear mixed models to analyse genotyped and imputed variants before combining association statistics from the two datasets using fixed-effects meta-analysis. **Results:** We observed a single signal at genome-wide significance, located in the human leukocyte antigen (HLA) class III region (OR=1.91, P=6.8 x 10^{-9}). Using conditional analyses, we demonstrated this was comprised of two independent signals, the first spanning HLA class I (HLA-B) and the second HLA class II (HLA-DQA1/B1). While the *IGHV4-61* signal was apparent, as previously reported in the Fijian Indian data (P=0.006), the signal was absent in the Northern Indian data (P=0.25), leaving only a negligible signal in the pooled analysis (OR=1.21, P=0.084). **Discussion:** By combining new data from India with previously reported data from Fiji, we provide the first insight into the genetic architecture of RHD susceptibility in South Asians with a promising finding in the HLA locus. To our knowledge, South-East Asian populations are poorly represented in existing HLA imputation reference panels and as such, the imputation of HLA classical alleles is likely to be inaccurate in these populations. To address these problems, we are building and validating a custom imputation panel optimised for South Asian populations and will provide an update on these efforts and its application to our GWAS at the meeting.

2676F

Updated genome-wide association study and functional annotation reveals new risk loci for mitral valve prolapse. *N. Bouatia-Naji[1], M. Yu[1], Z. Alsalman[1], J.J. Schott[2], T. Le Tourneau[2], P. Galan[3], A. Hagege[1], X. Jeunemaitre[1], C. Dina[2].* 1) Paris Cardiovascular Research Center U970 HEGP Research Center, INSERM, Paris, France; 2) INSERM UMR 1087, CNRS UMR 6291, Institut du Thorax, Centre Hospitalier Universitaire (CHU), Nantes University, Nantes, France; 3) Nutritional Epidemiology Research Group, Sorbonne-Paris-Cité, UMR University of Paris 13/Inserm U-557/INRA U-1125/CNAM, Bobigny, France F-93017, Bobigny, FRANCE.

Mitral valve prolapse (MVP) is common degenerative valve disease, a risk factor for arrhythmia, atrial fibrillation and sudden death and the leading indication for surgical repair of mitral regurgitation. We have previously reported 6 risk loci for MVP and aimed to use high-resolution imputation to improve our genetic map density and increase chances to find additional risk loci. To achieve these goals, we used genotypes from two European case control cohorts to generate ~40 million genotypes. We used SHAPE-IT for phasing and Minimac 3 for imputation through the online Michigan Imputation Server. We applied MAF specific imputation quality cut-offs as $R^2 > 0.80$ if MAF>0.05, or $R^2 > 0.95$ if 0.01<MAF=<0.05). We performed the association analysis of ~5.2 million genotyped or imputed SNPs using SNPTEST in two case control sets and performed a GWAS meta-analysis using METAL that included 1,590 MVP cases and 2,548 controls. We identified 22 GWAsig SNPs (P<5×10^{-8}). Compared to our previous GWAS based on 1000G imputations, we replicated the association signal located on Chr2 near TNS1 (rs6708579, P = 1.2 ×10^{-8}) and Chr22 near MN1/PITPNB (rs11703195, P= 2.1×10^{-8}) providing further support with additional SNPs to these risk loci. Most interestingly, we identified two new loci on Chr1 (MAF=0.19, OR=1.70, P=4.3×10^{-18}) intronic to *SPAG17* encoding a central pair protein part of cilia structure that we have recently demonstrated to be important to valvular development; and on Chr19 (MAF=0.39, OR=1.37, P=6.6×10^{-10}) near UQCRFS1 encoding a ubiquinol cytochrome c reductase, highly expressed in heart and involved in cardiac muscle contraction. ENCODE annotation of 156 SNPs with P<10^{-5} showed that 43 locate in DNaseI hypersensitive sites and 28 are transcription factors binding sites variants. In summary, our updated GWAS provided improved genetic information on known loci prior to functional exploration and identified new loci with potential candidate genes to be confirmed through replication in additional MVP cohorts.

2677W

Circulating cholesteryl ester transfer protein (CETP) concentration: A genome-wide association study followed by Mendelian randomization on coronary artery disease. *D. Mook-Kanamori[1,2], L.L. Blauw[1,3], R. Li-Gao[1], R. Noordam[4], R. de Mutsert[1], S. Trompet[4,5], J.F.P. Berbée[3,6], Y. Wang[3,6], T. Christen[1], D. van Heemst[4], F.R. Rosendaal[1], J.W. Jukema[5], P.C.N. Rensen[3,6], K. Willems van Dijk[3,6,7].* 1) Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands; 2) Public Health and Primary Care, Leiden University Medical Center, Leiden, Netherlands; 3) Internal Medicine, Div. Endocrinology, Leiden University Medical Center, Leiden, Netherlands; 4) Internal Medicine, Div. Gerontology and Geriatrics, Leiden University Medical Center, Leiden, Netherlands; 5) Cardiology, Leiden University Medical Center, Leiden, Netherlands; 6) Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, Netherlands; 7) Human Genetics, Leiden University Medical Center, Leiden, Netherlands.

Background/aims: Cholesteryl ester transfer protein (CETP) facilitates the net flux of cholesteryl esters from high-density lipoproteins (HDL) towards (very-) low-density lipoproteins ((V)LDL), coupled to a net flux of triglycerides from (V)LDL to HDL. Therefore, inhibition of CETP has been regarded a promising therapeutic strategy to attenuate dyslipidaemia and ultimately prevent the development of cardiovascular disease (CVD). The genetic basis of CETP concentration is largely unknown and therefore we performed a genome-wide association study on circulating CETP. Subsequently, causal effects of variation in CETP concentration on circulating lipids and cardiovascular disease risk were assessed using Mendelian Randomization. **Methods:** A genome-wide association discovery and replication study on serum CETP concentration were embedded in the Netherlands Epidemiology of Obesity (NEO) study. Based on identified variants, Mendelian randomization was conducted on serum lipids (NEO study) and coronary artery disease (CAD) (CARDIoGRAM-plusC4D consortium). **Results:** In the discovery analysis (N=4,248), we identified three independent variants (P<5×10^{-8}; rs247616, rs12720922 and rs1968905) all mapped to *CETP*, that were replicated in a separate subpopulation (N=1,458). Per-allele increase (SE) in serum CETP was 0.32 (0.015) μg/mL for rs247616-C, 0.35 (0.019) μg/mL for rs12720922-A, and 0.12 (0.02) μg/mL for rs1968905-G. Combined, these three variants explained 16.4% of the total variation in CETP concentration. One μg/mL increase in genetically-determined CETP concentration strongly decreased high-density lipoprotein (HDL) cholesterol (-0.23 mmol/L (95% CI: -0.26, -0.20)), but only moderately increased low density lipoprotein (LDL) cholesterol (0.08 mmol/L (95% CI: 0.00, 0.16)). Also, one μg/mL increase in genetically-determined CETP concentration was only marginally associated with an increased risk of CAD (odds ratio: 1.08 (95% CI: 0.94, 1.23)). **Conclusion:** This is the first large study identifying independent variants that largely determine serum CETP concentration, with strong effects on HDL-cholesterol and marginal effects on LDL-cholesterol and CAD risk. In line with CETP inhibitor trials failing to show a reduction in cardiovascular events, our genetics-based study suggests that potential effects of CETP inhibition on cardiovascular disease prevention are minimal.

2678T

A genome-wide association study identifies novel genetic signatures associated with thiazide diuretics adverse metabolic events. *M.H. Shahin[1], A. Akoluk[1], S. Singh[1], C.W. McDonough[1], Y. Gong[1], A.B. Chapman[2], S.T. Turner[3], J.G. Gums[1], A.L. Beitelshees[4], E. Boerwinkle[5], R.M. Copper-De-Hoff[1], J.A. Johnson[1].* 1) Department of Pharmacotherapy and Translational Research, Center for Pharmacogenomics, College of Pharmacy, University of Florida, Gainesville, FL; 2) Department of Medicine, The University of Chicago, Chicago, IL; 3) Division of Nephrology and Hypertension, Department of Medicine, College of Medicine, Mayo Clinic, Rochester, MN; 4) Department of Medicine, University of Maryland, Baltimore, MD; 5) Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, TX.

Thiazide diuretics, including hydrochlorothiazide (HCTZ) and chlorthalidone (CLT), are among the most commonly prescribed first-line classes of antihypertensives in the U.S with more than 100 million prescriptions annually. However, the use of these drugs has been associated with adverse effects, including hypertriglyceridemia. Given the heritable nature of triglyceride levels, genetics may also be influencing the adverse drug effect on triglycerides. Herein, we conducted a genome-wide association study (GWAS) to identify novel genetic biomarkers associated with changes in fasting triglyceride levels in response to thiazide diuretics. For the discovery analysis, we included 186 Caucasians treated with CLT in the Pharmacogenomic Evaluation of Antihypertensive Responses-2 (PEAR-2) study. A GWAS was conducted with adjustment for age, sex, waist circumference, baseline triglycerides, and principal components 1 and 2. Genetic signals with a p-value <1E-05 from the GWAS analysis were tested for replication in 228 Caucasians treated with HCTZ. Results from the discovery analysis revealed 12 independent signals associated with changes in triglycerides in response to CLT (at a p-value <1 E-05). From those signals, rs7877356 in Purinergic Receptor P2Y12 (*P2RY12*) was significantly replicated in the same direction of response in an independent group of hypertensive Caucasians treated with HCTZ (p=8.6E0−5). In PEAR-2, rs78777356 G/A genotype carriers had a significant increase in triglycerides compared to A/A genotype carriers (G/A=57.6 mg dl−1 vs. A/A=11 mg dl−1, p=3.3E-06). Consistently, we also found that G/A carriers had a significant elevation in triglycerides compared to A/A genotype carriers when treated with HCTZ (G/A=32.5 mg dl−1 vs. A/A=3.51 mg dl−1, p=8.6E0−5). The meta-analysis p-value of the rs78777356 between the discovery and replication cohorts reached genome-wide significance (p=1E-09). Additionally, we replicated the region in 142 African-Americans treated with CLT, where we found an association between rs4146770 and changes in triglyceride levels (p=1.5E-04). In conclusion, this study reveals a novel genetic signal, within *P2RY12*, with clinically relevant effect on the thiazide-induced changes in triglycerides. These findings may represent a novel mechanism underlying thiazide-induced adverse metabolic effects that warrant further investigation.

2679F

Genome-wide association study of transposition of the great arteries. *D. Skoric-Milosavljevic[1], R. Tadros[1], N. Lahrouchi[1], F.V. Tjong[1], L. Beekman[1], F. Bajolle[2], C. Preuss[3], E.M. Lodder[1], W.S. Kerstjens-Frederikse[4], J.A. Goodship[5], H.J. Cordell[5], U. Bauer[6], G. Andelfinger[3], J. Breckpot[7], P. Bouvagnet[8], S. Mital[9], G. Blue[10], D.S. Winlaw[10], M.W.T. Tanck[11], D. Bonnet[2], B.J. Mulder[1], B.D. Keavney[12], A.V. Postma[1], C.R. Bezzina[1].* 1) Department of Clinical and Experimental Cardiology, Academic Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, the Netherlands; 2) Hôpital Necker-Enfants-Malades, Université Paris Descartes , 149 Rue de Sèvres, 75015 Paris, France; 3) Cardiovascular Genetics, Department of Pediatrics, CHU Sainte-Justine, Université de Montréal, 3175 Ch de la Côte-Sainte-Catherine, Montreal, QC H3T 1C5, Canada; 4) Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands; 5) Institute of Genetic Medicine, Newcastle University, Newcastle Upon Tyne NE1 3BZ, United Kingdom; 6) National Register for Congenital Heart Defects, 13353 Berlin, Germany; Competence Network for Congenital Heart Defects, National Register for Congenital Heart Defects, DZHK, 13347, Berlin, Germany; 7) Center for Human Genetics, University Hospitals Leuven, Catholic University Leuven, Herestraat 49, 3000 Leuven, Belgium; 8) Laboratoire Cardiogénétique, INMG, CNRS UMR 5310, INSERM U1217, Université Lyon 1, Lyon, France; 9) Division of Cardiology, The Hospital for Sick Children, University of Toronto, 555 University Ave, Toronto, Ontario M5G 1X8, Canada; 10) The Children's Hospital at Westmead, University of Sydney, 170 Hawkesbury Rd, Westmead NSW 2145, Australia; 11) Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands; 12) Division of Cardiovascular Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester M13 9PL, UK; Manchester Heart Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester M13 9WL, United Kingdom.

Background: Congenital heart disease is the most common congenital defect, with an estimated prevalence of 1%. Transposition of the great arteries of the dextro type (D-TGA) is one of the most critical lesions that results in neonatal death if untreated. The role of genetic factors in predisposition for D-TGA is largely unknown. Complex inheritance may account for the sporadic presentation of the disorder. In this study we conducted a case-control genome-wide association study (GWAS) to explore the role of common genetic variants in the determination of risk for D-TGA in the context of a complex inheritance paradigm. Methods: D-TGA cases of Western-European descent were included and genotyped on Illumina OmniExpress Beadchip or Illumina Human660W arrays. After stringent quality control 245,999 single nucleotide polymorphisms (SNPs) in 969 cases and 4,347 ethnically matched controls remained for analysis. An additional >7 million SNPs were imputed using the Haplotype Reference Consortium panel (minor allele frequency > 0.05). SNP-disease association was assessed for genotyped SNPs and imputed SNP dosages by logistic regression analysis including sex and population structure as covariates. Loci that passed a suggestive threshold of statistical significance (p ≤ 1x10-5) were analyzed for expression quantitative trait locus (eQTL) effects in tissue sets from the GTeX resource. Furthermore, we performed a GO enrichment analysis on genes located within the 500 Kb upstream and downstream of the association signal. Results: No locus reached the genome-wide statistical significance threshold of p<5x10-8. Nineteen loci displayed an association signal that passed the suggestive threshold of <p<1x10-5. Six of these loci displayed an eQTL effect, some for genes well known to influence cardiac outflow tract formation (e.g. *WNT5A* (p=1.5x10-7) and *ALDH1A2* (p=6x10-6)). GO enrichment analysis identified an over-representation of genes involved in heart development among the 151 genes at the 19 loci (p=0.01). Conclusion: This is the first study exploring the role of common variants in a large cohort of D-TGA cases. These data indicate a possible role for common genetic variation in the determination of risk for D-TGA. Additional patients are currently being genotyped to extend the dataset. This work is expected to contribute to our understanding of the genetic architecture of D-TGA and to uncover new molecular mechanisms that underlie the disorder.

2680W

GWAS-driven pathway analyses and functional validation reveals *GLIS1* to associate with mitral valve prolapse. *M. Yu[1], C. Dina[2], N. Tucker[3], F. Delling[4], S. Slaugenhaupt[5], R.A. Levine[6], A.A. Hagège[7], J.J. Schott[2], X. Jeunemaitre[1,8], D. Milan[3], R. Norris[9], N. Bouatia-Naji[1].* 1) INSERM UMR 970 Paris Cardiovascular Research Center, Paris, France; 2) INSERM UMR 1087, CNRS UMR 6291, Institut du Thorax, Centre Hospitalier Universitaire (CHU), Nantes University, Nantes, France; 3) Cardiovascular Research Center, Massachusetts General Hospital, Charlestown, Massachusetts, United States; 4) Framingham Heart Study, Framingham, Massachusetts, United States; 5) Center for Human Genetic Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States; 6) Cardiac Ultrasound Laboratory, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, United States; 7) AP-HP, Department of Cardiology, European Georges Pompidou Hospital, Paris, France; 8) AP-HP, Department of Genetics, European Georges Pompidou Hospital, Paris, France; 9) Cardiovascular Developmental Biology Center, Children's Research Institute, Medical University of South Carolina, Charleston, South Carolina, United States.

Mitral valve prolapse (MVP) is the most common indication for surgical repair of mitral regurgitation. MVP is a risk factor for arrhythmia, atrial fibrillation and sudden death. We have previously performed genome-wide association study (GWAS) in 1412 MVP cases and 2439 controls that identified 6 GWASig loci for MVP. However, stringent statistical thresholds set in GWAS may exclude genuinely associated loci. Additional genetic predisposition loci and mechanisms for MVP may deserve follow-up using pathway-based analyses. We first analyzed GWAS data using iGSEA4GWAS to detect associated pathways by applying gene set enrichment analyses. We found *GLIS1*, a GLI-related Kruppel-like zinc finger transcription factor to be the best ranked gene in 8 enriched gene sets (FDR<0.05). We also investigated 355 suggestively associated SNPs (P<10^{-6}) using DEPICT, which uses predicted functions to prioritize causal genes and highlight systematically enriched pathways. We identified 5 enriched gene sets (P<10^{-4}) involved in endothelial cell migration (e.g.GO:0043542), a highly relevant mechanism to valve development. In addition, we found that 26 tissue types are enriched for MVP loci (FDR<0.05), including those part of the cardiovascular system. Interestingly, among 13 significantly highlighted loci (P<0.05), we found *SMG6*, *LMCD1* and *PBX1* genes already part of the 6 confirmed loci and *GLIS1*. In the GWAS, *GLIS1* located on Chr1 is significantly associated with MVP risk (best SNP rs1879734: OR=1.30, P=7.1×10^{-6}). Follow-up analyses indicate positive replication in four additional case control studies (~1400 cases and ~6400 controls, overall OR=1.23, P=1.2×10^{-7}). GLIS1 is part of the Hedgehog signaling, which is an important pathway for endothelial to mesenchymal transition and cilia organization during heart development. Using immunohistochemistry we found that Glis1 is expressed during valve morphogenesis in mice (E.13 and E.17 stages) and maintained in the adult. Glis1 is detected in nuclei from endothelial and valvular interstitial cells. *Glis1* knockdown by morpholino caused significant atrioventricular (AV) regurgitation in zebrafish embryos. In summary, our pathway enrichment investigation of GWAS data provides new candidate genes and pathways that deserve further investigation. We also provide genetic and functional evidence supporting *GLIS1* as a new locus and candidate gene for MVP, which involved Hedgehog signaling for the first time in valve development and degeneration.

2681T

A novel LDL-lowering missense variant in *B4GALT1* identifies novel biological connection between protein glycosylation and cardiovascular risk factors in human. *M. Montasser[1], C. Van Hout[2], G. Della Gatta[2], M. Puurunen[2], J. Reid[2], J. Overton[2], A. Baras[2], A. Economides[2], M. Healy[2], N. Zaghloul[1], C. Sztalryd-Woodle[1], E. Streeten[1], B. Mitchell[1], S. Taylor[1], J. O'Connell[1], A. Shuldiner[1,2].* 1) Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland USA; 2) Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc, Tarrytown, NY USA.

Cardiovascular disease (CVD) is the leading cause of death worldwide. Elevated low density lipoprotein cholesterol (LDL) and fibrinogen are major independent risk factors for CVD. Understanding their genetic basis may identify novel therapeutic targets to lower their levels and treat or prevent CVD. Isolated founder populations can enable discovery of novel disease-associated variants enriched in these populations through genetic drift. Although very rare in the cosmopolitan populations, these novel variants can inform biology relevant to all humans. To identify novel rare drifted alleles associated with CVD risk factors, we performed whole exome (n= 4,725) and genome sequencing (n=1113) on well phenotyped Amish subjects. We identified a strong novel association (p = 3.3E-18) between a missense SNP (N352S) in *B4GALT1* and LDL. Each 352S allele is associated with a 14.7 mg/dl lower LDL, and has a frequency of 6% in the Amish while extremely rare in the general population. In addition, this SNP was associated with a 20% decreased fibrinogen (p= 5.0E-4), a two-fold increase in aspartate transaminase (AST) (p= 3.0E-8), and a 50% increase in creatine kinase (CPK) (p=0.02). These observations are consistent with coagulopathy and myopathy phenotypes reported in patients homozygous for rare protein-truncating mutations in *B4GALT1*. Notably, the 13 Amish 352S homozygotes appear to be healthy adults without any obvious phenotypic sequelae. The enzyme encoded by *B4GALT1* plays a critical role in processing of N-linked oligosaccharide moieties in glycoproteins. The N352S substitution resides very close to the catalytic site of the enzyme. Alterations in glycosylation have been reported to impair intracellular trafficking of various glycoproteins including the LDL receptor, and may also affect various circulating glycoproteins such as apolipoprotein B and fibrinogen. As expected for a missense SNP we found no difference in the protein expression between the wild type and the mutant forms expressed in cultured Cos-7 cells, however, we expect, and are currently testing, that the 352S allele will cause lower enzymatic activity. We have knocked down *B4GALT1* in zebrafish and we will assess the associated traits in the knockdown model, and also assess the impact of N352S on these traits. More functional assays are underway to fully understand the biological impact of the protein glycosylation pathway on CVD risk factors and its therapeutic opportunities.

2682F

Whole genome cardiac DNA methylation fingerprint and gene expression analysis provide new insights in the pathogenesis of chronic Chagas disease cardiomyopathy. *C. Chevillard[1], L. Laugier[1], A.F. Frade[2,4,5], F.M. Ferreira[2,4,6], M.A. Baron[2,4], P.C. Teixeira[2,4], S. Cabantous[1], L.R.P. Ferreira[2,4,6], L. Louis[7], V.O.C. Rigaud[2,4], F.A. Gaiotto[8], F. Bacal[9], P. Pomerantzeff[9], E. Bocchi[10], J. Kalil[2,3,4], R.H.B. Santos[8], E. Cunha-Neto[2,3,4].* 1) Aix Marseille Université. Génétique et Immunologie des Maladies Parasitaires. Unité Mixte de Recherche S906; INSERM. U906. 13385 Marseille. France; 2) Laboratory of Immunology. Heart Institute (InCor) School of Medicine. University of São Paulo, São Paulo, Brazil; 3) Division of Clinical Immunology and Allergy. School of Medicine, University of São Paulo, São Paulo, Brazil; 4) Institute for Investigation in Immunology (iii), INCT, São Paulo, Brazil; 5) Department of Bioengineering , Brazil University, Rua Carolina Fonseca, 235 (Campus II), Vila Santana, 08230-030, Itaquera, São Paulo, SP, Brazil; 6) Health Sciences, University of Santo Amaro, São Paulo, Brazil; 7) Aix Marseille Université, Génétique médicale et génomique fonctionnelle (Plateforme Génomique et Transcriptomique). Unité Mixte de Recherche S910; INSERM U910, 13385 Marseille. France; 8) Division of Pathology, Heart Institute (InCor). University of São Paulo School of Medicine. São Paulo, Brazil; 9) Division of Heart Transplantation. Heart Institute (InCor). University of São Paulo, School of Medicine. São Paulo, Brazil; 10) Heart Failure Unit, Heart Institute (InCor), University of São Paulo. School of Medicine, São Paulo, Brazil.

Chagas disease, caused by the protozoan *Trypanosoma cruzi*, is endemic in Latin America and affects 10 million people worldwide. Approximately 12,000 deaths attributable to Chagas disease occur annually due to Chagas disease cardiomyopathy (CCC). Little is known about the role of epigenetic modifications in pathological gene expression patterns in CCC patients' myocardium Whole genome DNA methylation and global gene expression were done on myocardial samples from end-stage CCC patients and control samples. 4720 genes were differentially methylated between CCC patients and controls, of which 399 were also differentially expressed. Reporter gene and in silico transcription factor binding analyses indicated promoter methylation modified expression of key genes. These analyses have shown that Sp1 is a major transcription factor for differentially methylated CpG regions. It may be a therapeutic target for restoring gene expression patterns. We found potassium channel genes KCNA4 and KCNIP4, involved in electrical conduction and arrhythmia. *KCNA4* encodes the potassium voltage-gated channel Kv1.4, and *KCNIP4* is a potassium channel-interacting protein that regulates Kv4.3 potassium channel function. Both modulate the duration and shape of the cardiac action potential. Potassium channel gene expression was found to be disturbed in hearts of *T. cruzi*-infected mice. In addition, *SMOC2* gene is highly expressed during wound healing, matrix assembly and remodeling, processes associated with fibrosis in CCC. Transgenic mice over or underexpressing the *Smoc2* gene displayed increased /decreased kidney fibrosis in a kidney damage model, by modulating fibroblast proliferation and extracellular matrix deposition. *PENK* mRNA, upregulated in CCC, is expressed mainly in the brain and heart and encodes preproenkephalin. Enkephalin is a ligand of δ opioid receptors important in pain processing. δ opioid receptors are present in myocardium and immune cells; opioid signaling induces heart protection against ischemic injury in experimental models. Finally, *RUNX3*, interacts with transcription factor T-bet to induce maximal IFN-γ production in T cells which are abundant in CCC myocardium. This may indicate an additional driving factor for the potent, unrestricted activation of pathogenic IFN-γ producing T cells in CCC myocardium. Results support that DNA methylation plays a role in the regulation of pathogenically relevant genes in CCC myocardium, and identify novel therapeutic targets in CCC.

2683W

Primary lymphedema: A novel association with 22q11.2 deletion syndrome. *M. Unolt[1,2], J. Barry[2], M.C. Digilio[3], B. Marino[1], A.S. Bassett[4], E. Oechlin[4], D. Low[2,5], J. Belasco[2,5], S. Kallish[5,6], K. Sullivan[2,5], E.H. Zackai[2,5], D.M. McDonald-McGinn[2,5], International 22q11.2DS Consortium.* 1) Department of Pediatrics, "Sapienza" University of Rome, Rome, Italy; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 3) Ospedale Pediatrico Bambino Gesù, Rome, Italy; 4) University of Toronto, Centre for Addiction and Mental Health, Toronto, Canada; 5) Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA; 6) Hospital of the University of Pennsylvania, Philadelphia, PA, USA.

BACKGROUND: Lymphedema is an abnormal accumulation of interstitial fluid and fibroadipose tissues. Primary lymphedema (LMP) is congenital, caused by aberrant lymphangiogenesis. Classifications of LMP are based on age at presentation: congenital, precox, tarda. Although most cases are sporadic, LMP may be familial (Hereditary Primary Lymphedema - LMPH) inherited mainly in an autosomal dominant fashion with heterogeneous mutations or in association with chromosomal abnormalities/syndromic diagnoses. **METHODS:** We identified 4 patients (1 male, 3 females) with 22q11.2DS and LMP via our International 22q11.2DS Consortium: 2 congenital; 1 precox; 1 tarda. **RESULTS:** All patients underwent comprehensive clinical, laboratory and imaging assessments to rule out other causes of lymphedema including a negative next generation sequencing panel performed on the infant male with LMP. All patients had *de novo* typical deletions and family histories were negative for lymphedema. Of note, the female patient with congenital LMP developed a nodular lymphocyte predominant Hodgkin lymphoma (NLPHL) in late adolescence and another female with LMP precox had severe pre-eclampsia during pregnancy. **CONCLUSIONS:** We report the novel association of LMP with 22q11.2DS. Importantly, animal models demonstrated *Tbx1* playing a critical role in lymphangiogenesis by reducing *Vegfr3* expression in lymphatic endothelial cells. Moreover, the *VEGFR3* pathway is essential for lymphangiogenesis with mutations identified in LMPH. This is especially notable as *VEGFR3* is also involved in tumor lymphangiogenesis/metastases which may explain the NLPHL in our patient. Finally, severe pre-eclampsia results from abnormal placental angiogenesis in association with mutations in the *VEGF* family of genes, which also may explain the preeclampsia in our patient. Accordingly, our findings provide crucial insight into clinical management, as well as, a window into understanding cellular mechanisms of lymphangiogenesis disorders.

2684T

Utilization of drugs with evidence for pharmacogenomic testing following percutaneous coronary intervention. *N. El Rouby[1], A. Alrwisan[2], JA. Johnson[1], AG. Winterstein[2], L. Cavallari[1].* 1) Pharmacotherapy and Translational Research, UF, Gainesville, FL; 2) Pharmaceutical Outcomes and Policy, UF, Gainesville, FL.

Genotype informed prescribing, an important facet of precision medicine, offers great promise of optimizing drug therapy. This can be facilitated by preemptive genotyping using multi-panel testing, which may be particularly valuable for patients with polypharmacy, and whose therapy is likely to be influenced by multiple genes. The Pharmacogenomic Knowledge Base (PharmGKB) grades the level of evidence (1A>1B>2A>3>4) for the effect of genotype on drug response, and the Clinical Pharmacogenetic Implementation Consortium (CPIC) prioritizes gene-drug pairs for integration into prescribing decisions (A>B>C>D). To determine the value of preemptive genotyping, we sought to identify the incidence of drugs with high level of pharmacogenetic evidence, defined as CPIC level A or B and PharmGKB level 1A through 2A, in a population of patients with incident percutaneous intervention (PCI) initiated on clopidogrel. We focused on PCI patients because 1) PCI is one of the most common settings for clinical pharmacogenetic implementation, and 2) patients undergoing PCI represent a high risk population with multiple co-morbidities often requiring polypharmacy. The dataset was created using in- and outpatient pharmacy billing records from employer-based health plans made available in Truven MarketScan, and included patients with continuous enrollment of ≥ 6 months who had a clopidogrel prescription fill within 7 days of PCI (n=122,339). Data were collected up to 5 years after PCI. Among post-PCI patients who were prescribed clopidogrel and were followed for one or three years, 51% and 62% were prescribed ≥1 additional drug with pharmacogenetic evidence, respectively. Among patients followed for five years, 68% had ≥1 additional drug, and 28% were prescribed ≥3 drugs with pharmacogenetic evidence. The most commonly prescribed drugs were opiates, proton pump inhibitors, and simvastatin, representing 31%, 22%, and 21% of drugs with pharmacogenetic evidence and whose responses are influenced by variants in *CYP2D6*, *CYP2C19*, and *SLCO1B1* genes, respectively. In conclusion, these data demonstrate that panel-based pharmacogenetic testing at the time of PCI has not only the potential to inform antiplateley therapy, but in the majority of patients, to inform therapy decisions for additional medications impacted by genotype in the first 1 to 5 years post-PCI.

2685F

Genome-wide association study reveals novel genetic markers associated with chlorthalidone blood pressure response. *S. Singh[1], Y. Gong[1], C.W. McDonough[1], K.R. Bailey[2], A.L. Beitelshees[3], E. Boerwinkle[4], A.B. Chapman[5], J.G. Gums[1], S.T. Turner[6], R.M. Cooper-DeHoff[1,7], J.A. Johnson[1,7].* 1) Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, Gainesville, Florida; 2) Division of Biostatistics, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota; 3) College of Medicine, University of Maryland, Baltimore, MD; 4) Human Genetics and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas; 5) Division of Nephrology, University of Chicago, Chicago, IL, USA; 6) Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA; 7) Division of Cardiovascular Medicine, Department of Medicine, University of Florida, Gainesville, Florida, USA.

Around 30% of the adult USA population suffers from hypertension, of which almost 50% of patients have uncontrolled blood pressure (BP). Interpatient variability in response to antihypertensive drugs contributes importantly to this statistic. Thiazide and thiazide-like diuretics remain among the most widely used anti-hypertensives. Although hydrochlorothiazide is the more commonly prescribed diuretic, studies show that chlorthalidone is a more potent BP lowering diuretic and may have additional benefits. However there are few data regarding the influence of genetics on variable BP response to chlorthalidone, and we sought to address this question. We conducted a genome-wide association analysis of the systolic BP (SBP) and diastolic BP (DBP) change post 8 weeks of chlorthalidone treatment in 184 European Americans from the Pharmacogenomic Evaluation of Antihypertensive Responses-2 (PEAR-2) using linear regression, adjusting for age, sex, baseline BP and principal components 1 and 2. Genes harboring the variants that met the suggestive level of significance ($p<1 \times 10^{-6}$) were further tested for differential expression using RNA sequencing data between responders and non-responders to chlorthalidone treatment. Rs13066950 in the C3orf35 region met the suggestive level of significance and was associated (p-value=6.57×10^{-7}, β=4.0, SE=0.806) with decreased SBP response to chlorthalidone. Furthermore this variant was an eQTL for *MLH1* (MutL Homolog1) and *GOLGA4* (Golgin A4) and was in high Linkage Disequilibrium (D' > 0.9) with multiple missense variants of *GOLGA4*. Upon evaluating the differential gene expression between the good responders and non-responders to chlorthalidone treatment, the expression of C3orf35 was down regulated in non-responders ($\log_2$(Fold Change)= -1.11, q-value = 0.032). Further, based on the publically available GTEx dataset (n=7051), we also found that for rs13066950 in the *C3orf35, GOLGA4* expression was down regulated in the carriers of the alternate allele compared to non-carriers, which correlates well with our findings. *GOLGA4* is purported to play a role in the Rab6/RabGTPases pathway that has been implicated in hypertension and atherosclerosis. Our study identified novel genetic markers that may contribute to the genetic basis of the variability of chlorthalidone BP response. Further studies are warranted to understand the exact mechanisms of these genes to help advance the personalized anti-hypertensive therapy approaches.

2686W

Exome-chip meta-analysis identifies association between variation in *ANKRD26* and platelet aggregation. *A.D. Johnson[1], M-H. Chen[1], L.R. Yanek[2], J.D. Backman[3], J.D. Eicher[1], J.E. Huffman[1], Y. Ben-Shlomo[4], A.D. Beswick[5], L.M. Yerges-Armstrong[3], A.R. Shuldiner[3], J.R. O'Connell[3], R.A. Mathias[6], D.M. Becker[2], L.C. Becker[7], J.P. Lewis[3], N. Faraday[8].* 1) Population Sciences Branch, NHLBI's Framingham Heart Study, Framingham, MA; 2) Johns Hopkins University School of Medicine, GeneSTAR Research Program, Department of Medicine, Division of General Internal Medicine, Baltimore, MD; 3) School of Medicine, Division of Endocrinology, Diabetes and Nutrition, and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 4) School of Social and Community Medicine, University of Bristol, Bristol, UK; 5) School of Clinical Sciences, University of Bristol, Bristol, UK; 6) GeneSTAR Research Program, Department of Medicine, Divisions of Allergy and Clinical Immunology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 7) GeneSTAR Research Program, Department of Medicine, Divisions of Cardiology and General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 8) GeneSTAR Research Program, Department of Anesthesiology & Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Previous genome-wide association studies (GWAS) have identified several variants associated with platelet function phenotypes; however, the proportion of variance explained by the identified variants is small. Rare coding variants, particularly those with high potential for impact on protein structure/function, may have substantial impact on phenotype but are difficult to detect by GWAS. The main purpose of this study was to identify low frequency or rare variants associated with platelet function using genotype data from the Illumina HumanExome Bead Chip. Three family-based cohorts of European ancestry, including ~4,000 total subjects, comprised the discovery cohort, and two independent cohorts, one of European and one of African American ancestry, were used for replication. Optical aggregometry in platelet-rich plasma was performed in all the discovery cohorts in response to adenosine diphosphate (ADP), epinephrine, and collagen. Meta-analyses were performed using both gene-based and single nucleotide variant association methods. The gene-based meta-analysis identified a significant association (P=7.13 x 10-7) between rare genetic variants in *ANKRD26* and ADP-induced platelet aggregation. One of the *ANKRD26* SNVs - rs191015656, encoding a threonine to isoleucine substitution predicted to alter protein structure/function, was replicated in Europeans. Aggregation increases of ~20-50% were observed in heterozygotes in all cohorts. Novel genetic signals in *ABCG1* and *HCP5* were also associated with platelet aggregation to ADP in meta-analyses, although only results for *HCP5* could be replicated. The SNV in *HCP5* intersects epigenetic signatures in CD41+ megakaryocytes suggesting a new functional role in platelet biology for HCP5. This is the first study to use gene-based association methods from SNV array genotypes to identify rare variants related to platelet function. The molecular mechanisms and pathophysiological relevance for the identified genetic associations requires further study.

2687T

Functional fine-mapping of coronary artery disease risk variants. *B. Liu[1,2], M. Pjanic[3], T. Nguyen[3], S.B. Montgomery[2,4], C. Miller[3], T. Quertermous[3].* 1) Department of Biology, Stanford University, Palo Alto, CA; 2) Department of Pathology, Stanford University, Palo Alto, CA; 3) Department of Medicine - Division of Cardiovascular Medicine, Stanford University, Palo Alto, CA; 4) Department of Genetics, Stanford University, Palo Alto, CA.

Coronary artery disease (CAD) is the leading cause of death globally. Its complex etiology is influenced by both genetic and environment risk factors. Large-scale population association studies have identified over 150 risk loci, explaining ~10% of the disease heritability. Functional interpretation of the risk-associated variants has been challenging as the majority of the variants lie outside of coding regions and are predicted to be regulatory. One of the key arbiters in coronary artery disease is the human coronary artery smooth muscle cell (HCASMC). These cells constitute the majority of the coronary artery vessel wall and have been shown to give rise to ~30% of the total cells in the atherosclerotic lesion. While these cells remain differentiated in the healthy tissue, they become highly proliferative and invasive in response to vessel injury and lesion expansion. Global changes in the HCASMC transcriptome and the underlying regulatory mechanisms responsible for these disease state transitions have not been fully characterized due to limited sample availability. Here, we have generated RNA and whole-genome sequencing datasets on 52 serum-stimulated HCASMC lines (mimicking the disease state) and ATACseq on a subset of 9 lines, representing one of the largest collections of these valuable samples to date. By comparing the transcriptomes of HCASMC lines and closely related tissue types in the GTEx project, we identify upregulated pathways and HCASMC-specific genes. In addition, we identify HCASMC-specific eQTLs as compared to related GTEx tissues. Around 30% of all eQTLs overlap chromatin-accessible regions, and eQTL specificity positively correlates with chromatin specificity. As expected, CAD GWAS variants are enriched in HCASMC ATACseq regions as compared to those in other tissues, further emphasizing the disease relevance of this critical vascular cell type. By combining eQTL and ATAC-seq, we discovered 35 SNPs whose eQTL signals and GWAS signals colocalize. This study represents the first comprehensive characterization of the HCASMC transcriptome from multiple donors, ultimately revealing insights into the genetic underpinnings of CAD.

2688F

Genetic variation in thromboxane A synthase 1 is associated with stroke risk that can be reduced by daily aspirin. *S. Zajic[1], J.P. Jarvis[1], M. Byrska-Bishop[2], P. Zhang[1], J. Leader[2], A. Brangan[1], K. Dent-Ferguson[1], M. Ritchie[2], D. Ledbetter[2], M. Christman[1], DiscovEHR Collaboration.* 1) Coriell Institute for Medical Research, Camden, NJ; 2) Geisinger Health System, Biomedical and Translational Informatics Institute, State College, PA.

We investigated genetic and environmental factors associated with stroke risk using genetic, demographic, medical history, and medication data from 4,501 participants in the Coriell Personalized Medicine Collaborative (CPMC). Variation in metabolism-related genes was assayed using the Affymetrix DMET Plus array, a targeted set of 1931 SNPs in metabolic enzymes and transporters. We report the identification of a genetic variant of thromboxane A synthase 1 (TBXAS1) (rs8192868, c.1348G>A, p.Glu450Lys, CYP5A1*7) as a novel risk allele for stroke. Daily aspirin use appears to reduce stroke risk associated with this variant. Risk of stroke (n = 83 cases) was significantly higher in individuals with this variant not on daily aspirin (p-value = 0.0188, odds ratio [95% CI] of 2.49 [1.20, 5.18]), with the effect of this variant on stroke risk of borderline significance (p-value = 0.0613, odds ratio [95% CI] of 1.76 [0.946, 3.27]) across the entire cohort, including daily aspirin users. This variant showed a significant association with daily aspirin use (Bonferroni-adjusted p-value = 0.00694 at a 0.05 level), consistent with the role of TBXAS1 immediately downstream of the site of action of aspirin. Heterozygous individuals were nearly twice as likely to be taking daily aspirin as compared to homozygous wild-type, with an odds ratio [95% CI] of 1.85 [1.41, 2.44]. These associations were supported by analysis of data from 40,519 individuals (including 425 stroke cases) in the exome-sequenced DiscovEHR cohort of Geisinger Health System and Regeneron Genetics Center. In that cohort, there was a significant association between stroke and this variant with p-value = 0.05 and odds ratio [95% CI] of 1.47 [1.00 – 2.17]. The genetic risk for stroke associated with this variant in non-aspirin users (odds ratio [95% CI] of 1.88 [1.07, 3.29]) was reduced in individuals on daily aspirin (odds ratio [95% CI] of 1.39 [0.856, 2.27]), consistent with the expectation that altered TBXAS1 function would be of reduced consequence in individuals taking daily aspirin, though the interaction term between this variant and daily aspirin use was not significant in the regression analysis. Genetic stroke risk associated with this variant appears to be reduced or eliminated by daily aspirin use in both cohorts. This biologically plausible association suggests personalization of daily aspirin therapy to individuals based on their genetics for primary or secondary stroke prevention.

2689W

A longitudinal transcriptome analysis identifies novel gene expression signatures for body mass index in monocytes. *C. Müller[1,2,3], M.F. Hughes[4], F. Ojeda[1], D. Börnigen[1,2], K.J. Lackner[5,6], P.S. Wild[6,7,8], D.A. Trégouët[9], S. Blankenberg[1,2], T. Zeller[1,2].* 1) Clinic for general and interventional Cardiology, University Heart Center Hamburg, Hamburg, Germany, Germany; 2) DZHK (German Centre for Cardiovascular Research) partner site Hamburg/Lübeck/Kiel, Germany; 3) Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Schleswig-Holstein, Germany; 4) UKCRC Centre of Excellence for Public Health Northern Ireland, Queens University Belfast, Northern Ireland; 5) Institute of Clinical Chemistry and Laboratory Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Germany; 6) DZHK (German Centre for Cardiovascular Research) partner site Rhine Main, Germany; 7) Preventive Cardiology and Preventive Medicine, University Medical Center of the Johannes Gutenberg-University Mainz, Germany; 8) Center for Thrombosis and Hemostasis, Mainz, Germany; 9) Sorbonne Universités, UPMC, INSERM, UMR_S 1166, ICAN Institute for Cardiometabolism and Nutrition, Paris, France.

Background and study aims: More than 600 million obese people worldwide are at higher risk of developing cardiometabolic disorders. Monocytes are cell types participating to cardiovascular disease progression whose regulation is affected by obesity. However, the underlying molecular mechanisms are not fully understood. We aimed to identify longitudinal gene expression changes of circulating monocytes in relation to weight gain/loss during 5 years follow-up in a large population of healthy subjects. **Methods:** mRNA was isolated from monocytes of 1,092 subjects of the Gutenberg Health Study, both at baseline study visit (BL) and at 5 years follow-up (FU). Whole-transcriptome gene expression was measured by *Illumina HT12 BeadChip*s. 12,170 genes were tested for association between mRNA changes (mRNA$_{FU}$ − mRNA$_{BL}$) and changes of body mass index (BMI$_{FU}$ − BMI$_{BL}$) by linear regression adjusted for age, gender and BMI$_{BL}$. Genes reaching a Benjamini-Hochberg based false discovery rate (FDR) ≤ 0.01 were considered as significant. Pathway and gene-set enrichment analyses were performed using *LYNX* (http://lynx.ci.uchicago.edu/) and by Fisher's exact test, respectively. **Results:** mRNA changes of 143 genes were significantly associated to longitudinal BMI changes. The strongest association (FDR = 1.7×10^{-6}) was observed for lipocalin-2 (*LCN2*) whose decreased mRNA levels were associated with increased BMI. Significant genes were ~6-fold enriched (p < 10^{-16}) in a signature separating classical CD14^{++}CD16^{-} from non-classical CD14^{+}CD16^{++} monocytes indicating a shift of monocyte composition during obesity development. In addition, BMI related genes were >3-fold enriched (p = 1.3×10^{-8}) in a signature reflecting the age of transcriptomes. *LYNX* analyses revealed enrichment of *Gene Ontology* term *immune response* (FDR = 0.034) and – amongst others – *REACTOME* pathway *Oxidative Stress Induced Senescence* (FDR = 4.2×10^{-5}). **Conclusion:** This is the first longitudinal population-based study investigating BMI changes in relation to transcriptomes in monocytes. The top gene *LCN2* is known to be involved in type 2 diabetes and cardiovascular disorders, underlining the validity of our analysis. Enrichment analyses revealed strong over-representation of aging-related genes, and suggest a shift of monocyte sub-populations towards non-classical monocytes. The latter finding is consistent with current studies reporting a shift in monocyte composition during obesity and cardiovascular disease development.

2690T

Implementing genome-based predictive and preventive medicine: The GeneRISK follow-up study. *E. Widen[1], J. Aro[1], K. Hotakainen[2], J. Partanen[3], P. Pollanen[4], I. Salonen[4], S. Ripatti[1].* 1) Institute for Molecular Medicine Finland, Helsinki, University of Helsinki, Finland; 2) Mehiläinen Oy, Helsinki, Finland; 3) Finnish Red Cross Blood Service, Helsinki, Finland; 4) CAREA - Kymenlaakso Social and Health Care Services, Kotka, Finland.

While genomic strategies have resulted in successful identification of numerous genetic loci modifying the risk for common complex disorders, the use of this information for disease prevention has remained minimal, in part because the very large number of associated variants imposes a considerable interpretation and data management challenge. To facilitate the translation of recent genomic discoveries, to guide intervention, and to empower individuals to use this information and actively take care of their health, we have developed a novel interactive graphical interface for patients and doctors to utilize genome data to predict and prevent cardiovascular disease (CVD). We have also set up a prospective population-based study, GeneRISK, where 10,000 participants will use this tool to receive and interpret information on their personal disease risk, including genomic information. We initiated the study in 2015, and have hitherto enrolled 6,807 subjects from Southern Finland aged 45-65 years. The CVD risk is estimated based on traditional factors and on a genomic risk score (GRS) computed from ~ 49,000 common variants. 5,128 individuals have received their disease risk information so far. 65% had BMI >25 kg/m2, 18% were smokers, and 70% had total cholesterol >5 mmol/l. 25% of them had a high (>10%) 10-year risk for severe ischemic heart disease. 15% of subjects in the high-risk group had been reclassified from a lower risk category because of a high GRS. As many as 42% of individuals at high risk were current smokers, and only 16-18% were on lipid-lowering therapy. This suggests that both by life style and medical intervention can be applied to lower the disease risk. Follow-up visits will be initiated in 2017, aiming at monitoring changes in risk factor levels, behavior and life-style. We will additionally assess the participants' use of medication and disease outcomes prospectively by regularly linking to key nation-wide health registers containing information on hospitalization, drug purchases and causes of death. Setting up and testing comprehensive procedures to integrate genomic and traditional health information for clinical practice is expected to enhance disease prevention and treatment and lead to better life quality of patients and healthy individuals. By promoting early detection and targeted prevention of CVD in the GeneRISK-study, we aim at providing a precedent for practical targeted genome-based prevention to be adapted broadly.

2691F

The effect of genetic variation in donors and patients on rejection after heart transplantation. *J. van Setten[1], B. Chang[2], N. de Jonge[1], M.V. Holmes[2], C.C. Baan[4], O.C. Manintveld[4], A.M.A. Peeters[4], F. Dominguez[5], K.K. Khush[3], P. Garcia-Pavia[5], J.W. Rossano[2], R.A. de Weger[1], J.H. Moore[2], B. Keating[2], F.W. Asselbergs[1] on behalf of the iGeneTRAiN consortium.* 1) University Medical Center Utrecht, Utrecht, Netherlands; 2) University of Pennsylvania, Philadelphia, USA; 3) Stanford University, Stanford, USA; 4) Eramus MC, Rotterdam, Netherlands; 5) Puerta de Hierro University Hospital, Madrid, Spain.

Introduction: Heart transplantation is the best therapeutic option for selected patients with end-stage heart failure. However, the immunological barrier between the donor and recipient still limits long-term survival. Immunosuppressive drugs are needed to avoid rejection, but cause an increased incidence of cancer and infections. Besides HLA, also other genetic factors play a role in graft rejection. We aim to identify genetic variants in the patient and in the donor that are involved in rejection after heart transplantation. **Methods:** iGeneTRAiN is a large-scale international consortium, consisting of over 30,000 solid organ transplant patients and donors. We included 1022 patients and 945 heart transplant donors from five different hospitals. A majority of samples has been genotyped on the Affymetrix Tx Array. This array is specifically tailored for the transplantation community. We tested over 10 million variants and included high-resolution data of the MHC region. We used a mixed models approach to take relatedness and ancestry into account. Donor and recipient age and sex, year of transplant, genetic distance between the donor and the recipient, and cohort were included as covariates. **Results:** Two loci, both located on chromosome 6, in the recipients DNA are significantly associated with rejection in the first year after transplantation. We did not identify novel variants for rejection after heart transplantation in the donor DNA. **Conclusion:** We identified two loci in the recipient that are associated with rejection. Our current sample of 945 did not yield novel loci in donor DNA. We aim to increase our sample of heart transplant donors and recipients in the near future. In addition, we will conduct cross-organ meta-analyses including lung, liver, and kidney transplants, maximizing statistical power to identify novel variants. We ultimately aim to translate genetic data into clinical applications such as more optimal genomic compatibility matching of D-R pairs and immune suppression therapy dosing.

2692W

Univariate and phenome-wide GWAS of correlated electrocardiographic traits offer novel insights into genetics of cardiac electrophysiology: the Population Architecture using Genomics and Epidemiology (PAGE) study. *A.R. Baldassari[1], S. Buyske[3,4], H. Highland[1], L. Hindorff[5], A.P. Reiner[6], R.C. Kaplan[7], E.E. Kenny[8], U. Peters[6], D. Darbar[9], K.E. North[1,2], D.E. Arking[10], R.V. Shohet[11], M. Graff[1], E.Z. Soliman[12], C.M. Sitlani[13], N. Sotoodehnia[13], C.J. Hodonsky[1], R.B. Gondalia[1], R. Tao[14], E.A. Whitsel[1], C. Kooperberg[6], C.L. Avery[1].* 1) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Carolina Center for Genome Sciences, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ; 4) Department of Genetics, Rutgers University, Piscataway, NJ; 5) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 6) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 7) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 8) Department of Genetics and Genome Sciences, Icahn School of Medicine at Mount Sinai, NYC, NY; 9) Division of Cardiology, University of Illinois College of Medicine, Chicago, IL; 10) McKusick-Nathans Institute of Genetic Medicine, John Hopkins University School of Medicine, Baltimore, MD; 11) Center for Cardiovascular Research, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI; 12) Epidemiological Cardiology Research Center, Wake Forest University School of Medicine, Winston-Salem, NC; 13) Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 14) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN.

Temporal measures on the electrocardiogram (ECG) assess distinct and heritable aspects of cardiac electrophysiology centrally important to cardiovascular health. The broad impact of several genetic loci across ECG segments suggests pleiotropy, although few studies have systematically evaluated evidence of shared effects. Using PAGE data on 33,967 participants (20% African American; 30% Hispanic/Latino; 50% Caucasian; 74% female), we therefore conducted a series of univariate genome-wide association studies (GWAS; 18,698,945 common variants imputed to the 1000 Genomes reference panel, phases 1 and 3), for each component of a temporally-decomposed average heartbeat: P-wave, PR-segment, QRS interval, ST-segment, T-wave and TP-segment. We then calculated combined phenotype effects using the adaptive sum of powered score tests. In univariate GWAS, we identified fifteen novel loci ($P < 8.3\times10^{-9}$, i.e. 5×10^{-8}/ 6 ECG traits). Among them was rs3211938 for the TP-segment phenotype, a nonsense *CD36* mutation only common in African populations (minor allele frequency [MAF]=0.12), which given the variant's known ties to high-density lipoprotein cholesterol (HDL-C) may support hypothesized links between HDL-C and cardiac electrophysiology. We additionally identified two novel common loci when evaluating the six ECG traits in aggregate: rs12175748 (Global MAF=0.46), situated near *HEY2*, which has roles in cardiac development; and rs7136054 (Global MAF=0.31), in a locus near *HMGA2* with suspected ties to aortic root size. Identification of novel and population-specific loci underscore the value of GWAS in ancestrally diverse populations, and of the comprehensive evaluation of pleiotropic phenotypes.

2693T

Putative loss-of-function (pLOF) genetic variants in arrhythmogenic cardiomyopathy-associated genes: Prevalence and EHR-based phenotype in 50,000 biobank participants. *C.M. Haggerty, A. Pulenthiran, T.N. Person, D.N. Hartzel, J.B. Leader, D.J. Carey, D.H. Ledbetter, S.A. Pendergrass, M.D. Ritchie, K. Manickam, M.F. Murray, B.K. Fornwalt on behalf of the DiscovEHR collaboration.* Geisinger Health System, Danville, PA.

Arrhythmogenic cardiomyopathy (ACM) is associated with genetic variants in desmosome genes. Incidental findings of pathogenic/likely pathogenic (P/LP) variants in these genes, and select non-desmosomal genes, are recommended for clinical return; however, their prevalence and associated phenotype in the general population are not entirely known. Missense variants are relatively common, have uncertain physiologic significance (absent other data, such as curation in ClinVar), and were recently reported to demonstrate no disease phenotype in electronic health records (EHR). In contrast, pLOF variants are less common and may have a greater impact on traits and diseases. This study characterized the prevalence of ACM pLOF variants coupled with EHR data from 50,726 subjects with exome sequencing. We identified all participants from the Geisinger-Regeneron DiscovEHR Collaboration with rare (MAF≤0.00075) variants annotated as "high" impact by Ensembl's Variant Effect Predictor in Refseq canonical transcripts, or with 2-star P/LP designation in ClinVar in the following genes: *DSC2, DSG2, DSP, PKP2, JUP*. Two-star P/LP ClinVar variants in *TMEM43* and *TGFβ3* were also included. We excluded all variants occurring in the terminal exon, and all benign or "uncertain significance" variants in ClinVar. Phenotype data for all patients carrying variants meeting these criteria were retrieved, inclusive of EHR diagnoses, electrocardiograms (EKG), and echocardiograms. Age and sex-matched controls were selected for Bonferroni-adjusted two-sample comparisons. We identified 123 subjects carrying 1 of the 64 variants meeting the specified criteria. Distribution of variants and patients among genes is shown in the Table. Most variants resulted in a frameshift (n=24), stop gain (18), or splice donor/acceptor change (19). The subjects were 60±17 years old, 63% female, and 87% alive. Based on diagnostic codes and chart review of recent physician notes, none of these individuals had a clinical diagnosis of ACM. Furthermore, comparison of available EKG and echocardiogram measures did not reveal significant quantitative differences with variant-negative control subjects. In conclusion, 1 in 412 subjects in the DiscovEHR cohort has a pLOF variant associated with ACM, but no ACM diagnoses. The phenotypic consequences remain to be determined from prospective clinical evaluation.

Distribution of ARVC variants		
	Variant count	Subject count
DSP	18	26
DSG2	13	35
DSC2	11	17
PKP2	18	34
JUP	3	10
TMEM43	1	1
TGFβ3	0	0

2694F

Breast cancer clinical trial of chemotherapy and trastuzumab: Potential tool to identify cardiac modifying variants of dilated cardiomyopathy. *N. Norton[1], D.J. Serie[2], J.E. Crook[2], B.M. Necela[1], B.C. Axenfeld[1], T.J. Dockter[3], G. Colon-Otero[4], E.A. Perez[4], E.A. Thompson[1].* 1) Cancer Biology, Mayo Clinic FL, Jacksonville, FL; 2) Health Sciences Research, Mayo Clinic, Jacksonville, FL; 3) Health Sciences Research, Mayo Clinic, Rochester, MN; 4) Hematology/Oncology, Mayo Clinic, Jacksonville, FL.

Doxorubicin and the ERBB2 targeted therapy, trastuzumab are routinely used in the treatment of HER2+ breast cancer. In mouse models, doxorubicin is known to cause cardiomyopathy and conditional cardiac knock out of Erbb2 results in dilated cardiomyopathy and increased sensitivity to doxorubicin induced cell death. In humans, these drugs also result in cardiac phenotypes, but severity and reversibility is highly variable. We examined association of decline in LVEF at 15,204 SNPs spanning 72 cardiomyopathy genes, in 800 breast cancer patients who received combination doxorubicin and trastuzumab. For 7,033 common SNPs (MAF>0.01) we performed single marker linear regression. For all SNPs, we performed gene-based testing with SKAT, SKAT-O and SKAT-common/rare under rare variant non-burden, rare variant optimized burden and non-burden tests and combination of rare and common variants respectively. Single marker analyses identified seven missense variants in *OBSCN* (p=0.0045-0.0009, MAF=0.18-0.50) and two in *TTN* (p=0.04, MAF=0.22). Gene based rare variant analyses, SKAT and SKAT-O, performed very similarly (*ILK, TCAP, DSC2, VCL, FXN, DSP* and *KCNQ1*, p=0.042-0.006). Gene-based tests of rare/common variants were significant at the nominal 5% level for *OBSCN* as well as *TCAP, DSC2, VCL, NEXN, KCNJ2* and *DMD* (p=0.044-0.008). Our results suggest that rare and common variants in *OBSCN*, as well as in other genes, could have modifying effects in cardiomyopathy.

2695W

***APOL1* coding variants are associated with incident cardiovascular disease in community-dwelling African Americans.** *C.A. Winkler[1], M. Irwin[2], N. Chaudhary[2], M. Cushman[3], N. Zakai[3], S. Limou[4], N. Pamir[5], A. Reiner[6], R. Naik[7], M. Sale[8], G. Nelson[1], M.M. Safford[9], H. Hyacinth[10], S. Judd[2], O. Gutierrez[11], J.B. Kopp[12], Reasons for Geographic and Racial Differences in Stroke (REGARDS).* 1) National Cancer Institute, Leidos Biomedical Research Inc., Frederick National Laboratory, Frederick, MD; 2) University of Alabama at Birmingham, Birmingham, AL; 3) University of Vermont, Colchester, VT; 4) University of Nantes, Nantes, France; 5) Oregon Health and Science University, Portland, OR; 6) University of Washington, Seattle, WA; 7) Johns Hopkins University, Baltimore, MD; 8) University of Virginia, Charlottesville, VA; 9) Weill Cornell Medicine, New York, NY; 10) Aflac Cancer and Blood Disorder Center, Children's Healthcare of Atlanta, Emory University, Atlanta, GA; 11) University of Alabama School of Medicine, Birmingham, AL; 12) NIDDK, NIH, Bethesda, MD.

Background: *APOL1* renal risk variants are strongly associated with chronic kidney disease in African Americans (AA), but associations with incident cardiovascular disease (CVD) are conflicting. **Methods**: We determined *APOL1* genotypes for 10605 AA participants with incident coronary disease (Coronary Heart Disease (CHD), n=323), stroke (n=331), and combined composite CVD endpoint (n=500) enrolled in the REASONS for Geographic and Racial Differences in Stroke (REGARDS). Analyses compared those with *APOL1* high-risk genotypes (2 risk alleles) to APOL1 low-risk genotypes (0/1 risk alleles) in Cox models adjusted for CVD risk factors and ancestry principal components. **Results**: *APOL1* HR participants were younger and more likely to have albuminuria than low-risk participants. The risk of incident stroke, CHD, or composite CVD endpoint did not differ significantly by APOL1 genotype in multivariable models. However, the association of *APOL1* high-risk genotype with the composite CVD outcome differed by diabetes status (P=0.01) in fully adjusted models. This association appeared to be driven by incident stroke. In contrast, APOL1 HR status was associated with a trend towards lower risk of CVD in diabetics. **Conclusion**: *APOL1* high-risk status is associated with higher incidence of incident stroke and CVD disease in individuals without diabetes, whereas APOL1 HR status appeared to lower risk in patients with diabetes.

Hazard ratios (95% CI) of incident CVD in APOL1 HR vs. LR status		
Model*	N$_{events}$/N$_{at\ risk}$	**HR (95% CI)**
No diabetes		
Stroke	197/6621	2.32 (1.33, 4.07)
CHD	174/6173	1.31 (0.08, 2.10)
Composite outcome	297/5755	1.67 (1.12, 2.47)
With diabetes		
Stroke	131/2660	0.55 (0.22, 1.37)
CHD	148/2329	0.53 (0.24, 1.14)
Composite outcome	200/2072	0.52 (0.26, 1.03)
*Adjusted for age, sex, smoking, medications, and ancestry eigenvectors		

2696T

The role of Kringle IV 2 copy number variation and SNPs on Lp(a) levels and cardio-metabolic risk. *S.E. Ruotsalainen[1], I. Surakka[1,2], S.M. Zekavat[2,3,4], P. Natarajan[2,3,4,5], M. Alver[9], A. Ganna[2,3,6,7,8], B. Handsaker[2,6,7], A. Correa[10], J. Wilson[11], V. Salomaa[12], T. Esko[2,9], M.J. Daly[6,15], B.N. Neale[2,6,7], S. McCarroll[2,6,7], S. Kathiresan[2,3,4,5], S. Ripatti[1,13,14].* 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142; 3) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 4) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 5) Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston 02114, MA, USA; 7) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 171 77, Sweden; 9) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; 10) Departments of Pediatrics & Medicine, University of Mississippi Medical Center, Jackson, MS; 11) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 12) National institute for Health and Welfare, Helsinki, Finland; 13) Department of Public Health, University of Helsinki, Helsinki, Finland; 14) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK; 15) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA.

Lipoprotein(a), Lp(a), is a highly heritable, independent causal risk factor for coronary heart disease. Lp(a) is influenced by 5.5kb kringle IV-2 copy number repeats (KIV2-CN), and single nucleotide polymorphisms (SNPs) at the LPA locus. The distinct phenotypic consequences of KIV2-CN and SNPs at the LPA locus are unknown. Here, we address this question using a combination of deep (<30x) WGS and array genotyping data from 6,440 Finns from FINRISK study and performed phenome-wide association analyses (Phe-WAS) of distinct LPA variant classes with 10 incident clinical outcomes and 100 metabolic biomarkers. In addition to directly genotyped KIV2-CN for 455 individuals, to impute KIV2-CN to the rest of the cohort we develop a novel LASSO-based approach to impute KIV2-CN using directly genotyped KIV2-CN from the imputation reference panel of 2,215 whole genome sequenced Finns. Genotyped and imputed KIV2-CN were highly correlated (r^2=0.6, p=1.06e-9) and we observed a strong negative association between Lp(a) and predicted KIV2-CN (-0.087 SD per CN, p=5.48e-281). While both KIV2-CN and SNPs in the LPA region are known to associate with Lp(a) levels individually, their effects have not been extensively compared together in a PheWAS. Here, we accomplish this by using 3 main weighted genetic components: 1) a genetic risk score (GRS) independent of KIV2-CN using 400 independent SNPs from LPA region, 2) a KIV2-CN score, and 3) a combined score (GRS+KIV2-CN) incorporating both (1) and (2). Using the 3 genetic components, Mendelian randomization was performed to infer the causality between Lp(a) and 10 incident clinical events. Interestingly, we found that the association between Lp(a) and cardiovascular disease is associated with GRS (p=6.46e-8, HR=1.30), but not the KIV2-CN repeats (p=0.6, HR=1,03). Furthermore, association of the 3 genetic components across 100 metabolites found 9 significant associations (p<0.0009) with the GRS, 6 with the KIV2-CN and 15 with the GRS+KIV2-CN. Biomarkers associated with KIV2-CN were almost entirely blood lipid related whereas those associated with the GRS, in addition to blood lipid related biomarkers, were related to phospholipid and fatty acid metabolism. In conclusion, we observe the distinct associations of KIV2-CN and SNP GRS from the LPA region with Lp(a) as well as with other cardiometabolic biomarkers and diseases. These findings can potentially highlight the various roles Lp(a) plays in metabolism and cardiovascular diseases.

2697F

Genome-wide trans-ethnic meta-analysis for a novel sleep apnea endophenotype. *H. Wang[1,2], B.E. Cade[1,2,3], A.M. Stilp[4], T. Sofer[4], S.A. Sands[1,2], S. Ancoli-Israel[5,6], J. Cai[7], M.P. Conomos[4], D.S. Evans[8], S.A. Gharib[9], X. Guo[10], W.C. Johnson[4], S. Mukherjee[11], W.S. Post[12], S.M. Purcell[1,3,13,14], J.I. Rotter[10], N.A. Shah[15], K.L. Stone[8], G.J. Tranah[8], L.J. Palmer[16], J.G. Wilson[17], D.J. Gottlieb[1,2,18], S.R. Sunyaev[3,19,20], S.R. Patel[1,2,21,22], R. Saxena[1,2,3,23], X. Lin[24], X. Zhu[25], S. Redline[1,2,21].* 1) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; 2) Division of Sleep Medicine, Harvard Medical School, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) Department of Biostatistics, University of Washington, Seattle, WA; 5) Departments of Medicine and Psychiatry, University of California, San Diego, CA; 6) Department of Veterans Affairs San Diego Center of Excellence for Stress and Mental Health, San Diego, CA; 7) Department of Biostatistics, University of North Carolina-Chapel Hill, NC; 8) California Pacific Medical Center Research Institute, San Francisco, CA; 9) Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Division of Pulmonary and Critical Care Medicine, University of Washington, Seattle, WA; 10) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute and Department of Pediatrics at Harbor-UCLA Medical Center, Torrance, CA; 11) Adelaide Institute for Sleep Health, Flinders Centre of Research Excellence, Flinders University, Adelaide, Australia; 12) Division of Cardiology, Johns Hopkins University, Baltimore, MD; 13) Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 14) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 15) Division of Pulmonary, Critical Care and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 16) School of Public Health, University of Adelaide, Adelaide, Australia; 17) Department of Physiology and Biophysiscs, University of Mississippi, Jackson, MS; 18) VA Boston Healthcare System, Boston, MA; 19) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 20) Division of Medical Sciences, Harvard Medical School, Boston, MA; 21) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA; 22) Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA; 23) Center for Human Genetic Research and Department of Anesthesia, Pain, and Critical Care Medicine, Massachusetts General Hospital, Boston, MA; 24) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 25) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Introduction: Obstructive Sleep Apnea (OSA) is a common disorder characterized by recurrent episodes of upper airway obstruction causing sleep fragmentation and hypoxemia. OSA increases the risk of cardiovascular outcomes including hypertension, heart failure, coronary heart disease and all-cause mortality. Its etiology is multi-factorial, with growing evidence that abnormalities in respiratory arousal threshold contribute to disease susceptibility. Respiratory arousability can be approximated by measuring the average length of apneas and hypopneas during sleep. A recent large observational study found that shorter event duration, driven by enhanced arousability or chemoreflex, predicted mortality. High event duration heritability has been reported in both African-and European-American family studies (h^2=0.6, 0.4 respectively), indicating a strong genetic basis of this trait. Despite its clinical promise, the underlying genetic etiology of event duration is still under studied. **Methods:** To search for novel genes associated with OSA, we performed a genome-wide trans-ethnic meta-analysis of event duration using polysomnography data collected from multiple large cohorts (N=18,420): including ARIC, CFS, CHS, FHS, HCHS/SOL, JHS, MESA, and MrOS-Sleep. Imputed SNPs in individual cohorts were analyzed using a linear mixed model, adjusting for age, sex, BMI, population structure, and family/cryptic relatedness. Fixed effect meta-analysis was performed across ethnic groups. Secondary analyses stratified by sex and population (within African- and European-Americans) were also performed. Replications were conducted on an independent European-Australian clinical cohort (WASHS; N=1,645). **Results:** Mean event duration spanned 18.5 to 26.9 seconds across cohorts. We observed shorter event duration in females than males and in African-Americans. We identified a genome-wide significant locus (led by rs204544; p=1.74×10⁻⁸; chr19:44449585; ZNF229) in meta-analysis that was replicated in WASHS (p=4.56×10⁻²). Three additional genome-wide significant associations were identified in sex-specific and population-specific analyses (p<5×10⁻⁸). **Conclusion:** This is the first GWAS ever performed on apnea/hypopnea length across populations. Elucidating the genetic determinants for this physiologically-based novel endophenotype may help identify molecular mechanisms influencing respiratory control relevant to OSA and other sleep and respiratory traits.

2698W

Fitness, physical activity, and cardiovascular disease: Longitudinal and genetic analyses in the UK Biobank Study. *E. Tikkanen[1], S. Gustafsson[2], E. Ingelsson[1]. 1) Stanford University, Stanford, CA; 2) Uppsala University, Sweden.*

Background: Exercise is inversely related with cardiovascular disease (CVD), but large-scale studies of incident CVD events are lacking. Moreover, little is known about genetic determinants of fitness and physical activity, and modifiable effects of exercise in individuals with elevated genetic risk of CVD. Finally, causal analyses of exercise traits are limited. **Material and Methods:** We estimated associations of grip strength, physical activity, and cardiorespiratory fitness with CVD and all-cause death in up to 502,635 individuals from the UK Biobank. We also examined these associations in individuals with different genetic burden on coronary heart disease (CHD) and atrial fibrillation (AF). Finally, we performed genome-wide association study (GWAS) of grip strength and physical activity, as well as Mendelian randomization analysis to assess the causal role of grip strength in CHD. **Results:** Grip strength, physical activity, and cardiorespiratory fitness showed strong inverse associations with incident cardiovascular events and all-cause death (for composite CVD; HR, 0.78, 95% CI, 0.77-0.80; HR, 0.94, 95% CI, 0.93-0.95, and HR, 0.67, 95% CI, 0.63-0.71, per SD change, respectively). We observed stronger associations of grip strength with CHD and AF for individuals in the lowest tertile of genetic risk ($P_{interaction}$=0.006, $P_{interaction}$=0.03, respectively), but the inverse associations were present in each category of genetic risk. We report 27 novel genetic loci associated with grip strength and 2 loci with physical activity, with the strongest associations in *FTO* (rs56094641, P=3.8×10^{-24}) and *SMIM2* (rs9316077, P=1.4×10^{-8}), respectively. By use of Mendelian randomization, we provide evidence that grip strength is causally related to CHD (OR = 0.56, 95% CI 0.44-0.70, P=1.4×10^{-7}). The results remained similar in sensitivity analyses **Conclusions:** Maintaining physical strength is likely to prevent future cardiovascular events, also in individuals with elevated genetic risk for CVD.

2699T

Large-scale genomic study of >26,000 MyCode participants uncovers novel loci for hemostasis. *J. Backman[1], S. McCarthy[1], M. Puurunen[1], S. Setia[2], D. Carey[2], D. Ledbetter[2], M. Ritchie[2], J. Reid[1], J. Overton[1], O. Gottesman[1], A. Baras[1], F. Dewey[1], T. Teslovich[1]. 1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY; 2) Geisinger Health System, Danville, PA.*

Venous thromboembolism (VTE), consisting of deep venous thrombosis and pulmonary embolism, is a recurrent and debilitating disease characterized by the formation of blood clots in veins. Family-based studies suggest that genetic variation is a major contributor to VTE risk. However, VTE has a complex etiology, and polymorphisms identified through GWAS account for ~5% of the heritable component of VTE, providing limited insight into genetic underpinnings of the disease. The identification of novel genetic variants that influence VTE risk may illuminate new therapeutic targets and guide the way to safer and more effective alternatives to current therapies for VTE prophylaxis and treatment. Endophenotypes such as activated partial thromboplastin time (aPTT) and the international normalized ratio of prothrombin time (INR), which are well-correlated with thrombosis and measure specific and essential components of coagulation, could be leveraged to gain a more complete understanding of the genetic components of VTE risk. In this study, we utilize exome sequences and genome-wide SNP genotypes for >26,000 MyCode participants as part of the DiscovEHR collaboration to evaluate the impact of rare and common variation on hemostatic measures ascertained from electronic health records. Single variant analysis of aPTT, an *ex vivo* measure of the intrinsic coagulation pathway, led to replication of variants in well-established coagulation loci (near *F5*, *KNG1*, *ABO*, *HLA-DQA1*, *F11*, *F12*) as well as the discovery of a novel association (p=1.6 x 10-7) with a rare (MAF<0.01) missense variant in *SLC14A1*, a glycoprotein that may influence coagulation through modulation of serum urea levels. Gene-level tests of rare loss-of-function and missense variants also identified strong associations (p=2.5 x 10-7) for aPTT with *SLC14A1* and a second gene with no established role in hemostasis (*TOP3B*; p=1.4 x 10-6). Analyses of 3 additional hemostatic measures (INR, Protein C, and Protein S) replicated associations in well-known loci (near *F5*, *F7*, *F2*, *F10*, *PROZ*, *PROCR*), uncovered additional putatively novel variants within these loci, and led to the discovery of 6 novel loci associated with coagulation phenotypes. These findings have potential to improve our understanding of the mechanisms contributing to hemostasis and thrombosis, and follow-up is currently underway to independently replicate these results and extend analysis to risk of clinically-important thrombotic events, including VTE.

2700F

Cross-exposure multivariate interaction tests. *J. Kim, P. Kraft, H. Aschard.* Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA.

Gene-environment (GxE) interaction has proven difficult to identify in human traits and diseases, and various strategies have been proposed to increase statistical power. Recent studies showed that using multi-SNP (single nucleotide polymorphisms) approaches, aggregating genetic information into genetic risk score (GRS) and testing for GRS-by-exposure interaction, is a promising alternative to standard single-SNP-by-exposure interaction. With increasing collection of extensive genetic, phenotypic and environmental dataset, multivariate interaction approaches could be extended in many ways. Here, we assessed performances of a range of multivariate interaction approaches through simulation and real data analysis including over 37,000 individuals. We considered three strategies, 1) an unweighted GRS, 2) a weighted GRS (wGRS) where SNPs were weighted based on the marginal effects, and 3) an omnibus test where all interaction parameters were modeled separately but tested jointly. We focused on binary outcomes, using a generalized linear model. We first assessed the robustness and power in simulation with varying parameters (e.g. number of SNP, minor allele frequency (MAF), and correlation and distribution of exposures). We applied those to a dataset comprising cohorts, the Nurses' Health Study and Health Professional Follow-up Study. We considered 4 exposures (smoking, total energy intake, physical activity, and healthy eating index), 4 diseases (obesity, type 2 diabetes, hypertension, and coronary heart disease (CHD)), and SNP identified in large-scale genome-wide association studies with the diseases. All approaches showed decreasing robustness with lower MAF, increasing correlation between exposures, and increasing number of predictors in the models. However, the decrease was minimal for GRS-based approaches but much stronger for the omnibus test when the number of SNP and exposures was large (e.g. Nsnp=100, Nexposure=10, Nsample=20K). Overall, wGRS tended to show better robustness over unweighted GRS, and GRS-based approaches showed higher power than single SNPs analysis. Preliminary analysis considering interactions between disease-specific wGRS and the exposures (64 combinations) identified a few nominally significant associations. The strongest signal was observed for wGRS of BMI-by-smoking interaction on CHD (P= 0.003). Further stratified analysis showed a stronger effect of BMI wGRS in smokers (b=0.87, P=0.00041) than in non-smokers (b=-0.32, P=0.11).

2701W

A Bayesian approach for detecting gene by environment interactions with common and rare variants. *S.M. Lutz[1], J.E. Hokanson[2], C. Lange[3].* 1) Biostatistics, University of Colorado, Anschutz Medical Campus, Denver, CO; 2) Epidemiology, University of Colorado, Anschutz Medical Campus, Denver, CO; 3) Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

For genetic association studies of common and rare variants, the number of genes and environmental factors is potentially large. Stratifying by environmental risks factors results in a decrease in sample size and can potentially and substantially decrease the power. A full model that includes all interactions and main effects can be unstable or non-identifiable, especially for sparse genetic data. As a result, we propose a Bayesian hierarchical mixture model that allows for all interactions to be investigated simultaneously. We propose a mixture prior with two components: a null distribution and an alternative distribution based on biological plausibility. We allow for a spike and slab prior that alternates between the two prior distributions. The proposed prior framework is very flexible and allows for substantive information to be incorporated into the analysis. Through simulation studies, we show that our Bayesian approach has an increased ability to detect gene by environment interactions for rare and common variants compared to the standard methodology which do not readily incorporate this information. We apply this approach to the COPDGene study, a case-control study of Chronic Obstructive Pulmonary Disease (COPD) in current and former heavy smokers to examine gene by gender interactions for subclinical atherosclerosis.

2702T

StructLMM: Resolving genetic effects due to environmental sample substructure. *R. Moore[1,2], F. Casale[2], I. Barroso[1], O. Stegle[2].* 1) Human Genetics, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom.

Genotype-environment interactions (GxE) are common and manifest in variation of genetic effects across individuals. Accounting for interactions due to exposures, such as diet, smoking status and other lifestyle covariates, within genetic analyses can increase power and provide insights into the interplay between genetic and environmental factors. Whilst there have been numerous attempts to identify GxE effects in population cohorts, existing analysis approaches predominantly consider individual environments one-by-one, or consider discretised subgroups. These strategies entail a severe multiple-testing burden and may miss out on more subtle sample substructure. To address this, we propose StructLMM, a method based on linear mixed models that allows for integrating tens or even hundreds of continuous and/or discrete environmental variables in a joint genetic analysis of GxE. Technically, this is achieved by defining a hierarchical prior on genetic effect sizes between individuals, which allows for generalising association tests and identifying interactions. Our method is computationally efficient, facilitating the analysis of up to hundreds of thousands of samples and hundreds of environments. We first validated our model using simulations, finding increased power for detecting both association and interactions. We then applied StructLMM to 83,028 individuals from UK Biobank, testing for associations and interactions between genetic variants and a broad range of lifestyle factors, affecting BMI. In addition to recovering well-known BMI risk loci such as FTO, MC4R and TMEM18, we identify one novel association downstream of ADAMEC1 and upstream of ADAM7. Our model also retrieves significant GxE for loci that have been linked to GxE, including FTO and SEC16B, the latter has only been reported as nominally significant, and novel interaction effects for ADAM7 variants. Finally, StructLMM can be used to dissect environmental factors that drive the identified GxE signals. We find that distinct environments underlie GxE of individual loci, and consistently find that a model that combines multiple environments, better explains the data than any single-environment model. This indicates that GxE is complex and tends to involve a combination of different environmental, thereby highlighting the value of our approach, enabling the exploration of GxE in deeply phenotyped population cohorts such as UK Biobank.

2703F

Colocalization of gene-psychosocial interactions identifies novel genomic regions for blood pressure among African Americans. *M.A. Richard[1], S.K. Musani[2], M. Sims[2], M. Fornage[1] on behalf of the CHARGE Gene-Lifestyle Interactions Working Group.* 1) Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center, Houston, TX; 2) Department of Medicine, University of Mississippi School of Medicine, Jackson, MS.

Psychosocial factors are becoming recognized as risk factors for cardiovascular disease, particularly among minority populations, but it is unclear to what extent they modify genetic associations with blood pressure (BP). Depression and perceived social support are interrelated and we hypothesized that BP associations identified by accounting for genetic interaction with either depression or social support are consistent with a shared causal variant. In the CHARGE Gene-Lifestyle Interactions Working Group, depression and social support were measured in up to 14,922 individuals of African ancestry in nine cohorts. We performed one degree of freedom (df) single nucleotide polymorphism (SNP) main effect tests and joint two df tests of the SNP main effect and SNP-psychosocial interactions for systolic BP, diastolic BP, mean arterial pressure, and pulse pressure. Bayesian colocalization was performed 500 kb up and downstream of 2,950 regions at least nominally associated with BP in the analyses of individual psychosocial factors (P<1E-5 for 1df SNP main effect or 2df test). We identified 21 genomic regions with posterior probabilities (PP) favoring a colocalized gene-psychosocial interaction: 1) the 2df test had strong evidence of colocalization ($PP_{2df}>0.8$), and 2) the 2df test improved the colocalization PP beyond the main effect analysis ($PP_{2df} - PP_{main} > 0.2$). Many colocalized signals map to genes previously implicated in neuropsychiatric traits, such as schizophrenia, bipolar disorder, suicide risk, and response to antipsychotic medications (*PZDZ2*, *NAV3*, *AIG1*, *EPAS1*, *NRCAM*, *SORBS1*, and *CACNA1C*). We additionally identified evidence of colocalized psychosocial interactions at known BP loci (*SLC39A13* and *L3MBTL4*) and cardiac loci (*RYR2* and *B3GAT2*). Thirteen of the colocalized regions are driven by low frequency or African-specific alleles and eight regions harbored common variation. By identifying shared genetic variants influencing BP in the context of psychosocial factors, these results may help to explain the possible biological connection between psychological and cardiovascular health. We have identified novel candidate genes for BP that may help guide the development of new treatment and prevention strategies aimed at reducing health disparities in African Americans. Supported by HL118305.

2704W

Race, sex, and age differences in GxE association: *EBF1* **gene-by-stress interaction on central obesity differs among Blacks, Whites and male, female in multiple harmonized datasets.** *A. Singh[1], M.A. Babyak[1], M. Sims[2], S.K. Musani[2], B.H. Brummett[1], R. Jiang[1], I.C. Siegler[1], E.R. Hauser[1], R.B. Williams[1].* 1) Duke University Medical Center, Durham, NC, USA; 2) University of Mississippi Medical Center, Jackson, MS, USA.

In a precision medicine framework, it is critical to consider several factors, such as, individual's genetic and molecular makeup; complex physiological aspects of race, sex, and age; and how these factors may interact with environmental factors. In our previous study [Singh et al., EJHG, 2015] we identified a novel CVD risk gene *EBF1* in MESA White samples, where a common variant (rs4704963) contributed to inter-individual differences in central obesity (hip, waist, and BMI) in the presence of chronic psychosocial stress. In a subsequent study [Singh et al., Gen Epi, 2015], we presented an algorithm to build a synthetic measure of stress using the proxy indicators of its components (items) and replicated the *EBF1* x stress interaction in additional datasets. In the current study, we evaluated race, sex, and age differences in the *EBF1* x stress interaction on central obesity traits in nine datasets (MESA, FHS, ARIC, CARDIA, WHI, Jackson Heart Study (JHS), Family Heart Study, CATHGEN, and Duke Caregiver Study). Using a MESA-like 5-component harmonized synthetic stress measure, we replicated (P< 0.05) the *EBF1* x stress interaction in White women of all datasets except CARDIA which was the youngest cohort (mean age 25 years). We were not able to replicate the *EBF1* x stress interaction on obesity traits in Black participants of any dataset except in JHS where we observed a significant (P=0.02) *EBF1* x stress interaction on waist circumference in Black women using an 8-component stress measure. We hypothesized that a MESA-like 5-component stress measure might not be sufficient to capture the chronic stress in everyday life among the Black participants; and that one or more additional stress component(s) or an existing component but evaluated differently (specifically, 'racism/discrimination' and 'not meeting basic needs') may have contributed in observing the relationship between stress, *EBF1*, and central obesity. The *EBF1* interaction P-values for individual items 'racism/discrimination' and 'not meeting basic needs' in JHS Black women were 0.036 and 0.053, respectively, suggesting that the 5-component synthetic MESA-like measure is missing elements related to psychosocial stress in Black participants. In conclusion, this study elucidates the race, sex, and age related differences in the *EBF1* x stress interaction and demonstrates the need for careful evaluation of environmental measures in different ethnicities in cross-ethnic gene-by-stress interaction studies.

2705T

Multi-ancestry genome-wide study incorporating gene-smoking interactions identifies 139 genome-wide significant loci for systolic and diastolic blood pressure. *Y.J. Sung[1], T.W. Winkler[2], L. de las Fuentes[1,3], A.R. Bentley[4], M.R. Brown[5], A.T. Kraja[6], K. Schwander[1], I. Ntalla[7], M.J. Caulfield[7,8], P. Elliot[9], K. Rice[10], P.B. Munroe[7,8], A.C. Morrison[5], L.A. Cupples[11,12], D.C. Rao[1], D.I. Chaseman[13,14] on behalf of the CHARGE Gene-Lifestyle Interactions Working Group.* 1) Div Biostatistics, Washington Univ, St Louis, St Louis, MO; 2) Department of Genetic Epidemiology, University of Regensburg, Regensburg, 93051, Germany; 3) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO, 63110, USA; 4) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892, USA; 5) Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, TX, 77030, USA; 6) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, 63110, USA; 7) William Harvey Research Institute, Clinical Pharmacology, Queen Mary University of London, London, London, EC1M 6BQ, UK; 8) NIHR Cardiovascular Biomedical Research Unit, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, EC1M 6BQ, UK; 9) School of Public Health, Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, Imperial College London, London, W2 1PG, UK; 10) Department of Biostatistics, University of Washington, Seattle, WA, 98195, USA; 11) Department of Biostatistics, Boston University School of Public Health, Boston, MA, 2118, USA; 12) NHLBI Framingham Heart Study, Framingham, MA, 01702, USA; 13) Preventive Medicine, Brigham and Women's Hospital, Boston, MA, 02215, USA; 14) Harvard Medical School, Boston, MA, 02115, USA.

Introduction: Genome-wide association analysis has advanced understanding of blood pressure (BP), a major public health priority with implications for vascular conditions such as coronary heart disease and stroke. Incorporating gene-environment interactions in genome-wide analysis represents an additional route for discovery of BP loci and extending our knowledge of its genetic architecture. **Methods**: We performed genome-wide meta-analyses of systolic and diastolic BP incorporating gene-smoking interactions in up to 610,091 individuals from five ancestries. Stage 1 meta-analyses examined ~18.8 million SNPs and small insertion/deletion (indel) variants among 129,913 individuals from four ancestries (European, African, Asian, and Hispanic). Ancestry-specific meta-analyses were performed using inverse-variance weighting with the 1 degree of freedom (DF) test of interaction effect and the 2 DF joint test of genetic main and interaction effects. The ancestry-specific results were then combined through meta-analysis to obtain evidence of "trans-ancestry" association. Stage 2 analyses examined the top variants selected from Stage 1 (with $P < 10^{-6}$) among 480,178 additional individuals from five ancestries. **Results:** Stage 1 meta-analyses identified 1,993 genome-wide significant ($P < 5 \times 10^{-8}$) and 2,466 suggestive ($P < 10^{-6}$) variants. Among these 114 genome-wide significant loci in Stage 1, 14 novel and 39 known BP loci were replicated in Stage 2 even after accounting for multiple testing. A joint meta-analysis of Stage 1 and Stage 2 results identified 69 additional novel loci and 17 additional known loci ($P < 5 \times 10^{-8}$). **Conclusion:** This first large-scale multi-ancestry investigation of gene-smoking interactions validated 56 known BP loci and identified 83 novel loci. Several novel loci were identified through African ancestry analysis, highlighting the importance of pursuing genetic studies in diverse populations. The novel loci show strong regulatory features and evidence for shared pathophysiology with cardiometabolic traits and addiction traits including smoking. In addition, they highlight a role in BP regulation for biological candidates such as modulators of vascular structure and function (*CDKN1B*, *BCAR1-CFDP1*, *PXDN*, *EEA1*), ciliopathies (*SDCCAG8*, *RPGRIP1L*), telomere maintenance (*TNKS*, *PINX1*, *AKTIP*), and central dopaminergic signaling (*MSRA*, *EBF2*).

2706F

GWAS of red cell distribution width identifies discovery associations at GCNT4, KCNJ3, and chr6p22.1 in admixed U.S. populations: The PAGE Study. *C.J. Hodonsky[1], K.N. Nishimura[2], M. Graff[1], R. Tao[3], S. Buyske[4], L.A. Hindorff[5], J. Kocarnik[6,7], Y. Li[8,9,10], D. Lin[8], R.J.F. Loos[11,12,13], B. Thyagarajan[14], C.L. Wassel[15], A.P. Reiner[7], K.E. North[1,9], C. Kooperberg[7], C.L. Avery[1].* 1) Epidemiology, University of North Carolina Gillings School of Public Health, Chapel Hill, NC; 2) Epidemiology Department, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Department of Biostatistics, Vanderbilt University, Nashville, TN; 4) Department of Statistics and Biostatistics, Rutgers University, New Brunswick, NJ; 5) National Institutes of Health, National Human Genome Research Institute, Office of Population Genomics, Bethesda, MD; 6) Department of Epidemiology, University of Washington, Seattle, WA; 7) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 8) Department of Biostatistics, University of North Carolina Gillings School of Public Health, Chapel Hill, NC; 9) Department of Genetics, University of North Carolina Chapel Hill, Chapel Hill, NC; 10) Department of Computer Science, University of North Carolina Chapel Hill, Chapel Hill, NC; 11) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY; 12) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York City, NY; 13) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York City, NY; 14) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 15) Department of Pathology and Laboratory Medicine, University of Vermont Larner College of Medicine, Burlington, VT.

   Background: Red cell distribution width (RDW) is one of several heritable indices associated clinically with cardiovascular disease outcomes, including stroke and myocardial infarction, and for which large-scale genetic characterization remains incomplete. We thus conducted a genome-wide association study (GWAS) of RDW in Population Architecture using Genomics and Epidemiology (PAGE II) participants genotyped on the Multiethnic Genotyping Array (MEGA). **Methods**: The Illumina MEGA platform was designed to improve common and low frequency variant discovery and fine-mapping in global populations. Approximately 19,000 PAGE participants (19% African American, 81% Hispanic/Latino, 65% female) with RDW measurement and MEGA genotyping were available. We conducted a GWAS of inverse-normal-transformed RDW and ~30 million genotyped and imputed (1000 Genomes Phase 3) autosomal variants in PAGE II participants. For discovery, we define a locus as +/-250kb of a published lead variant. Linear regression models were adjusted for age, sex, study, study center, self-reported race/ancestry, family structure (when applicable), and ten ancestral principal components. **Results**: We identified 13 genome-wide significant ($p < 5 \times 10^{-8}$) associations, including three discovery loci: *KCNJ3* (chr2q24.1, PAGE coded allele frequency [CAF]=0.002, variant specific to African haplotypes), *GCNT4* (chr5q13.3, PAGE CAF=0.28; 1000 Genomes population CAFs range from 0.02 [AFR] to 0.41 [EUR]); and an indel in the histone region (chr6p22.1, CAF=0.09). Six published RDW loci (*SLC12A7*, *SLC12A2*, *PSMB5*, *DNAJA4/WD61*, *HBA1/2*, and *TMPRSS6*) and four loci previously associated with other correlated RBC traits (*TFRC*, *HFE*, *HBS1L/MYB*, and *HBB*) also displayed genome-wide significant associations with RDW in our study population. **Discussion**: Our identification of novel associations underscores the benefits of performing GWAS in ancestrally diverse populations, using arrays that better capture global genetic variation. This opportunity for improved discovery and fine-mapping is highlighted in allele frequency variation by ancestry of all three lead discovery variants, particularly the rare, African-specific SNP at the *KCNJ3* locus. Characterization of the *DNAJA4/WD61* locus is also merited, as *WD61* has a known role in hematopoiesis, and several variants in the *DNAJA4* coding region are trans-eQTLs for *BCL11A*, a beta-globin transcription factor. *In silico* functional characterization of all significant loci is ongoing.

2707W

Cadmium exposure, *MT* polymorphisms and subclinical cardiovascular disease in American Indians: The Strong Heart Family Study. *R. Hou[1], M. Tellez-Plaza[2], G. Chittoor[1], K. Haack[3], P. Balakrishnan[4], S. Laston[5], J.F. Deen[6], L.G. Best[7], J.W. MacCluer[3], R.B. Devereux[8], P.M. Okin[8], K.E. North[9], N. Franceschini[9], J.G. Umans[10], A. Navas-Acien[4,11,12,13], S.A. Cole[3], V.S. Voruganti[1].* 1) University of North Carolina at Chapel Hill, UNC Nutrition Research Institute, Kannapolis, NC; 2) Institute of Biomedical Research Hospital Clinic of Valencia, Valencia, Spain; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY; 5) South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX; 6) Division of Cardiology, Seattle Children's Hospital, University of Washington, Seattle, WA; 7) Missouri Breaks Industries Research Inc., Eagle Butte, SD; 8) The Greenberg Division of Cardiology, Department of Medicine, Weill Medical College of Cornell University, New York, NY; 9) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 10) Medstar Health Research Institute, Hyattsville, MD; 11) Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 12) The Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Medical Institutions, Baltimore, MD; 13) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

   Background: Exposure to cadmium (Cd) has been associated with all cause- and cardiovascular disease (CVD) mortality in the United States. The major sources of Cd exposure are food intake and cigarette smoking. Metallothioneins (MTs) play an important role as scavengers of excessive Cd, however, the effect of variant MT-encoding genes on Cd toxicity is unknown. American Indians have premature, exaggerated CVD mortality compared to the general population and have a higher Cd burden. **Aims:** To assess common genetic effects on urinary Cd (a measure of total body burden) and CVD outcomes, and to assess the association between urinary Cd and single nucleotide polymorphisms (SNPs) in MT-encoding genes in American Indian participants of the Strong Heart Family Study (SHFS). **Methods:** 2000 SHFS participants aged &gt 14 years were stratified by study site location (Dakotas, Oklahoma or Arizona) and 19 MT gene SNPs were genotyped. Association was assessed using measured genotype analysis accounting for family relationships. We also estimated correlations between urinary Cd and subclinical CVD measured by electrocardiogram, echocardiogram and carotid ultrasound. **Results**: Significant genetic correlations were found between urinary Cd and QRS duration (rhog (SE) = 0.46 (0.22), p &lt 0.05) in the Dakota and Oklahoma study centers. When stratified by sex, dietary zinc status or smoking status , the genetic correlation between urinary Cd and QRS duration was significant only in women (rhog (SE) = 0.53 (0.29), p &lt 0.05) and men and women smokers (rhog (SE) = 0.54 (0.26), p &lt 0.05) in the Dakotas. Significant genetic associations were found between urinary Cd and rs8044791 (β (SE) = 0.29 (0.09), p = 0.002) and rs2301234 (β (SE) = -0.12 (0.05), p = 0.02 in the Dakotas only. **Conclusions**: Urinary Cd is influenced by *MT* SNPs and is associated with prolonged QRS duration in American Indian women in the SHFS. Given that QRS duration is an independent predictor of CVD events in American Indian women, assessment of Cd exposure may be useful as a unique CVD risk factor in difficult to assess populations with a high lifetime risk of CVD.

2708T

Genome-wide association study of mitochondrial DNA copy number: The Cohorts for Heart & Aging Research in Genetic Epidemiology (CHARGE). *R.J. Longchamps[1], J.A. Lane[2], M.L. Grove[3], K.S. Lawson[3], E. Boerwinkle[3], N. Pankratz[2], D.E. Arking[1], CHARGE Aging and Longevity Working Group.* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 3) School of Public Health, Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX.

Mitochondria, the key organelle involved in energy production, have long been proposed to play a critical role in human disease and aging. A considerable amount of inter-individual variation exists in the number of mitochondrial genomes per individual, also known as mitochondrial DNA copy number (mtDNA CN). As an easily accessible marker for mitochondrial function, mtDNA CN provides us with the ability to assess the role of the mitochondria in disease. Recent work has found mtDNA CN declines with age and is associated with frailty, coronary artery disease (CAD), sudden cardiac death, and overall mortality. Here, we report interim results from a genome wide association study (GWAS) to identify genetic determinants of mtDNA CN, as well as efforts to determine the causality of mtDNA CN in CAD utilizing Mendelian Randomization (MR). Interim analyses consist of 41,908 individuals (18.2% Black, 2.8% Hispanic, 1.8% Chinese), with an expectation of an additional 52,310 samples by September. mtDNA CN was either estimated from microarray raw probe intensities via the Genvisis software package or from an optimized qPCR Taqman assay. Our final mtDNA CN phenotype is represented as the standardized residuals from a race stratified linear regression adjusting for location and age at which DNA was collected, sex, white blood cell (WBC) count, platelet (PLT) count as well as qPCR plate (qPCR) or PCs generated from nuclear probe intensities (microarray). GWAS were also race stratified and adjusted for age, location, sex, WBC and PLT count as well as GWAS PCs. Meta-analysis was performed using the Metal software package. Interim meta-analyses in whites (n = 32,300) identify one genome-wide significant locus (rs10432360), lying downstream of *ZEB2,* which encodes a zinc finger protein responsible for transcriptional repression. Additionally, we observe an enrichment of loci with p-values < 5e10-7 (3 expected, 6 observed). No loci reached genome-wide significance in other ethnic groups, however power was low due to our current sample size. Interim MR analysis in whites reveals no significant causal relationship between mtDNA CN and CAD (beta = -0.016, p-value = 0.17), however the direction of effect suggests increased mtDNA CN may mitigate CAD risk. Doubling of our sample size may identify additional genome-wide significant loci and further elucidate whether there is a causal relationship between mtDNA CN and CAD.

2709F

"Genotype-phenotype in Marfan syndrome patients with causative mutations in the calcium binding region of the 43 cbEGF-like domains in fibrillin-1 gene (*FBN1*)". *J.A. Aragon-Martin[1], Y.B.A. Wan[1], L.N.A. Belkacem[1], A. Samoto[1], A. Masoha[1], Y. Isekame[1], C.S. Gracias[1], P. Mazadza[1], T. Thayalacumar[1], A. Chandra[2], D.G. Charteri[3], A. Saggar[4], M. Jahangiri[5], A.H. Child[1].* 1) Cardiovascular and Cell Sciences Research Institute, St George's University of London, London, England, United Kingdom; 2) Southend University Hospital, Southend on Sea, United Kingdom; 3) Vitreoretinal Surgery, Moorfields Eye Hospital, London, United Kingdom; 4) Regional Genetics Unit, St. George's Healthcare NHS Trust, Cranmer Terrace, London, United Kingdom; 5) Department of Cardiothoracic Surgery, St. George's Healthcare NHS Trust, Cranmer Terrace, London SW17 0RE, United Kingdom.

Background: Marfan syndrome (MFS) is a phenotypically heterogeneous autosomal dominant monogenic connective tissue disorder due to mutations in *FBN1*. There are 43 cbEGF-like domains that contain a calcium binding region that is highly conserved in FBN1 protein. Calcium ions play a key role in the stability of these domains and increase resistance to proteolysis degradation. This study aims to observe a relationship between the genotype and phenotype in MFS patients. Methods: Patients (n=109) fulfilling the Ghent Criteria and with a mutation in the calcium binding region were entered in the study. Clinical data was obtained from a dedicated Marfan Clinic at St George's Hospital and the genotypes from the *FBN1* mutation-screening database in the Sonalee Laboratory. A genotype-phenotype database was created with this data. Results: This study found that MFS phenotypes may arise due to mutations in any of the calcium binding regions of the 43 cbEGF-like domains: – prevalence of TAA in all cb-Regions (p= 0.00021), cb-Region C (p= 0.00005), cb-Region Exon 52 (p= 0.06061), cb-Region Exon 61 (p= 0.02857) and cb-Region Exon 62 (p= 0.00191); and – prevalence of EL in cb-Region Exon 27 (p= 0.04762). Data showed that phenotypes found in calcium binding regions were EL (35%) and TAA (65%). Males were more likely to have TAA (59%) and Females EL (52%) phenotypes. Conclusions: The study shows **genotype-phenotype** relationship with mutations in **all** cb-Regions, **cb**-Region **C**, cb-Region **Exon 52**, cb-Region **Exon 61** and cb-Region **Exon 62** with **TAA**. When looking at cb-Region **Exon 27** the strongest genotype-phenotype is with **EL**. The variability in phenotypes arising from mutations in the calcium biding region could be due to a variable interaction between the amino acids in this region with the calcium molecule in the protein. Also, it could be due to other factors playing a role, such as environmental or gene modifiers that might increase or reduce the severity of the phenotype. Instability of the calcium binding region due to mutations in this part are more likely to express a TAA than an EL phenotype but the risk of TAA is still very high. Mutations in the cb-Region are not exclusive to either phenotype (TAA or EL) but rather a variable degree of expressions. Due to genetic and phenotypic variability in *FBN1* mutations it is essential to create an international database with single patient genotype-phenotype data to elucidate their relationship. Corresponding author: jaragon@sgul.ac.uk.

2710W

Family based method for the discovery of rare high penetrance sequence variants. *G. Sveinbjornsson[1,3], E.F. Olafsdottir[2], R.B. Thorolfsdottir[1], D.O. Arnar[2], H. Holm[1], D.F. Gudbjartsson[1,3], K. Stefansson[1,2].* 1) DeCode Genetics/Amgen, Inc, Reykjavik, Iceland; 2) Faculty of Medicine, University of Iceland, Reykjavik, Iceland; 3) School of Engineering and Natural Sciences, University of Iceland Reykjavik, Iceland.

Recent advances in whole-genome sequencing has provided opportunities to discover rare, disease causing variants. However, current methods for testing rare variants segregating in families are insufficient. Here we describe a family-based method to test rare sequence variants for co-segregation with disease using data from the deCODE genetics sequencing study. We construct a score for each variant propagating in the pedigree based on the predicted functional effect of the variant and a likelihood ratio based on its co-segregation with the disease. Depending on the probability that an individual is affected given sex, year of birth, county of birth, and lifespan, affected carriers have a positive impact on the log likelihood ratio and unaffected carriers have a negative impact. Analogous to parametric linkage analysis, we assume an effect of the variant, but different from parametric linkage analysis we do not search for linkage with unknown linkage variants but test particular variants with a known set of carriers. To estimate the significance of a variant we use genome-wide simulations where we sample founders from a set of 30K whole genome sequenced Icelanders outside of the family. Using electronic medical records and the Icelandic genealogy we applied our method to a pedigree with a high load of individuals with dilated cardiomyopathy. We found a disease causing missense variant, p.Phe145Leu (genome-wide corrected P = 5.2×10^{-5}, MAF 0.007%) in *NKX2-5*. For verification and deeper phenotyping we reviewed medical records of all individuals carrying the mutation revealing a high penetrance of a serious heart disease with variable expressivity.

2711T

Association of genetic risk score with childhood obesity-related traits: The Santiago Longitudinal Cohort Study (SLCS). *G. Chittoor[1], AE. Justice[2], E. Blanco[3], M. Graff[2], Y. Wang[2], C. Albala[4], JL. Santos[5], B. Angel[4], B. Lozoff[6], VS. Voruganti[1], KE. North[2], S. Gahagan[3].* 1) Department of Nutrition, UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA; 2) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 3) Division of Academic General Pediatrics, Child Development and Community Health at the Center for Community Health, University of California at San Diego, San Diego, CA, USA; 4) Department of Public Health Nutrition, Institute of Nutrition and Food Technology (INTA), University of Chile, Santiago, Chile; 5) Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; 6) Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI, USA.

Background: GWAS of obesity-related traits have identified multiple single nucleotide polymorphisms (SNPs) associated with obesity, although most findings have been revealed in European descent populations (EUR). At the same time, the use of genetic risk scores (GRS) to summarize phenotype – SNP relationships for the prediction of disease and clinical evaluation is growing rapidly. Yet, the applicability and appropriateness of using EUR tag SNPs in GRS' constructed in ancestrally diverse populations is a subject of great importance. Therefore, we aim to investigate the association of body mass index (BMI; age and sex standardized BMI scores – BMIz) with GRS derived from EUR tag SNPs compared to population-specific tag SNPs selected from the 1 MB interval of these known loci in the Santiago Longitudinal Cohort Study (SLCS; N=770; Mean Age=10.03y; 52% boys). **Methods:** BMIz were calculated using the CDC reference (2007–2010) in SLCS participants. We identified the most significant SNP positioned within 1 MB of each of 15 GWAS findings for BMIz in our SLCS cohort. In addition to finding a better tag SNP at 12 of 15 loci (3 tag SNPs were exactly same as in EUR GWAS), we identified 7 independent association signals (22 SNPs total). We estimated simple-count GRS two ways: 1) GRS_EUR: by summing the number of BMIz increasing risk alleles from the 15 known GWAS SNPs; and, 2) GRS_SLCS: by summing the number of independent BMIz increasing risk alleles from the 22 SNPs in 15 loci in our SLCS cohort. Multiple linear regression of GRS and BMIz were performed adjusting for sex, and population substructure using principal components. **Results:** This diverse ancestry GRS, was constructed using 22 SNPs from known obesity loci (e.g. *SEC16B, OLFM4, TFAP2B, FTO*), with EAF ranging from 1% to 49%. Both GRS' were significantly associated with BMIz (p<0.0001), but with a better overall model fit for BMIz–GRS_SLCS (R^2=13.7%) than BMIz–GRS_EUR (R^2=4.3%). Importantly, GRS_SLCS explained 10.08% of the total phenotypic variation in BMIz and GRS_EUR explained only 0.54%. **Conclusion:** We constructed an ancestry specific GRS in a Chilean cohort with extensive admixture. The estimation of GRS–BMIz score associations using published tag SNPs displayed much poorer performance than using ancestry appropriate tags for our population of interest. Our study demonstrates the importance of selecting ancestry appropriate tag SNPs in future studies of disease prediction and clinical evaluation.

2712F

New genetic variants unveiled using a predicted-VAT mass phenotype. *T. Karlsson, M. Rask-Andersen, W.E. Ek, A. Johansson.* Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden.

Purpose: An excess of visceral adipose tissue (VAT) is associated with increased risk of cardiovascular diseases, type 2 diabetes, and mortality[1], more so than other measures of fat mass like body-mass index and waist circumference[2]. However, as measurements of VAT mass, e.g., by dual energy X-ray absorptiometry (DXA), are both expensive and relatively time-consuming, no large-scale genome-wide association (GWA) study has yet been carried out. This study aims to investigate the effect of genetic variants on *predicted* VAT mass using data from the UK Biobank. **Methods:** To construct a nonlinear (interaction) multiple regression model for prediction, DXA-scan-measured VAT mass was regressed on a number of predictors, including age, menopausal status (women), body size and bioelectrical impedance measures. After stratification for sex, the models were calibrated on a training data set, including *4,198* individuals from the UK Biobank cohort. VAT mass was then predicted for all Caucasian participants in the UK Biobank. Finally, we performed GWA studies on *predicted* VAT mass, both for females and males separately and for the combined data set of *116,137* individuals with genotype data available at the time. We used linear regression, adjusted for age, sex (combined set), batch number and principal components. **Summary of results:** Female and male prediction models both had an adjusted $R^2 = 0.76$ while the root-mean-squared-error (RMSE) of the predicted residuals (leave-one-out cross-validation) was *RMSE = 0.28* kg for females and *RMSE = 0.47* kg for males. In the GWA study of the combined data set, *38* loci were detected ($p < 5 \times 10^{-8}$). As a confirmation of the prediction model, the majority of the previously reported loci were associated with BMI and/or obesity. However, *nine* loci were *not* found in the GWAS catalogue. The bulk of the potentially novel loci were found in close connection to genes, such as *RTN4RL1*, *ADARB1*, *ARNTL*, and *PAX2*. Moreover, a few gender-specific new loci were also detected, including one potential locus in *CLYBL* on chromosome 13, for which the effect on predicted VAT mass in males was in the opposite direction ($p = 1.6 \times 10^{-7}$) to that in females. References: [1] *Després, J.-P. et al. BMJ 2001, 322, 716* [2] *Smith, J. D. et al. J Clin Endocrinol Metab. 2012, 97, 1517.*

2713W

Towards precision therapy in hypertension: Genome-wide association study reveals genetic variants associated with uncontrolled blood pressure on thiazide diuretic/beta-blocker combination therapy. *O. Magvan-jav[1,2], Y. Gong[1], C.W. McDonough[1], A.B. Chapman[3], S.T. Turner[4], J.G. Gums[1], K.R. Bailey[5], E. Boerwinkle[6], A.L. Beitelshees[7], T. Tanaka[8], M. Kubo[8], C.J. Pepine[9], R.M. Cooper-DeHoff[1,9], J.A. Johnson[1,9].* 1) Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, College of Pharmacy, Gainesville, FL; 2) College of Medicine, University of Florida, Gainesville, FL; 3) Section of Nephrology, Department of Medicine, University of Chicago, Chicago, IL; 4) Division of Nephrology and Hypertension, Department of Medicine, Mayo Clinic, Rochester, MN; 5) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 6) Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, TX; 7) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD; 8) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 9) Division of Cardiovascular Medicine, Department of Medicine, University of Florida, College of Medicine, Gainesville, FL.

Background: Heart disease and stroke are the leading causes of death worldwide, and hypertension is the primary risk factor. The majority of hypertensive individuals require combination antihypertensive therapy to achieve adequate blood pressure control. Understanding the genetic underpinnings of differential response to common antihypertensive drugs may contribute to realizing the goal of precision therapy for hypertension. This study aimed to identify genetic variants associated with having uncontrolled or controlled blood pressure on combination therapy with a thiazide diuretic and a beta-blocker, two commonly prescribed antihypertensive drug classes. **Methods and Results:** A genome-wide association study of uncontrolled BP on combination therapy was conducted among 314 White participants of the Pharmacogenomic Evaluation of Antihypertensive Responses (PEAR) clinical trial. Multivariable logistic regression analysis was used. Genetic variants meeting a suggestive level of significance (p<1.0E-05) were tested for replication in the INternational VErapamil SR Trandolapril STudy (INVEST). We also examined genome-wide variant associations with systolic and diastolic blood pressure response on combination therapy and tested for replication. We discovered a variant, rs261316, at chromosome 15 in the gene *ALDH1A2* associated with a decreased probability of having uncontrolled BP on combination therapy (OR: 0.39, 95% CI: 0.26-0.59, p=8.64E-06). This variant was replicated in INVEST (OR: 0.54, 95% CI: 0.39-0.74, p=0.001) and approached genome-wide significance in the meta-analysis of discovery and replication cohorts (OR: 0.46, 95% CI: 0.35-0.61, p=8.60E-08). Other genes in the region surrounding rs261316 (*ALDH1A2)* include *AQP9* and *LIPC*. **Conclusion:** A variant in the gene *ALDH1A2* may be associated with a decreased probability of uncontrolled blood pressure following treatment with a thiazide diuretic/beta-blocker combination. **Clinical Trials Registration:** URL: http://www.clinicaltrials.gov. Unique identifiers: NCT00246519; NCT00133692.

2714T

Low frequency and rare variants of *RBFOX1* are associated with blood pressure. *K.Y. He[1], X. Li[1], B.E. Cade[2,3], J. Liang[1], H. Wang[2,3], X. Lin[4], S. Redline[2,3,5], X. Zhu[1], NHLBI TOPMed BP Working Group.* 1) Department of Population Quantitative Health Science, Case Western Reserve University, Cleveland, OH; 2) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; 3) Division of Sleep Medicine, Harvard Medical School, Boston, MA; 4) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 5) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA.

Our previous combined linkage and association analysis of exome array data in the Cleveland Family Study (CFS) and four cohorts with unrelated individuals (N = 57,234) identified rare, functional variants in *RBFOX1*, an evolutionarily conserved RNA-binding protein that regulates tissue-specific alternative splicing, to be associated with SBP and PP. In the present study, 487 European Americans (EA) and 507 African Americans (AA) from CFS have been whole-genome sequenced (WGS) as a part of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. We divided the low frequency and rare variants (defined as minor allele frequency < 5%) of *RBFOX1* into three groups: 1) functional coding variants that lead to an amino acid change, 2) synonymous variants, and 3) non-coding variants. The WGS identified 6 variants in group 1, 4 variants in group 2, and 163 variants in group 3 among CFS European Americans; and 12, 4, 269 variants in CFS African Americans, respectively. Using group 1 variants, gene-based analysis using linkage information verified our previous results that rare, coding variants in *RBFOX1* are associated with SBP (p-value = 0.00332), DBP (p-value = 0.0267), and PP (p-value = 0.0399) in CFS European Americans. WGS enabled us to identify 3 additional variants in EA that were previously not found in exome array. In African Americans, we detected novel association at the gene-level using functional coding variants (SBP p-value = 0.097; DBP p-value = 0.00663). However, the rare variants associated with BP traits in the two populations are mostly population-specific in this gene, suggesting that rare variant replication at a gene-level is more feasible and necessary than at a variant-level. Replication analysis of *RBFOX1* low frequency and rare variants in other TOPMed cohorts are currently underway.

2715F

Incorporating multiple sources of biological knowledge into association analysis of whole genome sequencing data identifies novel trait-associated rare variants. *Y. Ma[1,2], P. Wei[1].* 1) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas; 2) Department of Biostatistics, The University of Texas School of Public Health, Houston, Texas.

Despite the ongoing large-scale population-based whole genome sequencing (WGS) projects such as the NIH NHLBI TOPMed program and the NHGRI Genome Sequencing Program, WGS-based association analysis of complex traits remains a tremendous challenge due to the large number of rare variants, many of which are non-trait-associated neutral ones. External biological knowledge, such as functional annotations based on the ENCODE and Roadmap Epigenomics Project, may be helpful in distinguishing causal rare variants from neutral ones. However, each functional annotation can only provide certain aspect of the biological functions. Our knowledge to select the informative annotations *a priori* is limited while incorporating non-informative annotations will introduce noise and lose power. We propose an adaptive test that incorporates multiple biological annotations and is adaptive at both the annotation and variant levels, thus maintaining high power even in the presence of noninformative annotations. In addition to extensive simulations, we illustrate our proposed test using the TWINSUK cohort (n=1,718) of UK10K WGS data (Walter et al, Nature 2015) based on six functional annotations: CADD, RegulomeDB, FunSeq, Funseq2, GERP++, and GenoSkyline. We identified a genome-wide significant genetic locus on chromosome 19 near gene *TOMM40* associated with LDL, which is replicated in the UK10K ALSPAC cohort (n=1,796). This novel and replicated LDL-associated locus was missed by existing rare variant association tests that either ignore external biological information or rely on a single source of biological knowledge. We have implemented the proposed test in an R package "FunSPU".

2716W

Low frequency and rare variants in multiple genes are associated with sleep related traits using whole genome sequencing data. *X. Zhu[1], J. Liang[1], B. Cade[2,3], K. He[1], H. Wang[2,3], R. Saxena[2,3,4,5], S. Sunyaev[6], X. Lin[7], S. Redline[2,3,8], NHLBI TOPMed Sleep working group.* 1) Department of Population Quantitative Health Scien, Case Western Reserve Univ, Cleveland, OH; 2) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; 3) Division of Sleep Medicine, Harvard Medical School, Boston, MA; 4) Broad Institute, Cambridge, MA 02142, USA; 5) Center for Human Genetic Research and Department of Anesthesia, Pain, and Critical Care Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; 6) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 7) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 8) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA.

A disease trait often can be characterized by multiple phenotypic measurements that can provide complementary information on disease etiology, physiology, or clinical manifestations. We previously performed analysis on the principal components of the phenotypes (PCPs) and principal components of the heritability (PCHs) for six obstructive sleep apnea hypopnea syndrome (OSAHS) related phenotypes using the data from the Cleveland Family Study, which include both African-American and European-American families. We demonstrated that principal components generally result in higher heritability and linkage evidence than individual traits and the PCHs can be transferred across populations. We identified the largest linkage peak for the third PCH (PCH3) on chromosome 8 with a maximum LOD score of 3.7 in CFS European Americans. In this study, 487 European Americans and 507 African Americans from CFS have been whole-genome sequenced (WGS) by the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. There are 144,565 variants identified in the 20Mb linkage region in CFS European Americans. We hypothesized that low frequency and rare variants contribute to the observed linkage evidence. We divided the low frequency and rare variants (defined as MAF<5%) into three groups: 1) functional coding variants, 2) functional non-coding variants and 3) other variants. Under the linkage peak, we observed 88 and 258 genes with at least 3 functional coding and non-coding low frequency or rare variants, respectively. We observed 1.5 to 2.3 fold enrichment of genes associated with PCH3 using gene-based SKAT or burden analysis with or without adjusting for BMI. By restricting to these low frequency or rare functional coding variants that segregate within the families with family-specific LOD scores >0.1, we identified 17 genes with at least 3 low frequency or rare variants. The enrichment was further improved to 7 times and we identified 6 out of the 17 genes with SKAT P<0.05. Three (*DLGAP2, CSMD1* and *ERI1)* of the 6 genes were further replicated in CFS African Americans (SKAT P<0.05). Further replication using additional TOPMed cohorts is underway. Our study suggests that analyzing multiple traits informed by linkage evidence and interrogation of low frequency and rare variants using WGS data can be fruitful.

2717T

Partitioning genome-wide summary statistics improves polygenic risk prediction. *S. Chun[1,2], M. Imakaev[1,2], N. Marziliano[3,4], M.F. Notarangelo[3], P.A. Merlini[4,5], D. Ardissino[3,5], S. Kathiresan[2,6,7,8], M. Orho-Melander[9], O. Melander[10], N.O. Stitziel[11,12,13], S.R. Sunyaev[1,2,8,14].* 1) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 2) Dept of Medicine, Harvard Medical School, Boston, MA; 3) Division of Cardiology Azienda Ospedaliero-Universitaria di Parma, Parma, Italy; 4) Associazione per lo Studio della Trombosi in Cardiologia, Pavia, Italy; 5) Division of Cardiology Niguarda Hospital, Milano, Italy; 6) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 7) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA; 8) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 9) Dept of Clinical Sciences in Malmö, Lund University, Clinical Research Center, Malmö, Sweden; 10) Dept of Clinical Sciences, Diabetes and Endocrinology, Lund University, University Hospital Malmö, Malmö, Sweden; 11) Cardiovascular Division, Dept of Medicine, Washington University School of Medicine, Saint Louis, MO; 12) Dept of Genetics, Washington University School of Medicine, Saint Louis, MO; 13) McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO; 14) Dept of Biomedical Informatics, Harvard Medical School, Boston, MA.

Traditional methods for polygenic risk prediction of complex disease have historically focused on the set of genetic markers associated with disease at a genome-wide level of significance. Newer approaches have been developed to incorporate information from all markers across the genome, however these methods typically require individual level genotypes or depend on accurately specifying the underlying genetic architecture in order to optimally apply shrinkage methods on estimated effect sizes. Here, we propose a novel partitioning-based risk prediction model to achieve non-parametric shrinkage that does not require explicitly modeling the genetic architecture. We partition markers by intervals of discovery association *p*-values and apply a piecewise constant interpolation of shrinkage curve by utilizing a small second cohort providing individual genotypes. Based on simulated genetic architecture of independent markers, we show that our new partitioning-based risk predictor outperforms *p*-value thresholding and even Bayesian shrinkage when the estimated prior deviates from the true genetic architecture. Further, we demonstrate the utility of this new approach by applying it to a myocardial infarction (MI) dataset, which is comprised of summary level data from >85,000 individuals and genotype level data from a sub-cohort (n=2,685). We assessed the ability of this method to predict MI in a completely independent prospective population-based cohort (n=6,708). Our new partitioning method produces significantly more accurate MI predictions compared with a genome-wide significant score (1.39-fold improvement in AUC; *P* = 0.001) and a thresholding-based polygenic score (1.44-fold improvement in AUC; *P* = 0.0001). Future method development in partitioning summary statistics may further improve polygenic risk prediction for common diseases like MI.

2718F

Geographic distribution of polygenic risk of complex traits and diseases in Finland. *S. Kerminen[1], A.S. Havulinna[1,2], I. Surakka[1], A. Palotie[1,3,4,5,6], M. Perola[1,2], V. Salomaa[2], M.J. Daly[1,3,4,7], S. Ripatti[1,8], M. Pirinen[1,8,9].* 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) National Institute of Health and Welfare, Helsinki, Finland; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; 6) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 8) Department of Public Health, University of Helsinki, Helsinki, Finland; 9) Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

Understanding sources of regional differences in disease incidence is important for targeting health care resources optimally. As many complex diseases have a considerable genetic component, it is important to quantify the role of genetics in geographic variation of disease risk. Genetic risk is often estimated using polygenic risk scores (PRS) derived from large GWAS meta-analyses. A challenge with this approach is that a complex interplay of population genetic differences within the target sample and the GWAS data can confound the results. Here we show that, after careful assessments of biases, PRS reveal geographic differences in genetic risk of several complex traits and diseases in Finland.   We estimated polygenic risk across Finland for five diseases: Crohn's disease (CD), coronary artery disease (CAD), rheumatoid arthritis (RA), schizophrenia (SCZ) and ulcerative colitis (UC), and for three traits: waist-hip ratio (WHR), body-mass index (BMI) and height, using summary statistics of publically available GWAS meta-analyses and a geographically well-defined sample of 2,376 individuals from the FINRISK Study. By spatially smoothing PRS across the country we observed clear geographic patterns for all quantitative traits and for CAD, RA and SCZ, but not for CD or UC. We quantified these patterns using a general linear model and observed the strongest effects for longitude on height (P = 2E-60) and WHR (P = 7E-20) and lower but non-zero effects (P<0.02) on CAD, RA, SCZ and BMI.   We performed several analyses to assess possible biases. First, we generated random PRS by permuting the effect sizes. Second, we assessed the geographic distribution of PRS within four risk variant bins that explained equal amount of total variance but included different numbers of variants. Third, we compared SCZ scores that were generated with and without Finnish samples included in the GWAS meta-analysis. Fourth, we tested whether the risk allele frequency differences between eastern and western Finland varied within MAF bins. In these analyses, possible biases did not explain away the geographic differences in PRS.   Our work provides multiple ways to assess possible biases related to PRS and demonstrates that there are geographic differences in genetic risk of several complex traits and diseases in Finland, including CAD and SCZ. Such information is important for future use of PRS for individualized risk prediction and for targeting the health care resources optimally.

2719W

An information theoretic approach to filtering false signals of pathogenicity across ancestrally diverse populations. *A.K. Manrai, I.S. Kohane.* Biomedical Informatics, Harvard Medical School, Cambridge, MA.

Over the past several years, large catalogues of genomic variation such as the Genome Aggregation Database (gnomAD) and 1000 Genomes Project (1000G) have become mainstays of clinical genomics. These resources enable researchers and genetic testing laboratories to filter both novel variants discovered in sequencing studies as well as benign variants previously implicated as pathogenic. Recently, we have shown how these resources are critical for ancestrally diverse populations that have been understudied in the primary literature (Manrai et al., *NEJM* 2016). Although there is widespread agreement about the need to increase diversity in clinical genomics research, there remains limited precise guidance or formalism on how best to accomplish this goal. Here, we start with a popular proposal, "proportional representation," in which ancestrally diverse populations are represented in proportion to their representation in the general population. We study this proposal in the context of using moderate to large control sequence databases (>1,000 to >100,000 individuals) to rule out variation that is too common in the general population to be causal for a particular disease of interest. We develop an information theoretic approach that uses the diminishing returns of increases in sample size to design more efficient schemes to allocate samples amongst a set of ancestrally diverse populations. We use theoretical analyses and simulations leveraging gnomAD and 1000G to demonstrate the optimality properties as well as limitations of this approach. Finally, we apply our approach to filter a set of >4,000 assertions regarding the causality of variants labeled pathogenic for the inherited heart disease hypertrophic cardiomyopathy (HCM) by at least one submitter to the ClinVar database.

2720T

Control of ethnically-stratified vascular risk factors in modeling of intra-cerebral hemorrhage. *S. Marini[1,2], U.K. Lena[1,2], C.J. Moomaw[3], F.D. Testai[3], S.J. Kittner[4], M.L. James[5], D. Woo[3], J. Rosand[1,2], C.D. Langefeld[6], C.D. Anderson[1,2].* 1) Center for Genomic Medicine, MGH, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Department of Neurology, University of Cincinnati College of Medicine, Cincinnati, OH; 4) Department of Neurology, Baltimore Veterans Administration Medical Center and University of Maryland School of Medicine, Baltimore, MD; 5) Neurology, Duke University Medical Center, Durham, NC; 6) Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest University, Winston-Salem, NC.

   Introduction: Presence of vascular risk factors (VRFs) and their net effects vary across ethnic populations. We hypothesize that use of DNA-based ancestry informative markers (AIMs), compared with self-reported race/ethnicity, results in better prediction of VRF status in U.S. minority populations and thus provides more accuracy in attributing risk of intracerebral hemorrhage (ICH) across racial/ethnic groups. **Methods:** Using data from the Ethnic/Racial Variations of Intracerebral Hemorrhage (ERICH) study (a multi-center case-control study of ICH in whites, blacks, and Hispanics), we utilized 15 ancestry-informative single nucleotide polymorphisms (SNPs) to perform principal component (PC) analysis using PLINK v1.9 variance-standardized relationship matrix dimension reduction. Logistic regression (adjusted for age) and tests of two independent samples (χ^2 and Mann-Whitney U tests) were used to compare PC-defined with self-identified race/ethnicity in association with VRFs. Results of sub-analyses within each population were further meta-analyzed to highlight differences between self-identified classifications. Finally, we compared two logistic models for ICH risk (one with self-reported ancestry, one with the AIMs), adjusting for age and hypertension. **Results:** Among 4,935 subjects (median age 61, interquartile range 52–72; cases 43.4%) 34.7% were blacks, 35.1% white, and 30.2% Hispanic. Percentages of hypertension, hypercholesterolemia, and diabetes differed across self-identified race/ethnicity as well as PCs (all p <0.001). However, even within each self-reported population, PC dispersals differed among those with and without each of the VRFs (p <0.001). Regression of each VRF with PCs, adjusting for age, confirmed independent associations of PCs with both hypertension and diabetes exposures (p <0.001). Akaike information criterion (6303 vs. 6321) and Davidson-MacKinnon J test (p <0.001) verified that PC-defined ancestry significantly improved the ICH risk prediction model over self-identified ancestry alone. **Conclusion:** AIMs provide a more detailed assessment of risk exposures in studies employing U.S. minority populations. Particularly among Hispanics and blacks, the use of PCs derived from AIMs adds value in controlling for genetic and environmental exposures unique to these populations. This more inclusive ancestry modeling improves accuracy in risk prediction.

2721F

Association of *SELP* variants and soluble P-selectin levels with type 2 diabetes mellitus: A case-control study. *R. Kaur[1], J. Singh[1], R. Kapoor[2], M. Kaur[3].* 1) Department of Molecular biology and biochemistry, Guru Nanak Dev univetsity, Amritsar, Punjab, India; 2) Carewell Heart & superspeciality hospital, Amritsar, Punjab, India; 3) Department of Human Genetics, Guru Nanak Dev, univetsity, Amritsar, India.

   Statement of purpose: P-selectin, an adhesion molecule, is a marker of endothelial dysfunction. Soluble P-selectin (sP-selectin) is proposed as a predictor of adverse cardiovascular events. The present study was aimed to determine *SELP* single nucleotide polymorphisms (SNPs) and sP-selectin levels in Type 2 diabetes mellitus (T2DM) patients. The future CVD risk, based on brachial-ankle Pulse wave velocity (baPWV), was also assessed in the patients. Methods used: The study subjects comprised of 100 T2DM patients and 100 age- and gender- matched healthy controls from north India. The subjects were screened for demographic as well as clinical characteristics. Two non-synonymous (ns) SNPs of P-selectin i.e. rs6136 (T715P) and rs6127 (N562D) were genotyped by PCR- RFLP. To evaluate the future mortality and CVD risk, patients were segregated on the basis of baPWV and Ankle Brachial Index, Summary of results: Serum P-selectin levels were significantly high in patients as compared to controls (P<0.01). Frequency of rs6127 polymorphism was found to be significantly high (p<0.01) in T2DM patients as compared to controls. The variant allele A was associated with 1.5 fold risk. For rs6136, frequency of heterozygous genotype was found to be significantly high in patients as compared to controls (p<0.01) and the variant allele was associated with 2.4 fold risk. Furthermore, both of these variants were in strong linkage disequilibrium (D'= 0.999). Out of total, 22% of patients were found to have very high risk, 43% subjects with high risk, while 34% subjects have moderate risk. In conclusion, the findings of the present study indicated significant association of *SELP* variants with risk towards T2DM. High P-selectin levels in these patients can be correlated with chronic inflammatory condition in T2DM. Furthermore, baPWV, being a sensitive technique, can offer novel method to assess future CVD risk in these patients.

2722W

Whole genome sequence reveals selection for muscle and cardiovascular functions in sport hunting dog breeds. *J. Kim, F. Williams, D. Dreger, B. Davis, E. Ostrander.* Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

   Modern dogs are distinguished among domesticated species by the vast breadth of phenotypic variation produced by strong and consistent human-driven selective pressure. The resulting breeds reflect the development of closed populations with well-defined physical and behavioral attributes. The sport hunting dog group has long been employed in assistance to hunters, reflecting strong behavioral pressure to produce dogs that can locate and pursue quarry over great distances and variable terrain. Comparison of whole genome sequences (WGS) data between sport hunting and terrier breeds, a group that represents the end of a continuum in both form and function, reveals that genes underlying cardiovascular and muscular functions are under greatest selection in sport hunting breeds, including *ADRB1*, *TRPM3*, *RYR3, UTRN* and *ASIC3*, rather than genes associated with physical attributes. Among the most interesting is the positively selected allele of *TRPM3* which is significantly associated with the increased racing speed in whippets, accounting for 11.6% of variance in racing performance. We also observed strong selection for a high impact mutation in a widely-conserved site in *CDH23*, a gene associated with hearing loss and human Usher syndrome, possibly as a mechanism to control startle reflex in so called "gun" dogs. Together these results provide strong evidence that sport hunting dogs have adapted to their roles by improving endurance, cardiac function, blood fluidity, and startle reflex, demonstrating how strong behavioral selection alters physiology, over form, to create breeds with distinct capabilities.

2723T

Heterogeneity in coronary artery disease GWAS results is associated with pan-tissue eQTL count. *K.W. Johnson[1,2], B.S. Glicksberg[1,2], K. Shameer[1,2], J.T. Dudley[1,2,3].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Institute for Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY.

   Results from Genome Wide Association Studies (GWAS) have primarily driven our current understanding of the contribution of genetics to coronary artery disease (CAD). Genetic characterization is most often performed via discovery of common single nucleotide polymorphisms (SNPs) in large case/control studies and results are ultimately collated with meta-analysis. Recently, a large meta-analysis (*CARDIoGRAMplusC4D Consortium*) using data from a number of genome-wide investigations was made public to enable re-analysis of CAD genomic data. In addition to identifying novel putative CAD-related loci, a central finding from the meta-analysis was the determination of between-study heterogeneity of SNP significance and effect sizes (I^2). In a previous work, we have shown that GWAS discoveries are biased toward evolutionarily conserved genomic positions. Cis-expression Quantitative Trait Loci (cis-eQTL) are evolutionarily conserved genomic regions that regulate the expression of one or more proximal genes. Here, we demonstrate a significant association between the number of tissues in which a SNP functions as a cis-eQTL and the GWAS study heterogeneity for the given SNP. First, we combined 2213 CAD SNPs which reached genome-wide significance ($p < 5 \times 10^{-8}$) with cis-eQTL results in 44 tissues from the Genotype-Tissue Expression (GTEx) consortium. Using negative binomial (NB) regression, we found the $-\log 10$(between-study heterogeneity P Value) significantly negatively predicted the number of tissues in which a SNP functions as an eQTL (NB regression, $p < 0.025$). After splitting the GTEx tissues into cardiovascular related (n=6) and non-cardiovascular related (n=39), we also found that this relationship holds in non-cardiovasular tissues ($p < 10^{-8}$, NB regression) but not in cardiovascular tissues ($p < 0.99$, NB regression). We believe this study demonstrates the utility of making summary statistics from GWAS publically available for re-use. We plan to extend this study by investigating heterogeneity patterns from other public disease-related meta-analyses with the goal of better characterizing GWAS results. In additional, tissue-specific eQTL searches may also lead to novel genes driving pathophysiology of complex diseases.

2724F

Origins and dynamics of the Brazilian population and sickle cell mutations reveal unexpected diversity. *Y. Guo[1], B. Custer[2], S. Kelley[2], E. Sabino[3], A.B. Proietti[4], D. Werneck[5], R. Afonso[6], P. Loureiro[7], C. Lobo[8], L. Preisse[1], M. Busch[2], G.P. Page[9], NHLBI Recipient Epidemiology and Donor Evaluation Study-III.* 1) RTI International, RTP, NC, USA; 2) Blood Systems Research Institute, San Francisco, CA, USA; 3) Universidade de São Paulo, São Paulo, Brazil; 4) Fundação Hemominas, Belo Horizonte, Brazil; 5) Fundação Hemominas, Juiz de Fora, Brazil; 6) Fundação Hemominas, Montes Claros, Brazil; 7) Fundação Hemope, Recife, Brazil; 8) Fundação Hemorio, Rio de Janeiro, Brazil; 9) RTI International, Atlanta, GA, USA.

   Background Sickle cell disease is relatively common in Brazil. Mutations within the HBB gene, specifically the 1 base apart Hemoglobin S mutation (HbS) and Hemoglobin C mutation (HbC), reach relatively high frequency due to balancing positive selection from the protection they appear to provide against certain types of malaria. The frequency differ markedly across Africa, with rates of HbC range from 12.5% (Burkina Faso and Ghana) to less than 1% (Angola and Mozambique) and HbS from 18% (Angola) to about 2% (other Malaria endemic regions across Africa). **Methods** The REDS-III Brazilian Sickle Cell Disease cohort (SCDC) is a cohort of ~2,800 sickle cell patients from six geographically diverse sites within Brazil which has been genotyped with a genome wide Affymetrix Axiom SNP array. The specific sickle cell mutation type was determined by both SNP and SSCP typing. Haplotypes were constructed using 1516 markers spanning >2 MB around HBB. **Results** Principal components analysis using 1000 genomes as reference support the finding that most of the African ancestry in Brazil is related to Luhya (Bantu), congruent with the slave trade data. The sites from the interior state of Minas Gerias had higher rates of Native South American ancestry, while samples from Rio De Janeiro had the highest African ancestry, and Recife and Sao Paulo had the highest levels of European Ancestry. High concordance (>97%) between the SNP and SSCP calling of the HbS and HbC alleles was observed. The Bantu and Benin haplotypes accounted for over 95% of the HbS chromosome in Brazil. The Benin haplotype averages over 1.4 MB in length, the Bantu haplotype around 1.6 Mb, and the HbC around 700 kb in length. As many as 33% of the sickle cell patients in Minas Gerias has a HbC mutation and the Benin HbS haplotype was 40-50%, while in Recife HbC frequency it is as low as 9% and Benin HbS haplotype only 20%. **Conclusions** The data suggest high flow of Africans from the Bite of Benin which is supported by the slave trade data which suggests people brought from there were first taken to Bahia and then into the interior. The lengths of the haplotypes suggest that the HbC is the oldest followed by the Benin and Bantu mutations. This could have implications for the spread of humans and malaria. For the first time, we can accurately call HbC and HbS mutations using SNP based markers and define specific HbS and HbC haplotypes using dense SNP markers.   .

2725W

Integrating biological age and transcriptome markers for predicting the functional recovery potential of patients undergoing mechanical circulatory support surgery. *G. Bondar[2], S. Deshmukh[1], S. Ramachandrula[1], M. Raghanivedha[1], S.C. Cherukuri[1], N. Agrawal[1], M. Deng[2].* 1) Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India; 2) Advanced Heart Failure/Mechanical Support/Heart Transplant, David Geffen School of Medicine at UCLA, Ronald Reagan UCLA Medical Center, 100 Medical Plaza, Suite 630, Los Angeles, CA 90095, USA.

   Heart failure (HF) is a syndrome that results from a mismatch between demand and supply of oxygenated blood in turn leading to systemic organ failure. At any given chronological age and HF stage, some patients are more frail, malnourished, and immunologically compromised, while others are less so. Frail patients with HF not only have poorer outcomes on Optimal Medical Management (OMM) but also have an increased risk for organ dysfunction (OD) and death even after HF therapies such as mechanical circulatory support (MCS)-surgery. For these reasons, we believe that mortality risk does not solely depend on a patient's chronological age. Our goals were: 1. Establish the impact of various clinical variables in predisposition of patients towards mortality risk. 2. Facilitate improved clinical management of patients through pre-operative risk prediction of postoperative organ function recovery and long term survival using a combination of the most prominent preoperative clinical parameters and transcriptome data. Strand NGS, a bioinformatics software was used for alignment and analysis of the RNA Seq data. In this pilot study, a total of 29 patients undergoing MCS surgery were monitored for a period of 9 days starting from a day before the surgery. During this period, blood samples were collected along with a comprehensive set of clinical variables. Based on medical records and clinician's assessment, each of these patients was labeled as high or low risk for OD. Data from 24 patients was used for building prediction models while the remaining 5 patients data were used for validating these models. Multivariate logistic regression was used to identify the smallest set of non-redundant clinical variables that impact risk outcome. Support Vector Machine (SVM) Algorithm was used to build the model using transcriptome data. Combinatorial model was built by combining clinical information and transcriptome data. We observed that Respiratory Rate (RR) and WBC count in combination with chronological age predict the outcome of MCS surgery. The SVM model built on transcriptome data alone had a risk prediction accuracy of 92%. By incorporating RR, WBC count and age into the model, prediction accuracy increased to 96%. This pilot study indicates that the functional recovery potential of a person indicative of biological age has a more prominent role in the recovery process than isolated chronological age and should therefore be weighted for risk prediction before MCS-surgery.

2726T

Integrated analysis using RNA-Seq and ChIP-Seq data to understand the regulation of cardiogenesis. *M. Toufiq, S. Deshmukh, S. Ramachandrula, S.C. Cherukuri.* Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India.

1% of all live births are affected with congenital heart disease in the west and this constitutes a major burden on public health organizations. Transcription factors like NKX2-5 and MEIS1 have been shown to play a critical role in vertebrate heart development. Identifying their expression and targets of these factors, along with the regulatory interactions will be a major step towards understanding the broader cardiac developmental processes. Numerous next-generation sequencing studies utilize multiple technologies to understand and answer complex biological questions. Here, using Strand NGS bioinformatics software, we provide an illustrative example on how to integrate data from different sources to identify differential expression profiles and infer their transcription factor binding sites (using RNA-Seq and ChIP-Seq data) in a combined analysis to define the regulation of cardiogenesis. As a part of this study, we re-analyzed publicly available datasets GSE44576. This RNA-Seq and ChIP-Seq data was generated using the Illumina platform to investigate the expression profiles and genome binding sites. Similar to the author's conclusions, we could identify the mechanism of transcriptional regulation during cardiac differentiation by successive binding of the two homeodomain transcription factors NKX2-5 and MEIS1 on *Popdc2* enhancer. ChIP-Seq helped in finding the binding domains of NKX2-5 and MEIS1, while RNA-Seq aided in finding their impact on the expression of other genes. In addition to these results, we also built architecture of pathway and NLP networks. Pathway Analysis displayed NKX2-5 playing a pivotal role in cardiogenesis and any change in its levels causing multiplying effects on heart development pathway. NLP was used to build a network with plausible interaction of genes mined from interaction DB and a proprietary database constructed by Strand using all the PubMed abstracts. Further, NLP showed that most proteins that interact with NKX2-5 are either enhancers or growth factors which might explain the mechanism of heightened impact on heart development. We intend to showcase Strand NGS as a go to tool for analyzing and interpreting multi-omics data with intuitive workflows and user friendly features.

2727F

Disease-specific variant pathogenicity prediction using machine learning methods improves interpretation in inherited cardiac conditions. *X. Zhang[1,2], N. Whiffin[1,2,3], K.L. Thomson[4,5], R. Walsh[1,2], R. Govind[1,2], H. Watkins[5,6], L. Bottolo[7,8,9], J.S. Ware[1,2,3].* 1) National Heart and Lung Institute, Imperial College London, London, UK; 2) Royal Brompton Cardiovascular Research Centre, Royal Brompton & Harefield Hospitals NHS Trust, London, UK; 3) MRC London Institute of Medical Sciences, Imperial College London, London, UK; 4) Oxford Medical Genetics Laboratory, Oxford University Hospitals NHS Foundation Trust, The Churchill Hospital, Oxford, UK; 5) Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 6) The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 7) Department of Medical Genetics, University of Cambridge, UK; 8) The Alan Turing Institute, London, UK; 9) MRC Biostatistics Unit, Cambridge Institute of Public Health, UK.

Accurate variant interpretation is an unmet need in clinical genetics. The ACMG-AMP guidelines provide a standardised qualitative classification framework to integrate diverse lines of evidence and support variant interpretation. While these guidelines represent an important step forward and a valuable platform for reproducible interpretation, limitations include a lack of transparent and quantitative measures of classification probability and confidence. Many bioinformatics tools have been developed to predict the functional effects of variants given multiple features and derive a likelihood-based score of pathogenicity. However, these tools are typically trained on variants across the genome and do not benefit from specific lines of evidence only available for a subset of well-characterised genes or diseases. We have previously shown that the addition of such gene and disease-specific evidence into a transparent Bayesian logistic regression framework improves variant interpretation. Here we develop a syndrome-specific variant interpretation tool to predict the probability of variant pathogenicity on three inherited cardiac conditions: long QT syndrome, Brugada syndrome and cardiomyopathies. The predictor integrates variant-level evidence, domain-level information, gene-level evidence derived from large disease cohorts and pathogenicity scores reported by various existing genome-wide variant effect predictors. Using state-of-the-art ensemble learning techniques, the prediction model was first built on a training data set of highly confident interpreted rare variants collected from ClinVar, alongside variants derived from healthy individuals confirmed by MRI scan. After training, the model was evaluated in independent disease cohorts with known causative variants and control data sets. Compared to the popular genome-wide tools, our disease-specific predictor significantly improves the classification of variants by reducing the predictive loss. In long QT syndrome, it improves at least by 45.13%. It shows the best performance in terms of the area under the curve as well as the measures of clinical relevance such as specificity and sensitivity. We present that a novel pathogenic variant classifier based on ensemble learning outperforms genome-wide methods by leveraging expert-curated gene and disease-specific data. The use of equivalent variant features and machine learning methods is highly transferable to variant interpretation in other genetic diseases.

2728W

A disease-specific and automated variant annotator enables fast and accurate clinical variant interpretation. *N. Whiffin[1,2,3], R. Walsh[1,3], E. Edwards[1,3], R. Govind[1,3], M. Ahmad[1,3], M. Edwards[3], X. Zhang[1,3], R. Buchan[1,3], W. Midwinter[1,3], H. Najgebauer[1,3], A.E. Wilk[1,3], D.J. Morris-Rosendahl[1,3], P.J.R. Barton[1,3], J.S. Ware[1,2,3], S.A. Cook[1,2,3].* 1) National Heart and Lung Institute, Imperial College London, London, United Kingdom; 2) MRC London Institute of Medical Sciences, Imperial College London, London, United Kingdom; 3) Royal Brompton and Harefield NHS Foundation Trust, London, United Kingdom.

Guidelines from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) aim to standardise variant interpretation by outlining the evidence types to be considered and how these might be weighted. Although the authors stress the importance of refining the framework for individual genes and diseases, to date disease-specific guidance is lacking. To address this shortcoming, we created CardioClassifier, an interactive web-tool that supports semi-automated variant interpretation for inherited cardiac conditions according to the ACMG/AMP guidelines. Our tool is the first to combine automation with expertly curated disease and gene-specific thresholds, using data derived from robustly refined allele frequency thresholds, regional mutational burden analyses, single variant association statistics and intricately phenotyped disease and control cohorts.   To quantitatively assess the added value of incorporating gene- and disease-specificity, we compared our disease-specific CardioClassifier tool to InterVar, a semi-automated generic tool. Using a set of 219 variants validated as Pathogenic (by multiple submitters) in ClinVar, we show that the sensitivity of CardioClassifier is two-fold greater than InterVar, without a loss of specificity.   Additionally, we compared the performance of CardioClassifer for 57 'gold-standard variants', previously manually-curated by the Clinical Genome Resource (ClinGen). After accounting for expected differences in *in silico* effect prediction algorithms and the availability of clinical data, we achieved perfect concordance for 53 (93%) of these variants.   To further increase the utility of CardioClassifier for clinical application, we developed an associated knowledge-base, incorporating segregation and functional data manually curated from the literature. The tool can be continually refined to align with consensus disease-specific guidance emerging from community efforts such as ClinGen and to incorporate new knowledge, and will support data sharing with community resources (ClinVar). Through creation of CardioClassifier we show that using automation to guide variant interpretation markedly reduces the time needed for annotation while increasing accuracy and consistency. We demonstrate the importance of disease-specific variant interpretation and we hope this stimulates the development of similar tools for additional diseases. CardioClassifier is available at cardioclassifier.org.

2729T

Data mining "normal" chromosome microarrays for gene discovery. *N. Walton[1], H. Nguyen[3], P. Jay[2].* 1) Genetics and Genomic Medicine, Washington University, St Louis, MO; 2) Pediatric Cardiology, Washington University, St Louis, MO; 3) Pediatric Cardiology, Rush University, Chicago, IL.

A chromosome microarray (CMA) is a commonly ordered clinical test to ascertain genetic etiologies of congenital heart disease (CHD). Laboratories report the CMA as "normal" if there are no deletions greater than 200kb or duplications greater than 500kb. Hundreds of smaller copy number variants (CNV's) go unreported. It is logical to think that even these unreported CNV's may be informative for research and clinical diagnosis if they overlap genes that cause CHD. We hypothesize that by using the raw data from "normal" CMA's in combination with detailed phenotypic information, that we may be able to discover new genetic causes of CHD. For this study, we analyzed data on 1,762 Patients who had a CMA performed for diagnostic purposes whose results were reported as normal. This included 319 patients with CHD and 1443 without CHD. Charts of all patients were manually reviewed and phenotyped. CHD patients were categorized into 28 different cardiac phenotypes and groups. We developed a computer algorithm that compared the number of deletions in a gene for patients with CHD phenotypes compared to patients without CHD. This algorithm was looped through every CHD phenotype in combination with 18,232 individual genes resulting in a table of 510,496 scores. The results from the analysis were then filtered based on p-value, odds ratio, CNV prevalence, and tolerance to loss of function (pLI, Shet) to produce a list of genes enriched for known CHD genes. Permutation analyses determined the number of known CHD genes that could be expected under the null hypothesis. We further assessed our results for enrichment of genes with de-novo variants in three large CHD exome sequencing cohorts and used a cohort of patients without CHD as a control. Our filtering approach produced greater than 30-fold enrichment of CHD genes that was statistically significant when compared to permuted data. Our method also produced a small set of candidate genes. We further validated our method by showing significant enrichment of genes with de-novo variants in the CHD exome cohorts while showing no significant enrichment in the control population. Candidate genes that overlap between our study and the exome cohorts may serve as good candidates for further study. Two of our "candidate" genes have since been published as CHD genes. The results of this analysis also suggests that there may be valuable diagnostic information regarding deletions in known CHD genes that goes unreported in clinical CMA's.

2730F

Postoperative risk prediction based on preoperative leuckocyte immunobiology. *S. Ramachandrula[1], G. Bondar[2], S. Deshmukh[1], N. Agrawal[1], S.C. Cherukuri[1], M. Deng[2].* 1) Strand Life Sciences Pvt. Ltd, Bangalore, Karnataka, India; 2) Advanced Heart Failure/Mechanical Support/Heart Transplant, David Geffen School of Medicine at UCLA, Ronald Reagan UCLA Medical Center, 100 Medical Plaza, Suite 630, Los Angeles, CA 90095, USA.

Mechanical Circulatory Support (MCS) devices are a therapeutic option for Advanced Heart Failure (AdHF) patients with challenging clinical profiles. Even after MCS surgery, some patients are still at an increased mortality risk due to Multi-Organ Dysfunction (MOD) Syndrome associated with immune cell dysfunction. The ability to precisely predict postoperative risk for each AdHF-patient before surgery would therefore be paramount in clinical decision-making and management. Our goal is to use leukocyte immune-biology information to develop a preoperative test which accurately predicts postoperative outcomes in individual AdHF-patients.   As a part of this pilot study, 29 patients undergoing MCS surgery were monitored for a period of 9 days starting from a day before the surgery. During this period, blood samples were collected along with a comprehensive set of clinical variables to assess MOD. Based on medical records and clinician's assessment, each of these patients was labeled as high or low risk. PBMCs were extracted from the blood samples of preoperative day and used for transcriptome, cytokine, and flow cytometry analysis. We then used this data from 23 patients to build a class prediction model that distinguished low risk patients from those of high risk; and the remaining 6 samples were set aside for testing the prediction model. Data from each platform was imported into Strand NGS, a bioinformatics software, using the most optimal data transformation and normalization options. After quality assessments and filters, high quality data points were used for statistical and fold change analysis. Feature selection methods were employed to find the features that best captured trends in the data and differentiated low-risk patients from those of high-risk. Multiple class prediction models were then built on patient data from each platform using Support Vector Machine, Neural Network, Decision Tree, and Naive Bayes algorithms. In order to leverage the potential of multi-dimensional data from different platforms, combinatorial, sequential, and weighted models were built using features from across platforms. We obtained the following prediction accuracies in classifying pre-operative patients into high or low risk. Such models when constructed using a larger cohort can be used to optimally predict the MOD risk to be faced by a new patient due to undergo MCS surgery.

SVM Model	Accuracy (%)
RNA-Seq	93
Flow Cytometry	91
Cytokines	92
Combinatorial	92

2731W

Genetic variants in familial abdominal aortic aneurysms identified by whole genome and exome sequencing. *A. IJpma[1,2], D. Heijsman[1], Y. Li[2], H. Brüggenwirth[1], I. van den Boogert[1], D. Majoor-Krakauer[1].* 1) Department of Clinical Genetics, Erasmus MC, Rotterdam, Netherlands; 2) Department of Bioinformatics, Erasmus MC, Rotterdam, Netherlands.

Introduction: Abdominal Aortic Aneurysm (AAA) is a frequent disorder with a prevalence of 5% in the elderly population. An Abdominal Aortic Aneurysm occurs when the aorta below the renal arteries expands to a diameter of 3cm or more. In the Netherlands, approximately 5000 AAA patients are hospitalized each year and around 750 people die annually due to AAA rupture. Approximately 20% of AAA patients are familial. In non-familial AAA, genetic predisposition likely also plays a role since known cardiovascular risk factors alone do not explain the localized weakening of the aortic wall. Our goal is to identify the genes that play a role in the formation of abdominal aneurysms. Methods: Our study population consists of approximately 1250 AAA patients. So far we sequenced 548 of these patients of which 365 have a family history of AAA. Complete Genomics whole genome sequencing (WGS) was performed in 3 families (15 individuals)and whole exome sequencing (WES) on the illumina platform using Agilent Haloplex and CRE sureselect exome capturing technology was performed in 71 families (175 individuals) and 358 single AAA patients. Burden analysis was used to identify genes enriched in our AAA population. Prioritization of resulting variants was performed according to the following gene sets: 1. Genes in the diagnostics thoracic or syndromic aneurysms (n=25) panel as applied in the Erasmus MC. 2. All genes in the exome dataset. Results: We present the detailed workflow of the analysis of the genomics data, including the results so far. In 73 out of 365 families a variant with predicted pathogenicity was identified in one of the diagnostic genes. Analysis of all genes in the exome dataset led to the identification of several candidate genes that show variants with predicted pathogenicity in more than one AAA family and that have not been linked to AAA before. We will discuss several candidate genes identified such as the enrichment for pathogenic variants we identified in the COL4A2 gene. Supported by Stichting Lijf en Leven. .

2732T

High-throughput discovery of deleterious cardiac sodium channel variants. *A. Glazer[1], B. Kroncke[1], K. Matreyek[2], T. Yang[1], D. Fowler[2], D. Roden[1].* 1) Vanderbilt University Medical Center, Nashville, TN; 2) University of Washington, Seattle, WA.

A major challenge in genomic medicine is classifying the pathogenicity of variants in Mendelian disease genes. Coding variants in the cardiac sodium channel, *SCN5A*, are a major cause of inherited arrhythmia syndromes including Brugada Syndrome (loss of function) and Long QT Syndrome (gain of function). Although pathogenic variants in *SCN5A* are considered actionable and reportable to the patient/clinician, the pathogenicity of most *SCN5A* variants is unknown or disputed. Here we describe a program to generate high-throughput *in vitro* functional data for *SCN5A* variants. We created a pilot mutant library in a 36 base pair region of *SCN5A* (including known disease-associated mutations) and confirmed by high-throughput sequencing that all 240 possible single amino acid changes were represented. Plasmids were tagged with a random 18-mer barcode and integrated into a HEK293T "landing pad" cell line, resulting in expression of a single variant in each cell. Patch clamp electrophysiology of cells with integrated wild-type channels confirmed a high level of sodium current with typical properties. We next developed an assay to detect cells with dysfunctional sodium channels. A cocktail of three drugs (veratridine, brevetoxin, and ouabain) was used to promote sodium influx and preferentially kill cells with functional channels. This assay successfully discriminated between wild-type channels (35% fitness), a nonsense variant (97% fitness), a partial loss of function variant (71% fitness), and a gain of function variant (16% fitness). High-throughput sequencing of the pre- and post-assay pools will be used to count the prevalence of each variant and identify variants with abnormal function. These experiments demonstrate proof-of-principle methods for a high-throughput *in vitro* screen of *SCN5A* variant function. These methods will ultimately be used across all possible *SCN5A* variants, which may help identify individuals at risk for Brugada Syndrome and Long QT Syndrome.

2733F

Integration of exome genetic variation into mass spectrometry peptide identification to effectively identify plasma proteome QTLs. *T. Solomon[1], J.D. Lapek[2], H. Matsui[3], W.W. Greenwald[4], K. Hindberg[5], N. Latysheva[5], S.B. Jensen[5], S.K. Braekken[5,6], D.J. Gonzalez[2], K.A. Frazer[3,5,7], E.N. Smith[7], J-B. Hansen[5,6].* 1) Biomedical Science Graduate Program, University of California San Diego, La Jolla, CA; 2) Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA; 3) Institute for Genomic Medicine, University of California San Diego, La Jolla, CA; 4) Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA; 5) K.G. Jebsen Thrombosis Research and Expertise Center, Department of Clinical Medicine, UiT - The Arctic University of Norway, Tromsø, Norway; 6) Division of Internal Medicine, University Hospital of North Norway, Tromsø, Norway; 7) Department of Pediatrics and Rady's Children's Hospital, University of California San Diego, La Jolla, CA.

Identifying genetic variation associated with plasma protein levels (protein quantitative trait loci, pQTLs) can provide insight into mechanisms underlying disease. Mass spectrometry can be used to estimate protein levels in a high-throughput manner, however variant peptides are challenging to detect using mass spectrometry and this systematic loss of variant peptide information may influence pQTL identification. Our aim was to integrate known coding variation measured by exome sequencing and genotyping arrays into mass spectrometry protein level estimation in order to more effectively identify and characterize genetic variants associated with plasma protein levels. We performed exome sequencing or exome array genotyping of 170 participants from the Tromsø Study; a longitudinal, single-center, prospective study of the inhabitants of Tromsø, Norway. Additionally, plasma was isolated at study entry and profiled for protein levels using TMT-multiplexed mass spectrometry. Mass spectra were mapped to peptides and proteins using Proteome Discoverer and the Gencode protein sequence database. Mixed models were used to test for associations between common genetic variants near the protein's locus (cis) and peptide or protein levels. We quantified levels of 5,826 peptides that map to 693 Gencode protein IDs and 675 gene loci. We identified common genetic variation associated with 152 of the 5,826 peptides. By examining the coding impact of the most associated SNP, we found that almost 30% of the variants that were associated with peptide levels were missense mutations that disrupt the peptide sequence and cause misidentification of the peptide. In some cases, these missense mutations have known functional effects on the peptide, such as rs854560 (L55M) in paraoxonase 1 (PON1), which decreases PON1 mRNA and plasma levels, while others may not be functional and may have confounded mass spectrometry protein level estimation. We are currently incorporating common and rare nonsynonymous coding variation into the Proteome Discoverer peptide database to determine if this can prevent peptide misidentification and improve pQTL identification. In summary, we have identified cis common variation associated with plasma peptide and protein levels and show that genetic variation often alters mass spectrometry estimates of peptide levels and thus may be important to take into account when identifying or assaying protein biomarkers.

2734W

Large-scale generation of iPSC-derived cardiomyocytes for functional genomic applications. *M.K.R. Donovan[1,2], A. D'Antonio-Chronowska[3], P. Benaglio[4], M. Cook[3], K. Farnam[3], M. D'Antonio[3], E.N. Smith[4], K.A. Frazer[3,4].* 1) Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA 92093, USA; 2) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; 3) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA; 4) Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA.

Regulatory genetic variation associated with molecular and physiological traits often shows effects restricted to a subset of cell types. Experimentally elucidating the functional molecular underpinnings of regulatory variants is challenging because of the lack of genetically diverse human cell models. Human induced pluripotent stem cells (iPSCs) can be derived into cell types that are relevant for disease, such as cardiomyocytes, and may be an effective system for the comprehensive functional characterization of genetic variants in relevant cell types. Here we used iPSCs from the iPSCORE (iPSC Collection for Omic Research) study to derive cardiomyocytes (iPSC-CMs) from 145 different individuals, all of whom had whole-genome sequence data. This collection included 89 individuals belonging to one of 28 families (which include 6 trios and 7 MZ twin pairs) and genetically related to at least one other person, as well as 56 singletons. To achieve large-scale and robust derivations of iPSC-CM, we developed a highly-standardized protocol, which yielded up to 5.7×10^8 iPSC-CM with an average purity of 87.5% and up to 99.5% of cTnT-positive cells measured by flow cytometry. To study how genetic variation influences molecular phenotypes in iPSC-CM lines derived from both the same and different individuals, we analyzed the transcriptome and epigenome (RNA-seq, ATAC-seq, ChIP-seq for H3K27ac) of a subset of 24 iPSC lines derived from 5 monozygotic (MZ) twin pairs. Comparison of these data to reference data revealed cell type-specific transcriptomic and epigenomic profiles that recapitulated those of fetal heart. We then examined genetic and non-genetic sources of variation in the molecular phenotypes by comparing iPSC-CM molecular phenotypes within and across the 24 iPSC lines derived from the 5 MZ twin pairs. We observed that non-genetic factors including the % of cTnT-positive cells and harvest density were associated with variability between clones from the same subject. However, genetic background accounted for most molecular variability in iPSC-CM and these genetic effects were independent of variation associated with non-genetic factors, suggesting that inherited genetic variants associated with molecular traits could be identified. Our study demonstrates that large collections of iPSC-CMs enabling molecular phenotype genetic studies can be created and supports the use of iPSC-derived cells as a model to functionally characterize regulatory genetic variants.

2735T

The iPSCORE resource: 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *E.N. Smith[3], A.D. Panopoulos[1,11], M. D'Antonio[2], P. Benaglio[3], R. Williams[3,4], S.I. Hashem[5], B.M. Schuldt[6], C. DeBoever[7], A.D. Arias[3], M. Garcia[2], B.C. Nelson[5], O. Harismendy[6,7], D.A. Jakubosky[8], M.K.R. Donovan[7], W.W. Greenwald[7], H. Matsui[2], D.T. O'Connor[5], G.W. Yeo[2,10], S.M. Evans[5], N.C. Chi[2,5,8], K. Jepsen[2], N. Nariai[3], F.J. Müller[6], L.S.B. Goldstein[10], J.C. Izpisua Belmonte[1], E. Adler[5], J.F. Loring[4], T. Berggren[9], A. D'Antonio-Chronowska[2], K.A. Frazer[2,3].* 1) Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA; 2) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA; 3) Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA; 4) Center for Regenerative Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA; 5) Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA; 6) Zentrum für Integrative Psychiatrie, Universittsäklinikum Schleswig-Holstein, 24105 Kiel, Germany; 7) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; 8) Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA; 9) Stem Cell Core, Salk Institute for Biological Studies, La Jolla, CA 92037, USA; 10) Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA; 11) Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA.

Large-scale collections of induced pluripotent stem cells (iPSCs) could serve as powerful model systems for examining how genetic variation affects biology and disease. Here we describe the iPSCORE resource: a collection of systematically derived and characterized iPSC lines from 222 ethnically diverse individuals that allows for both familial and association-based genetic studies. Participants were recruited to include families, twins, and individuals of diverse ethnicity to enable genetic studies investigating the segregation of traits. Using high-throughput RNA-sequencing and genotyping arrays, we show that the iPSCORE lines are pluripotent with high genomic integrity (no or low numbers of somatic copy-number variants). We differentiated a subset of iPSCs from a family to cardiomyocytes and examined how the donor's genetic background was associated with gene expression variation in derived cell lines. We show that iPSC-derived cardiomyocytes demonstrate gene expression patterns that cluster by genetic background, and can be used to examine variants associated with physiological and disease phenotypes. Finally, we examined and annotated how individuals in the iPSCORE resource carry SNPs associated with diverse genome-wide association studies (GWAS) phenotypes. The iPSCORE collection contains representative individuals for risk and non-risk alleles for 95% of SNPs associated with human phenotypes through genome-wide association studies. Our study demonstrates the utility of iPSCORE for examining how genetic variants influence molecular and physiological traits in iPSCs and derived cell lines.

2736F

Finding biomarkers for thromboembolism leading to stroke. *D.B. Dogini[1], F.H. Menezes[2], A.T. Guillaumon[2], W.M. Avelar[3], C.S. Rocha[1], I. Lopes-Cendes[1], BRAINN - Brazilian Institute of Neuroscience and Neurotechnology.* 1) Medical Genetics Dept, School of Medical Sciences - UNICAMP, Campinas, Sao Paulo, Brazil; 2) Surgery Dept, School of Medical Sciences - UNICAMP, Campinas, Sao Paulo, Brazil; 3) Neurology Dept, School of Medical Sciences - UNICAMP, Campinas, Sao Paulo, Brazil.

The world is facing an epidemic of stroke. Severe atherosclerotic tightening of the internal carotid artery is found in 20% to 30% of patients with ischemic stroke. However, features associated with a symptomatic plaque including the degree of vessel stenosis, plaque characteristics, ulceration, inflammatory cell infiltration, and others are still poor predictors for the risk of thromboembolism leading stroke. In order identify better biomarkers for carotid disease due to atherosclerotic plaque we apply microarray technology to generate gene expression profile of carotid plaques (CP). In the present pilot study we used 3 CP from patients who underwent endarterectomy and, as control we used 3 healthy extracranial vessels from patients who underwent neurological surgery. Microarray experiments were performed using the Human Genome U133 Plus 2.0 array (Affymetrix™). Data was acquired by the GeneChip Scanner 3000 (Affymetrix™) and analyzed using DNA-Chip Analyzer (dChip). Background correction, summarization and normalization were performed by RMA function and differential gene expression was analyzed using RankProd (FDR $p < 0.05$). These analyses were performed in R environment. We found 4027 genes which were differentially expressed (up and down regulated) in CP as compared to controls. Some upregulated genes found are involved with negative regulation of transcription (GDF15 - growth differentiation factor 15; TCF4 - transcription factor 4; LMCD1 - LIM and cysteine-rich domains 1 and others), cell morphogenesis (NOX4 - NADPH oxidase 4; LST1 - leukocyte specific transcript 1), angiogenesis (PGF - Placental growth factor; ARHGAP24 - Rho GTPase activating protein 24; COL15A1 - collagen, type XV, alpha 1). Some downregulated genes were found to be involved with regulation of cell growth (FAM107A - family with sequence similarity 107 member A; SOCS3 - suppressor of cytokine signaling 3; IGFBP6 - insulin-like growth factor binding protein 6), vasculogenesis (HEY1 - hairy enhancer-of-split related with YRPW motif 1), blood vessel development (PPAP2B - phosphatidic acid phosphatase type 2B; FOXO1 - forkhead box O1). Our results suggest the involvement of a wide variety of biological processes in the formation and maintenance of CP, which need to be better explored in further studies.

2737W

Danon Disease: A lysosomal hypertrophic cardiomyopathy model created by CRISPR editing LAMP2 in iPSC and fibroblasts. *C. McKinney[1,2], K. Baumgarner[1,2].* 1) Genetics and Biochemistry, Edward Via College of Osteopathic Medicine, Spartanburg, SC; 2) Gibbs Research Institution, Spartanburg, SC.

Danon disease, an X-linked genetic disorder, is a dysfunction of LAMP2 protein in cardiomyocyte lysosomes. The disorder affects both sexes but more severely affects males who show an earlier onset of cardiac symptoms resulting in shorter lifespans or cardiac transplant. This disease manifests as a cardiac hypertrophy that may also affect cells of the skeletal muscle. In some patients, mild intellectual disability is also reported. Danon disease is caused by documented mutations in the LAMP2 gene with mutations resulting in frameshifts and several small and large deletions that produce LAMP2 knockouts. Cardiac cells accumulate vacuoles, also seen in the mouse LAMP2 knockout, that appear to contribute to muscle weakness in the heart. LAMP2 protein has three isoforms, differing at the exon 9 sequence, that function as lysosomal membrane proteins and also appear to participate in the cell's autophagy pathway. LAMP2A is a known participant in chaperone mediate autophagy but functional analysis LAMP2B and LAMP2C function remains ambiguous although it maybe involved with the immune system. *We hypothesized that creating cardiomyocytes lacking LAMP2 isoforms and then a specific LAMP2B knockout would allow investigation of the role of LAMP2 in these cells.* Utilizing CRISPR/Cas9 gene editing and iPSC reprograming methods, we created a BJ fibroblast LAMP2 knockout of all three isoforms by placing an insertion early in exon 2. This is confirmed by genomic cleavage assays and Sanger sequencing of the mutant LAMP2 gene. We are reprogramming the LAMP2 KO fibroblasts to iPSC and then cardiomyocytes. Using a single stranded oligo and gRNAs, a second CRISPR introduced mutant in LAMP2 exon 9B models a KO of the LAMP2B isoform. This gives us the opportunity to decipher LAMP2B function in cardiomyocytes. We have reprogrammed the control BJ fibroblast line into cardiomyocytes and shown that they acquire a beating phenotype and express cardiac-specific biomarkers (troponin T2, GATA4 and Nkx2.5). These analyses confirm the cardiac nature of these reprogrammed cells. To further characterize defects in Danon mutated BJ cardiomyocytes compared to our isogenic control line, we plan to assess cardiomyocyte function by MEA techniques, autophagy and lysosomal/mitochondria assays. The study outcome yields two LAMP2 mutant cell lines that probe functional cardiac deficits leading to Danon disease.

2738T

Chromosome 22q11 microdeletion: Modifiers of the cardiovascular phenotype identified by whole exome sequencing. *G. Repetto[1], B. Rebolledo[1], F. Benavides[1], H. Poggi[1], A. Cuiza[1], P. Alvarez[2].* 1) Center for Genetics and Genomics, Clin Alemana- Univ Desarrollo, Santiago, Chile; 2) Hospital Dr Roberto del Rio, Santiago, Chile.

Chromosome 22q11.2 microdeletion syndrome (22q11DS) is the most common microdeletion syndrome in humans, with an estimated frequency of 1/4000. The deletion is caused by non-homologous recombination between low copy repeats (LCRs) in the region, and there are 3 common deletion sizes: 3Mb (between LCRs A-D), 2 Mb (A-C) and 1.5 (A-B). The syndrome shows marked variable expressivity, that is not associated with the deletion size or to parental origin, and suggests the presence of modifiers elsewhere in the genome. Approximately 50-60 % of patients have congenital heart disease (CHD), usually affecting the cardiac outflow tract. CHD is a feature associated with earlier diagnosis and higher case fatality rate. The aim of this study was to search for genetic modifiers of the cardiovascular phenotype in the remaining 22q11 locus or outside of the deletion region, through whole exome sequencing (WES). We included 60 patients with 22q11DS and known cardiac anatomy: 25 cases with conotruncal severe CHD: tetralogy of Fallot (TOF), TOF with pulmonary atresia (TOF+PA) or interrupted aortic arch (IAA), and 35 controls with normal cardiac phenotype. WES was performed at the Northwest Genome Center (NWGC), using Roche Nimblegen V2 Exome capture and 75 bp paired-end sequencing. After quality control, 22 cases and 33 control samples were analysed. We focused on the identification of non-synonymous variants present only in cases. Analysis of the likelihood of pathogenicity for each SNV was carried out using five standard prediction metrics : SIFT, Polyphen2 HVAR, MutationTaster, MutationAssessor and FATHMM. We found 16 variants with potential pathogenicity in cases. Among them, was a variant on KDM6B, involved in histone acetylation and one on CLTCL1, a gene located within the 22q11 deletion that participates in intracellular endosomal trafficking. Variants in histone modifier genes have been previously reported as associated both with non-syndromic and 22q11 deletion-related CHD. These results suggest that, in a small number of patients with the deletion, variants in genes in the remaining 22q11 region or elsewhere in the genome may be associated with the presence of CHD. Funded by FONDECYT Chile grant 1130392 and 1171014.

2739F

AIF-1 in association with TLR-2 induces proinflammatory response in monocytes after ischemia reperfusion (IR). *D.Olga. McDaniel[1], AllenA. Simeone[1], LarryS. McDaniel[2].* 1) Department of Surgery, University Mississippi Med Ctr, Jackson, MS; 2) Department of Microbiology and Immunology, University Mississippi Med Center, Jackson MS.

Inflammatory responses fundamentally influence the short and the long term performance of cardiac allograft after transplantation. Genes associated with innate immunity are prime activators of early inflammatory responses to an allograft and that may lead to host-induced inflammation and organ dysfunction after transplantation. We hypothesized that endogenous substances or damage-associated molecular patterns (DAMPs), released after allograft ischemia-reperfusion (IR) such as allograft inflammatory factor 1 (AIF-1) by interacting with cardiac resident TLR-2 could promote activation of innate immune responses causing phenotypic switching of macrophages from the M2 to M1 subtype and in turn induces a proinflammatory response. AIF-1 and TLR-2 mRNA transcripts were significantly increased in a time-dependent manner in a rat model of the left anterior descending (LAD) artery occlusion after IR. AIF-1 was specifically detected in monocytes by the level of mRNA expression. However, a minimal level of anti-AIF-1 stain was detected by IHC. Functional activity of AIF-1 was confirmed in an *in vitro* model using human coronary vascular smooth muscle cells and monocytes, treated with IFN-g, as well as using HEK293 cells transfected with h-TLR-2 or TLR-4 in which was determined by production of proinflammatory cytokine such as IL-18. AIF-1 treated cells demonstrated 1.25 to 4.5 fold, dose dependent increase in IL-18 secretion in HEK-293 transfected cells with TLR-2. However, a significantly less increase was observed in cells transfected with TLR-4. To investigate the impact of TLR-2 on activation of human HEK293 by AIF-1, cells were incubated with a monoclonal antibody (mAb) to TLR-2 for 1 hour before treatment with AIF-1. The IL-18 production was significantly reduced with mAb to TLR2. Using AIF-1 protein sequence in a 3D protein database, we found a strong homology between domain 2 of human AIF-1 (sdi: 160028) and cardiac Troponin C (1IH0 A). It has been shown that the rate of Troponin release by cardiomyocytes after IR has direct effects on myocardial protection during cardiac allograft preservation. Thus, characterization of AIF-1 as an endogenous ligand for the TLR-2 and elucidation of the mechanisms of an induced inflammation within the myocardial cells has the potential for the development of therapeutic inhibitors for clinical application and further to improve transplantation outcomes.

2740W

Association of rare recurrent copy number variants in next generation sequencing data from family trios with congenital heart defects. *Y. Liu, X. Chang, A. Desai, L. Tian, Y. Guo, D. Li, P. Sleiman, H. Hakonarson.* Center for Applied Genomics CAG, The Children's Hospital of Philadelphia CHOP, Philadelphia, PA.

Congenital Heart Defects (CHD) are common, affecting approximately 1% of newborn children in the United States. A significant number of CHD defects are attributed to alter Copy Number Variation (CNV), usually loss, as reported in previous studies. In this project, we used whole exome sequencing (WES) data generated by the PCGC (CHOP being one of the site), including discovery data set of 2109 individuals from 760 trios, and an independent replication set of 4940 individuals from 1770 trios. The candidate targets uncovered were further validated, using OMNI-2.5M SNP array chip data from 2099 individuals, including 922 CHD patients, as well as in whole genome sequencing (WGS) data generated on 150 individuals from 50 trios. The RNA sequencing (RNA-seq) data used were generated from cardiac tissues obtained from 57 patients, and included left/right ventricle, aorta, right atrium and papillary tissues. The genes harboring CNVs of interest were then tested for expression alternations. We identified five loci harboring CNVs in *LIMS, GCC2, RANBP2, TTC3, MAP3K7CL,* and *FXR1*. Three out of the five loci included *de novo* CNVs. Two of the CNV loci associated with developmental delays. All six genes are highly expressed in human heart/coronary/artery/aorta, and highly expressed in mouse 14.5 days embryo heart except *MAP3K7CL*.Taken together, we have uncovered five novel CNV loci all of which alter expression of the respective candidate genes involved. Raw data is publicly available through the database of genotypes and phenotypes (dbGaP) .

2741T

The effects of missense mutations causing PRKAG2 cardiomyopathy on expression levels of selected genes involved in AMPK pathway. *E. Komurcu-Bayrak, M.A. Kalkan, G. Celebi, N. Coban, B. Ozsait-Selcuk, B. Vural.* Department of Genetics , Istanbul University, Aziz Sancar Institute of Experimental Medicine, Istanbul, Turkey.

Background: Missense mutations in *PRKAG2*, the gene for the gamma 2 regulatory subunit of adenosine monophosphate-activated protein kinase (AMPK), cause cardiac hypertrophy and electrophysiological abnormalities, a condition named as PRKAG2 cardiomyopathy. In our previous study, we reported a novel *PRKAG2* missense mutation (E506K) in three generations of a family with ventricular hypertrophy with normal systolic function and ventricular pre-excitation. The mutation affects a residue in the fourth CBS (cystathionine h-synthase) domain with unknown function, although thought to alter AMPK activity. The aim of this study was to determine the expression profiles of genes involved in the AMPK signaling pathway in vitro. Methods: The *PRKAG2b* E506K, E506Q and R531G mutations were constructed by site-directed mutagenesis technique in pcDNA3.1/V5-His-TOPO expression vector and then transfected into HEK293 cells by lipofectamine system. The PRKAG2 expression was determined by immunofluorescence staining with anti-His-Tag antibody in tagged wild-type and stable mutated cells. The expression of *PRKAG2, PRKAA2, PRKAB2, PFKM, SLCA5A1, SLCA2A4* and *PFKFB3* genes were determined by quantitative RT-PCR in wild-type and stable mutated cells. Results: We found that the expression level of *PRKAG2b* transcript variant (short form composed of 328 amino acid residues) was higher than *PRKAG2d* transcript variant (intermediate form composed of 444 amino acid residues) in the adult human heart and HEK293 cells using variant specific long PCR. Therefore, the 2b transcript encoding the short form was cloned into the expression vector. The mRNA levels of *PRKAG2, PRKAA2* and *PRKAB2* encoding AMPK isoforms showed no significant increase in *PRKAG2* E506K, E506Q and R531G in mutated cells compared to wild-type cells. In contrast, the mRNA expression levels of *PFKM, SLCA5A1, SLCA2A4* and *PFKFB3* were increased by 1.5–5 fold in mutated cells according to wild-type cells. And also we localized PRKAG2 with His-tag by immunofluorescence in the cytoplasm of the mutated cells. Conclusion: In this study, we found the relationship between *PRKAG2* missense mutations and over-expressed four genes involved in AMPK signaling pathway, however further study is required to analyze the AMPK activity. This study was supported by The Scientific And Technological Research Council Of Turkey (Project number: 115S137).

2742F

Investigation of microRNA expression in coronary artery disease. *N. Coban[1], D. Pirim[2], A.F. Erkan[3], F. Guclu-Geyik[1], B. Ekici[3], N. Erginel-Unaltuna[1].* 1) Genetics, Institute for Experimental Medicine, Istanbul, Turkey; 2) Department of Molecular Biology and Genetics, Mus Alparslan University, Mus, Turkey; 3) Department of Cardiology, Medical Faculty, Ufuk University, Ankara, Turkey.

BACKGROUND AND AIM:Although patients with coronary artery disease (CAD) have a high mortality rate, the pathogenesis of CAD is still poorly understood. During the past decade, microRNAs (miRNAs) have emerged as new, potential diagnostic biomarkers in several diseases, including CAD. The aim of this study was to investigate the expression profiles of miRNAs in Turkish patients with CAD and controls (≤30% stenosis). METHODS:Ninety nine patients with angina or acute myocardial infarction who underwent coronary angiography were divided into 2 groups: normal coronary arteries (coronary lesion with ≤30% stenosis) and critical disease (≥1 coronary lesion with ≥70% stenosis). Blood samples were drawn before coronary angiography. Gensini and SYNTAX scores and myocardial blush grade (MBG) were assessed. The Agilent's microarray analyses were performed to compare plasma miRNA profile of selected individuals with CAD (n=12) and without CAD (n=12). MiRDB target prediction tool was utilized to identify miRNA target genes involved in atherosclerosis pathway for differently expressed miRNAs. Expressions of selected miRNAs were analyzed by using real-time PCR in 57 patients with CAD and of 42 controls. RESULTS:Microarray results showed 2 miRNAs (miR-26a, miR-19a) were upregulated and 3 miRNAs (miR-18a, miR-130b and miR-584) were down-regulated (fold change>1.5, p<0.05) in patients with CAD. Based on miRDB target prediction scores, we identifed *TIMP2, ADM, LRP2, GRIA1* and *PHACTR1* genes were strong candidates (target score≥89) for miR-18a, miR-26a, miR-19a, miR-130b and miR-584, respectively. Currently, expression profiles of miR-18a, miR-26a, miR-19a, miR-130b and miR-584 are being analyzed by qRT-PCR in 57 patients with CAD and of 42 controls. CONCLUSIONS:Our findings highlight significantly different pattern of miRNA expression in Turkish CAD patients. These miRNAs might serve as biomarkers of CAD development and progression in Turkish patients and warrant further attention.

2743W

An integrated genetic-epigenetic prediction model for coronary heart disease. *M. Dogan[1,2,3], I. Grumbach[5,6], J. Michaelson[2,3], R. Philibert[2,3,4].* 1) Cardio Diagnostics, LLC, Coralville, IA; 2) Biomedical Engineering, University of Iowa, Iowa City, IA; 3) Psychiatry, University of Iowa, Iowa City, IA; 4) Behavioral Diagnostics, LLC, Coralville, IA; 5) Internal Medicine, University of Iowa, Iowa City, IA; 6) Iowa City Veterans Affairs Healthcare System, Iowa City, IA.

Heart disease is the leading cause of death in the United States. The most common type of heart disease is coronary heart disease (CHD), which is responsible for over 300,000 deaths annually. An improved method for detecting CHD could have a substantial clinical impact given that sudden cardiac death is the initial presentation in 15% of patients with CHD. Building on the idea that systemic effects of CHD risk factors are a conglomeration of genetic and environmental factors, our study was designed to use machine learning techniques and integrate genetic, epigenetic and phenotype data from the Framingham Heart Study to build and test a Random Forest classification model for predicting symptomatic CHD. Genome-wide SNP and DNA methylation data were profiled using the Affymetrix GeneChip HumanMapping 500K and Illumina Infinium HumanMethylation450 BeadChip arrays, respectively. The training set consisted of 1545 individuals (851 females and 694 males) of Northern European ancestry. Of these individuals, 58 females and 115 males were diagnosed with CHD. The test set consisted of 142 individuals (54 females and 88 males) with 22 females and 49 males diagnosed with CHD. Our classifier built using the training dataset consisted of four DNA methylation sites, two SNPs, age and gender. This classifier was capable of predicting CHD status of the test set with an accuracy, sensitivity and specificity of 78%, 0.75 and 0.80, respectively. To determine if this integrated model may be a viable alternative to a model consisting of only conventional CHD risk factors (age, gender, systolic blood pressure, total cholesterol, HDL cholesterol, diabetes and smoking status), another Random Forest model was built on the training set and tested on the test set. In contrast, this model had an accuracy and sensitivity of only 65% and 0.41, respectively, but a specificity of 0.89 in the test set. Regression analyses of the methylation signatures of individual clinical risk factors illustrate the strong role of pathways moderated by smoking in CHD pathogenesis. These results demonstrate the capability of an integrated approach to effectively predict symptomatic CHD status. They also suggest that future studies of biomaterial collected from longitudinally informative cohorts that are specifically characterized for cardiac disease at follow-up could lead to the introduction of sensitive, readily employable integrated genetic-epigenetic models for predicting onset of future CHD.

2744T

Epigenetic modulation in the pathogenesis and treatment of inherited aortic aneurysm conditions. *B.E. Kang[1,4], D. Bedja[2], R. Bagirzadeh[1,4], H.C. Dietz[1,3,4].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Howard Hughes Medical Institute, Bethesda, MD.

   The transforming growth factor β (TGFβ) aortic aneurysm syndromes are unified by striking clinical overlap in the craniofacial, skeletal, cutaneous and cardiovascular systems and a common mechanism, as evidenced by a uniform tissue signature for high TGFβ signaling in the aorta. Primary mutations occur in the genes encoding regulators of extracellular TGFβ bioavailability (*FBN1* or *BGN*), TGFβ ligands (*TGFB2* or *TGFB3*), TGFβ signaling effectors (*TGFBR1*, *TGFBR2* or *SMAD3*), or regulators of the TGFβ transcriptional response (*SKI*). Discordance in the predicted consequence of the underlying mutations and lack of precise pathogenic understanding has engendered ambiguity regarding the most promising treatment strategies. Shprintzen-Goldberg syndrome (SGS) shares most features of Marfan and Loeys-Dietz syndromes (MFS and LDS) with the added finding of intellectual disability. The underlying mutations in the Sloan-Kettering Institute proto-oncoprotein SKI occur in the Smad-binding domain, preventing co-recruitment of SKI to regulatory elements in TGFβ target genes, unopposed histone acetyltransferase (HAT) activity of EP300/CREB-binding protein (CBP), and an amplified TGFβ transcriptional response. Mice heterozygous for a severe SGS mutation (*Ski^{G34D/+}*) phenocopy the human condition; conditional targeting specifically to vascular smooth muscle cells *(Ski^{iVSMC:G34D/+})* causes highly penetrant aortic root (AoR) aneurysm and dissection. This occurs in association with all of the histologic and functional hallmarks of MFS and LDS, including aortic wall thickening with collagen accumulation and elastic fiber fragmentation, an overt mRNA signature for a TGFβ synthetic repertoire (TSR), and a therapeutic response to angiotensin receptor blockers. In consideration of mechanism, we hypothesized therapeutic potential for HAT inhibitors (HATi). Remarkably, systemically-administered C646 (a selective EP300/CBP HATi) prevented aneurysm progression in *Ski^{iVSMC:G34D/+}* mice in association with preservation of aortic wall architecture, prevention of excessive H3K27 acetylation (both globally and at TGFβ target gene enhancers), and normalization of the TSR. C646 also normalized TGFβ target gene expression in SGS patient fibroblasts, with comparable efficiency as a TGFβ type I receptor (Alk5) kinase inhibitor. These data unambiguously implicate the TGFβ transcriptional response in the pathogenesis of AoR aneurysm and highlight novel therapeutic strategies aimed at epigenetic modulation.

2745F

Gender transcriptome signatures for congenital heart defects (CHD) children based on next generation sequencing technologies of cardiac tissues. *C. Kim, Y. Li, Y. Guo, X. Chang, L. Tian, H. Hakonarson.* The Center Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

   Congenital heart defects (CHD) are the most common form of birth defects, affecting nearly 1%, or ~40,000 children in United States annually. Previous studies show that the gender mortality rates are different, more specially, males are more likely to die from congenital heart defects than females, and that genetic factors may be associated with those differences. In this study, we compared the gender transcriptome differences using RNA sequencing based on 125 cardiac samples of seven tissue types from 113 CHD patients of European American (EA) ancestry. We identify a set of differentially expressed genes and showed tissue specific expression profiles. The target genes were examined for structural variations and if highly expressed in developing mouse heart at 14.5 days. Those signatures were further filtered by structural variations based on whole exome sequencing from 96 blood samples from CHD patients, a subset of the 113 CHD patients. Integrative analysis uncovered four genes: *ELN, MYH7, MYOM1,* and *RYR2* as the transcriptome signatures for CHD gender differences among the tissue samples evaluated. Those four genes were differentially expressed between the genders and are all highly expressed in the developing mouse heart and they are nominally associated with CHD related structural variations. We conclude that genes that are significantly differentially expressed between genders may contribute to the differences in mortality observed between genders with CHD.

2746W

Interpreting genetic variation in coronary artery disease (CAD). *I. Selvarajan[1], M. Laakso[2,3], MU. Kaikkonen[1].* 1) Department of Biotechnology and Molecular Medicine, University of Eastern Finland, A.I.Virtanen Institute for Molecular Sciences, Kuopio, Itä-suomi, Finland; 2) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland; 3) Department of Medicine, Kuopio University Hospital, Kuopio, Finland.

Genome-wide association studies (GWAS) have identified many significant and suggestive loci for coronary artery disease (CAD). It is also known that almost 90% of the disease causing genetic variants are present at the non-coding region. It is important to interpret the variants present at the enhancer level that regulate gene expression. Analysis of enhancer activity and gene expression correlation is an important method for discovering relevant transcription factors and their specific functionality, which can be used to gain better understanding of transcriptional regulation and disease states. Hence, we aim to functionally characterize of SNPs associated with CAD by focusing our search for causal SNPs to enhancers of the disease relevant cell types, namely endothelial cells, smooth muscle cells and macrophages of the vessel wall as well as liver and adipose tissue. The enhancers of these cell types harbour 32% of all CAD-SNPs. By identifying the target genes of the candidate enhancers and establishing causal relationships between the enhancer activity and gene expression we hope to obtain a more complete picture of the gene regulatory events driving the disease progression. Shedding light into the molecular mechanisms of these variants will bring a complete picture of gene regulatory programs and causal events driving the progression of atherosclerosis and has the potential to improve risk prediction, biomarker identification and personalized medicine in the clinical practice.

2747T

Epigenetic regulation of PAR-4-mediated platelet activation: Understanding the mechanistic links between smoking and cardiovascular disease. *N. Timpson[1,2], L. Corbin[1,2], A. Taylor[1,3], S. White[5], C. Williams[6], J. Teasdale[4], M. Jones[4], M. Bond[4], M. Bosch[6], M. Harper[11], L. Falk[1], A. Groom[1], B. Nordestgaard[8,9,10], A. Tybaerg-Hansen[8,9,10], C. Relton[1,2], S. Bojesen[8,9,10], A. Mumford[7], A. Poole[6].* 1) MRC/UoB IEU/SSCM, Bristol Univ, Bristol, United Kingdom; 2) School of Social and Community Medicine (SSCM), University of Bristol, Bristol, UK; 3) UK Centre for Tobacco and Alcohol Studies and School of Experimental Psychology, University of Bristol, Bristol, UK; 4) School of Clinical Sciences, University of Bristol, Bristol, UK; 5) School of Healthcare Science, Manchester Metropolitan University, Manchester, UK; 6) School of Physiology, Pharmacology and Neuroscience, University of Bristol, Bristol, UK; 7) School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK; 8) Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark; 9) Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 10) The Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen University Hospital, Copenhagen, Denmark; 11) Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, UK.

We have used a set of complementary experiments to show that *F2RL3* hypomethylation a robust epigenetic index of smoking) results in upregulated F2RL3/PAR4 expression and increased platelet responsiveness. This suggests a mechanism for the cardiovascular effect of cigarette smoke exposure. In a sample of acute myocardial infarction cases and controls from the Copenhagen City Heart Study there was evidence of association between death and methylation at *F2RL3* after adjustment using methylation at the independent locus *AHRR* as an objective measure of smoking (HR per 10% decrease in methylation: 1.21, 95%CI:1.12,1.31). In a two-step Mendelian randomization analyses using the same sample and in UKBiobank, we are able to show that smoking intensity is likely to be causal in affecting methylation at *F2RL3* and that there is evidence that differential methylation at *F2RL3* is causally associated with cardiovascular events involving thrombosis (OR per standard deviation decrease in methylation: 1.29, 95%CI:1.08,1.44). Having established a likely causal pathway we interrogated the biological mechanisms underpinning it. Cell modelling provided evidence of the effect of exposure to cigarette smoke extract on differential methylation (up to -2.3%) and expression at *F2RL3* (5.4 fold increase in *F2RL3* mRNA levels (95%CI:3.9,7.6). Functional analysis of the expression of PAR-4 supports the importance of a methylation regulated recognition site for CCAAT-enhancer binding factor (CEBP) at *F2RL3*. Recruitment and collection of fresh blood in non-smoking participants selected based on differences in methylation at *F2RL3* was undertaken and shows association between methylation and platelet function given exposure to PAR-4 specific agonists (integrin activation (EC_{50}, p=0.001) and P-selectin exposure (EC_{50}, p=0.05) high versus low methylation) and in comparison to the negative control (PAR-1). Lastly and importantly for clinical interpretation of these results, we are able to show evidence that smoking-induced hypomethylation of the *F2RL3* CpG sites is likely to persist for decades after cessation with methylation levels returning to levels similar to never smokers around 20 years after quitting. Overall, observations suggest that the epigenetic regulation of *F2RL3* could be a route between smoking and coagulation response. PAR4 antagonists are an emerging class of anti-platelet drugs able to reduce risk of thrombosis, but cause less bleeding than existing anti-platelet drugs.

2748F

Silencing the miR-30c-5p pathway attenuates cocaine-induced increases in blood pressure and aortic stiffness. *W. Zhu, H. Wang, J. Wei, G.C. Sartor, M. Bao, C.T. Pierce, C.R. Wahlestedt, D.M. Dykxhoorn, C. Dong.* University of Miami, Miami, FL.

   Introduction: Cocaine abuse increases the risk of hypertension and aortic stiffness; however, the underlying molecular mechanisms remain elusive. We previously detected that cocaine exposure upregulates miR-30c-5p expression, which targets on inhibiting malic enzyme 1 (Me1) expression, leading to reactive oxygen species (ROS) acceleration in the mouse aorta. **Goal:** We aimed to test *in vivo* a functional role of the miR-30c-5p-Me1-ROS pathway in mediating the effect of cocaine on increasing blood pressure (BP) and aortic stiffness. **Methods:** C57BL/6 mice were injected (ip) daily with cocaine (20mg/kg body weight) or saline for 10 consecutive days. In parallel, mice were injected via tail vein daily for 5 days with lentiviral vectors (2x10^5 IU per mouse) encoding the miR-30c-5p antagonist (miR-Zip30c-5p) or no gene (vector Ctr) under driven by a smooth muscle cell-specific promoter, and subsequently exposed to cocaine. Additional mice were injected with N-Acetyl cysteine (NAC, 250mg/kg body weight), a scavenger of free oxygen radicals, one day prior and daily during cocaine injection. BP and pulse wave velocity (PWV) for aortic stiffness were measured. MiR-30c-5p and Me1 expression was determined by RT-qPCR. Superoxide (O_2^-) was detected by dihydroethidium. **Results:** Relative to saline treatment, cocaine treatment led to a significant increase in systolic and diastolic BP over the entire treatment course. However, the cocaine-induced increases in BP were significantly reduced from day 3 to 10 in animals that were pretreated with miR-Zip30c-5p but not the vector Ctr. Treating mice with NAC was also able to largely inhibit the increase in BP seen upon cocaine exposure. Similarly, pretreatment with miR-Zip30c-5p or NAC reduced the cocaine-increased PWV by almost 20%. Concomitantly, in contrast to a significant reduction in Me1 expression upon cocaine exposure alone, pretreatment with miR-Zip30c-5p showed only a modest decrease in Me1 expression. Moreover, compared to cocaine treatment alone, pretreatment with miR-Zip30c-5p and NAC reduced the level of cocaine-induced increases in O_2^- by 56% and 45%, respectively. **Conclusion:** Blocking the miR-30c-5p-Me1-ROS pathway, either by genetically silencing the miR-30c-5p or pharmacologically impairing ROS production with NAC, suppresses the cocaine-increased BP and aortic stiffness. This miR-30c-5p-related pathway provides a novel mechanistic base for potential therapeutic interventions to treat cocaine abuse-induced CVD.

2749W

Epigenome-wide association study of the previous number of strokes in participants from the Vitamin Intervention for Stroke Prevention (VISP) clinical trial. *N.M. Davis-Armstrong[1], W.M. Chen[2,3], M.S. Brewer[1], S.R. Williams[4], M.M. Sale[2,3], B.B. Worrall[3,4], K.L. Keene[1,5].* 1) Department of Biology, East Carolina University, Greenville, NC; 2) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 3) Department of Public Health Genomics, University of Virginia, Charlottesville, VA; 4) Department of Neurology, University of Virginia, Charlottesville, VA; 5) Center for Health Disparities, East Carolina University, Greenville, NC.

   To evaluate epigenetic influences associated with recurrent stroke and related traits, we performed an epigenome-wide association study (EWAS) assessing differential methylation using DNA extracted from whole blood and Illumina HumanMethylation450 BeadChip arrays in 180 ischemic stroke patients from the Vitamin Intervention for Stroke Prevention (VISP) clinical trial. VISP was a multi-center, double-blind, randomized, controlled clinical trial designed to determine whether daily supplementation of folic acid, vitamin B6, and vitamin B12 reduced recurrent cerebral infarction. Following quantile normalization, filtering, and quality control, a total of 473,864 autosomal CpG sites were analyzed. A multiple linear regression approach was used to assess the association between the number of strokes prior to VISP enrollment and the degree of methylation (beta value), stratified by race (European descent, EA n=104; African descent, AA n=76). Two differentially methylated CpG sites exceeded genome-wide significance ($p \leq 1.055 \times 10^{-7}$) in the EA stratum. A locus, cg22812874, in the ankyrin repeat and SOCS box containing 10 gene (*ASB10*) demonstrated the most significant difference in methylation ($p=3.4 \times 10^{-9}$). The second significantly associated CpG site, cg00340919, is located in intron 4 of the tetratricopeptide repeat domain 37 gene (*TTC37*; $p=8.74 \times 10^{-8}$). It was observed that a lower number of prior strokes were associated with higher methylation beta values for both CpG sites. *ASB10* is upregulated by tumor necrosis factor alpha and interleukin-1 beta cytokines, both of which contribute to neurotoxicity in rats after an ischemic event, as well as being primary inflammation markers and risk factors of ischemic stroke in humans. The *TTC37* gene product is highly expressed in vascular tissues and has been implicated in trichohepatoenteric (THE) syndrome. Platelet abnormalities, such as large platelet size and thrombocytosis are commonly observed in THE syndrome patients and have been implicated as risk factors for stroke. In conclusion, we have identified two novel epigenetic modifications that influence stroke recurrence.

2750T

Enhancer signature of dilated cardiomyopathy. *D. Hemerich[1,2], M. Harakalova[1], J. Pei[3], E. Nagyova[1], I.R. Efimov[4], B. Boukens[4], C. Cheng[3], R. Weger[5], A. Vink[5], V. Tragante[1], M. Mokry[6], F.W. Asselbergs[1,7,8].* 1) University Medical Center Utrecht, Utrecht, Utrecht, Netherlands; 2) CAPES Foundation, Ministry of Education of Brazil, Brasilia, Brazil; 3) Department of Nephrology, UMC Utrecht, Utrecht, NL; 4) Department of Biomedical Engineering, George Washington University, Washington DC, US; 5) Department of Pathology, University Medical Center Utrecht, Utrecht, NL; 6) Department of Pediatrics, Wihelmina Children's Hospital, Utrecht, NL; 7) Institute of Cardiovascular Science, UCL, London, UK; 8) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, NL.

   Cardiovascular diseases are a leading cause of death worldwide. Dilated cardiomyopathy (DCM) is the most common type of cardiomyopathies and a major cause of heart failure and mortality. Despite the identification of multiple highly penetrant causal genes, the regulatory processes that mediate biological mechanisms of DCM remain incompletely understood. Distal regulatory sequences, such as enhancers, have been shown to act on gene activation in a cell-type specific manner. Disruption in such sequences is believed to contribute significantly to the aetiology of traits. We assessed the hypothesis that distal enhancers play a significant role in heart functions and the identification of heart enhancers in disease state can shed a light into the mechanisms of DCM. To identify enhancers and their role in progression of human heart disease, we profiled epigenomic (H3K27ac ChIP-seq) and gene expression (RNA-seq) of eight normal human heart samples and six DCM patients. Our analysis identified thousands of both hyper (10,668) and hypo-acetylated (5,951) regions. Significantly enriched biological processes and disease ontology include regulation of cardiac muscle contraction, growth and cell-differentiation, abnormal cardiac development and remodelling. Additional analysis revealed the enrichment of JNK MAPK, PDGFRα and Notch signalling pathways, known to be involved in cardiac remodelling and whose mutations result in heart defects. Motif analysis using public data on DNase I hypersensitive sites of left ventricle identified 274 transcription factors enriched on hypo-acetylated regions, and 143 unique motifs enriched on hyper-acetylated peaks. Further investigation on the target genes of these putative enhancers can facilitate the understanding of disease biology and the establishment of an epigenetic signature of dilated cardiomyopathy.

2751F

Identification of eQTLs for platelet and hemostasis related genes in platelets and leukocytes within the Framingham Heart Study. *J.E. Huffman[1,2,3], J.D. Eicher[1,2], K. Tanriverdi[4], J.E. Freedman[4], C.J. O'Donnell[1,2,3,5], A.D. Johnson[1,2].* 1) Population Science Branch, National Heart, Lung, and Blood Institute, Framingham, MA; 2) The Framingham Heart Study, Framingham, MA, USA; 3) Center for Population Genomics, Boston VA Healthcare System, Jamaica Plain, MA; 4) Department of Medicine, University of Massachusetts Medical School, Worcester, MA; 5) Cardiology Section Administration, Boston VA Healthcare System, West Roxbury, MA.

   Background: Platelets and hemostatic factors are important in thrombosis, inflammation, and wound healing. Their functions and signaling pathways interact in the tight control of these processes. Consequently, platelets and hemostatic factor measures have been associated with cardiovascular disease, myocardial infarction, stroke, and venous thrombosis. Identifying key genetic regulators of platelets and hemostatic factors and their mechanisms are important to understanding their contribution to disease. **Aim**: To identify expression quantitative trait loci (eQTLs) for platelet and hemostatic factor associated genes. **Methods:** Platelet and/or hemostasis related transcripts were selected based on known biological function or prior GWAS findings. RNA expression levels were measured by qRT-PCR in leukocytes and platelets from ~1,600 Framingham Heart Study participants. *Cis*-eQTLs (±1 Mb of gene) and *trans*-eQTLs were identified corrected for age, sex, and other technical covariates using 1000Genomes imputed genotypes. A Bonferroni significance threshold was used for *cis*-eQTLS based on the number of SNPs within the *cis* region and a $p<5E-08$ was used for *trans*-eQTLS. eQTLs were annotated using megakaryocyte histone marks from the BLUEPRINT epigenome project, conditional analyses performed to identify independent loci, and disease associations identified using GRASP and GWAS Catalog. **Results:** Significant *cis*- and *trans*-eQTLs were identified in both leukocytes and platelets (Table 1). Of the *trans*-eQTLs, SNPs within the well-characterized *ARHGEF3* region were associated with expression of 14 transcripts including *VWF* and *FADS2*. Additionally, we observed intriguing associations of: 1)SNPs near *GP1BA*, a critical platelet receptor for vWF, with platelet *CCL5* expression and 2)SNPs in *AP2B1*, a clathrin-dependent endocytotic factor associated with platelet count in prior GWAS, with platelet *ITGA2B* expression. Significant *cis*-eQTLs were enriched for activating histone marks (p<0.0005) or those identifying promoters or enhancers (p<0.0002) in megakaryocytes from BLUEPRINT. **Conclusion:** We identified numerous *cis*- and *trans*-eQTLs of platelet and hemostatic factor related transcripts. Such insights into their regulatory architecture may help reveal how they influence thrombosis and hemostasis, and contribute to human disease.

Table 1: Summary of RNA Transcripts Investigated			
	Leukocytes	**Platelets**	**Overlapping**
Transcripts Analyzed	138	82	77
Significant *cis*-eQTLs	56	29	16
Significant *trans*-eQTLs	41	8	2

2752W

Characterization of experimentally validated heart disease genes using functional genomic information and 3d genome structure. *R. Gill, B. Pei, Y. Hsu.* Amgen, Cambridge, MA.

   Introduction The majority of disease-associated SNPs are present in non-coding regions, and some reside within putative regulatory elements, such as enhancers. Understanding the relationships between these enhancers and their target genes could identify genes directly relevant to disease patho-physiology, and distinguish causal variants from other disease-associated SNPs. Evidence of physical interaction between enhancers and target genes, including topological features derived from Hi-C data, are critical to delineating disease-relevant regulatory circuits, but need to be combined with histone and ATAC-seq signatures of active enhancers and promoters along with gene expression, from disease-relevant cellular contexts, in order to accurately predict regulatory relationships. Method We hypothesize that the cardiac fibroblast represents a heart disease-relevant cellular context, and generated functional genomic data, including high-resolution (1.6-kb) Hi-C, ATAC-seq, histone ChIP-seq, and RNA-seq data from human primary fibroblasts from the atria and ventricles. All data processing was performed on DNA Nexus. We trained a deep neural network (DNN) to integrate multiple topological and functional genomic features of 58 pairs of heart disease genes and proximal (within 100 kb) enhancers with (22 positive pairs) and without regulatory effects (36 negative pairs), based on experimental validation in E11.5 murine developing heart and blood vessels. As input, we used pairwise information from enhancers and promoters, including normalized contact frequency, signatures of active enhancers and promoters from histone marks and open chromatin, and correlation between pairs of ATAC-peaks, to learn their weighted contribution to positive and negative enhancer-gene pairs. Our prediction set consisted of genome-wide combinations of enhancers and promoters active in cardiac fibroblasts, based on ChIP-seq and ATAC-seq data. Results and Summary We first examined topologically associated domain (TAD) boundaries and contact frequencies between enhancer-gene pairs. As expected both positive and negative pairs of heart disease genes and enhancers resided in the same TAD, and positive pairs had increased significant interactions. The DNN will learn higher-level representations of the inputs important for classification, such as hidden features of Hi-C data predictive of regulatory relationships, while suppressing irrelevant information. We will present additional findings at the meeting.

2753T

Functional analysis of *β-globin* locus control region hypersensitive site 2-associated proteins and noncoding RNA. *A. Gurumurthy, P. Wassel, S. Thakur, M. Hossain, J. Stees, K. Gavrilova, J. Bungert.* UNIVERSITY OF FLORIDA, GAINESVILLE, FL.

   The hemoglobin molecule is a tetramer composed of two α and two β-globin chains. The *β-globin* locus undergoes dynamic conformational changes in differentiating erythroid cells which are thought to be important for proper globin gene expression. The Locus Control Region (LCR) consists of several DNase I hypersensitive sites (HSs) that bind to erythroid and ubiquitously expressed transcription factors and function as enhancers. In addition, RNA Polymerase II (Pol II) synthesizes non coding, enhancer RNAs (eRNA) from the LCR HSs that may regulate β-globin gene expression. We aim to further understand the role of LCR HS2 and its associated eRNA in regulating *β-globin* expression. We created stable HS2 eRNA knockdown cell lines and observed reductions in *β-globin* RNA levels. The data suggest that LCR-associated eRNAs may participate in long-range regulation of *β-globin* gene expression. We also found that eRNAs remain associated with the *β-globin* gene locus. Previous studies have shown that the Integrator complex is involved in processing eRNAs. We hypothesize that termination of eRNA transcription releases Pol II which is then transferred to the globin gene promoters. To study this hypothesis, we created stable INT11 knockdown cell lines. INT11 is the endonucleolytic subunit of the Integrator complex. Furthermore, to identify proteins associated with HS2 in the context of intact cells, we sought to specifically purify LCR-HS2 associated chromatin using a two-step purification strategy. We designed two zinc finger DNA-binding domains, binding to the 5′ and 3′ flanking region of LCR HS2, respectively. The ZF-DBDs are associated with different tags for pull-down assays. By functionally assessing LCR HS2 associated proteins and non-coding RNA, we aim to gain better insight into the regulation of the *β-globin* locus which may lead to new strategies to treat hemoglobinopathies.

2754F

RNA-seq of human heart tissue identifies shared and divergent expression signatures of heart failure. *M.E. Sweet[1], A. Cocciolo[2], D. Slavov[2], S.L. Graw[2], K.L. Jones[3], T.B. Reece[4], A.V. Ambardekar[2], M.R. Bristow[2], L.M. Mestroni[1,2], M.R.G. Taylor[1,2].* 1) Human Medical Genetics and Genomics, University of Colorado, Aurora, CO; 2) Cardiovascular Institute and Adult Medical Genetics Program, University of Colorado, Aurora, CO; 3) Department of Pediatrics, Section of Hematology, Oncology, and Bone Marrow Transplant, University of Colorado, Aurora, CO; 4) Department of Cardiothoracic Surgery, University of Colorado Hospital, Aurora, CO.

Heart failure occurs when disease or injury weakens the heart muscle and its ability to pump blood effectively. Current treatment paradigms are based on symptom-severity classification, with little regard to etiology, reflecting a lack of knowledge of underlying heart failure mechanisms and highlighting a critical barrier to more effective treatments. We used RNA-seq and pathway analysis in 64 explanted left ventricle human heart tissues to identify transcriptome signatures common to heart failure as well as unique to specific heart failure subsets. We identified differentially expressed genes (DEGs) between dilated cardiomyopathy (DCM, n=37) and non-failing controls (NF, n=14) and between ischemic cardiomyopathy (ICM, n=13) and NF. Between DCM and NF, there were 3,649 DEGs, and between ICM and NF, there were 4,150 DEGs (FDR ≤ 5%). These comparisons share 2,691 DEGs, all with fold-changes in the same direction for both comparisons. These genes represent the expression signature common to heart failure with reduced ejection fraction and, using Ingenuity Pathway Analysis (IPA), are enriched for pathways suggesting dysregulation of energy metabolism (Mitochondrial Dysfunction, $p = 5e^{-16}$; Oxidative Phosphorylation, $p = 2e^{-14}$) and apoptosis (Protein Ubiquitination, $p = 1e^{-11}$). There are 688 DCM-specific genes and 1,189 ICM-specific genes. Using unsupervised hierarchical clustering, the expression of these 1,877 genes clusters the samples into their disease groups. Pathway analysis at p ≤ 0.003 demonstrates that the majority of pathways are disease-specific, with DCM and ICM sharing only two pathways. DCM-specific genes are enriched for pathways suggesting dysfunctional cell-cell adhesion (Germ Cell-Sertoli Cell Junction Signaling) and detoxification (NRF2-mediated Oxidative Stress Response). ICM-specific pathways are enriched for cytoskeletal (Actin Cytoskeleton Signaling, Integrin Signaling, Epithelial Adherens Junction Signaling, Rac Signaling, RhoA Signaling, Signaling by Rho Family GTPases) and immune system pathways (Antigen Presentation Pathway, CD40 Signaling, CD28 Signaling in T Helper Cells, JAK/Stat Signaling, IL-8 Signaling, and IL-6 Signaling), suggesting an immune response and fibrosis of damaged tissue. Our results demonstrate that there are both common heart failure pathways as well as disease-specific signatures in DCM and ICM that could potentially aid in development of future precision medicine treatments.

2755W

The communal relation of MTHFR, MTR, RFC gene polymorphisms and hyperhomocysteinemia as plausible risk of congenital septal defects. *S.B. Sunayana[1], K. Srujana[1], K. Nageswara Rao[2], G. Sandhya Devi[3], D. Sujatha[4], K. Rajender Rao[4], V. Sreedevi[1], M. Hema Prasad[1].* 1) Environmental Toxicology, Institute of Genetics, Hyderabad, India; 2) Care Hospital, Banjara Hills, Hyderabad, India; 3) Indo American Cancer Institute, Banjara Hills, Hyderabad; 4) National Institute of Nutrition, Hyderabad, India.

Introduction: Congenital septal defects (CSD) are the non-infectious and highly heterogeneous developmental anomaly in newborns with genetically complex traits, involving both the genetic and environmental factors. The present study is aimed to investigate the contribution of genetic determinants in severity of disease. Cytogenetic analysis using fluorescence in situ hybridization (FISH) and comet assay were performed to identify the microdeletions in chromosome 22 and to measure the oxidative DNA damage in the lymphocytes, respectively. Single nucleotide polymorphisms (SNPs) of Methylenetetrahydrofolate reductase (MTHFR) (677C>T), methionine synthase reductase (MTRR; 66 A>G) and reduced folate carrier (RFC) (80 A>G) were analyzed to identify the role of folic acid and homocysteine metabolism in CSD children. **Methods:** One hundred and sixty two children with non-syndromic CSD, were studied with age matched controls. Blood samples were collected from the Department of Pediatric Cardiology, Care Hospital, Hyderabad, with prior clearance from the Institutional Ethics Committee and written consent from the parents. The lymphocytes were isolated from fresh blood for comet assay as well as for the identification of 22q11.2 microdeletions, using specific FISH probes on G-Banded glass slides. Plasma and red blood cells (RBCs) were used for the estimation of homocysteine and folic acid. Genomic DNA was isolated for PCR and restriction fragment length polymorphism analysis of SNPs 677C> T, A66G and A80G genes. **Results:** The mean values of RBC folate are less in patients (225.5±146.5 ng/ml) compared to controls (368.5 ± 130.9 ng/ml). An increased comet tail length was observed patients (4.3 ± 1.5 μm) compared to the controls (2.74 ± 1.2 μm), demonstrating the severity of oxidative DNA damage. Microdeletions of 22q11.2 and significant DNA damage were observed in CSD patients. An increased frequency of MTHFR 677C> T, MTRR 66 A>G and RFC 80A> G gene polymorphisms were observed in CSD children when compared to controls. **Conclusions:** Increased homocysteine, low RBC folate levels and genetic polymorphism have shown increased oxidative DNA damage in children with congenital septal defects.

2756T

Fibulin-4a inhibits vascular and enhances cardiac cell fate by inhibiting transforming growth factor beta signaling. *Z. Urban, S.M. Khatri.* Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Fibulin-4 (FBLN4) mutations cause autosomal recessive cutis laxa type 1B, a disease characterized by arterial tortuosity, aneurysms, developmental emphysema, bone fragility and loose, inelastic skin. To study the developmental functions of FBLN4, we used zebrafish as a model. Zebrafish have two fibulin-4 genes, *fbln4a* and *fbln4b*. This study focused on *fbln4a*, which is more similar to human FBLN4 than *fbln4b*. Gene expression studies showed that *fbln4a* mRNA was expressed in the adaxial cells adjacent to the notochord during somite formation, and in the myosepta, head and the heart at later stages. Knockdown of *fbln4a* by an antisense morpholino oligonucleotide (MO) caused pooling of blood at the caudal vein plexus or around the heart, vascular hemorrhage in the head, reduced circulation and heart rate, bent notochord, rounded somites and reduced embryo length. All of these features were rescued by co-injection of fbln4a mRNA with MO. Early cardiac and vascular progenitor markers showed an expansion of the heart field and reduction of the vascular/blood fields in the anterior lateral plate mesoderm in knockdown embryos. These early developmental changes led to a randomization of heart laterality, and variably reduced or enlarged heart size. The lumens of both the segmental and axial blood vessels were irregular. Inhibition of transforming growth factor beta receptor 1 (Tgfbr1) kinase activity with a small molecule rescued all of the cardiovascular and connective tissue anomalies in knockdown embryos. By varying the period of treatment, we identified late gastrulation and early segmentation as the critical periods of rescue by Tgfbr1 inhibition. Simultaneous knockdown of transforming growth factor 2 (*tgfb2*) or 3 (*tgfb3*) with *fbln4a* also rescued the knockdown phenotype. We conclude that Fbln4a inhibits Tgfb2 and/or Tgfb3 signals emanating from the notochord to determine the size of cardiac and vascular progenitor pools during late gastrulation and early segmentation. These results identify a previously unrecognized function of fibulin-4 in addition to its well-known role in elastic fiber assembly.

2757F

Genetic causes of heterotaxy identified by whole exome sequencing. *A. Sridhar[1], J. Lajiness[1], C. Crawford[1], L. Elmore[1], M. Tariq[2], S. Ware[1]. 1) Indiana University School of Medicine, Indianapolis, IN; 2) University of Tabuk, Tabuk, Kingdom of Saudi Arabia.*

Heterotaxy (HTX) is a rare congenital anomaly syndrome accompanied by thoracic and/or abdominal situs abnormalities resulting from abnormal left-right patterning in the developing embryo. Although multiple inheritance patterns have been described in conditions that share features with HTX such as isolated congenital heart defects (CHD) and Primary Ciliary Dyskinesia (PCD), the genetic basis of HTX syndrome is rarely identified in a clinical setting. The purpose of this study is to define the frequency of pathogenic variants in genes known to affect laterality as well as identify novel candidate genes using whole exome sequencing in a large HTX cohort. We analyzed whole exomes of 285 subjects with HTX-spectrum disorders of which 37 (13%) had positive family history. In order to investigate potential pathogenic variants, we evaluated 172 genes compiled from various HTX, PCD and laterality defect clinical panels. Nearly 11% of the cohort had molecular findings that were likely or definitely pathogenic for PCD (5.6%), ciliopathies (3.2%, including multiple individuals with Bardet-Biedl syndrome), or other genetic syndromes (2.1%, including multiple participants with Oculo-Facio-Cardio-Dental syndrome). *DNAH5, DNAH11* and *GPR98* showed highest counts with more than 40 rare heterozygous variants. We also analyzed 5 trio samples utilizing an inheritance-pattern based analysis to identify novel candidate genes. We identified 20 genes that had recessive, compound heterozygous or de novo mutations. Of these genes, *SPIRE1* and *FMN2*, involved in the dorso-ventral axis formation, and *SYNE1* and *PKD1*, involved in ciliogenesis/cilia function, are interesting candidates for functional studies. This study highlights the utility of whole exome studies using inheritance-based models to identify novel genes and variants that begin to elucidate the mechanisms of a complex congenital disorder such as heterotaxy.

2758W

A mutation in the *LMOD1* actin-binding domain segregating with disease in a large British family with thoracic aortic aneurysms and dissections. *Y. Wan[1], M.A Simpson[2], J.A Aragon-Martin[1], D.P.S Osborn[3], E. Regalado[4], D.C Guo[4], C. Boileau[5,6,7,8], G. Jondeau[5,6,7,9], L. Benarroch[5], Y. Isekame[1], J. Bharj[3], J. Sneddon[10], E. Fisher[11], J. Dean[12], M.T. Tome Esteban[1,13], A. Saggar[14], D. Milewicz[4], M. Jahangiri[13,15], E.R Behr[1,13], A. Smith[16].* 1) Molecular and Clinical Sciences Research Institute, St George s University of London, London, London, United Kingdom; 2) Division of Genetics and Molecular Medicine, Kings College London, United Kingdom; 3) Cell Biology and Genetics Research Centre, St. George's, University of London, United Kingdom; 4) Department of Internal Medicine, University of Texas, USA; 5) Laboratory of Vascular Translational Science, INSERM, Paris, France; 6) Centre de référence pour le syndrome de Marfan et apparentés, APHP hopital Bichat, Paris, France; 7) Université Paris Diderot, Paris France; 8) Department de Genetique, APHP, Hopital Bichat Paris, France; 9) Service de Cardiologie, APHP, Hopital Bichat, Paris, France; 10) East Surrey Hospital, Redhill, United Kingdom; 11) The Park Surgery, Albion Way, Horsham, United Kingdom; 12) Clinical Genetics, NHS Grampian, Foresterhill, Aberdeen, United Kingdom; 13) Cardiology Academic Group, Cardiology Department, St George's Hospital, United Kingdom; 14) Clinical Genetics Unit, St George's, University of London, United Kingdom; 15) Department of Cardiothoracic Surgery, St. George's Healthcare NHS Trust, London, United Kingdom; 16) Academic Department of Vascular Surgery, Cardiovascular Division, BHF Centre of Research Excellence, Kings College London, United Kingdom.

We describe a mutation in *LMOD1,* which predisposes individuals to thoracic aortic aneurysms and dissections in a large multi-generation British family. Exome variant profiles for the proband and two distantly related affected relatives were generated and a rare protein-altering, heterozygous variant was identified, present in all the exome-sequenced affected individuals. The allele c.1784T>C, p.(V595A) in *LMOD1* is located in a known actin-binding WH2 domain and is carried by all living affected individuals in the family. *LMOD1* was further assessed in a consecutive series of 98 UK TAAD patients and one further mutation was found, yielding an incidence of ~2% in our study group. Assessment of *LMOD1* in international TAAD cohorts discovered nine other missense variants of which three were assessed as likely pathogenic. Validation of *LMOD1* was undertaken using a zebrafish animal model. Knockdown of both *lmod1a* and *lmod1b* paralogs using morpholino oligonucleotides showed a reproducible abnormal phenotype involving the aortic arches. Injection of the human *LMOD1* c.1784T>C, p.(V595A) mutation demonstrated a likely dominant negative effect and illustrated a loss of function cause. Mutations found in the WH2 actin-binding domain of *LMOD1* may delay actin polymerization and therefore compromise actin length, dynamics and interaction with myosin in the smooth muscle contraction pathway.

2759T

Homozygous *EEF1A2* mutation causes dilated cardiomyopathy, failure to thrive, global developmental delay, epilepsy and early death. *P.B. Agrawal[1], S. Cao[1], L.L. Smith[1], S.R. Padilla-Lopez[2], B.S. Guida[2], E. Blume[1], C.A. Brownstein[1], A.H. Beggs[1], M.C. Kruer[2].* 1) Boston Children's Hospital, Boston, MA.02115, USA; 2) Barrow Neurological Institute, University of Arizona College of Medicine, Phoenix, AZ 85013 USA.

Eukaryotic elongation factor 1A (EEF1A), the second most abundant protein in the cell, is encoded by two distinct isoforms, *EEF1A1* and *EEF1A2*; whereas *EEF1A1* is expressed almost ubiquitously, *EEF1A2* expression is limited such that it is only detectable in skeletal muscle, heart, brain and spinal cord. Currently, the role of EEF1A2 in normal cardiac development and function is unclear. There have been several reports linking *de novo EEF1A2* mutations to neurological issues in humans. We report a pair of siblings carrying a homozygous missense mutation p.P333L in *EEF1A2* who exhibited global developmental delay, failure to thrive, dilated cardiomyopathy and epilepsy, ultimately leading to death in early childhood. A third sibling also died of a similar presentation, but DNA was unavailable to confirm the mutation. Functional genomic analysis was performed in *S. cerevisiae* and zebrafish. In *S. cerevisiae*, there was no evidence for a dominant-negative effect. Previously identified putative *de novo* mutations failed to complement yeast strains lacking the *EEF1A* ortholog showing a major growth defect. In contrast, introduction of the mutation seen in our family led to a milder growth defect. To evaluate its function in zebrafish, we knocked down *eef1a2* expression using translation blocking and splice-site interfering morpholinos. EEF1A2-deficient zebrafish had skeletal muscle weakness, cardiac failure and small heads. Human *EEF1A2* wild-type mRNA successfully rescued the morphant phenotype, but mutant RNA did not. Overall, EEF1A2 appears to be critical for normal heart function in humans, and its deficiency results in clinical abnormalities in neurologic function as well as skeletal and cardiac muscle defects.

2760F

Isoforms and eQTLs of the myocardial infarction gene *PHACTR1*. *V. Codina-Fauteux[1,2], M. Beaudoin[1], G. Lettre[1].* 1) Montreal Heart Institute, Montreal, Canada; 2) Université de Montréal, Montreal, Canada.

Coronary artery disease (CAD), including myocardial infarction (MI), is the major cause of death and disability worldwide. Traditional risk factors include: sex, age, smoking, hypertension, diabetes, dyslipidemia, etc. Genetic variation also influences the risk to develop CAD/MI. Large GWAS have identified >60 SNPs associated with this disease, including an intronic SNP (rs9349379) in *PHACTR1* located on chromosome 6. We previously showed that this SNP is an expression quantitative trait locus (eQTL) for *PHACTR1* expression in human right coronary artery (hRCA) – a result since then replicated in the GTEx resource – and that it abrogates a MEF2 binding site (Beaudoin et *al*., ATVB, 2015). Here, we present our effort to define the different *PHACTR1* isoforms in various human cells and tissues. Specifically, we explored which isoforms are expressed in brain, heart, hRCA, and several other cells (HUVEC, HCAEC, HAEC, teloHAEC, HCASMC, THP-1) using a combination of RT-PCR, 5' and 3' Rapid Amplification of cDNA Ends (RACE), as well as long molecules sequencing (PacBio). Integration of these results led to the identification of four isoforms expressed from the same open reading frame and that share the same stop codon. We note, however, that these isoforms all contain multiple possible alternative initiation codons. In the brain, we found one major "long" isoform that encodes a protein of 580 amino acids. We also identified two ubiquitously expressed "intermediate" isoforms (488/557 amino acids), differentiated by alternative splicing of a 207-bp exon. Finally, we found in THP-1, heart, hRCA and brain a "short" isoforms (144 amino acids) that starts at exon 14 of the long isoform, a result consistent with a recent report (Reschen et *al.*, Atherosclerosis, 2016). We then tested if genotypes at rs9349379 are associated with the expression of the two intermediate and short *PHACTR1* isoforms in 36 hRCA (the long isoform is not expressed in this tissue). Genotypes were associated with the intermediate + alternative exon isoform (r^2=0.164, P=0.014) but not with the other two transcripts (intermediate – alternative exon (r^2=0.064, P=0.135); short isoform (r^2=0.003, P=0.733)). This result suggests that *PHACTR1*-rs9349379 is unlikely to mediate its effect on CAD/MI risk by regulating the expression of the immune-specific short isoform in hRCA.

2761W

Integration of GWAS and local genetic effects on gene expression (eQTL/ASE) highlights genes with kidney function and disease. *C. Qiu[1], S. Huang[1], J. Park[1], Y. Ko[1], Y. Park[2], M. Palmer[3], C. Boustany[4], S. Pullen[4], C. Brown[2], K. Susztak[1].* 1) Department of Medicine, Renal Electrolyte and Hypertension Division, University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 3) Pathology and Laboratory Medicine at the Hospital of the University of Pennsylvania, Philadelphia, PA; 4) Boehringer Ingelheim Pharmaceuticals, Inc. Ridgefield, CT 06877, USA.

    Chronic kidney disease (CKD) is a common complex disease affecting 10% people of the population. Genome-wide association studies (GWAS) have identified more than 100 genomic regions that are significantly and reproducibly associated with kidney function. Identifying genetic variations associated with quantifiable cellular phenotypes such as gene expression levels (eQTLs) is a powerful approach to identify functional variants. We have earlier generated eQTL data for 96 whole kidney samples and integrated these results with GWAS signals and identified 7 potential target genes. Here we examined whether we can improve upon our GWAS target gene identification by generating kidney compartment specific eQTL and allele-specific expression (ASE) datasets. Human kidney tissue samples and associated clinical data were collected from unaffected portion of nephrectomies and manually microdissected into key kidney compartments, glomerulus and tubule fractions. Whole transcriptome sequencing was performed separately and genotyping was done using Affymetrix arrays. Histological analysis and in silico cell type composition analysis was performed to assure consistency of cell type composition of the analyzed samples. Finally 121 of samples that were of European descent were used for the analysis. We identified cis-eQTLs and associated ASE regions variation in kidney compartments and compared our results with publicly available 44 human tissue data from Genotype Tissue Expression consortium to test kidney-specificity. Then we integrated CKD-GWAS and eQTL findings by multiple genetic colocalization methods, including Coloc and RTC. We identify 4081 eGenes in the tubule compartment and 4913 in glomeruli, the list of eGenes showed 91% overlap. The eSNPs were enriched on kidney specific regulatory regions such as promoters and enhancers. By overlapping kidney eGenes with the GWAS catalogue, we found statistically significant enrichment for blood metabolite and kidney specific traits. Using meta-analysis method, we have identified 507 kidney tissue specific eGenes, which were enriched for kidney function and disease traits. By integrating kidney compartment specific eQTL and ASE and kidney disease GWAS variants, we identified 17 likely GWAS target genes. Integration of GWAS data with compartment-specific eQTLs and ASE analysis can improve upon identification for potential causal genes for kidney function and disease when compared to whole tissue samples.

2762T

Contribution of AMD risk variants to the genetic architecture of choroidal thickness in the Amish. *N. Restrepo[1], Y. Song[1], R. Laux[1], L. Adams[2], D. Fuzzell[1], L. Caywood[2], V. Horst[3], T. MacKay[3], D. Dana[3], M. Nittala[4], S.V. Sadda[4,5], W. Scott[2], . Stambolian[3], M.A. Pericak-Vance[2], J. Haines[1].* 1) Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Departments of Ophthalmology and Genetics, University of Pennsylvania, Philadelphia, PA; 4) Doheny Image Reading Center, Doheny Eye Institute, Los Angeles, CA; 5) Department of Ophthalmology, University of California –Los Angeles, CA.

    Purpose: To understand the role that genetics plays in risk and progression of AMD. Parsing AMD into quantitative endophenotypes, each with a distinct genetic basis, will improve prediction of risk and increase our understanding of the genetic architecture of AMD. Here we assess the contribution that known AMD risk SNPs contribute to the genetics of choroidal thickness (CT) in an Amish cohort. **Methods:** We sampled 579 related individuals from Amish families with early/intermediate AMD cases in Pennsylvania, Ohio, and Indiana. Individuals underwent a health history and ophthalmologic exam including color fundus photography and SD-OCT scans. Individuals were genotyped for 52 known AMD risk-associated SNPs. We performed a genetic association study of individuals from AMD pedigrees ascertained because at least one individual was diagnosed with AMD. CT was treated as both a continuous and a dichotomous trait (thin vs. normal). Tests were performed with R-Package "GWAF" assuming an additive genetic model and a Wald chi-square test for right, left, and mean eye measures. Continuous CT was analyzed under a Generalized Linear Mixed Effects model (GLMM) and dichotomous CT under Generalized Estimating Equations (GEE) model. **Results:** Of the SNPs tested, two were consistently identified across statistical algorithms and across models. Rs1142 on chr. 7 is located in the 3' UTR of the *SRSF protein kinase 2* (*SRPK2*) gene. When considering CT as a dichotomous trait, rs1142 is statistically significant ($p < 0.001$) across GLMM and GEE regardless of AMD status. Modeling CT as a continuous trait, rs1142 is significant ($p = 0.005$) only after excluding AMD cases from the pedigrees. Rs2230199 is a missense variant in *complement component 3* (*C3*) and is significantly associated with CT as a dichotomous trait under the GEE model and with nominal associations in GLMM ($p < 0.05$). **Conclusions:** The association of a known AMD-risk variant in *C3* with CT suggests that the CT endophenotype shares pathophysiology with AMD. Additionally, variants in C3 (i.e., rs2241394) are associated with risk of polypoidal choroidal vasculopathy, which shares characteristics of wet AMD, in the Japanese. The.

2763F

Type 1 diabetes progression is correlated with changes in the co-expression relationships of immune response genes. *I. Braenne, R. Chen, A.W. Manichaikul, S. Onengut-Gumuscu, S.S. Rich, W. Chen, C.R. Farber, TEDDY Study Group.* Center of Public Health Genomics, University of Virginia, Charlottesville, VA.

Type 1 diabetes (T1D) is an autoimmune disease caused by destruction of beta-cells in the pancreas. The etiology of T1D is unknown, but likely due to effects of multiple genetic and environmental factors. The Environmental Determinants of Diabetes in the Young (TEDDY) study aims to identify factors influencing T1D through the long-term monitoring of children at high genetic risk of T1D. Here, we applied Weighted Gene Co-expression Network Analysis (WGCNA) to whole blood gene expression profiles taken every 3-6 months from TEDDY participants with the goal of identifying network alterations influencing T1D progression. We generated a co-expression network using gene expression profiles collected approximately six months before T1D diagnosis (N=60). From this network, we identified two co-expression modules whose behavior was highly associated with T1D progression. The first module contained 1437 genes and its eigengene was positively correlated with age at T1D diagnosis (slower progression) ($r=0.50$, $p=6 \times 10^{-5}$). This module was enriched for "immune response" genes ($p=5 \times 10^{-21}$) and the pathway "Reactome neutrophil degranulation" ($p=2.15 \times 10^{-38}$). The second significant module contained 36 genes and its eigengene was negatively correlated with age at T1D diagnosis (faster progression) ($r=-0.54$, $p=1 \times 10^{-5}$). This module was enriched for the KEGG pathway "B-cell signaling" ($p=3 \times 10^{-6}$). Over time, we observed significant dynamic changes in the strength of gene-gene relationships (connectivity). In particular, we observed consistent increases in connectivity at time points preceding T1D onset. These network changes were not due to differences in individual gene expression levels or apparent blood cell composition. The hub genes exhibiting the largest changes in connectivity as a function of time were *CXCR5*, *MMP25* and *IFIT3*. *CXCR5* belongs to the chemokine receptor family and is a putative biomarker for T1D. *MMP25* is a matrix metalloproteinase known to play a role in immune system function. *IFIT3* is related to the interferon gamma signaling pathway, a pathway suggested to play a role in T1D progression. The results of this study suggest that changes in gene connectivity play a role in the etiology of T1D. Further dissection of network changes may identify specific factors and mechanisms influencing T1D progression.

2764W

Understanding progression and subtypes of prediabetes with metabolomics and genomic profiling in Starr County Mexican Americans. *G. Jun[1], C. Evans[2], G. Bell[3], C. Burant[2], C. Hanis[1].* 1) Human Genetics Center, UTHealth School of Public Health, Houston, TX; 2) University of Michigan; 3) University of Chicago.

Biological and metabolic changes occur long before the diagnosis of type 2 diabetes (T2D). Understanding the metabolic landscape in the early stages is important to understand progression to and subtypes of diabetes. Currently there are three different criteria defining prediabetes: fasting glucose, 2-hr glucose, and HbA1c. All three predict diabetes but they differ in underlying biological and metabolic mechanisms. In Starr County, only 12.3% of those with prediabetes meet all three criteria. It has been long known that branched-chain amino acids play an important role in insulin regulation and glucose metabolism and recent studies show that metabolomics profiling provide important information about diabetes and its progression. Metabolomic and lipidomic profiles were assayed on 155 samples from 127 unrelated females as a pilot analysis at the University of Michigan. 28 individuals were profiled at two time points 10 years apart with selection based on BMI change profiles. Results show distinctive amino acid signatures in normal and prediabetic individuals. Moreover, we observed differences in amino acid levels between subtypes of prediabetes. Average arginine levels of individuals with prediabetes by impaired fasting glucose were much higher than normal, but arginine levels in prediabetes by impaired glucose tolerance (2-hour glucose) was noticeably lower than normal. Next we tested for quantitative associations of metabolites with glucose phenotypes. The top metabolites associated with fasting glucose and HbA1c were from amino acids including lysine, tryptophan, arginine, and isoleucine, while 2-hr glucose associated signals were mainly from fatty acids including palmitate, palmitoleic acid and oleic acid. We performed genetic association analyses of these metabolites and found genome-wide significant signals for kynurenine (rs2973808) and elaidic acid (chr11:61602453:D). We also observed additional signals for unknown metabolites that are highly correlated with prediabetes-associated metabolites. We also performed multi-dimensional discriminant analyses to take full advantage of the untargeted metabolomics data. K-nearest neighbor imputation was done on 3,560 metabolites with ≤30% missing rates and 78 metabolites were selected by Elastic Net with multi-response Gaussian model. Discriminant analyses on selected metabolites show clear separation between subtypes of prediabetes, suggesting that metabolomic profiles help identifying subtypes of prediabetes.

2765T

Preliminary evidence suggests that a 6.7 kb deletion polymorphism in *LILRA3* is associated with Type 1 Diabetes. *C. Maroteau[1], M.K Siddiqui[1], P. Appelby[2], C.N.A Palmer[1].* 1) Division of Molecular and Clinical Medecine, University of Dundee, Dundee, ANGUS, United Kingdom; 2) Health Informatics Centre (HIC) / FARR Scotland, Ninewells Hospital and Medical School, Dundee, ANGUS, United Kingdom.

The 6.7 kb deletion polymorphism in *LILRA3* is well characterized in the East Asian population and shows evidence of natural selection. The deletion has known associations with autoimmune traits such as multiple sclerosis, Sjorgen's syndrome, Systemic Lupus Erythematosus. We examine exome sequence data from 1,373 individuals to determine the frequency of the deletion, and to identify variants strongly associated with the deletion in a European population. Finally, we perform a PheWAS on a sub-population for whom detailed clinical data was available. We find the deletion occurs in 63% of our population, as opposed to Northeast Asian populations where the frequency is variable, but often above 75%. Previous studies in Northeast Asian populations have also found variants in neighbouring *LILRB2* to be in total linkage with the deletion, however we observe that the R^2 is 0.24. Further examination of variants associated with the deletion showed a SNP in *HLA-DRB5* (rs41559420) was strongly associated. This locus has known associations with the development of Type 1 diabetes. A preliminary PheWAS showed a strong association between the deletion and insulin use (n = 212), but no other diabetic medication, suggesting a type 1 diabetic phenotype. We present evidence that suggests *LILRA3* might play a role in the etiology of Type 1 diabetes. The relationship between *LILRA3*, *HLA-DRB5* and Type1 diabetes is worthy of further exploration. Furthermore, disease profiles of autoimmune traits juxtaposed with the *LILRA3* deletion frequency should be examined across populations that show diverse frequency.

2766F

Smoking-by-genotype interaction in type 2 diabetes. *P. Wu[1], D.V. Rybin[1], J.G. Wilson[2], M. Feitosa[3], N. Franceschini[4], J. Dupuis[1,5], J.B. Meigs[6,7,8], J.L. Vassy[9,10,11], C.T. Liu[1] on behalf of the CARe Consortium and the CHARGe Gene-Lifestyle Interactions Consortium.* 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 2) Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS; 3) Division of Statistical Genomics, Washington University in St. Louis, St. Louis, MO; 4) Department of Epidemiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA; 6) Department of Medicine, Harvard Medical School, Boston, MA; 7) Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA; 8) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA; 9) Division of General Medicine and Primary Care, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 10) VA Boston Healthcare System, Boston, Massachusetts, USA; 11) Harvard Medical School, Boston, Massachusetts, USA.

Background: Smoking is an important behavioral risk factor for type 2 diabetes (T2D), but not all smokers develop T2D. It is unknown whether genotypes explain some of this variability. We performed a genome-wide gene-by-environment interaction study (GEWIS) to identify SNPs that exhibit interaction with baseline smoking status (ever vs. never) on incident T2D. **Method:** We performed ethnicity-stratified analyses using Cox proportional hazards model to evaluate whether SNP-by-smoke interactions additionally contributing to the incident T2D. We implemented several strategies and determined that there was an association if a SNP was significant in one of 3 scenarios: (1) SNP-by-smoking interaction, (2) joint test of interaction and main effect, excluding SNPs with significant main effects (3) SNP effect in only one smoking status-stratified analysis. In the discovery stage, we conduct cohort-specific association analysis and performed inverse variance weighted fixed-effect meta-analysis to analyze 5 cohorts, including 17,823 European American and 4,820 African American, from Candidate-gene Association Resource (CARe) Consortium and evaluated the association for the SNPs from the gene centric IBC 50k. We then followed up 701candidate SNPs (p-value < 10-3 in one of three scenarios) from the discovery stage in 14 cohorts from the CHARGE Gene-Lifestyle Interactions Consortium ($N_{European}$=35831, $N_{African}$=2690). We selected 2x10-6 as the chip-wide significance threshold for the combined analysis. **Results:** We identified and replicated four loci significantly associated with incident T2D. Two loci, *RTN4* (EA) and *FBN1* (AA), were identified in joint analysis. Four loci reached chip-wide significance threshold in smoke-stratified analysis: *RTN4* (EA) in never smokers, *TCF7L2* (EA), *FBN1* (AA) and *CUBN* (AA) in ever smokers. *TCF7L2* has been associated with T2D and fasting glucose. Genetic variation in the cubilin (*CUBN*) has been associated with albuminuria, chronic kidney disease and end-stage renal disease in blacks. Prior study also showed that hyperinsulinaemia/hyperglycaemia-induced the changes of mRNA expression of Fbn1 in rat. And *FBN1* has been found to be responsible for cardiovascular disease. **Conclusion**: We have identified loci that could have been detected only in a stratified analysis. Our findings provide evidence for genetic interactions of smoking exposure on the risk of on genes that may explain some of the T2D genetic risk heterogeneity.

2767W

Genome-by-environment interactions have a major impact on obesity. *C. Amador[1], C. Xia[1], R. Nagy[1], A. Campbell[2,3], D. Porteous[2,3], C. Hayward[1], P.K. Joshi[4], J.F. Wilson[1,4], P. Navarro[1], C. Haley[1,5].* 1) MRC Human Genetics Unit, Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 2) Centre for Genomic and Experimental Medicine, Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 3) Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetic and Experimental Medicine, University of Edinburgh, Edinburgh, United Kingdom; 4) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 5) Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom.

Obesity has a large genetic component: estimates from twin studies for the heritability of BMI are in the range of 48-60%. Large-scale genome-wide association studies have now identified many genetic variants associated with BMI, but almost all the detected individual variants have a very small effect on BMI. Non-genetic variation is attributed to environmental causes, with factors such as lack of exercise and a high-calorie diets often blamed for the obesity epidemic. Several studies have shown that specific genetic variants have an effect on BMI that depends on interactions with such lifestyle variables. However, the proportion of the trait variation that is explained by interactions with individual genetic variants is even smaller than the direct effects of the variants themselves, making them very difficult to detect. Here we take a new approach and study the overall contribution of interactions between the genome and environment, by estimating the total variance that genome-by-environment interactions (GxE) explain. To do this, we constructed separate similarity matrices for individuals based on their genomes and their lifestyles, and we combined those two matrices to create a joint similarity matrix reflecting the genome-by-environment interactions. This novel method measures the effect of sharing both genes and lifestyles, over and above their separate effects. Using 15,000 genotyped subjects from the Generation Scotland cohort, we estimated that GxE explains up to 15% of the observed variance for obesity related traits. This GxE variance is in addition to the previously detected variance explained by genetics and common environment. We replicated some of the results in UK Biobank (with 150,000 genotyped individuals). GxE is a potential source of phantom heritability measured in twin studies and our work suggests the effect is substantial and explains a large part of the long standing missing heritability problem. Even more importantly, better analysis of the interactions between genes and environment improves our understanding of trait architecture and will lead to enhanced methods of disease prediction that will facilitate more effective interventions to tackle obesity.

2768T

Pleiotropic associations of adiposity-related genetic risk scores. *Z. Fairhurst-Hunter[1,3], C.M. Lindgren[2], M. McCarthy[1], Z. Cheng[3], R. Walters[3], M. Holmes[3], China Kadoorie Biobank Collective and UK Biobank.* 1) Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom; 2) The Big Data institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford; 3) Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford.

Aim: To assess, in 2 different ancestries, how genetic risk scores (GRSs) for adiposity traits associate with various measures of adiposity and body shape **Methods:** Externally weighted GRSs were constructed for the adiposity traits: body mass index (BMI); waist and hip circumferences (WC, HC); waist-hip ratio (WHR); and BMI-adjusted WC, HC and WHR (WCadjBMI, HCadjBMI, WHRadjBMI), using published SNPs associated with the respective adiposity trait at $P<5\times10^{-8}$. Residuals for each adiposity trait, derived from adjustment for: age, sex, and population stratification, were tested for association with each GRS, in 32,211 individuals from China Kadoorie Biobank(CKB) and 118,727 individuals from UK Biobank(UKB).   **Results:** All scores were significantly associated with their corresponding adiposity trait, but these associations were consistently stronger in Europeans (EUR) than in East Asians (EAS). In EUR, a unit increase in each GRS was associated with an increase in the corresponding trait of 0.81 SDs or greater. By contrast, in EAS the only GRS to attain an effect size greater than 0.8 with the corresponding trait was the GRS for BMI (β =0.84).   Despite limited overlap between the GRSs in terms of loci (average overlap 16%), each GRS was associated with multiple other adiposity traits. In the extreme the GRSs for HC, WHRadjBMI and WCadjBMI were significantly associated ($P<5.\times10^{-5}$) with every adiposity trait tested.   Furthermore, GRSs often had effects on non-corresponding traits similar to or greater than the effect on their corresponding trait. For example, a unit increase in the GRS for WC was associated with an increase in WC of β=0.73 in CKB and β=0.84 in UKB, but had slightly stronger effects on BMI of β=0.76 in CKB and β=0.89 in UKB.   **Summary:** GRSs for individual adiposity traits showed pleiotropic effects across multiple traits. This may hamper interpretation of studies using such scores to assess the relationship of different adiposity traits with disease risk. In order to assign risk to a certain adiposity distribution, the pleiotropic effects of the scores should be addressed. Additionally, reduced performance of many of the GRSs in EAS suggests differences in genetic architecture for body shape between EUR and EAS, highlighting the importance of performing GWAS in diverse populations.

2769F

Investigation of the association between ITLN1 gene A326T polymorphism and in subjects with type 2 diabetes mellitus and obese: In the TARF study. *F. Geyik[1], A. Onat[2], G. Can[3], N. Coban[1], E. Komurcu-Bayrak[1], N. Erginel-Unaltuna[1].* 1) Department of Genetics, Aziz Sancar Institute of Experimental Medicine, Istanbul University, Istanbul,Turkey; 2) Department of Cardiology, Cerrahpasa Medical Faculty, Istanbul University, Istanbul, Turkey; 3) Department of Public Health, Cerrahpasa Medical Faculty, Istanbul University, Istanbul, Turkey.

   Background: Omentin is a novel adipokine expressed predominantly in visceral adipose tissue. Plasma omentin levels are negatively correlated with obesity, insulin resistance, and type 2 diabetes (T2DM). The aim of the study is to evaluate the frequency of ITLN-1 (intelectin 1) gene A326T polymorphism and assessment of its relation with serum omentin level, obesity, T2DM and other biochemical parameters in a large cross-sectional population-based random sample of Turkish Adults (TARF Study). **Methods:** Genotyping was performed using LightSNiP assay (TIB Molbiol) in 2291 adult individuals (mean age 51.9 ±11.7; 49.1% men). Analyses were performed in men and women separately. Stratified analysis according to T2DM and obesity, was performed. **Results:** The frequency of ITLN1 rare allele (T) was 27.9%, and the genotype distribution was in Hardy-Weinberg equilibrium. In both genders, A326T polymorphism T allele carriers were not at increased risk for obesity and T2DM. When the study population was stratified according to T2DM, homozygotes for T allele were strongly associated with elevated fasting triglyceride (*p=0.025*) and C-reactive protein levels *(p=0.029)* in diabetic men. Non-diabetic men with homozygote T allele had elevated serum omentin levels (*p=0.033*). ITLN1 A326T polymorphism was related to trigliceride (p=*0.022*) and lipoprotein(a) (*p=0.014*) levels in obese women. **Conclusion:** The ITLN1 A326T polymorphism is associated with serum omentin levels in a sex- and diabetes-specific manner. Presence of obesity in women and diabetes in men seems to be required for A326T polymorphism to induce markedly elevated triglycerides.

2770W

Fine-mapping and characterization of GWAS loci harboring extensive allelic heterogeneity. *C. Spracklen[1], A.U. Jackson[2], A. Iyengar[1], S. Vadlamudi[1], H.M. Stringham[2], Y. Wu[1], M.E. Cannon[1], M. Civelek[3], K. Currin[1], C. Fuchsberger[4], A.E. Locke[5], R. Welch[2], P.S. Chines[6], N. Narisu[6], A.J. Lusis[7], J.K. Kuusisto[8], F.S. Collins[6], M. Boehnke[2], M. Laakso[8], K.L. Mohlke[1].* 1) Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI; 3) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 4) Center for Biomedicine, European Academy of Bozen/Bolzano (EURAC), University of Lubeck, Bolzano, Italy; 5) McDonnell Genome Institute, Washington University School of Medicine, Saint Louis, MO; 6) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 7) Department of Medicine, University of California, Los Angeles, Los Angeles, CA; 8) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.

   Interpreting GWAS loci can be challenging if multiple functional variants exist. Using adiponectin levels, a complex trait, we sought to fine-map loci containing multiple signals. In a GWAS of 19M variants in 9,262 nondiabetic individuals from the Metabolic Syndrome in Men (METSIM) study, we identified 3 significant loci, *CDH13, ADIPOQ,* and *IRS1*. At each locus, we performed stepwise conditional analyses to identify additional signals. Distinct signals were defined as $P_{cond}<1\times10^{-5}$ within 1 Mb of the index variant with pairwise linkage disequilibrium (LD) between variants $r^2>.01$; signals were defined as independent if $r^2\leq.01$. We annotated signals using eQTLs in adipose tissue from 770 METSIM subjects, ATAC-seq data from adipose tissue of 3 subjects and SGBS preadipocyte cells, and Epigenomics Roadmap adipose nuclei data. Two adiponectin-associated loci (*CDH13, ADIPOQ*) contained 2 and 8 distinct signals, respectively, 2 and 4 of which are independent. Joint conditional analyses were performed with GCTA for comparison, identifying 2 (*CDH13*) and 6 (*ADIPOQ*) of the signals from the stepwise approach. At *CDH13*, the 1st previously-identified signal contained the lead adipose eQTL variant for *CDH13*, as well as a proposed functional variant and a methylation QTL. Within the new 2nd *CDH13* signal, we examined a candidate regulatory element containing rs4782722 and rs12444113 and observed >1.6 fold allelic differences (P=0.001) in transcriptional reporter activity. At *ADIPOQ*, each of the 8 distinguishable signals consisted of 1-15 variants ($r^2>0.8$), including rs62625753, a coding variant (G90S; $P_{init}=3\times10^{-3}$, $P_{cond}=6\times10^{-4}$) predicted to be deleterious (SIFT) and probably damaging (PolyPhen). The lead adipose eQTL for *ADIPOQ* was the 8th ranked signal (rs35469083; $P_{init}=1\times10^{-44}$, $P_{cond}=2\times10^{-3}$). The trait-raising alleles of signals 1 & 2 and of signals 1, 4 & 8 share haplotypes with frequency >0.37 (D'≥0.94; $r^2=0.05-0.33$), which may explain the strength of signal 1. All 8 signals contain ≥1 variant in a putative enhancer. Additional independent eQTL signals (FDR<1%) were identified for *CDH13* and *ADIPOQ* that were not coincident with the other GWAS signals. Accounting for multiple signals resulted in a 1.6-fold increase in variance explained over the lead signals alone (5.9 vs 9.4%). Taken together, fine-mapping, annotation, and experimental validation identified multiple distinguishable association signals, including potentially novel coding and functional variants.

2771T

Gender-, genotype- and ethnic-specific effects of sugar-sweetened beverages on serum uric acid concentrations. *X. Zhang, B. Mass, R. Hou, V.S. Voruganti.* Nutrition, University of North Carolina at Chapel Hill, Nutrition Research Institute, Chapel Hill, NC.

Increased consumption of sugar-sweetened beverages (SSB) has been associated with elevated serum uric acid (SUA) concentration, which in turn increases the risk for gout, kidney and cardiovascular disease. However, it is not clear whether an individual's response in SUA to SSB is dependent on one's genotype and ethnicity. The aim of this study was to investigate the acute effects of SSB on SUA in adults, aged 30-50yrs, of Caucasian (C), African-American (A) or Hispanic (H) descent and determine whether the SUA response is dependent on their gender, genotype and ethnicity. After an overnight fast, 61 subjects (C=37, A=20, H=4) consumed SSBs containing 60 grams of fructose and 15 grams of glucose. Blood samples were collected at 0-, 30-, 60-, 120- and 180-minute post-baseline for the analysis of SUA. Six single nucleotide polymorphisms (SNPs) from uric acid transporters, rs2231142 of *ABCG2*, rs1183201 of *SLC17A1* and rs16890979, rs737267, rs6449213 and rs3775948 of *SLC2A9*, were genotyped. Baseline SUA levels were significantly higher in men (6.2 (0.3) mg/dL) than in women (5.2 (0.1) mg/dL, p &lt 0.01). Following SSB consumption, SUA increased 0.65 mg/dL within 30 minutes and then gradually returned to the baseline concentration by 180 minutes. The changes in SUA throughout five time points were significantly different between men and women (p &lt 0.009). Genotype-specific analysis showed that rs3775948 was significantly associated with changes in SUA at all time points between GG and GC (p = 0.03 (0min); p = 0.03 (30min); p = 0.01(60, 120 and 180min) and between GC and CC (p = 0.04 (60min), p = 0.02 (120min and 180min) genotypes in women but not in men. Ethnic-specific analysis showed that baseline SUA was similar in all ethnicities [(C-5.6 (0.21 mg/dL); A-5.4 (0.4) mg/dL; H-5.2 (0.24) mg/dL, p = 0.31]. However, significant differences were found between C and A post SSB consumption [(0-30min, p &lt 0.05) (30-60min, p &lt 0.05), (60-120 min, p &lt 0.05), (120-180min, p &lt 0.05)]. The present pilot data demonstrate that SUA response to SSB depends on individual genders, genotypes and ethnicity and they need to be confirmed in larger population studies. Overall, our work underscores the importance of gender, genotype and ethnic-specific responses to dietary challenges in intervention studies.

2772F

Long-term response to oral eliglustat in treatment-naïve adults with Gaucher disease type 1: Final efficacy and safety results from a phase 2 clinical trial after 8 years of treatment. *H. Lau[1], E. Lukina[2], N. Watman[3], M. Dragosky[4], E. Avila Arreguin[5], H. Rosenbaum[6], Y. Wu[7], S.J.M. Gaemers[7], M.J. Peterschmitt[7].* 1) New York University School of Medicine, New York, NY; 2) National Research Center for Hematology, Moscow, Russia; 3) Hospital Ramos Mejia, Buenos Aires, Argentina; 4) IMAI-Research, Buenos Aires, Argentina; 5) Instituto Mexicano del Seguro Social Hospital de Especialidades, Col. La Raza, Mexico; 6) Rambam Medical Center, Haifa, Israel; 7) Sanofi Genzyme, Cambridge, MA.

Background: We report final long-term efficacy and safety results of a Phase 2 trial (NCT00358150, Sanofi Genzyme) in previously untreated adult patients with Gaucher disease type 1 (GD1) after 8 years of treatment with eliglustat, an oral substrate reduction therapy. **Methods:** Adult GD1 patients who had splenomegaly with thrombocytopenia and/or anemia received 50 or 100 mg eliglustat tartrate (equivalent to 42 or 84 mg eliglustat) twice daily, dosed by plasma trough levels. Efficacy outcomes included changes in hemoglobin, platelets, spleen and liver volumes, disease-related biomarker levels, skeletal manifestations, and achievement of therapeutic goals for anemia, thrombocytopenia, splenomegaly, and hepatomegaly (Pastores. *Semin Hematol.* 2004; Lukina. *Blood.* 2010). **Results:** Of 26 enrolled patients, 19 completed the trial and 7 withdrew: 2 due to asymptomatic nonsustained ventricular tachycardia detected during 36-hour ECG safety monitoring on Day 1 (plasma drug levels were undetectable); 1 after 1 year due to progression of a bone lesion retrospectively identified at baseline; 1 administrative withdrawal after 2 years; and 3 due to pregnancy. After 8 years of eliglustat, mean±SD hemoglobin and platelet count increased by 2.1±1.7 g/dL (from 11.3±1.6 to 13.4±1.3 g/dL) and 110% (from 67.5±21.1 to 130.7±59.8 x10^9/L), respectively. Mean spleen and liver volumes decreased by 68% (from 17.3±10.4 to 5.1±3.5 multiples of normal [MN]) and 31% (from 1.6±0.5 to 1.1±0.3 MN), respectively. All patients met ≥3 of 4 long-term therapeutic goals (spleen, 100% of patients; liver, 100%; hemoglobin, 93%; platelets, 53%) by 7–8 years. Median chitotriosidase levels decreased by 84%, CCL-18 by 82%, and glucosylsphingosine by 88%; plasma GL-1 normalized. Total mean lumbar spine bone mineral density increased by 0.12 g/cm^2; mean Z-score increased by 0.88 (from -1.27±1.02 to -0.39±1.13) and mean T-score by 0.95 (from -1.64±1.07 to -0.69±1.31). Eliglustat was well-tolerated. All quality of life measures (SF-36, fatigue severity score, Gaucher disease severity score) showed improvement over time. Most adverse events in this long-term trial were mild or moderate in severity (98%, 342/348) and considered unrelated (94%, 328/348) to treatment. No new safety concerns emerged. **Conclusion:** Clinically meaningful improvements in hematologic, visceral, biomarker, and bone parameters continued or were maintained in previously untreated GD1 patients treated with eliglustat for 8 years.

2773W

Long-term results of ENGAGE: A phase 3, randomized, double-blind, placebo-controlled, multi-center study investigating the efficacy and safety of eliglustat in adults with Gaucher disease type 1. *P. Mistry[1], E. Lukina[2], H. Ben Turkia[3], S. Shankar[4], H. Baris Feldman[5], M. Ghosn[6], A. Mehta[7], S. Packman[8], H. Lau[9], M. Petakov[10], S. Assouline[11], M. Balwani[12], S. Danda[13], E. Hadjiev[14], A. Ortega[15], Y. Wu[16], S.J.M. Gaemers[16], M.J. Peterschmitt[16].* 1) Yale University School of Medicine, New Haven, CT; 2) National Research Center for Hematology, Moscow, Russia; 3) Hôpital La Rabta, Tunis, Tunisia; 4) Emory University, Atlanta, GA; 5) Rabin Medical Center, Petach Tikvah, and Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 6) Hôtel-Dieu de France University Hospital, Beirut, Lebanon; 7) The Royal Free Hospital, London, England; 8) University of California, San Francisco, School of Medicine, San Francisco, CA; 9) New York University School of Medicine, New York, NY; 10) Clinical Center of Serbia, University of Belgrade School of Medicine, Belgrade, Serbia; 11) Jewish General Hospital, Montreal, Quebec, Canada; 12) Icahn School of Medicine at Mount Sinai, New York, NY; 13) Christian Medical College, Vellore, India; 14) University Hospital Alexandrovska, Sofia, Bulgaria; 15) OCA Hospital, Monterrey, Mexico; 16) Sanofi Genzyme, Cambridge, MA.

Background: Eliglustat, an oral substrate reduction therapy, is a first-line treatment for adults with Gaucher disease type 1 (GD1) who have poor, intermediate, or extensive CYP2D6-metabolizer phenotypes (>90% of patients). ENGAGE (NCT00891202, Sanofi Genzyme), a randomized, double-blind, placebo-controlled, Phase 3 trial, investigated efficacy and safety of eliglustat in untreated adults with GD1. In the 9-month primary analysis period (PAP), all visceral and hematologic efficacy parameters improved significantly in eliglustat-treated but not placebo-treated patients (Mistry, *JAMA*. 2015). **Methods:** Of 40 patients in the PAP, 39 entered the open-label extension phase, during which all received eliglustat. Long-term efficacy and safety data were analyzed with respect to time on eliglustat (including the PAP), which varied from 2.3 to ≥4.5 years depending on when randomization occurred (randomization spanned 2 full years), initial randomization group (placebo or eliglustat), and protocol-mandated switches to commercial eliglustat when it became available. **Results:** Overall, 7 patients were switched to commercial eliglustat, 2 discontinued due to pregnancy, and 4 chose to withdraw. Among the 14 patients with 4.5 years of data, mean spleen and liver volumes decreased by 66% and 23%, respectively, and mean hemoglobin level and platelet count increased by 1.4 g/dL and 87%, respectively. With regard to long-term therapeutic goals for these 4 parameters (Pastores, *Semin Hematol*. 2004), among the 33 patients with ≥2.5 years of data, the majority met all 4 goals, and 91% met at least 3 goals. Mean spine T-score increased by 21%, and median levels of the biomarkers glucosylceramide, glucosylsphingosine, chitotriosidase, and MIP-1β decreased by 79%, 84%, 82%, and 71%, respectively. Eliglustat was generally well-tolerated; 99% of all 559 adverse events (AEs) and all 7 serious AEs were mild or moderate; 83% were considered unrelated to treatment. No patients discontinued due to AEs, and there were no new safety concerns. **Conclusions**: In summary, final and complete results of the ENGAGE trial demonstrated continuation and maintenance of improvements in hematologic, visceral, biomarker, and bone measures after 4.5 years of treatment with eliglustat. Eliglustat was generally well-tolerated with no AE-related withdrawals. These results are consistent with the 4-year results of the eliglustat open-label Phase 2 trial in treatment-naïve patients (Lukina, *Blood Cells Mol Dis*. 2014).

2774T

Diagnostic of a worldwide cohort of Gaucher patients: Glucosylsphingosine levels in blood reflects the severity of GBA mutations. *A. Rolfs[1,2], A.K. Giese[1], S. Zielke[1], S. Eichler[2], C. Cozma[2].* 1) Albrecht-Kossel-Institute, University of Rostock, Rostock, Germany; 2) Centogene AG, Rostock, Germany.

Gaucher disease (GD) is an autosomal recessive, rare genetic disorder characterized by a deficient beta-glucocerebrosidase that leads to deposition of glucocerebroside in cells especially of the macrophage-monocyte system. Gaucher phenotypes vary widely among affected individuals from perinatal lethal form to an asymptomatic type. The rate of proper early diagnosis of GD patients is low. Standardized, simple and highly reproducible workflows for the diagnosis are crucial, especially since more than 15 years a highly effective treatment is available. We have developed and systematically validated a high-throughput workflow for the simple testing of GD patients: beta-glucocerebrosidase enzymatic activity, glycosylsphingosine (lyso-Gb1) quantification in DBS followed by GBA gene sequencing from the sample blood sample. Here, we report data from over 1,000 GBA variants in over 640 Gaucher individuals. Determination of lyso-Gb1 is performed by LC/MRM-MS. Lyso-Gb1 was investigated in three GBA cohorts comprising: (A) 133 homozygotes GBA cases; (B) 226 compound heterozygotes GBA patients and (C) 201 Gaucher carriers. The lyso-Gb1 levels found Gaucher patients were clinically divided in: very mild (12ng/ml -25.0 ng/mL), mild (25.1 - 50 ng/mL), moderate (50.1 - 200 ng/mL) and severe (>200 ng/mL). Lyso-Gb1 was proven to have a sensitivity and a specificity of 100% . The clinical severity of the mutations and their location could be correlated with the lyso-Gb1 in the homozygous cases, e.g. c.1295G>T is correlated with mild lyso-Gb1 values while c.1060G>A and c.518C>A with extremely high lyso-Gb1 values (> 600 ng/mL). The most common mutations: were c.1226A>G (30.1%) and c.1448T>C (24.7 %). c.1448T>C is a severe mutation correlating to a massive increase of lyso-Gb1 up to 1,250 ng/mL. lyso-Gb1 for c.1226A>G can vary from very mild to moderate. Lyso-Gb1 concentrations in blood can be used for the easy and early diagnosis of Gaucher patients and for treatment monitoring. Literature: Rolfs et al., 2013.

2775F

Obesity is a systemic regulatory outcome and mainly controlled by several tissues. *R. Hao, T. Yang, Y. Rong, S. Yao, S. Dong, Y. Guo.* School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China.

   Abstract It is normally believed that the abnormal or excessive accumulation of adipose tissue is the direct reason for obesity. With the growing evidence that other tissues, apart from adipose, could have essential effects on the incidence of obesity, it is necessary to comprehensively understand the roles of other tissues in affecting obesity. Here, we included 549 participants with 20 different tissue types to find the different importance of each tissue in affecting obesity. Firstly, in each tissue, we employed both Spearman's correlation test and weighted correlation network analysis (WGCNA) to identify genes with expression changes along with BMI distribution. Subsequently, we performed enrichment analyses with obesity-associated genes and pathways to see different regulation patterns among tissues. In addition, we compared obesity genes identified by genome-wide association studies (GWAS) with BMI-related genes to find the overlapping proportion in each tissue. Finally, we integrated preceding results to indicate 4 categories to represent varying degrees of effect on obesity. Statistical analyses revealed diverse BMI-related gene numbers and enrichment patterns among tissues. Comparison between all the BMI-related genes and GWAS findings showed over half overlap. Ultimately, 5 tissues showed abundant numbers of both BMI-related genes and enriched obesity-associated pathways, and successfully enriched with obesity-associated genes and prominently embraced GWAS genes. We further refer them as strong obesity-associated tissues, including adipose, brain, esophagus, muscle and pituitary. The BMI-related genes in these 5 tissues together could recognize 53.89% of GWAS identified obesity genes. In conclusion, our research suggests tissue-specific functions in pathogenesis of obesity. Also, we propose that 5 tissues could play the central roles in contributing to obesity development.

2776W

Genetic polymorphism of APOA5 gene is associated with metabolic syndrome in Koreans. *S.W. Oh[1], E. Shin[2], H. Kwon[3], EK. Choe[1], SY. Choi[1], SH. Choi[1], CH. Lee[1].* 1) Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, South Korea; 2) DNA Link, Inc., Seoul, South Korea; 3) Seoul National University Hospital, Seoul, South Korea.

   Background/Aims: Metabolic syndrome (Mets) which is caused by obesity and insulin resistance, is well known for its predictive capability for the risk of diabetes mellitus and cardiovascular disease. The development of Mets is associated with multiple genetic factors, environmental factors and lifestyle. Methods: We performed a genome-wide association study to identify genetic factors related to Mets in large Korean population based samples of 1,362 subjects with Mets and 6,061 controls using the Axiom® Korean Biobank Array 1.0. We replicated the data in another sample including 502 subjects with Mets and 1,751 controls. Results: After adjusting for age and sex, rs662799 located in the APOA5 gene, a locus previously reported, was validated in our population ($p = 2.85 \times 10^{-10}$ in the discovery set and $p = 3.19 \times 10^{-3}$ in the replication set). We investigated the regional plotting for SNPs which were aggregated regionally with distances less than 100 base pair apart and had P value less than 10^{-3}. 3 SNPs (rs8139657, rs5998275 and rs6518762) in the RFPL2 and C22orf42 gene were in strong linkage disequilibrium (LD) (R^2 =0.966, 0.983 and 0.963) and showed significant associations with Mets in the replication set ($p = 1.5 \times 10^{-2}$, $p = 7.77 \times 10^{-3}$ and $p = 1.90 \times 10^{-2}$). Conclusions: We confirmed that the APOA5 gene is associated with Mets in Korean. Additional studies for investigating the association of RFPL2 and C22orf42 gene and Mets are necessary.

2777T

Gene-level differential methylation analysis. *H. Xu[1], S. Li[1], X. Xu[2], X. Wang[2], X. Wang[2], S. Su[2], V. George[1].* 1) Dept Biostatistics & Epidemiology, Augusta University, Augusta, GA; 2) Georgia Prevention Institute, Department of Pediatrics, Augusta University, Augusta, GA.

DNA methylation is an important mechanism involved in many complex diseases. The biological and clinical significance results in increasing interests in the methods for differential methylation analysis in genome-wide association studies. Gene-level differential methylation analysis aims to identify differential methylated genes between groups. Limited statistical approaches are available for gene-level differential methylation analysis. We developed a statistical approach for gene-level differential methylation analysis based on clustered data analysis. This approach allows potential correlations between methylation levels from CpG sites within the defined genes. We compared our approach with IMA using two Illumina 450k methylation data sets, our own data on obesity and GSE54503. Analyses of these data indicate that our approach not only enables the identification of promising genes but also controls false positive rate efficiently.

2778F

Newborn screening for six lysosomal storage diseases in a cohort of Mexican patients: Three-year findings from a screening program in a closed Mexican health system. *J.I. Navarrete[1], A.E. Limón[1], M.J. Gaytán[1], J. Reyna[1], G. Wakida[1], R. Delgado[2], C. Cantú[3], H. Cruz[3], D. Cervantes[1].* 1) Hospital Central sur de Alta Especialidad Petroleos Mexicanos, Mexico City, Mexico; 2) Hospital Regional de Villahermosa Tabasco Petróleos Mexicanos; 3) Laboratorio Genomi.k Monterrey Nuevo León, México.

To evaluate the results of a lysosomal newborn screening(NBS) program in a cohort of 20,018 Mexican patients over the course of 3 years in a closed Mexican Health System (Petróleos Mexicanos Pemex Health Services). Using dried blood spots(DBS), we performed a multiplex tandem mass spectometry enzymatic assay for six lysosomal storage disorders (LSDs) including Pompe, Fabry, Gaucher, Mucopolysaccharidosis type I, Niemann-Pick Type A and B and Krabbe disease. Screen positive cases were confirmed using leukocytes enzymatic activity and DNA molecular analysis. Results From July 2012 to April 2016, 20.018 patiens were screened; 20 patients were confirmed to have an LSD phenotype (99.9 in 100,000 newborns) final distributions included 11 Pompe disease, 5 Fabry disease, 2 MPSI and 2 Niemann Pick type A/B. We did not find any Gaucher or Krabbe patients. A final frequency of 1 in 1001 LSD newborn phentypes was stablished. Discussion NBS is a major public health achievemente that has decreased the morbidity and mortality of inborn errors of metabolism. The introduction of NBS for LSD presents new challenges. This is the first multiplex Latin.american study of six LSDs detected through NBS:.

2779W

Obesity revisited: Evidence of genetic predisposition for metabolically healthy obesity. *L.O. Huang[1], R.J.F. Loos[2,3,4,5], T.O. Kilpeläinen[1,5].* 1) Section for Metabolic Genetics, Novo Nordisk Foundation Center for Basic Metabolic Research, København Ø, København , Denmark; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; 3) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York, United States of America; 4) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York United States of America; 5) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America.

Obesity has evolved into a global pandemic, which constitutes a major threat to public health. The majority of obesity-related health care costs are due to cardiometabolic complications, such as insulin resistance, dyslipidemia, and hypertension, which are risk factors for type 2 diabetes and cardiovascular disease. However, many obese individuals, often called metabolically healthy obese (MHO), seem to be protected from these complications. Recent large-scale genomic studies have provided evidence that some genetic variants show an association with increased adiposity but a favorable cardiometabolic profile, an indicator for the genetic basis of MHO. We want to identify genetic variants that are strongly associated with the MHO phenotype, defined by a range of trait pairs. We carried out cross-phenotype bivariate genome-wide meta-analyses on pairs of adiposity trait, measured by body fat percentage (BFP) and BMI, and cardiometabolic trait such as high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglycerides (TG), fasting insulin (FI), fasting glucose (FG), and systolic blood pressure (SBP), to confirm association pattern reminiscent of MHO. In the preliminary analysis of the BFP-HDL pair, we replicated some known loci, such as *IRS1, GRB14, CCDC92/DNAH10, TOMM40/APOE, FAM13A,* and *KFL14,* from previous GWAS. Furthermore, we identified additional loci using this approach, including *ABCA1, UBASH3B,* and *PCIF1,* which do not have genome-wide significant univariate p-values, but yielded genome-wide significant bivariate p-values. The bivariate analysis on the BFP-HDL trait pair has identified 85 independent loci exhibiting the MHO feature. Similarly, BFP-LDL, BFP-TG, BFP-FI, BFP-FG, and BFP-SBP trait pairs implicated over 150 loci, which include shared loci*, IRS1, CCDC92/DNAH10,* and *TOMM40/APOE*, with BFP-HDL, and additional ones, such as *SARS, CELSR2,* and *PPARG*. Analysis on BMI-associated trait pairs identified over 300 loci, some coincide with BFP-pair implicated loci, while others are novel. Many of these known and novel loci could locate in or near effector genes that implicate pathways involved in adipogenesis, fat distribution, insulin signaling, and insulin resistance. These important findings will likely contribute to the better understanding of the genetic aspects of the mechanisms that underpin MHO, which is crucial to define appropriate public health action points and to develop effective intervention measures.

2780T

Socioeconomic deprivation amplifies genetic susceptibility to obesity and its comorbidities. *K.A. Kentistou[1,2], P.K. Joshi[1], J.F. Wilson[1,3].* 1) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK; 2) Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, Scotland; 3) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.

Obesity's rapid inflation to epidemic proportions has often been attributed to a complex interplay between environmental and genetic factors[1]. It is associated with increased morbidity mediated by predisposition to type 2 diabetes, cardiovascular disease and various cancers. Body mass index (BMI) is a clinically important surrogate of obesity and is a representative measure of body fat composition. Thus far, genome-wide association studies (GWAS) have linked thousands of variants to obesity and its proxies[2]. Using the GIANT GWA meta-analysis[3], we constructed a polygenic risk score (PGRS) for BMI using all independent SNPs (p<=1) from 118,214 UK Biobank participants, to investigate the way in which the effect of genetic susceptibility to obesity might differ by socieconomic stratum. We observed that low socioeconomic status increased the effect of the PGRS on BMI by 25% (+/-0.05), between the top and bottom strata, beyond any effects of scale (p<2.2e-16). Although the size of the PGRS effects were stronger in women, their overall amplification was similar for both sexes. The effect of the genetic instruments on self reported prevalance of heart disease in the same cohort in the most socioeconomically deprived tertile was more than double that of the least deprived one (118% increase, +/-0.16, p=0.08), although no difference was observed between the two sexes. Even though non-significant, we also found a substantial increase in the effect of the PGRS on diabetes prevalence of 58% (+/-0.34, p=0.115). Our findings suggest that gene-environment interactions have the potential to explain a significant proportion of trait variance in obesity and its comorbities, and highlight the potential efficacy of targetted interventions and public health policies.   1. Barness, L.A., Opitz, J.M. and Gilbert☐Barness, E., 2007. Obesity: genetic, molecular, and environmental aspects. *American Journal of Medical Genetics*, *143*:3016-3034. 2. Burdett, T., Hall, P.N., Hasting, E., Hindorff, L.A., Junkins, H.A., Klemm, A.K., MacArthur, J., Manolio, T.A., Morales, J., Parkinson, H. and Welter, D., 2016. The NHGRI-EBI Catalog of published genome-wide association studies. Available at: www.ebi.ac.uk/gwas. 3. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J. and Croteau-Chonka, D.C., 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*:197-206.

2781F

Effects of the interaction between a *CREBRF* missense variant and body mass index on type 2 diabetes risk in Samoans. *E.M. Russell[1], N.L. Hawley[2], G. Sun[3], H. Cheng[3], T. Naseri[4], M. Sefuiva Reupena[5], I. Miljkovic[6], R. Deka[3], D.E. Weeks[1,7], S.T. McGarvey[8,9], R.L. Minster[1].* 1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, CT; 3) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH; 4) Ministry of Health, Government of Samoa, Apia, Samoa; 5) Bureau of Statistics, Government of Samoa, Apia, Samoa; 6) Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 7) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 8) International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, RI; 9) Department of Anthropology, Brown University, Providence, RI.

Studying genetically isolated groups, such as Samoans, may provide insight into the etiology of common complex diseases. Obesity and type 2 diabetes (T2D) are highly prevalent in Samoans (41% of men, 65% of women and 16% of men, 18% of women, respectively, in our cross-sectional sample). The minor allele of a missense variant in *CREBRF*, rs373863828 (c.1370G>A, p.Arg457Gln), is associated with increased body mass index (BMI) in the Samoan population. This variant is common in Samoans (minor allele frequency [MAF]=0.259), but is rare in other non-Pacific Islander populations (MAF < 0.0001). While BMI is a risk factor for T2D, the A allele is associated with lower than expected odds of T2D. We hypothesized that, in addition to main effects, there is a statistical interaction between the *CREBRF* variant and BMI affecting T2D odds. After log transforming BMI, we first used logistic regression to test this hypothesis. The interaction between genotype and BMI was not significantly associated with T2D (p=0.830); however, both BMI and genotype were significantly associated with T2D as main effects in an analysis that did not include an interaction term (p=8.31×10^{-9} and p=1.97×10^{-8}, respectively). To further explore the relationship, we divided the sample into three genotype-specific subgroups (GG, GA, AA), and we used logistic regression to test for association between BMI and T2D within each subgroup. There was an association between BMI and T2D in the GG (n=1513) and GA (n=1116) subgroups (p=3.68×10^{-5}, OR=4.749; and p=3.23×10^{-5}, OR=8.234, respectively); however, in the AA subgroup (n=232) there was no significant association between BMI and T2D (p=0.619, OR=1.773). This may suggest that the AA subgroup has a weaker association between BMI and T2D risk; however, this could also be the result of low power due to the small sample size of this subgroup. We used path analysis to further model the relationship between age, sex, the *CREBRF* variant and BMI. Our model indicates that sex has a moderate effect on BMI (p < 0.001, path coefficient ($P_{BMI,Sex}$)=−0.260) while genotype (Gt) and age have small effects (p < 0.001, $P_{BMI,Gt}$=0.128; and p < 0.001, $P_{BMI,Age}$=0.103, respectively). Age has a moderate effect on T2D (p < 0.001, $P_{T2D,Age}$=0.281) while genotype and BMI have small effects (p < 0.001, $P_{T2D,Gt}$=−0.102; and p < 0.001, $P_{T2D,BMI}$=0.108, respectively). These analyses suggest that the *CREBRF* variant does not synergistically alter the effect of BMI on the odds of T2D.

2782W

Genetic determinants of glycemic response to metformin in the Million Veteran Program. *C. Roumie[1,8], A. Giri[2], C. Kovesdy[7], D.R. Velez Edwards[1,2,3], J. Hellwege[1,2,5], O. Wilson[1,4], E. Torstenson[1,2,5], K.A. Birdwell[4], C. Chung[1], E. Siew[1,4], M. Matheny[1,6], T. Edwards[5], A. Hung[1,4] on behalf of the VA Million Veteran Program.* 1) Tennessee Valley Healthcare System Nasville VA; Nashville , TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center Nashville TN; 3) Department of Obstetrics and Gynecology, Vanderbilt University Medical Center; 4) Division of Nephrology & Hypertension, Vanderbilt University Medical Center; 5) Division of Epidemiology, Department of Medicine Vanderbilt University Medical Center; 6) Department of Bioinformatics, Vanderbilt University medical Center; 7) Nephrology Section, Memphis VA Medical Center, Memphis TN; 8) Department of General Internal Medicine and Public Health Vanderbilt University Medical Center, Nashville TN.

Glycemic response to metformin, the first line therapy for type-2 diabetes, is variable and heritable. Only two genome-wide significant loci have been identified to date. A better understanding of the genetic architecture may help differentiate non-responders from responders. We used data from 8,282 veterans [6,553 self-reported European Americans (EA) and 1,729 self-reported African Americans (AA)] enrolled in the Million Veteran Program (MVP) to perform a trans-ethnic genome-wide association study (GWAS) of glycemic response to metformin. Adults with normal kidney function, who initiated and took metformin for at least 180 days and had available pre (baseline) - and on-metformin treatment Hba1c measures were included in the study. The lowest recorded on-treatment Hba1c measures within 18-months of baseline Hba1c were considered for each individual. Glycemic response to metformin was computed as absolute reduction in Hba1c (baseline minus on-treatment), and relative reduction in Hba1c (absolute change divided by baseline). Linear regression models were used to evaluate the association between markers imputed to the 1000 Genomes cosmopolitan reference panel and response to metformin-Hba1c traits while adjusting for important clinical covariates and 5 principal components. We report two novel genome-wide significant loci, one for relative reduction in EAs (rs55755349: intron in *FAM107B*, 0.39% reduction for allele T, P=1.4x10^{-9}) and another for absolute reduction in trans-ethnic meta-analysis (rs4253564; 0.17 unit reduction for allele GT; P=2.4x10^{-8}). Variants within 10KB of rs55755349 have been associated with platelet function. Large platelet aggregation is characteristic in hyperglycemia. Rs4253564 lies within 500KB of well-known type-2 diabetes associated gene, *ZMIZ1*. Previously reported GWAS significant SNP rs8192675 at *SLC2A* was also nominally significant with consistent direction of effect in AAs (P=0.02), but not in EAs. Interaction with BMI was also observed at this SNP in AAs, with largest effect in obese AAs (P-value=0.009). We also identify another SNP rs11309000 (P=4.9x10^{-6}) within <500Kb of rs8192675. We identify several other suggestive loci (P< 1x10^{-6}) with ties to type-2 diabetes and Hba1c. This trans-ethnic GWAS finds two novel loci associated with glycemic response to metformin in a well-defined cohort of EA and AA veterans. We further extend evidence for previously reported main- and interaction- effects at *SLC2A* to AAs as well.

2783T

Principal component-based prediction of complex traits by using support vector machine approach. *X. Li, X. Zhu.* Department of Population and Quantitative Health Science, Case Western Reserve University, School of Medicine, Cleveland, OH.

The use of genetic information for predicting complex disease risk has played a major role in precision medicine. In this work, we have developed a technique for genetic risk prediction using GWAS data. Our method, a two-step principal component based support vector machine approach, focuses on relaxing the assumption on the effect to accommodate different genetic architecture and takes into account linkage disequilibrium. Unlike traditional prediction approaches, such as the genetic risk score which relies on pruning and setting an arbitrary P-value threshold, the linear mixed model which assumes that the SNP effects follow a normal distribution or the Bayesian approach that requires specifying a prior, the major benefit of such an approach is that we no longer need to make assumptions on the distribution of the input data. We compared our method with the best linear unbiased predictor (BLUP) and demonstrated the improvement in risk prediction performances through Welcome Trust Case Control Consortium (WTCCC) data. Area under the ROC curve (AUC) are 75% for type 1 diabetes, 66% for rheumatoid arthritis, 64% for Coronary Artery Disease and 61% for Crohn's disease, consistently higher than 67%, 61%, 57% and 60% from BLUP, respectively. Our proposed method can substantially increase the accuracy of risk prediction across a range of diseases with different genetic architecture.

2784F

Trans-ethnic meta-analysis of rare variants in sequencing association studies. *J. Shi, S. Lee.* Biostatistics, University of Michigan, MI.

The rapid technological advances in high-throughput sequencing platforms have made it possible to test for rare variant associations to accelerate our knowledge of complex trait genetics. A practical approach to identify rare variants is to aggregate studies that have already been conducted through a trans-ethnic meta-analysis, as it aims to include samples from as many studies as possible, even if they come from different ancestries. With the increased sample size, trans-ethnic meta-analysis is expected to be more powerful at detecting novel loci. Unfortunately, under the presence of genetic effect heterogeneity among participating studies in multiethnic meta-analysis, existing approaches may be unsatisfactory because they only assume varying genetic effects between studies, but do not take into consideration that studies from the more closely related ancestries can be more homogeneous than those that are more distinctly related. To fully take advantage of the strengths of multi-ethnic meta-analysis, we propose a unified score test under a modified random effects model framework for rare variants associations. The proposed method is capable of not only accounting for the expected heterogeneous genetic effects among studies, but also flexibly modeling varying levels of heterogeneity according to the relatedness between the populations. Specifically, we adopt the kernel regression framework to construct the modified random effects model, and incorporate the genetic distances across ancestry groups into modeling the heterogeneity structure of the genetic effect coefficients. In addition, we use the adaptive variance component test to achieve robust power regardless of the degree of heterogeneity. The copula method is employed to approximate the asymptotic distribution of the proposed test, which enables rapid estimation of p-values without the need for Monte Carlo simulations. To evaluate performance of our proposed method, we conducted a series of simulations with varying assumptions on genetic effect heterogeneity. Results show that our proposed method controls type I error rates at stringent significance levels and improves power over existing approaches under the presence of heterogeneity. We will further apply our method to the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia data, a multiethnic sample of 12,940 individuals which focuses on exome sequence variations.

2785W

A novel approach to analyze the mediation model when the mediator is a censored variable. *J. Wang[1], S. Shete[1,2].* 1) Dept Biostatistics, UT MD Anderson Cancer Ctr, Houston, TX; 2) Dept Epidemiology, UT MD Anderson Cancer Ctr, Houston, TX.

A mediation model is a widely used statistical method which studies the direct and indirect effects of an initial variable on an outcome by including one or more mediators. In realistic studies, investigators might observe censored data instead of the complete data. The most current approaches for analyzing mediation models with censored variables focus on the censored outcome variable. However, the mediator in mediation model can also be a censored variable. In this study, we proposed a novel approach to analyze the mediation model when the mediator is a censored variable, where we suggested using the accelerated failure time (AFT) model to evaluate the effect of the initial variable on the censored mediator and a multiple imputation approach to assess effects of the initial variable and mediator on the outcome variable. We also adapted a measure of the indirect effect for the mediation model with a censored mediator, which can assess the indirect effect at both the group and individual levels. We conducted simulation studies to investigate the performance of the proposed approach and compared it to that of the existing approaches, including a naïve approach, complete-case analysis, and the To-bit mediation model. The proposed approach, based on AFT and the multiple imputation approach, accurately estimates the coefficients of different paths, indirect effects and percentages of the total effects mediated; while the existing approaches provided biased estimations. We applied the existing and proposed approaches to investigate the indirect effect of the age at menopause on the association between SNPs and type-2 diabetes. Our findings showed that there was no statistically significant indirect effect of age at menopause on the association between SNPs and type-2 diabetes.

2786T

Genome-wide association analysis in the UK Household Longitudinal Study offers insights into the genetic architecture of health-related biomarkers. *K. Kuchenbaecker[1], B. Prins[1], Y. Bao[2], M. Smart[2], D. Zabaneh[3], G. Fatemifar[4], J. Luan[5], N. Wareham[5], R. Scott[5], J. Perry[5], C. Langenberg[5], M. Benzeval[2], M. Kumari[2], E. Zeggini[1].* 1) Wellcome Trust Sanger Institute, Hinxton, UK; 2) Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, UK; 3) MRC Social, Genetic & Developmental Psychiatry Centre, IoPPN, KCL, London; 4) Institute for Health Informatics, UCL and the Farr Institute of Health Informatics, London, UK; 5) MRC Epidemiology Unit, University of Cambridge and Clinical Medicine, Institute of Metabolic Science, Cambridge, UK.

Levels of serum biomarkers can reflect health status and are associated with complex diseases, such as cardiovascular, metabolic and auto-immune disorders. We aimed to gain additional insights into the genetic architecture of biomarker traits and identify novel genetic associations.   We performed genome-wide association analyses for a panel of twenty serum biomarkers involved in liver and kidney function, and reproductive health. A total of 9,961 individuals from the UK Household Longitudinal Study were genotyped using the Illumina CoreExome array and variants imputed to a dense reference panel of 1000 Genomes Project and UK10K haplotypes. The association analyses were carried out using linear mixed models to account for relatedness and betas estimated per biomarker standard deviation. Heritability and genetic correlations were estimated using GREML as implemented in GCTA.   We demonstrate polygenic heritability of the majority of biomarkers included in this study. Alkaline phosphatase and testosterone had the highest array heritability estimates with h^2=27.7% (standard error (SE): 0.040) and h^2=27.1% (SE: 0.084), respectively. The lowest estimate was observed for ferritin (h^2=6.1%, SE: 0.037). We identify previously unreported genetic correlations between several of the biomarkers. Polygenic factors for triglyceride levels were negatively correlated with the ones for dihydroepiandrosterone sulphate (DHEAS), a precursor to sex-hormones (genetic correlation rg=-0.53 p=4.0x10[-4]). There was a positive genetic correlation between C-reactive protein and fibrinogen levels (rg=0.60 p=3.2x10[-8]), which could be due to shared inflammation pathways. The genetic correlation between creatinine and urea (rg=0.56 p=1.2x10[-5]) is a highly biologically plausible finding as both markers are increased in blood when glomerular filtration rate declines, reflecting impaired kidney function. We also identify five novel, replicating associations at stringent genome-wide significance with levels of alanine transaminase, an indicator of liver damage, with DHEAS, and with glycated haemoglobin (HbA1c) levels, a biomarker used to diagnose and manage type 2 diabetes. Our study offers insights into the genetic architecture of both well-known and less well-studied biomarkers.

2787F

PLEIOVAR, testing for association between multiple traits and multiple variants. *O. Meirelles[1], D. Schlessinger[1], L. Ferrucci[2], F. Cucca[3], J. Ding[1].* 1) Laboratory of Genetics, National Institute on Aging, 251 Bayview Blvd. Baltimore, MD,21224, USA; 2) Longitudinal Studies Section, National Institute on Aging, 251 Bayview Blvd. Baltimore, MD,21224, USA; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Cagliari - 09042 Monserrato CA Sardegna, Italy.

Genome-Wide Association Studies (GWAS) explore associations between DNA variants, usually single nucleotide polymorphism (SNPs), and a trait. Recently, analyses have been extended to assess association between multiple SNPs and a trait, or between a single SNP and multiple traits. We add an assessment of pleiotropic effects of groups of SNPs, from regions defined by each gene and the DNA in close proximity. PLEIOVAR, the assessment program, is based on principal component analysis of SNPs and traits. The level of association is measured by a summary statistic whose significance is evaluated using a Chi-squared distribution. We applied the program to trait values in 5,715 individuals in the SardiNIA population study. For lipids -- LDL, HDL and triglycerides – and for glycemic traits -- fasting glucose, fasting insulin and HBA1c -- PLEIOVAR identified both previously reported pleiotropic associations and additional signals of moderate effect size that had not been seen with standard GWAS. The analysis yields a candidate set of novel joint associations with multiple traits for studies of underlying functional interactions. We have replicated our results applied to the same traits in 1,206 individuals in the InCHIANTI study cohort.

2788W

Gene-depressive symptoms interactions identify novel lipid loci in multi-ethnic cohorts. *S.K. Musani[1], M.A. Richard[2], A.R. Bentley[2], M. Brown[3], M. Sims[1], M. Fornage[2], CHARGE Gene-Lifestyle Interactions Working Group.* 1) Jackson Heart Study, University of Mississippi Medical Center, Jackson, MS; 2) Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas Health Science Center, Houston, TX; 3) Center for Research on Genomics and Global Health, NHGRI, NIH, Bethesda, MD.

Background: Genetic and psychosocial factors, such as depression and depressive symptoms, are known to contribute independently to variation in cardiovascular traits including plasma lipids, but little is known about interactions between genetic and psychosocial factors. We hypothesized that depressive symptoms modulate the effect of genetic variants on plasma lipids. Hence, we performed a trans-ethnic genome-wide meta-analysis using 1000G imputed single nucleotide polymorphisms (SNP) and interaction with the presence of depressive symptoms on 3 plasma lipids: fasting high density lipoproteins (HDL), fasting low density lipoproteins (LDL) and fasting triglycerides (TG). **Methods**: Presence of depressive symptoms was defined using standard screening questionnaires. Our analyses represented 4 ancestries: 46,177 (European); 14,499 (African); 4,453 (Hispanic) and 2,592 (Asian). In the interaction analyses, each study calculated the joint 2 degree-of-freedom (df) test of the SNP effect and the SNP × depression symptoms interaction effect, as well as the 1 df SNP × depressive symptom interaction effect. **Results**: In trans-ancestry meta-analyses we replicated over 43 known lipid loci and identified 10 novel loci where variants were significantly (p< 5×10[-8]) associated with HDL (LOC105375070, *WAC, MACROD2*); LDL (*CREB5, CREB3L2, TRIM5, TRIM6,* LOC105370312, *ZAN790-AS1) and TG* (LOC105375070) using the 2df joint test. Variants in the *MACROD2* locus also showed significant (p< 5×10[-8]) interaction with depressive symptoms using the 1 df test. African Ancestry specific meta-analyses identified 2 additional loci: variants in *KNCIP* and LOC105370312 were significantly associated with LDL while European specific meta-analyses identified 3 additional loci (*GABBR2, MAF* and *PIK3R2*) associated with all three lipid traits. Each of these ancestry-specific loci were identified using the 2 df test. Some of these genes have been reported to be associated with cancer metastasis (*CREB5, MAF*); neurocognitive phenotypes (*WAC, MACROD2, KNCIP4*), regulation of calcium concentration (*GABBR2*) and risk of HIV infection (*TRIM5, TRIM6*). Further work is ongoing to validate these loci and to understand the biological mechanisms underlying each of the associations.

2789T

Targeted sequencing of 109 genes in the eMERGEseq panel uncovers novel variants and genes influencing triglyceride levels. *X. Fan, M.S. Safarova, K. Ding, J.E. Olson, S.N. Thibodeau, D.J. Schaid, I.J. Kullo.* Mayo Clinic, Rochester, MN.

   BACKGROUND: Triglyceride (TG)-rich lipoprotein levels increase the risk of coronary heart disease. We investigated whether rare variants and genomic regions (sets of variants) in 109 medically relevant genes included in the eMERGE (electronic MEdical Records and GEnomics) sequencing panel, influence TG levels.   **METHODS**: We performed targeted sequencing in 2,547 individuals participating in the Return of Actionable Variants Empirical (RAVE) study. We excluded 5 individuals without a lipid panel and 16 non-European ancestry individuals based on principal component analyses. The mean age was 52±9, 57% were female, mean body mass index (BMI) was 30±6 kg/m^2, mean TG level was 265±274 mg/dL and 813 (35%) individuals had TG levels > 250 mg/dL. Variant and gene associations with log(TG) adjusted for age, sex, BMI and the first two principal components were conducted using linear regression in PLINK and kernel regression in sequence kernel association test (SKAT).   **RESULTS**: Five missense variants passed a genome-wide threshold of significance ($P<5\times10^{-8}$) (Table): rs3135506 (*APOA5*) and rs1260326 (*GCKR*) were both common variants previously known to be associated with TG; a rare variant (chr20:44540064C>A) in *PLTP* that inhibits VLDL lipidation; chr16:15802684T>C (*MYH11*) and rs200652865 (*CORIN*) were novel loci. At the gene level, *APOA5*, *CORIN* and *KCNQ1* were also significantly associated with TG. The lead variants (variant with the lowest *P*-value in each gene) in *APOA5* and *CORIN* were significant at variant level. Two intronic variants in *KCNQ1* were reported to be associated with TG in a Chinese cohort. Minor allele frequency (MAF) stratified analyses indicated that rare variants in *CORIN* and *KCNQ1* drove the signal. *CORIN* remained associated with TG after removing the lead variant in conditional analysis. In contrast, after removing the lead variants in *APOA5* and *KCNQ1*, these loci were not associated with TG. **CONCLUSIONS**: In this large-scale targeted sequencing study, we found five variants and three genes to be associated with TG levels including two variants and *CORIN* that have not been reported previously to be associated with TG. These findings suggest that TG levels are influenced by multiple variants in different genes.

Summary of significant variants associated with TG levels				
Variant	**Gene**	**MAF**	**Beta**	***P*-value**
rs1260326	*GCKR*	0.408	0.16	3.60E-11
rs200652865	*CORIN*	0.000	4.76	3.80E-17
rs3135506	*APOA5*	0.074	0.27	5.30E-9
16:15802684T>C	*MYH11*	0.001	2.33	6.40E-9
20:44540064C>A	*PLTP*	0.000	4.83	1.80E-9

2790F

Examining the causal relationship between vitamin D and serum metabolic measures: A Mendelian randomization study. *T. Dudding[1], L.J. Corbin[1], P. Wurtz[2], M. Mannikko[3], S. Mannisto[4], V. Salomaa[4], M. Perola[5], J. Kettunen[2], T. Lehtimäki[6], O. Raitakari[7], M. Ala-Korpela[1,2], R. Martin[1], N.J. Timpson[1].* 1) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 2) Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu, Finland; 3) Centre for Life Course Epidemiology, Institute of Health Sciences, University of Oulu, Finland; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Public Health Genomics, Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 6) Department of Clinical Chemistry, University of Tampere, Finland; 7) The Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Finland.

   Background Circulating vitamin D is associated with skeletal health, cancer risk and progression and cardio-metabolic diseases but intervention studies give conflicting results. Few epidemiological studies have investigated vitamin D associated intermediates that could elucidate causal pathways to disease. This study aims to identify and estimate the size of causal associations between serum vitamin D concentration and 158 metabolic measures independent of confounders and reverse causality using Mendelian randomization. **Participants and methods** 20,650 participants from 6 population-based cohorts were included in the study. 158 metabolic measures were obtained from a nuclear magnetic resonance platform and where available, serum 25 hydroxyvitamin D (25OHD) concentrations were measured using a blood assay. Genetic variants reliably associated with 25OHD concentrations were extracted from genome wide association data. Two-sample Mendelian randomization using participant-level genetic and metabolite data for the second stage regressions was conducted. Observational 25OHD – metabolic measure associations were assessed using linear regression. **Results** In observational analyses, a one standard deviation (SD) unit increase in log-25OHD was associated with a decrease in larger subclasses of very-low-density lipoprotein (VLDL) (e.g. very large VLDL particle concentration: -0.05 SD (95% CI: -0.06, -0.03) $P=9\times10^{-9}$). Mendelian randomization showed consistent but imprecise estimates in support of a causal relationship in the same direction (very large VLDL particle concentration: -0.04 SD (95% CI: -0.11, 0.03) $P=0.3$). Serum 25OHD concentration was strongly associated with degree of fatty-acid unsaturation, omega-3 and docosahexanaenoic acid concentration in observational analyses but not in causal analyses. Observational associations between serum 25OHD and amino acids were small or absent but suggestive inverse causal associations were identified between 25OHD and several amino acids, including branched-chain amino acids (Leucine: -0.07 SDs (95% CI: -0.14, -0.01) $P=0.02$). **Conclusions** Mendelian randomization estimates are consistent with previous findings that vitamin D reduces VLDL. Associations of vitamin D with healthy fatty acid profiles are unlikely to be causal. Novel potentially causal inverse associations between vitamin D and branched-chain amino acids may have therapeutic potential where branched-chain amino acids are causally related to health outcomes.

2791W

Enriched loss-of-function variants associating with lipids in Finns. *P. Helkkula[1,2], I. Surakka[1,2], V. Salomaa[3], S. Ripatti[1,2,4].* 1) Institute for Molecular Medicine Finland, Helsinki, Finland; 2) Unit of Public Health Genomics, National Institute for Health and Welfare, Helsinki, Finland; 3) Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom.

Population isolates, such as Finland, show an enrichment of loss-of-function (LoF) variants with potential impact on phenotypes and disease risk. Previous examples include two LoF variants in *LPA,* causing marked reductions in lipoprotein(a), and show more than tenfold enrichment in Finns compared to non-Finnish Europeans (NFEs). We considered protein-truncating and splice site variants associated with lipids in Finland. Our study consisted of 20,643 autosomal LoF and splice site variants, with 8,353 with minimum twofold enrichment in Finns. We tested their association with LDL-C, HDL-C and triglycerides in population-based FINRISK cohort (N=23,975) which has been genotyped and imputed with a whole-genome and high-coverage Finnish-specific imputation reference panel. We found five LoF variants with strong effects on lipids. A protein-truncating variant (*rs760351239;* MAF=0.0015) in *ANGPTL8* is strongly associated with lower triglyceride levels (TG-effect=-0.613, *P*=2.46e-08). The variant is more than 80 times enriched in Finns compared to NFEs. It was also associated with other lipids (HDL-effect=0.603, *P*=1.92e-08; non-HDL-effect=-0.539, *P*=1.414e-06; LDL-effect=-0.416, *P*=2.51e-04; TC-effect=-0.296, *P*=9.04e-03) and protective of T2D (OR=0.29, *P*=0.036). Angptl8 is a feeding-induced hepatokine with an essential role in postprandial triglyceride trafficking, due to its lpl-inhibiting action. Hypotriglyceridemia has previously been observed in murine *ANGPTL8* knockouts. We also observed three LoF variants with substantial HDL-increasing effects in genes *LIPG* (*rs200435657*, MAF=0.0017, effect=0.638, *P*=4.69e-11), *SLC12A3* (*rs765555468*, MAF=0.0013, effect=1.00, *P*=1.51e-17) and *CETP* (*rs751916721*, MAF=0.0014, effect=1.09, *P*=3.11e-20). The variants are over 200, 50 and 47 times enriched in Finns, respectively. *LIPG* encodes endothelial lipase, which hydrolyzes triglycerides and phospholipids, but especially HDL-C. *SLC12A3* is the target of thiazide diuretics used for treating high blood pressure, but the way it influences HDL-C levels is unknown. *CETP* mediates reverse cholesterol transport, exchanging cholesteryl ester in HDL for triglycerides in ApoB lipoproteins. These examples illustrate the potential in using population isolates to discover downstream health effects of enriched variants interrupting protein coding. These types of findings may also point to potential intervention targets.

2792T

Do blood lipid levels influence bone mineral density? Findings from a Mendelian randomization study. *J. Zheng[1], M.J. Brion[2], J. Kemp[2], N. Warrington[2], C. Borges[1], G. Hemani[1], T.G. Richardson[1], Z. Qiao[2], P.C. Haycock[1], G. Davey Smith[1], J.H. Tobias[1,3], D.M Evans[1,2].* 1) Medical Research Council (MRC) Integrative Epidemiology Unit, School of Social & Community Medicine, University of Bristol, Bristol, United Kingdom; 2) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia; 3) Musculoskeletal Research Unit, University of Bristol, Bristol, UK.

Treatment with statins is associated with increased bone mineral density (BMD) and reduced fracture risk. Previous laboratory studies suggest that direct effects of statins on osteoblasts and osteoclasts underlie this relationship. However, few studies have investigated an alternative possibility that the beneficial effects of statins on bone might also be mediated by lowering levels of low density lipoprotein cholesterol (LDL-C). In order to investigate the relationship between plasma lipids and BMD, we performed a two-sample Mendelian randomization (MR) study. We utilized 184 single nucleotide polymorphisms (SNPs) robustly associated with plasma lipid levels and 239 SNPs associated with heel BMD derived from quantitative ultrasound in 142,487 individuals from the UK Biobank Study as genetic instruments. We performed univariate MR analyses on LDL-C, high density lipoprotein cholesterol (HDL-C) and triglyceride levels, and multivariable MR to account for genetic pleiotropy among the different lipid fractions. To test whether the effect of statins on BMD was mediated by lowering lipid levels, MR was repeated with and without SNPs in the *HMGCR* region, which was used to mimic the effects of statin use. We also performed bi-directional MR to examine reverse causality between BMD and blood lipids. Analyses were conducted using the MR-Base R package (https://github.com/MRCIEU/TwoSampleMR). Univariate MR using SNPs associated with LDL-C, HDL-C or triglyceride levels provided evidence for a causal effect of LDL-C on BMD (β=-0.077; P=2x10^{-5}; SD change in BMD per SD change in LDL-C) and the results were robust to pleiotropy. Multivariable analysis suggested that the effect of LDL-C on BMD was independent of HDL-C and triglycerides. MR analyses involving SNPs in the *HMGCR* region were consistent with a causal effect of LDL-C on BMD (β=-0.110; P=0.016) as were MR analyses excluding these SNPs (β=-0.102, P=3x10^{-5}). Bidirectional MR analyses provided no evidence for a causal effect of BMD on blood lipids. We are following up these results in the full Biobank sample to ensure that they are free from possible biases due to ascertainment. Our results suggest lower LDL-C is causally related to increased BMD. This suggests the effect of statins on BMD is at least partly due to their LDL-C lowering effect. Further studies are justified to explore the mechanisms by which lower LDL-C improves BMD, and to examine the potential role of modifying blood lipids in treating osteoporosis.

2793F

Adaptive multi-trait association test using GWAS summary data. *B. Wu, B. Guo.* Division of Biostatistics, University of Minnesota, Minneapolis, MN.

Genetics hold great promise to precision medicine by tailoring treatment to the individual patient based on their genetic profiles. Toward this goal, many large-scale genome-wide association studies (GWAS) have been performed in the last decade to identify genetic variants associated with various traits and diseases. They have successfully identified tens of thousands of diseases related variants. However they have only explained a small proportion of the overall trait heritability for most traits and are of very limited clinical use. This is partly owing to the small effect sizes of most genetic variants, and the common practice of "testing association between one trait and one genetic variant at a time" in most GWAS, even when multiple related traits are often measured for each individual. There is increasing evidence showing that many genetic variants can influence multiple traits simultaneously, and we can gain great power by testing association of multiple traits simultaneously. Ideally we can reanalyze those existing GWAS genetic and phenotype data using the multi-trait association approach. However due to privacy concerns and various logistical considerations, it is generally very hard to access the individual-level GWAS phenotype and genotype data, which creates a barrier to use these existing data to further identify novel genetic variants. We discuss several of our recently developed statistical methods for adaptive multi-trait association test that only needs summary association statistics publicly available from most GWAS. We first show that the principal component (PC) based association test can be readily reconstructed from the summary statistics. PC based test has optimal power when the underlying multi-trait signal can be captured by the first PC, and otherwise it will have suboptimal performance. We then construct an adaptive test by optimally weighting the PC based test and the omnibus chi-square test to achieve robust performance under various scenarios. A class of novel tests explicitly accounting for potential pleiotropy effects are also developed. We develop efficient numerical algorithms to compute the analytical p-value for all our proposed adaptive tests without the need of resampling or permutation. We illustrate the competitive performance of proposed methods through application to several GWAS meta-analysis summary data. .

2794W

Genetic variation associated with telomere length in African American children with and without asthma. *M. White[1], A. Zeiger[1], J. Witonsky[1], M. Contreras[1,2], O. Risse-Adams[1,3], C. Eng[1], D. Hu[1], S. Huntsman[1], S. Oh[1], E. Burchard[1,4].* 1) Department of Medicine, University of California, San Francisco, San Francisco, CA; 2) SF BUILD, San Francisco State University, San Francisco, CA; 3) Lowell Science Research Program, Lowell High School, San Francisco, CA; 4) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA.

Telomere length is a biomarker of biological aging and has been associated with the progression and presence of several chronic diseases including asthma. Telomere length varies by race/ethnicity with African Americans having shorter average telomeres compared to European-descent individuals. Genome-wide association studies (GWAS) have identified several genetic variants associated with telomere length; however, these studies have been largely limited to adult populations of European descent. Genetic associations discovered in European populations do not always translate to other populations and it is possible that there may also be both population-specific and age-specific genetic determinants of telomere length that have been overlooked. It is also possible that individuals with asthma may have unique genetic associations with telomere length compared with healthy controls. Our aim was to address this gap in knowledge by providing insights and preliminary results on the genetic determinants of telomere length in African American children with and without asthma. Absolute telomere length was measured using qPCR using DNA isolated from whole blood collected from 1442 African American children with and without asthma (asthma = 879, controls = 563) and available GWAS data. Linear regression models were adjusted for age, percentage of African ancestry, maternal education level, and insurance status. Analyses were stratified by asthma case control status as well as by gender. We evaluated polymorphisms previously associated with telomere length in adults as well as performed a discovery-only GWAS and meta-analysis in our pediatric African American study population. The previously reported GWAS-identified variants in adults were not associated with telomere length in our pediatric sample. Several novel polymorphisms in various genes were found to be genome-wide significant (p-value < 5 x 10[-8]) in our discovery-only GWAS and meta-analysis. Our study of the genetic determinants of telomere length in African American children revealed several novel associations between telomere length and genetic polymorphisms from multiple genes that were genome-wide significant. Further large-scale studies of children from multiple populations will be needed to determine if our findings are population-specific or age-specific.

2795T

Genome-wide haplotype-based association study reveals novel non-HLA susceptibility loci for primary biliary cirrhosis in Japanese cohorts. *C. Im[1], Y. Sapkota[2], M. Nakamura[3], K. Tokunaga[4], Y. Yasui[1,2].* 1) School of Public Health, University of Alberta, Edmonton, Alberta, Canada; 2) Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN; 3) Clinical Research Center, National Hospital Organization, Nagasaki Medical Center, Nagasaki, Japan; 4) Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan.

Primary biliary cirrhosis (PBC) is a chronic, progressive autoimmune disease of the liver with a strong hereditary component. While robust PBC susceptibility loci have been discovered through single SNP association testing, exploration of the joint effects of SNPs organized on the same chromosome, or haplotypes, on PBC pathogenesis is warranted. Haplotype investigations may be more powerful in mapping disease loci and can reflect *cis*-interactions that regulate gene expression. To this end, we conducted a gene-based search of haplotype SNP patterns associated with PBC risk in a cohort of 1,937 Japanese individuals (901 cases, 1,036 controls) and replicated selected signals in an independent Japanese cohort (480 cases, 469 controls). Haplotype phase was computationally estimated using SHAPEITv2.79 for unphased, genome-wide SNP genotype data. After mapping SNPs to RefSeq genes and corresponding 500-kb windows at 5'/3' ends, our analysis included ~272K SNPs mapped to 15,105 gene regions. We identified gene-based haplotype patterns consisting of 3 SNPs associated with PBC using logic regression, an adaptive regression methodology that detects best combinations of predictors associated with an outcome. We then applied a permutation-based selection statistic to 53 HLA (human leukocyte antigen) and 8,878 non-HLA unique haplotype patterns, selecting patterns with the top 1% of selection statistics for replication. A total of 24 haplotype patterns met Bonferroni-corrected p-value thresholds in replication (P<0.05/# of selected haplotypes). Replicated haplotype signals not only corroborated previous GWAS signals, but also implicated additional SNPs with regulatory functions in relevant cell types. For example, *TNFS15-TNFSF8* is an important PBC susceptibility locus in Japanese cohorts. One of the strongest non-HLA haplotype signals involved 3 intergenic SNPs residing near *TNFSF15* (9q32-9q33.1; OR=1.7, P=3.0x10^{-21}, combined sample), each overlapping with promoter and enhancer histone mark peaks in liver and immune cell epigenomes; 2 SNPs were significantly associated with *TNFSF8* expression in whole blood. A novel non-HLA haplotype signal involving SNPs mapped to *UMAD1* introns (7p21.3; OR=15.2, P=3.9x10^{-9}, combined sample) includes SNPs overlapping enhancer histone peaks in liver and memory T$_h$ cells. In summary, our study demonstrates the utility of haplotype association analyses in characterizing PBC susceptibility loci.

2796F

Precisely controlled differential gene expression system to investigate the effect of eQTL. *X. Lu[1,2], X. Chen[1], C. Schroeder[1], C. Forney[1], J. Harley[1,2,4], M. Weirauch[1,3], L. Kottyan[1,2].* 1) Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Immunobiology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH; 3) Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) US Department of Veteran Affairs Medical Center, Cincinnati, OH.

Genome-wide association studies and large-scale sequencing studies identify many non-coding genetic variants that increase disease risk. At least 60% of these loci have been associated with genotype-dependent expression of nearby genes. It remains to be determined that 50-100% expression differences are biologically relevant in specific immune cell types. Here, we developed a system combining the Clustered regularly interspaced short palindromic repeats (CRISPR) Homology-directed Repair technique with Tet-inducible Controlled Gene Expression technique to precisely control gene expression. This system allows us to mimic the differential gene expressions caused by genetic variants. We focus on the expression quantitative trait loci (eQTL) at the ETS1 lupus-risk locus. Patients with lupus have 50% less peripheral blood mononuclear cell mRNA expression of ETS1 than people without SLE, and people with the risk haplotype at the ETS1 locus have 50% less mRNA expression than people with the non-risk haplotype. Since ETS1 is a transcription factor, differential ETS1 expression could have easily measurable consequences on transcription factor binding and downstream gene expression. There are nearly 800 genes that are known to be dysregulated in subjects with SLE. Our analyses of these genes using publically available ChIP-seq data sets from B cell line GM12878 and epithelial cell line K562 indicates that 191 and 151 of these genes, respectively, have ETS1 binding sites proximal to the transcription start site (enrichment of 3-4 fold compared to randomly chosen gene sets, p<10^{-46}). We hypothesize that these changes in ETS1 levels are important in 1) ETS1 binding throughout the genome, 2) expression of ETS1 downstream target genes (such as BLIMP1), and 3) immunological dysfunction and hyperactivity of B cells. We test these hypotheses by analyzing the binding of ETS1 (ChIP-seq) and the changes to the transcriptome of cells with 2-fold differences in ETS1 expression (RNA-seq). The investigation of the ETS1 eQTL will allows us to make important progress in the field of human genetics and especially complex genetic disease etiology.

2797W

Transcriptional risk scores link GWAS to eQTL and predict complications in Crohn's disease. *U.M. Marigorta[1], L.A. Denson[2], J.S. Hyams[3], S. Kugathasan[4], G. Gibson[1], RISK Consortium.* 1) Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA; 2) Division of Pediatric Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Division of Digestive Diseases, Hepatology, and Nutrition, Connecticut Children's Medical Center, Hartford, CT; 4) Division of Pediatric Gastroenterology, Emory University School of Medicine, Atlanta, GA.

Differential gene expression profiling can be used to uncover the mechanisms by which GWAS associations contribute to pathology. Given that most GWAS hits are regulatory in nature and transcript abundance is in a sense closer to the phenotype, we hypothesize that risk scores based on gene expression could constitute an alternative to classical genetic risk scores (GRS). While GWAS have been very successful in identifying thousands of genetic variants associated with disease, the predictive performance of GRS is limited by the amount of heritability they explain, which is usually low. Intuitively, summation of the z-scores of risk-allele associated gene expression, namely a Transcriptional Risk Score (TRS), should provide a more accurate estimate of disease risk. We integrate summary-level GWAS and eQTL data with RNA-Seq from the RISK study, an inception cohort of pediatric Crohn's disease (CD) designed to identify factors that increase risk of a complicated course of disease. Ileal biopsies from 215 complication-free CD patients and 35 controls were profiled at diagnosis with RNA-Seq. After careful monitoring for 3 years, 27 of the CD patients progressed to stricturing or penetrating disease. We show that TRS based on genes regulated by IBD variants not only outperform GRS in distinguishing CD from healthy samples, but also serve to identify patients who in time will progress to complicated disease. Furthermore, our dissection of eQTL effects may be used to distinguish genes whose association with disease is either through promotion or protection, thereby linking statistical association to biological mechanism. The TRS approach constitutes a potential strategy for personalized medicine that enhances inference from static genotypic risk assessment.

2798T

***HLA-DQ* variants interact with pregnancy to modify risk of multiple sclerosis among women of European ancestry.** *C. Adams[1], X. Shao[1], H. Quach[1], S. Bos[2], H. Harbo[2], I. Brorson[2], L. Alfredsson[3], T. Olsson[3,4], I. Kockum[3], P. Strid[3], A. Langer-Gould[5], C. Schaefer[6,7], L.F. Barcellos[1,6,7].* 1) School of Public Health, Division of Epidemiology, University of California, Berkeley, Berkeley, CA; 2) Department of Neurology, Oslo University Hospital, Oslo, Norway; Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 3) Department of Clinical Neuroscience and Center for Molecular Medicine, Karolinska Institutet at Karolinska University Hospital, Stockholm, Sweden; 4) Centre for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden; 5) Kaiser Permanente, Southern California, Department of Research & Evaluation and Southern California Permanente Medical Group, Los Angeles Medical Center, Neurology Department; 6) Kaiser Permanente Division of Research, Oakland, CA; 7) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS [MIM 126200]) is a demyelinating autoimmune disease with unknown etiology. Genetic and environmental risk factors have been identified, but their mechanisms are not understood. Human leukocyte antigen (HLA) genes contribute substantially to MS risk, and previous genome-wide association studies have established 110 non-HLA genetic variants associated with MS risk. MS predominantly affects women and first symptoms typically appear during childbearing years. Pregnancy is known to affect MS progression, but its role in MS etiology is unknown. It is hypothesized the immunologic changes which occur during pregnancy contribute to the risk of MS, however, studies of the marginal effect of pregnancy on MS risk have mixed results. Gene-environment (GxE) interactions can provide insight into the unexplained heritability of MS and complex pathways involving genes and exposures. This study investigated the interaction between MS genetic risk factors and pregnancy on MS risk among women using a case-only GxE study design. We hypothesized interactions between established MS genetic risk factors and pregnancy are associated with MS risk. All cases were adult females with confirmed disease selected from Kaiser Permanente California (KP), Norwegian, and Swedish populations (n=2,500). Pregnancy exposure was defined as first live-birth pregnancy before onset of first MS symptoms. Genotype data were obtained through whole-genome profiling and imputation. The case-only study design allows for estimation of GxE interaction effects when control data are not available, but cannot estimate marginal effects. Generalized linear mixed models with a logit link function were used to test for interaction between pregnancy exposure and genetic risk variants with random effects for population of origin, correcting for multiple tests ($q<0.2$). Initial results demonstrated evidence for interaction between pregnancy and *HLA-DQ* variants. Among women who were pregnant prior to MS onset, each *HLA-DQA1*01:01* allele and *HLA-DQB1*04:02* allele increased risk of MS 1.57-fold (95% CI: 1.22-2.11) and decreased risk of MS 0.77-fold (95% CI: 0.65-0.92), respectively. Our results provide evidence that biologic pathways for MS risk include the interaction between class II HLA genes and pregnancy. Further studies are needed to investigate genetic interaction with additional reproductive factors and MS risk as well as possible biologic mechanisms linking pregnancy and genetic risk.

2799F

Trans-ethnic GWAS identifies genetic variants associated with white blood cell counts in the Population Architecture using Genomics and Epidemiology (PAGE) Study. *K.K. Nishimura[1], M. Graff[2], J. Haessler[1], C. Hodonsky[2], R. Tao[3], S. Bien[1], C. Schurmann[4], G. Nadkarni[4], S. Buyske[5], C. Haiman[6], L. Le Marchand[7], R.J.F. Loos[4], T. Matise[8], K.E. North[2], U. Peters[1], C.S. Carlson[1], A.P. Reiner[1], C.L. Avery[2], C. Kooperberg[1].* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of North Carolina Chapel Hill, Chapel Hill, NC; 3) Vanderbilt University Medical Center, Nashville, TN; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Statistics and Biostatistics, Rutgers University, New Brunswick, NJ; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 7) Cancer Research Center, University of Hawaii, Honolulu, HI; 8) Department of Genetics, Rutgers University, New Brunswick, NJ.

Background: White Blood Cell (WBC) counts are often used by clinicians to understand the nature of an ongoing inflammatory process. WBC counts are moderately heritable, yet can also be highly variable between healthy individuals. Some of this variability has previously been linked to certain genetic ancestral backgrounds. We conducted a large, multi-ethnic genome-wide association study (GWAS) in the Population Architecture using Genomics and Epidemiology (PAGE) Study to gain a deeper understanding of ancestry-specific genetic variants associated with WBC phenotypes. **Methods:** 28,534 PAGE individuals of African American, Hispanic/Latino, Asian, Native American, or Other ancestry with WBC measurements were genotyped on the Multiethnic Genotyping Array (MEGA) and subsequently imputed into the 1000 Genomes Phase 3 data. WBC outliers (+/- 4 standard deviations) were excluded, and WBC counts were log transformed. Linear regression models were run using the SUGEN software, and adjusted by age at WBC measurement, sex, BMI, study, current smoking status, self-identified race/ethnicity, study center (where applicable), family/household membership (where applicable), and 10 ancestral principal components (PCs). Additional analyses including race-ethnicity stratified, conditional analyses using previously reported known loci, and SNPxPC interaction models were also conducted. **Results:** We identified a new independent signal in the known locus 13q21.33/*DACH1* (rs17088512, P=6.94E-09, PAGE coded allele frequency (CAF)=0.12), which remained genome-wide significant after conditioning on all previously known loci on the chromosome (P-conditional=3.64E-08). The PAGE lead variant is rare in 1000 Genomes Europeans (CAF=0.01), but common in those with African (CAF=0.32) and Admixed American (CAF=0.04) ancestry, populations which comprise the majority of the PAGE study population (P-African American=1.17E-05; P-Hispanic/Latino=4.13E-04). **Discussion:** For traits that show variability by genetic ancestry, it is especially important to perform GWAS in ancestrally diverse study populations. Trans-ethnic GWAS studies can identify genetic traits that are linked to certain genetic ancestries, such as the rs17088512 variant, which is frequent in both African Americans and Hispanic/Latinos. Additional analyses including rare variant and gene burden tests, fine mapping, and GWAS of WBC subtypes are ongoing to gain a deeper understanding of WBC phenotypes in non-European populations.

2800W

Finding genomic variants regulating the exon-skipping. *R. Liu[1,2,3], A. Byrnes[1,2], M. Yu[3], M. Daly[1,2], H. Huang[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA, USA; 2) The Broad Institute, Cambridge, MA, USA; 3) Huazhong Agricultural University, Wuhan Hubei, China.

Multiple different mature mRNA isoforms can be generated from a single gene locus through alternative splicing. Abnormality in alternative splicing has been linked to many human disorders. To date, the extent to which genomic variations are involved in regulating the alternative splicing has not been systematically evaluated. Here, using RNA-seq data from GTEx, we present a study to find genomic variants regulating exon-skipping and isoform balance in a genome-wide fashion. We examined RNA-seq reads crossing each splice junction in the genome and looked for high quality exon-skipping events by comparing the number of the exon-skipping and non-exon-skipping reads. For each sample, we tested the imbalance of different RNA isoform levels using the binomial test, and converted the resulting p-values to z-scores. To find variants regulating the exon-skipping, we tested the association between the z-scores and variant genotypes across individuals. All variants within 1Mbp around the exon junction were tested. We applied this approach to the RNA-seq data from the EBV-transformed lymphocytes cell from the GTEx project (sample size: 83). We only considered exon-skipping events that have >1 junction read from each isoform in >25% samples. The initial analysis identified 44,016 such events, among which 919 events are significantly associated with at least one regulatory variant (p<10[-10], threshold determined as 0.05 / 44,016 events /10,000 variants tested per event). One interesting example is the *SP140* exon 7 skipping event. We found rs28445040 to be the most significantly associated SNP with this skipping event (p= 1.19x10[-11]). This observation is consistent with an earlier exon splicing assay which also found rs28445040 regulating the skipping of *SP140* exon 7 (Matesanz et al., Human Molecular Genetics, 2015). In addition, the *SP140* locus has been implicated by GWAS as a risk factor for the inflammatory bowel diseases (IBD). A recent IBD fine-mapping study mapped this association to a credible set of 31 variants including rs28445040, suggesting that the skipping of exon 7 in *SP140* could be the disease-causing mechanism. In summary, we have designed a novel approach and demonstrated its utility to identify variants regulating exon-skipping events. Results from this approach could potentially help to interpret the functional mechanism of non-coding disease associations.

2801T

Variance component selection with microbiome taxonomic data. *J. Zhou[1], J. Zhai[1], J. Kim[2], H. Zhou[2].* 1) University of Arizona, Tucson, AZ; 2) University of California, Los Angeles.

Pulmonary complications remain a major cause of morbidity in subjects infected with Human Immunodeficiency Virus (HIV). The fundamental question raised recently was whether respiratory microbiome plays a role in these complications. In order to identify the bacterial taxa that are associated with a response, e.g., lung function, microbiome data are summarized as the counts or composition of the bacterial taxa at different taxonomic levels. Previous methods consider variable selection of taxonomic data in regression analyses and taxonomic data at different levels are considered as fixed covariates. Due to high-dimensional features of metagenomic information, different penalization schemes have been adopted. On the other hand, the association of microbiome composition and clinical phenotypes was assessed by testing the nullity of variance components, where phylogenetic tree information and distances measures between communities can be incorporated into the model. By combining these two methods, in this paper, we consider regression analysis by treating bacterial taxa at different level as multiple random effects. We propose a variance component selection scheme of high-dimensional taxonomic clusters with Lasso penalization. Our methods seamlessly coupled different distance measures with associated taxonomic selection. Extensive simulations demonstrate the superiority of our methods vs existing methods. Finally, using a longitudinal HIV lung microbiome cohort, We identify two interesting genera associated with Forced Vital Capacity (FVC), which shed lights on biological mechanisms.

2802F

Genome-wide association study identifies susceptibility loci for primary non-response to anti-TNF therapy in patients with inflammatory bowel disease. *T. De[1], C. Alarcon[1], E. Smithberger[2], M. Horvath[2], S. Kwan[2], M. Young[2], S. Adhikari[2], J. Kwon[2], M. Perera[1].* 1) Department of Pharmacology, Northwestern University, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL.

BACKGROUND: Tumor necrosis factor (TNF) α is one of the major pro-inflammatory cytokines in the underlying pathogenesis of mucosal inflammation in inflammatory bowel disease (IBD), and are the primary therapeutic target for IBD. Anti-TNF drugs have been shown to induce clinical response in 60% of the patients, and long-term maintenance of remission in a large number of patients that respond to the drug. However, approximately 30% of the patients fail to respond to anti-TNFs (primary non-responders). Individual genetic differences could contribute to lack of efficacy. We therefore conducted the first genome-wide association study to identify the genetic variants associated with underlying differences in response to anti-TNFs. **METHODS:** The Discovery Cohort comprised of 109 cases that were primary non-responders, and 504 controls that responded to anti-TNF therapy. Population stratification was investigated through principal component analysis and only Caucasian patients were considered for this study. Association between each SNP and non-response to anti-TNFs was tested by logistic regression analysis using SNPTEST v2.5.2. **RESULTS:** Among the clinical covariates, combination or symptomatic therapy with azathioprine (p = 0.007), methotrexate (p = 0.007) and antibiotics (p<0.0001) were found to influence the response to anti-TNFs. Therefore, the association analysis was carried out after adjusting for these covariates, and the first principal component. A variant on chromosome 5 was found to be significantly associated with non-response to anti-TNFs (OR: 2.07, 95% CI:1.46-2.94, p = 2.43E-07). Bioinformatics analysis showed the variant to be an expression quantitative trait loci (eQTLs) for *MFAP3* and *FAM114A2* genes, in blood and duodenum mucosa. The findings from this study will be replicated in an independent Cohort of 63 cases and 339 controls. **CONCLUSION:** Our study may identify a clinically significant variant that can be used to tailor drug therapy to those patients that are most likely to gain benefit from the therapy and reduce the drug related risks in those that are unlikely to respond to anti-TNF drugs.

2803W

Testing for colocalization of causal variants underlying obstructive sleep apnea and immune-related phenotypes. *S. Akle[1,2], S. Chun[2,3], B. Cade[4], R. Saxena[4,5], S. Redline[4], S. Sunyaev[2,3].* 1) Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Division of Genetics, Brigham and Women's Hospital, Boston, MA; 3) Division of Medical Sciences, Harvard Medical School, Boston, MA; 4) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; 5) Center for Human Genetic Research and Department of Anesthesia, Pain, and Critical Care Medicine, Massachusetts General Hospital, Boston, MA.

Obstructive Sleep Apnea (OSA) is a complex genetic disease manifested by the loss of upper airway patency during sleep. The causal pathways and mechanisms underlying OSA have not been fully understood. Recently, genome-wide association studies on OSA have uncovered multiple genetic loci and suggested that immune-related genes may play a pleiotropic role in its ethology. To examine this hypothesis further, we applied Joint Likelihood Mapping (JLIM) to test for colocalization of genetic association signals between OSA and previously published immune-related cell count traits in blood of healthy individuals. Specifically, JLIM contrasts the likelihood of shared pleiotropic causal effect from that of distinct causal effects for two traits. To minimize the possibility of ancestry mismatch between cohorts, we restricted the analyses to only Caucasian subjects (n=1463) and controlled for age, sex and population structure covariates. Among 1796 previously known genome-wide significant associations to 14 non-compound immune related blood cell traits in myeloid and lymphoid lineages, 729 had nominal association to the Apnea-Hypopnea Index (AHI), the mostly widely used clinical indicator of OSA, in 200-kb windows (P < 0.01). We ran JLIM to colocalize AHI and each of the 14 immune related blood traits in 729 tests. At nominal JLIM P-value threshold of 0.01, we find significant colocalization between OSA and 2 blood cell traits. Although we could not find any individually significant hits after multiple-testing correction, we can find signal of colocalization between AHI and two blood cell traits (P =0.038). The blood cell count traits that colocalize with AHI are: monocyte count and neutrophil count. This lends credence to previous suggestions that inflammation and other immune related processes might have a pleiotropic role to play in sleep related pathologies. .

2804T

Association study of *R3HDM1* variants with aspirin exacerbated respiratory disease and FEV1 decline after aspirin provocation. *J. Kim[1], Y. Cho[2], C. Park[3], H. Shin[1,4].* 1) Research Institute for Basic Science, Sogang University, Seoul, Republic of Korea; 2) Endocrine Institute, Yonsei University, Seoul, Republic of Korea; 3) Division of Allergy and Respiratory Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea; 4) Department of Life Science, Sogang University, Seoul, Republic of Korea.

Aspirin exacerbated respiratory disease (AERD) is a rare clinical condition characterized by a severe decline in forced expiratory volume in one second (FEV1) when taken with non-steroidal anti-inflammatory drugs (NSAIDs) including aspirin. Our previous genome-wide association study has identified several candidate genes, including *R3HDM1 (R3H domain containing 1)*, as potential susceptibility loci for AERD. To validate the potential association of *R3HDM1* variants with AERD and FEV1 in asthmatic patients, we performed an extended study using additional single nucleotide polymorphisms (SNPs), different genetic models, and haplotype analysis in a larger cohort. In the logistic analysis, *R3HDM1* polymorphisms showed potential association signals for the development of AERD (*P* = 0.01). In additional linear regression analysis, *R3HDM1* polymorphisms and haplotype showed potential associations with FEV1 decline after aspirin provocation in asthmatics. Although further replications in other populations and functional evaluations are required, our findings suggest that *R3HDM1* polymorphisms might be associated with AERD and FEV1 decline, or its related symptoms.

2805F

Statistical framework for biological interpretation and improvement of genetic association studies. *M. Artomov[1], A.A. Sergushichev[2,3], A.A. Loboda[2], M.N. Artyomov[3], M.J. Daly[1].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Computer Technologies Department, ITMO University, Saint Petersburg, 197101, Russia; 3) Department of Pathology & Immunology, Washington University in St. Louis, St.Louis, MO 63110, USA.

   Analysis of the genetic risk factors provides only unstructured pieces of information about the biology of a disorder. Generally, after identification of the associated loci massive follow-up studies are required to, first, prove the causal relationship, and, most importantly, understand the molecular mechanism of causality. Which locus should be prioritized for protein-level studies is currently determined based on empirical knowledge of protein function. Integration of the experimentally proven individual proteins functionality is then aimed to identify pathways affected by disease. Alternatively to this extensive approach, we developed a statistical framework that integrates genetic association data from multiple sources (GWAS, RVAS, etc.) and finds the protein-protein network returning the best cumulative association score. Using Bayesian model association results are then refined with evidence of the specific gene appearance in the best network. Our method provides a ranked list of genes prioritized based on both association strength and integration in the functional pathway. Such approach is essential for understanding biology of the disorders where it is impossible to build adequate animal model – autism, schizophrenia and other neuropsychiatric diseases. We used ranked list of genes from focal segmental glomerulosclerosis (FSGS) study (PMID: 26901816) to test our approach. None of the genes originally reached significance threshold for genetic association (0.05/1642 genes represented in reference interactome = 3×10^{-5}). Resulting most associated network consists of five genes – *WNK4, COL4A4, DLG5, KAT2B, UBC.* We used two schemes of permutation to estimate probability of each of these genes to appear in the best scoring network. According to our approach - posterior p-values of genetic association were calculated. Two genes became significant – *COL4A4, WNK4* and had more than 10 fold improvement in association signal: *COL4A4* – is a known risk gene for FGSG; mouse model (PMID: 26901816) demonstrated that mice with non-functional *WNK4* develop histologically confirmed FSGS, thus proving its causal role. At the same time mouse model ruled out *DLG5* and *KAT2B* as causal genes. Our model successfully predicts functionally-relevant genes from genetic association studies and could also be applied for identification of functional candidates in polygenic GWAS loci.

2806W

Integrated clinical genome database on hepatitis B-related diseases for genome-wide association: Project goals and utilization of materials and genomic information in the ToMMo biobank. *S. Teraguchi[1], N. Nishida[2], M. Sugiyama[2], K. Kojima[1], Y. Kawai[1], H. Kudo[1], R. Yamashita[1], N. Minegishi[1], K. Tokunaga[3], M. Nagasaki[1], M. Mizokami[2].* 1) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Genome Medical Sciences Project, National Center for Global Health and Medicine, Chiba, Japan; 3) Department of Human Genetics, The University of Tokyo, Tokyo, Japan.

   Hepatitis B is an infectious disease caused by the hepatitis B virus (HBV), which affects the liver. Chronic hepatitis B infection, as well as hepatitis C infection, is known to lead to liver cirrhosis, and eventually to liver cancer. While hepatitis C can be well treated with the recent innovative drugs, effective drugs are lacking for hepatitis B. This lack of effective treatment for hepatitis B is partially due to the diversity of the disease, depending on virus genotypes and host genotypes; understanding of such diversity is urgent for the development of effective treatments against HBV. To this end, we are constructing an integrated clinical genome database on hepatitis B-related diseases under "the program for an integrated database of clinical and genomic information" of the Japan Agency for Medical Research and Development. A number of hospitals and biobanks in Japan, including the Tohoku Medical Megabank Organization (ToMMo) of Tohoku University, are involved in this project. Here, we present recent advances on this project, with particular focus on the utilization of the biobank materials and genomic information of the ToMMo biobank. ToMMo developed highly accurate Japanese population reference panels based on whole-genome sequences of healthy Japanese individuals. Using these reference panels, we imputed the genome-wide SNVs from SNP array data from healthy and HBV-infected individuals. We also performed genome wide association analysis for the imputed data. These results will be registered in the integrated clinical genome database regarding hepatitis B-related diseases. ToMMo also provided serum samples from two thousand volunteers for this project. HBV antigen/antibody tests were performed on these samples to refine the clinical information. We performed integrated analysis with the resulting HBV status and the associated WGS data of these volunteers.

2807T

Trans-ethnic Bayesian meta-analysis detects novel replication evidence for multiple loci for inflammatory bowel disease in African Americans.
R.Y. Cordero[1], S. Kugathasan[2], D.P.B. McGovern[3], S.R. Brant[4], C.L. Simpson[1]. 1) Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN; 2) Department of Pediatrics and Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 3) F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA; 4) Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine and Department of Epidemiology, School of Public Health, Johns Hopkins University, Baltimore, MD.

Background: Inflammatory bowel disease (IBD) is an immune-mediated chronic intestinal disorder with major phenotypes; ulcerative colitis (UC) and Crohn's disease (CD). Multiple studies have identified over 240 IBD susceptibility loci. Recently, trans-ethnic analyses have gained ground in detecting additional loci. However, most studies have centered on European (EUR) and East Asian (EAS) populations. The prevalence of IBD in non-EUR including African Americans (AAs) has risen in recent years. Here we present the first attempt to identify loci in AAs using a trans-ethnic Bayesian approach. **Methods**: We conducted trans-ethnic meta-analysis using GWAS and Immunochip data from EUR meta-analysis of 38155 IBD cases and 48485 controls; EAS Immunochip study of 2824 IBD cases and 3719 controls, and our recent AA IBD GWAS of 2345 cases and 5002 controls. To account for heterogeneity in allelic effects between diverse ancestry groups while allowing for similarity in allelic effects between closely related populations, we adopted a Bayesian model as implemented in MANTRA. We employed a $\log_{10}$ Bayes' factor (BF) threshold of 20 as strong evidence for association in AAs. **Results:** We identified IBD loci at *IL23R* (BF 59.8) and *PTGER4* (BF>31.8), and at *CARD9, NKX2-3, BRWD1, ZNF365* and *MST1* with $20.1<BF<27.1$. We identified strong signals for CD again at *IL23R* and *PTGER4*, but also *ATG16L1* (BF 34.7). However, there was evidence of heterogeneity in allelic effect for *ATG16L1,* demonstrating that unlike in the GWAS analysis, which did not find evidence for association with this gene, variants in *ATG16L1* do have some effect on risk of disease, albeit much smaller than in EUR. For UC we identified loci at *RNF186* (BF 26.2) and IL23R. A comparison of AA with the much smaller EAS cohort identified BF>20 loci only for CD at *TNFSF15* (BF 33.8). Only the *IL23R, PTGER4* and *TNFSF15* loci had significant evidence for locus replication in our parent IBD GWAS in AAs. **Conclusion:** Our Bayesian trans-ethnic analysis identified strong replication evidence for multiple IBD loci in AAs with the majority of these loci demonstrating no compelling evidence for locus replication in the parent GWAS. In particular, for *ATG16L1*, prior replication studies only showed minimal association evidence. These results highlight the value of using prior association evidence from much larger studies in other populations to enable strong evidence for replication in smaller cohorts of understudied diverse populations.

2808F

Relationship of genetic and clinical factors and prevalence of CKD in a Japanese population: J-MICC Study. *R. Fujii[1], A. Hishida[2], M. Nakatochi[3], N. Furusho[4], H. Ikezaki[4], R. Okada[2], S. Kawai[2], T. Sasakabe[2], K. Tanaka[5], C. Shimanoe[5], S. Suzuki[6], A. Hosono[6], T. Koyama[7], N. Kuriyama[7], R. Ibusuki[8], I. Shimoshikiryo[8], H. Uemura[9], S. Katsuura-Kamano[9], N. Takashima[10], TC. Turin[10], K. Kuriki[11], K. Endoh[11], H. Mikami[12], H. Nagase[12], I. Oze[13], H. Ito[13], M. Kubo[14], T. Kondo[1], M. Naito[2], K. Wakai[2], J-MICC Study Group.* 1) Department of Pathophysiological Laboratory Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan; 2) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 4) Department of General Internal Medicine, Kyushu University Hospital, Fukuoka, Japan; 5) Department of Preventive Medicine, Faculty of Medicine, Saga University, Saga, Japan; 6) Department of Public Health, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 7) Department of Epidemiology for Community Health and Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan; 8) Department of International Island and Community Medicine, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan; 9) Department of Preventive Medicine, Institute of Health Biosciences, the University of Tokushima Graduate School, Tokushima, Japan; 10) Department of Health Science, Shiga University of Medical Science, Otsu, Japan; 11) Laboratory of Public Health, School of Food and Nutritional Sciences, University of Shizuoka, Shizuoka, Japan; 12) Division of Epidemiology, Chiba Cancer Center Research Institute, Chiba, Japan; 13) Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Nagoya, Japan; 14) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Chronic kidney disease (CKD) is a common health concern throughout the world, as well as in Japan. The majority of genome-wide association studies (GWAS) for renal traits have conducted in European populations with a few studies in an East Asian population. Furthermore, clinical relevance of combination of CKD risk factors including genetic information remains unclear. Therefore, we investigated a cross-sectional relationship of clinical and genetic factors and prevalence of CKD in a Japanese cohort study. Clinical and genetic information were the part of the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study, which is one of the largest cohort studies in Japan. The present study consists of 14,539 randomly selected participants from 12 J-MICC Study sites (Chiba, Sakuragaoka, Shizuoka-Daiko, Okazaki, Aichi, Takashima, Kyoto, Tokushima, Fukuoka, Kagoshima and Kyushu-KOPS). The study protocol and procedure was approved by the Ethics Committee of Nagoya University Graduate School of Medicine and all participating institutions. Written informed consent was obtained from all participants. Genotyping was performed using the Illumina OmniExpressExome Array. After a quality check, genotype imputation was conducted based on the 1000 Genomes reference panel. A total of 11,184 participants and 7,094,228 variants was available for our analyses. Estimated glomerular filtration rate (eGFR) was calculated based on serum creatinine, age, and sex using the Japanese equation. Participants in stages 3–5 CKD (eGFR < 60 ml/min/1.73 m^2) was defined as case in this study. We computed a genetic risk score (GRS) based on 36 eGFR-associated SNPs identified by the present population-based GWAS in Japanese. The mean age (SD) of this population was 54.9 (9.3), and the proportion of men was 45.9%. The frequency of individuals with CKD was 8.5%, which was relatively low comparing with other studies in Japan. In multivariable analysis, the odds ratio (OR) of having stage 3-5 CKD was observed as 10 GRS increased (OR: 1.16, 95% confidence interval: 1.10-1.23). The C-statistic was significantly increased in the adjusted model with GRS, comparing with the model without GRS (0.719 vs. 0.724, $P_{\text{difference}}$ < 0.01). We confirmed that increment of GRS was associated with increased risk of CKD and GRS significantly improved the discrimination of CKD. Further studies with longitudinal datasets are needed to assess clinical utility of GRS for predicting onset and prognosis of CKD.

2809W

A genome- and phenome-wide association study of diverticular disease using electronic health records. *Y.J. Joo[1], WK. Thompson[1], FT. Lin[1], JA. Pacheco[1], LJ. Rasmussen-Torvik[1], L. Rasmussen[1], J. Denny[2], D. Carrell[3], J. Grafton[3], PA. Peissig[4], KM. Borthwick[5], M. Andrade[6], J. Pathak[7], A. Veerappan[8], A. Gawron[9], GC. Tromp[10], M. Smith[1], RL. Chisholm[1], MG. Hayes[1], AN. Kho[1].* 1) Northwestern University, Chicago, IL; 2) Vanderbilt University, Nashville, TN; 3) Kaiser Permanente Washington, Seattle, WA; 4) Marshfield Clinic, Marshfield, WI; 5) Geisinger Health System, Danville, PA; 6) Mayo Clinic, Rochester, MN; 7) Weill Cornell Medical College, New York, NY; 8) VA San Diego Healthcare System, San Diego, CA; 9) University of Utah Health, Salt Lake City, UT; 10) Stellenbosch University, Stellenbosch, South Africa.

    Diverticulosis is one of the most common functional gastrointestinal disorders affecting almost half of all Americans before the age of 60. Its pathogenesis is unknown, but if untreated, it can develop into diverticulitis with life-threatening complications such as peritonitis or colonic perforation. As part the Electronic Medical Records and Genomics (eMERGE) network, we performed a genome-wide association study (GWAS) to identify common variants associated with diverticular disease and a phenome-wide association study (PheWAS) of the top variants to identify potential pleiotropic effects. We developed and validated an electronic health record (EHR)-based phenotyping algorithm to identify cases and controls for diverticulosis and diverticulitis using EHR data from all adult eMERGE subjects. (n=38,828) We identified two associations with diverticulosis approaching genome-wide significance: rs12935426 (G allele, OR=0.86, P=1.2E-7) in *PHLPP2*, in European ancestry (EA) participants and rs12428560 (G allele, OR=3.61, P=8.5E-8) near *LINC00448* in African ancestry (AA) participants. Three genome-wide significant variants were associated with diverticulitis: chr2:80804829:I (G insertion, P=3.7E-8) near *CTNNA2* in EA, rs6016299 (T allele, OR=0.51, P=2.5E-8) between *LINC01370* and *MAFB* in AA, and rs11062710 (allele G, MANTRA Bayes factor = 5.28) in *PRMT8* in a meta-analysis of EA and AA. We selected the top 12 SNPs associated with diverticulosis and diverticulitis for PheWAS, in which we identified several clinical traits associated with these variants, including and eye and skin disorders; such as cataract (rs10423674, p=8.8E-05), glaucoma (rs10423674, p=2.4E-04), disorders of optic nerve and visual pathways (rs6028974, p=1.5E-03; rs6016299, p=2.1E-03), degenerative skin conditions and other dermatoses (rs10423674, p=1.6E-03) atopic/contact dermatitis (rs10423674, p=4.30E-03), and skin cancer (chr2.80804829.i, p=1.1E-02), suggesting pleiotropy of these variants with diverticular disease and the nervous and integumentary systems. These associations, combined with the observation of calcium metabolism related genes (e.g. *COMP*, *CRACR2A*, *RYR3*) among the most significantly associations, suggests that signal transduction of the central nervous system, interruption of the neural cell growth in the colon, and malfunction of the collagen component of the epithelial membrane may play a role in the development of diverticular disease.

2810T

Gene-based pathway analysis for osteoporosis: Insights from genomic-wide association. *K.J. Su[1,2], W. Zhu[2], C. Xu[1,2], H. Shen[1,2], HW. Deng[1,2].* 1) Global Biostatistics and Data Science, Tulane University, New Orleans, LA; 2) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA.

    Background: Osteoporosis is a chronic human complex disease with decreased bone strength and increased risk of fractures, which has been characterized by low bone mineral density (BMD). To date, genome-wide association studies (GWASs) have successfully identified a large number of susceptibility loci associated with low BMD. However, these identified loci only explain a small proportion of the BMD variation and provided limited insights into the pathophysiological mechanisms of osteoporosis. To further explore the pathologic mechanism of osteoporosis, we performed gene-based pathway analyses for BMD based on the largest meta-analysis of GWASs in the bone field. **Methods:** First, we examined BMD-associated genes by analyzing the summary statistics of the sequencing project of the GEnetic Factors for OSteoporosis Consortium (GEFOS.seq) with BMD of the femoral neck (FN-BMD), lumbar spine (LS-BMD), and forearm (FA-BMD), respectively. To improve the robustness of the identified candidate genes, we selected significant genes identified by two different gene-based association analysis strategies (the all-SNP test and the best-SNP test). Next, these BMD-associated genes were subject to pathway analysis through Enrichr and annotated by Gene Ontology (GO) enrichment analysis. Additionally, to gain detailed information of pathway analysis, we integrated a transcriptomic data from bone biopsy and conducted a differential expression analysis (DEA) to identify key mechanism of genes within each pathway. **Results:** We identified 709 genes nominally significantly associated with FN-BMD, 693 genes with LS-BMD, and 463 genes for FA-BMD. The pathway analysis revealed a total of 16 significant pathways (FDR < 0.05) for BMD, such as Wnt signaling pathway, RANKL/RANK signaling pathway, and Hedgehog signaling pathway. Three bone-related biological processes terms were identified on both the FN-BMD and LS-BMD (FDR < 0.05), including skeletal system development, ossification, and bone development. Notably, the difference in the expression of BMD associated genes in bone biopsy showed a gene *JUN* which could potentially provide valuable clues for the significant functional pathways. **Conclusion:** This study not only identified several known bone-related pathways, but also explored more potential risk pathways that may provide new insights into the pathological mechanism of osteoporosis.

2811F

Integrated pediatric bone density phenotypes and genetic regulation of the developing skeleton. *J.A. Mitchell[1,2], A. Chesi[3], S.E. McCormack[2,4], D.L. Cousminer[3], H.J. Kalkwarf[5], J.M. Lappe[6], V. Gilsanz[7], S.E. Oberfield[8], J.A. Shepherd[9], A. Kelly[2,4], B.S. Zemel[1,2], S.F.A. Grant[2,3,4].* 1) Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia; 3) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia; 4) Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia; 5) Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati; 6) Division of Endocrinology, Department of Medicine, Creighton University, Omaha; 7) Department of Radiology, Children's Hospital Los Angeles, Los Angeles; 8) Division of Pediatric Endocrinology, Diabetes, and Metabolism, Department of Pediatrics, Columbia University Medical Center, New York; 9) Department of Radiology, University of California San Francisco, San Francisco.

Background Osteoporosis is a complex disease with potential developmental origins in childhood. Common and low frequency variants have been associated with pediatric areal bone mineral density (aBMD), but phenotyping approaches used thus far have considered skeletal sites in isolation. We therefore aimed to determine if genetic variants associate with principal component (PC)-derived pediatric aBMD loading scores that integrate information across four skeletal sites. **Methods** Our sample comprised 1,293 children of European ancestry enrolled in the Bone Mineral Density in Childhood Study (52% female). Participants underwent dual energy X-ray absorptiometry for up to 7 annual study visits. Sex and age-specific aBMD Z-scores, adjusted for height, were calculated for total hip, femoral neck, spine and distal radius. PC analysis generated new integrated aBMD phenotypes by transforming the four correlated bone Z-scores into linearly uncorrelated variables (PC loading scores). Linear mixed effects models, adjusted for age, pubertal status, BMI-Z, dietary calcium and physical activity, were used to test associations between genetic score (percentage aBMD-lowering alleles carried at 63 GWAS-implicated loci) and loading scores. GWAS was applied to baseline data to identify loci associated with loading scores. **Results** Four principal components (PC1-PC4) were identified that explained 68.1%, 18.6%, 10.5%, and 2.8% of the variance, respectively. Higher PC1 loading score indicated higher bone Z-scores across all four sites. Genetic score was associated with lower PC1 loading score (β=-0.05, P=3.9x10^{-10}); from GWAS, rs114260199 (*LMO2/CAPRIN1*, P=3.9x10^{-8}) and rs75321045 (*ZMAT4*, P=2.5x10^{-8}, females) were associated with PC1. Higher PC2 loading score indicated higher distal radius-Z only. Genetic score was not associated with PC2; from GWAS, rs67991850 (*CPED1, P$=2.5x10^{-11}*) was associated with PC2. Higher PC3 loading score indicated higher spine-Z only. Genetic score was not associated with PC3; from GWAS, rs58649746 (*RAB11FIP5, P$=4.8x10^{-9}, females) was associated with PC3. **Conclusion** We identified non-site-specific (PC1), distal radius-specific (PC2) and spine-specific phenotypes (PC3) in children. An established genetic score associated with the non-site-specific phenotype only. Novel variants near *LMO2/CAPRIN1*, *ZMAT4,* and *RAB11FIP5* associated with non-site specific or spine specific phenotypes. These genetic variants may ultimately provide insight into childhood bone accrual.

2812W

Penalized regression for detecting rare variant effects under extreme phenotype sampling for continuous traits. *C. Xu[1,2], J. Fang[1,3], H. Shen[1,2], YP. Wang[1,3], HW. Deng[1,2].* 1) Center of Genomics and Bioinformatics, Tulane University, New Orleans, LA; 2) Department of Global Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 3) Department of Biomedical Engineering, Tulane University, New Orleans, LA.

Rare variant (RV) association study is a method to identify SNPs with minor allele frequency < 0.01, which may contribute to specific disease. The method is expected to play an important role in explaining the missing heritability in GWAS and uncovering novel gene-phenotype relationships. However, the detection and association power for RV is low in random samples. Several recent studies have shown that extreme phenotype sampling (EPS) is beneficial for the RV detection studies. By enriching the causal variants in the extreme phenotypic samples within the top and bottom percentiles, EPS can boost the study power compared to random sampling with the same sample size. Currently, there are two types of statistical methods for RV study in EPS: case-control method, which ignores the quantitative data of the extreme samples and tests the genes/regions individually in case/control groups; and likelihood method, which models the non-normality for the quantitative data but still tests the association individually. However, many disorders are caused by multiple genetic factors. Therefore, it is desirable to simultaneously model the joint effects of genetic markers, which may increase the power of current genetic association studies and identify novel disease-associated genetic markers under EPS. The challenge of the simultaneous analysis is: the number of genetic markers (e.g., p~10,000) is typically greater than the sample size (e.g., n~1,000) in a study. The standard linear model would be inappropriate for this p>n problem due to the rank deficiency of the design matrix. An alternative solution is to apply a penalized regression method – the least absolute shrinkage and selection operator (LASSO). The application of LASSO to genetic association study under random sampling has been widely studied. Here, we investigated a penalized regression model (LASSO-HT) by combining LASSO hypothesis testing with rare variant collapsing to jointly detect genes/regions carrying RV effects under EPS. Our comprehensive simulation study indicates LASSO-HT outperforms other state-of-the-art methods with the correct type I error and relative low FDR. Both simulation and real data analysis of bone mineral density show the power of LASSO-HT is comparable with, even better than, under certain conditions, existing methods. The comparison of area under receiver operating characteristic curve demonstrates the overall performance of LASSO-HT exceeds or tightly matches others for most cases.

2813T

Alzheimer's Disease Sequencing Project: Case-control analyses of over 10,000 whole exomes. *J.C. Bis[1], X. Jian[2], B. Kunkle[3], K. Hamilton[3], W. Bush[4], W. Salerno[5], D. Lancour[5], Y. Ma[6], Y. Chen[6], A. Destefano[6], J. Dupuis[6,7], J. Farrell[6], Y. Zhao[8], L. Qu[8], C. Bellenguez[9,10,11], J.C. Lambert[9,10,11], S. van der Lee[12], A. Renton[13], E. Marcora[13], M. Kurki[14], A. Palotie[14], M. Daly[14], C. van Duijn[12], E. Boerwinkle[5,15], E. Martin[3], G. Schellenberg[8], S. Seshadri[6,7], A. Naj[8], M. Fornage[2], L. Farrer[6] on behalf of the Alzheimer's Disease Sequencing Project Case-Control Work Group.* 1) University of Washington, Seattle, WA; 2) University of Texas Health Science Center at Houston, Houston, TX; 3) University of Miami, Miami, FL; 4) Case Western Reserve University, Cleveland, OH; 5) Baylor College of Medicine, Houston, TX; 6) Boston University, Boston, MA; 7) The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA; 8) University of Pennsylvania, Philadelphia, PA; 9) Université Lille 2, Lille,France; 10) Institut Pasteur de Lille, Lille, France; 11) Inserm, Lille, France; 12) Erasmus University Medical Center, Rotterdam, the Netherlands; 13) Icahn School of Medicine at Mount Sinai, New York, NY; 14) Broad Institute, Cambridge, MA; 15) University of Texas School of Public Health, Houston, TX.

   Background: Responding to a 2012 Presidential Initiative to fight Alzheimer's disease (AD), the National Institute on Aging (NIA) and National Human Genome Research Institute (NHGRI) jointly developed a project to analyze the genomes of well-characterized individuals with or without clinical AD using NINCDS-ADRDA criteria. This effort, the Alzheimer's Disease Sequencing Project (ADSP) has the overarching goals of identifying new (1) genes involved in AD, (2) genomic variation contributing to AD risk or protection, and (3) potential avenues for therapies and AD prevention. This presentation highlights findings from the ADSP Discovery Phase Case-Control Study, which analyzed whole exome sequence (WES) variation from unrelated participants of the Alzheimer's Disease Genetics Consortium (ADGC) and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. **Methods:** We selected European-ancestry cases and dementia-free controls based on age, sex, *APOE* genotype, and pathology, along with additional unrelated cases from AD families and Caribbean Hispanic controls to enrich the risk profile and diversity of the study. We analyzed individual common variants (minor allele count > 10) using logistic regression and aggregated rare (minor allele frequency < 5%) variants into genes using SKAT-O with filters based on predicted functional impact or loss-of-function. All models included adjustment for sequencing center and population structure. Sensitivity analyses adjusted for age, sex, and *APOE* genotype. Analyses were performed separately by ancestry and were combined using seqMeta/R. Statistical significance was established using a Bonferroni correction. **Results:** After QC, the analysis included 5,740 cases and 5,096 controls (218 cases and 177 controls were Hispanic) and approximately 1.5 million SNVs and 50,000 indels. Significant associations from single variant and gene-based tests included common and rare variants in several AD risk genes (*TREM2*, *MS4A6A*, *ABCA7*, *PILRA*), as well as potentially novel findings in *GAS2L2*, *OPRL1*, *ZNF655*, *NSF* and a long non-coding RNA. Some findings show preliminary evidence of replication in ongoing efforts among independent WES and imputed datasets. **Conclusions:** WES association analyses of ADSP case-control data replicated several known AD genes and identified novel associations; efforts to replicate these findings and characterize the biological mechanisms of the identified genes in AD are currently underway.

2814F

Tissue-specific genetic regulated expression in late-onset Alzheimer's disease: The Alzheimer's Disease Genetics Consortium (ADGC). *H.-H. Chen[1], the ADGC[2], Y. Zhao[3], S. van der Lee[4], K. Hamilton-Nelson[5], L. Petty[1], B. Kunkle[5], A. Partch[3], O. Valladares[3], C. Reitz[6], G. Beecham[5], E. Martin[5], L-S. Wang[3], J. Haines[7], R. Mayeux[6], L. Farrer[8], M. Pericak-Vance[5], G. Schellenberg[3], W. Bush[7], A. Naj[3], J.E. Below[1].* 1) School of Medicine, Vanderbilt University , Nashville, TN, USA; 2) the Alzheimer's Disease Genetics Consortium; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) Erasmus University Medical Centre, Rotterdam, Netherlands; 5) Miller School of Medicine, University of Miami, Miami, FL, USA; 6) Columbia University, New York, NY, USA; 7) School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 8) School of Medicine, Boston University, Boston, MA, USA.

   As average life expectancy has increased, so has the global prevalence of Alzheimer's disease (AD) which is the leading cause of dementia in the United States. AD is highly heritable (estimated $h^2=0.4$-0.94), with 19 confirmed common susceptibility loci, and multiple rare loci including TREM2 recently identified. Little is known about how these and other genetic factors influence expression patterns, which may influence AD etiology. To improve understanding of the functional genetic etiology of AD we collected 25 GWAS datasets including 13,042 cases and 13,674 controls from the Alzheimer's Disease Genetics Consortium, imputed to ~39M variants as rare as MAF=0.0004 with the HRC r1.1 reference panel using Minimac3. High quality variants with MAF>0.05 and imputation R^2>0.8 were used to predict gene expression in 45 different tissues using PrediXcan, an approach that infers the genetically regulated portion of gene expression from common expression quantitative trait loci identified in the Genotype-Tissue Expression Project (GTEx). Logistic regression was applied to test the association between each tissue-specific predicted gene expression level and AD with adjustment for sex age, and ancestry-based principal components. Meta-analysis across all studies was performed using METAL and identified 47 unique genes across all tissues (17 genes showed cross-tissue effects). These signals reached Bonferroni-corrected genome-wide significance ($P<2.5×10^{-6}$) and include APOC1 (β =-1.931, p-value=$2.3×10^{-61}$), GEMIN7 (β =0.91, p-value=$1.0×10^{-32}$), and AMMECR1L (β =3.78, p-value=$2.1×10^{-19}$). We further adjusted models using nearby previously-reported sentinel risk variants for AD. Across all tissues, 44 unique genes reached genome-wide significance after correction for known signals, and 16 of them located outside of known APOE region. In conclusion, multiple novel and known genes show predicted expression that associates with AD in ADGC, illustrating that gene regulatory models may be useful in the discovery of new AD loci.

2815W

Convergent evidence for *LRP2BP* in resilience to Alzheimer's disease.
D. Felsky[1,2], R. Cheng[3], J.A. Schneider[4,5], D.A. Bennett[4,5], J.H. Lee[3], P.L. De Jager[1,2]. 1) Center for Translational and Systems Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Sergievsky Center and Taub Institute, Columbia University Medical Center, New York, NY; 4) Department of Neurological Sciences, Rush University Medical Center, Chicago, IL; 5) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL.

Early-onset familial Alzheimer's disease (eFAD) commonly results from mutations in *PSEN1* and is characterized by onset before age 65. To date, the contribution of common genetic variation to the clinical and pathological trajectories of those with *PSEN1* mutations is not well known. We implemented an integrative approach to identifying and characterizing the molecular-genetic modifiers of cognitive decline in *PSEN1* eFAD families. First, genetic linkage and association analyses in 45 *PSEN1*-G206A mutation carrier families with total 305 members found 12 candidate loci linked and associated with AD and age at onset of AD. These loci were carried forward for multi-omics analyses in two large aging cohorts: The Religious Orders Study and Memory and Aging Project (ROS/MAP). Cognition was measured longitudinally before death and autopsy was performed postmortem for AD neuropathology. Of 1,291 subjects with postmortem data, 1,116 had genome-wide genotype data, and 599 had RNA-sequencing performed on postmortem brain. Variants within 100kb of each candidate gene were evaluated against pathology and cognitive decline. Multiple SNPs tagging four haplotypes near *SNX25* were associated with cognitive decline and AD pathology at nominal significance ($1>p>1.3\times10^{-3}$). Interestingly, the top statistical signal for this region fell in proximity to the transcription start site of the adjacent gene *LRP2BP* rather than within *SNX25* itself. After FDR correction, RNA expression of *LRP2BP* but not *SNX25* was significantly associated with neuritic plaque ($p_{FDR}=1.3\times10^{-4}$), PHF-tau ($p_{FDR}=0.0019$), and neurofibrillary tangle ($p_{FDR}=4.6\times10^{-5}$) burdens. EQTL analyses were then performed in ROS/MAP and validated using GTEx. In ROS/MAP, the top pathology-associated *LRP2BP* variant was strongly associated with *LRP2BP* expression after correction ($p=8.6\times10^{-14}$, $p_{FDR}<0.0001$). In GTEx, this same variant is an eQTL for *LRP2BP* in subcutaneous adipose ($p=3.0\times10^{-5}$), tibial nerve ($p=3.3\times10^{-6}$), and thyroid ($p=1.6\times10^{-6}$). Consistently, the same allele associated with higher *LRP2BP* expression was associated with lower levels of neuropathology, and higher *LRP2BP* expression was associated with lower levels of neuropathology and slower cognitive decline. LRP2BP binds LRP2, which is important for axonal regeneration and clearance of β-amyloid across the blood brain barrier. Our results suggest that genetically-determined levels of *LRP2BP* modulate pathological load in eFAD and may be a novel target for intervention.

2816T

Cell free single stranded DNA concentration in CSF as biomarker to diagnose Alzheimer's disease status.
J.D. Gonzalez Murcia[1], L.A. Stanley[1], M. Ferrel[1], M.T.W. Ebbert[1,2], H. Zetterberg[3], J.S.K. Kauwe[1]. 1) Biology, Brigham Young University, Provo, UT; 2) Mayo Clinic, Jacksonville; 3) University College London, London.

Alzheimer's disease (AD) is a fatal neurodegenerative condition and is the leading cause of dementia in the elderly (Jonsson et al., 2013). AD is identified as a proteopathic disease due to the extensive accumulation of amyloid plaques and neurofibrillary tangles (Guerreiro et al., 2013; Jonsson et al.,2013). Currently there is not adequate treatment or prevention. It has been postulated that earlier and more accurate diagnosis of AD could be a key to better therapeutic development. The aim of this study is to evaluate if the concentration of cell free single stranded DNA (cfssDNA) is correlated with the diagnosis of AD. The study was performed by obtaining 30 CSF non-demented control samples from Sweden and 30 CSF samples of AD cases from Sweden. All samples were clinically ascertained following established protocols for AD diagnosis. We proceed to extract single stranded DNA from all CSF samples and measured using Qubit assay technology. During this process we screened for dsDNA, RNA, and ssDNA. We then proceeded to analyze measurements using a T test pair analysis, a nominal regression analysis, and then a Receiver Operating Characteristic (ROC) curve. Result from Qubit assay shows detectable concentration only for ssDNA. In addition, we observed from paired T-test analysis, that cfssDNA concentration in AD samples is significantly lower than non-demented control samples. (2 tail p-value 0.0011). Nominal regression analysis shows that cfssDNA concentration can predict AD status and the power of prediction increases when age included in the model (p-value <0.0001). Finally, ROC curve has an area under the curve of 0.82. We conclude that cell free single stranded DNA concentration in CSF is significantly lower in AD cases than non-demented controls among Swedish samples. Because of this finding cell free single stranded DNA can be added as a possible biomarker for AD.

2817F

Genome-wide search for genetic loci perturbing gene co-expression networks in Alzheimer's disease. *L. He[1,2], Y. Park[1,2], L. Hou[1,2], D. Jose[1,2], M. Kellis[1,2].* 1) MIT, Cambridge, MA; 2) Broad institute, Cambridge, MA.

Genome-wide association studies (GWAS) have successfully identified thousands of single nucleotide polymorphisms (SNPs) associated with human complex diseases. However, the underlying biological and pathological mechanisms of the majority of these SNPs are still elusive. Gene co-expression network structure is an important intermediate phenotype that has the potential to improve our understanding of bridging the relationship of genetic variations and complex diseases. Expression quantitative trait loci (eQTLs) have shown local genetic effects on differential expression of single genes; however, it is yet to be examined how SNPs globally affect disease-related gene co-expression networks. In this work, we develop statistical methods to genome-widely identify SNPs that significantly perturb the co-expression structure of a gene module of interest, termed as netQTLs. We applied the proposed method to genotype and gene expression datasets from the ROS/MAP project to detect which SNPs perturb Alzheimer's disease (AD)-related gene modules and which modules are perturbed by AD-related GWAS loci. In addition to the confirmation of well-controlled type I error rate, we identified novel SNPs affecting the co-expression structure of AD-related gene modules. We found that many SNPs have large effects on the overall module structure, but are not necessarily show significant marginal association with single gene expression. We also identified multiple gene modules in immune systems and circadian clock pathways that are significantly associated with both AD and AD-related GWAS loci. Our results suggest that other than eQTLs, the investigation of genetic effects on the overall module perturbation using netQTLs may provide additional insights into the relationship between genetics and gene expression.

2818W

Genome-wide rare variant imputation and tissue-specific transcriptomic analysis identify novel rare variant candidate loci in late-onset Alzheimer's Disease: The Alzheimer's Disease Genetics Consortium (ADGC). *A.C. Naj[1], J.E. Below[2], Y. Zhao[1], H.-H. Chen[2], S.J. van der Lee[3], K.L. Hamilton-Nelson[4], L. Petty[2], B.W. Kunkle[4], A. Kuzma[1], O. Valladares[1], C. Reitz[5], G.W. Beecham[4], E.R. Martin[4], L.-S. Wang[1], J.L. Haines[6], R. Mayeux[5], L.A. Farrer[7], M.A. Pericak-Vance[4], G.D. Schellenberg[1], Alzheimer s Disease Genetics Consortium.* 1) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 2) Vanderbilt University Medical Center, Nashville, TN, USA; 3) Erasmus University Medical Centre, Rotterdam, Netherlands; 4) University of Miami, Miami, FL, USA; 5) Columbia University, New York, NY, USA; 6) School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 7) School of Medicine, Boston University, Boston, MA, USA.

The International Genomics of Alzheimer's Project (IGAP) GWAS identified 19 susceptibility LOAD loci in addition to *APOE*, however the majority of these were common (minor allele frequency (MAF)>0.05). The Haplotype Reference Consortium (HRC) released a dense reference panel (64,976 haplotypes/39,235,157 SNPs) allowing imputation of rare variants (MAF>0.00008). ADGC imputed 33 GWAS datasets to HRC to identify novel rare variant associations and genetically-regulated gene expression patterns contributing to LOAD. We imputed 14,743 cases and 15,871 controls to the HRC r1.1 reference panel using Minimac3 on the University of Michigan Imputation Server. Logistic regression on individual variants with MAF>0.01 was performed in PLINKv1.9 (GLMM/R for family-based variants) using imputed genotype probabilities and meta-analyzed in METAL, while variants with MAF≤0.01 were analyzed using score-based tests and meta-analyzed in the SeqMeta/R package; both analyses adjusted for age, sex, and population substructure. Gene-based association was performed using SKAT-O and gene-based testing of expression regulation in LOAD was performed using PrediXcan. Preliminary analyses of ~39.2 million genotyped or imputed SNVs identified single variant associations of $P<5\times10^{-8}$ in 5 known IGAP LOAD candidate loci (*APOE*, *BIN1*, the *MS4A* region, *PICALM*, and *CR1*), and multiple suggestive associations ($P<10^{-5}$) at each of 13 additional loci:Â rs13155750 in *MEF2C* (OR(95% CI): 1.13(1.08,1.20); $P=5.09\times10^{-6}$) and rs755951 in *PTK2B* (OR(95% CI): 1.11(1.06,1.16); $P=5.61\times10^{-6}$) [known IGAP locus]. Novel associations include signals at *LILRA5* (19:54821819; OR(95% CI): 1.14(1.08,1.20); $P=4.84\times10^{-7}$), involved in innate immunity pathways; and at *SMOX* (rs1884732; OR(95% CI): 1.11(1.06,1.17); $P=5.17\times10^{-6}$), involved in catabolism of polyamines, levels of which are altered in AD brains. Rare variant and gene-based analyses demonstrated significant *APOE* and *TREM2* associations, while gene-based testing identified strong but marginal association for *SORL1* ($P=5.55\times10^{-6}$). PrediXcan analyses identified significant strong genetically-regulated gene expression in LOAD for *MS4A4A* ($Q=4.26\times10^{-26}$), *BIN1* ($Q=1.70\times10^{-18}$), and *FBXO46* ($Q=4.25\times10^{-8}$). Several novel candidate loci for LOAD have been identified using high-quality imputation of low-frequency variants in ADGC, reinforcing the utility of high-density imputation panels, and providing a resource to newly identify genes with perturbed expression in LOAD.

2819T

Complex disease prediction: A framework to integrate SNP and imaging data. *B. Zhao[1], J. Yang[1], J. Zhang[1], Y. Shan[1], C. Huang[1], T. Li[2], J. Ibrahim[1], H. Zhu[2].* 1) Department of Biostatistics , University of North Carolina at Chapel Hill , Chapel Hill, NC; 2) Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX.

   Introduction: In many recent complex disease studies, both whole genome single nucleotide polymorphism (SNP) data and imaging data have been used to help identify subjects at risk and predict disease progression. It is of a great practical interest to explore how to integrate the genetics and imaging information into one prediction model. **Methods and Results**: We propose a flexible framework to predict disease using both SNPs and brain volumetric data. We assess the performance of framework using more than 1100 subjects with Alzheimer's disease (AD), mild cognitive impairment (MCI), as well as healthy controls (NC) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We compare the integrating model with models only with either SNPs or brain volumes via cross validation. Both binary outcome (AD, NC) and three-level outcome (AD, MCI, NC) cases are considered and several types of prediction models are applied, including generalized linear model, random forest, support vector machine, and Xgboost. We show that additional prediction accuracy can be achieved through integration, but the improvement might be limited. As an illustration, we use the ADNI-1 dataset to predict the outcomes of subjects in the ADNI-GO/2. The integration can improve the prediction accuracy from 80% to 82% in binary outcome prediction, and from 54% to 61% in three-level outcome prediction. In addition, we illustrate that unobserved batch effects may exist in multi-phase multi-site brain volumetric datasets. We propose a tool to detect and control these hidden effects when testing and estimating the primary variable(s) (such as disease status, polygenic risk score). In ADNI study, the brain regions frequently reported to be associated with Alzheimer's disease (Left/Right Hippocampus, Left/Right Entorhinal, and Left/Right Amygdala) can clearly stand out from other regions after adjusting for the unobserved batch effects. **Conclusion:** We have proposed a framework to predict complex disease by integrating both SNPs and brain volumetric data. It can be widely used for disease prediction in large-scale neuroimaging and genetics studies.

2820F

Mitochondrial variants associated with increased risk of late-onset Alzheimer's disease. *T.J. Zhou[1], X. Zhang[1,2], J.J. Farrell[2], J. Chung[2], A.C. Naj[3], K.L. Lunetta[1], L.A. Farrer[1,2], Alzheimer s Disease Genetics Consortium (ADGC).* 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 2) Section of Biomedical Genetics, Boston University School of Medicine, Boston, MA, USA; 3) Department of Biostatistics, Epidemiology, and Informatics; Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.

   Background: Alzheimer's disease (AD) is a neurodegenerative disease with complex, multifactorial etiology. Mitochondrial (MT) dysfunction has been implicated in AD. However, the association between the total array of MT variants and AD has not been fully explored in large datasets. **Methods:** The Alzheimer's Disease Genetics Consortium (ADGC) recently genotyped 8,706 AD cases and 7,002 healthy controls of European ancestry from five cohorts using the Illumina Exome-Chip 1.0 which contains 226 MT variants. Each cohort was analyzed separately for variants with minor allele count (MAC) $\geq$ 10 and call rate $\geq$ 0.95 (94 variants on average), and the results were meta-analyzed using the seqMeta/R package. In gene-based tests, only single nucleotide variants (SNVs) with minor allele frequency (MAF) < 0.05 were considered, and genes with $\geq$ 2 SNVs and cumulative MAC $\geq$ 10 were tested. Associations with AD status were tested using single variant (GLM) and gene-based (SKAT-O) tests in seqMeta, controlling for age, sex, and principal components of ancestry. Bonferroni-corrected thresholds were used to determine statistical significance for single variant ($\alpha = 8.6 \times 10^{-4}$) and gene-based ($\alpha = 4.5 \times 10^{-3}$) associations. **Results:** The single variant analyses resulted in only one significant association between increased AD risk and a missense variant (rs2853826/Thr114Ala; OR = 1.043; P = 2.5×10^{-4}; MAF = 0.22) in the gene *MT-ND3*, which encodes NADH dehydrogenase 3 involved in oxidative phosphorylation of ATP. No genes passed the Bonferroni-adjusted significance level in the gene-based analysis, however *MT-TP* yielded the most significant association (P = 7.9×10^{-3}). **Conclusion:** This study identified one common MT variant significantly associated with an increased risk of AD. A replication study of independent cohorts from the International Genomics of Alzheimer's Project (including datasets from the Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer's Disease Initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium) is currently underway. Additionally, a pathway-based aggregation test is also underway to elucidate potential biological mechanisms.

2821W

Genetic association study on white matter microstructure by integrating multiple neuroimaging datasets. *J. Zhang[1], T. Li[2], C. Huang[1], J. Yang[1], Y. Wang[1], J.G. Ibrahim[1], H. Zhu[1,2].* 1) Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX.

   Introduction: In recent decades, a number of large-scale neuroimaging cohort studies have been launched to improve our understanding of the structure and function of human brain. Studying the environmental and genetic influences is an essential topic to unravel the biological basis for brain development and degeneration, which could eventually inspire novel inventions in neuro-psychiatric disease preventions, diagnoses and clinical treatments.   **Method:** In this study, we conducted a genome-wide association analysis of diffusion tensor image using the tract-based spatial statistics (TBSS) of 17 brain regions by merging data from two studies of comparable age group, the Philadelphia Neurodevelopmental Cohort (PNC) study (8-21yrs) and the Pediatric Imaging Neurocognition and genetics study (3-21 yrs). The integration allows us to improve the statistical power to detect genetic variants with relatively weak effect. Previous analysis has confirmed the consistency of the TBSS data structure as well as the covariate effect across studies. Factor analysis is performed on the merged TBSS data and one common factor is extracted as outcome. Subsequently, GWAS analysis and gene-level analysis are performed on the common factor for European ancestry and African American ancestry respectively. **Results:** No single variant reached the genome-wide significance level in European ancestry. In gene-level analysis, the Tetraspanin2 (TSPAN2) gene remains significant after adjusting for multiple comparisons, which is known to be involved in the process of brain development. For African American ancestry, 6 LD independent loci exceeded the conventional GWAS p-value threshold of 5E-08. The top variant rs2624415 (chr5:17221636) is of particular interest, as it is located within the BASP1 gene, a brain abundant membrane attached signal protein coding gene that plays an important role in regulating the neurite growth and the maturation of the neural synapses.

2822T

Investigating the underlying genetic basis of the co-occurence of epilepsy and psychiatric disorders. *H.O. Heyne[1,2], F.K. Satterstrom[1,2], J. Kosmicki[1,2], E. Wigdor[1,2], A.D. Børglum[3,4,5], T.M. Werge[3,6,7], P.B. Mortensen[3,4,8], E. Robinson[1,2], D. Lal[1,2], M.J. Daly[1,2], iPSYCH consortium.* 1) Analytical and Translational Genetics Unit, Massachusetts General Hospital , Boston, MA, US; 2) Broad Institute, Stanley Center for Psychiatric Research, Cambridge, MA, US; 3) iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; 4) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 5) Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark; 6) Mental Health Centre Sct. Hans, Institute for Biological Psychiatry, Capital Region of Denmark, Roskilde, Denmark; 7) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 8) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark.

   Individuals with neurodevelopmental disorders (NDDs), such as intellectual disability and autism are frequently comorbid with epilepsy. Vice versa, individuals with epilepsy display more neurodevelopmental or psychiatric comorbidities than individuals from the general population. Here, we investigated, whether this phenotypic co-occurrence is reflected by shared rare and common genetic variants. We screened for genes enriched for de novo variants (DNV) in comparison to expectation by a mutational model (Samocha et al., 2015) in ca. 10,000 individuals ascertained for having an NDD of developmental delay, autism, intellectual disability or epileptic encephalopathy. In a subset of ca. 1,300 individuals with NDDs comorbid but not ascertained for epilepsy, the spectrum of de novo variants was strongly enriched for genes associated with epilepsy syndromes. A lower but significant de novo burden in known epilepsy genes was also present in NDDs without epilepsy but not observed in individuals without NDDs. We confirmed similar patterns with ultra-rare variant burden analyses in the Danish population iPSYCH study - comparing ca. 11,000 individuals diagnosed with autism, ADHD, schizophrenia or bipolar/affective disorder to ca. 5,200 matched controls. Epilepsy but also intellectual disability independently predicted the presence of damaging ultra rare variants in genes associated with epilepsy. Next, we moved from rare to common variants and investigated whether individuals with autism overinherited common variants (pTDT) associated with epilepsy from their parents (n= 4003). Finally, we further describe the phenotypic spectrum of individuals carrying genes shared between epilepsy and other neuropsychiatric disorders.

2823F

Small posterior fossa in Chiari malformation affected families is significantly linked to 1q43-44 and 12q23. *A. Musolf[1], S.C.W. Ho[2], K.A. Long[1], P. Zhuang[2,3], H. Sun[1,4], B.A. Moiz[1], E.G. Mendelevich[5], E.I. Bogdanov[5], J.E. Bailey-Wilson[1], J.D. Heiss[2].* 1) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD; 3) Neuro-Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD; 4) Laboratory of Medical Genetics, Harbin Medical University, Harbin, China; 5) Department of Neurology & Rehab, Kazan State Medical University, Kazan, Russia.

The posterior fossa is a cranial cavity at the base of the skull. When it is small, it causes the cerebellum and brainstem to be pushed downward, resulting in a Chiari malformation. Chiari malformations cause neck pain, balance issues, decreased motor skills and headaches in those affected. We have posterior fossa measurements and whole exome sequence data on individuals from 9 extended families from the United States and Russia that have a family history of Chiari malformations. We performed parametric linkage analyses using an autosomal dominant inheritance model with a disease allele frequency of 0.01 and penetrance for small posterior fossa of 0.8 for carriers and 0.1 for non-carriers. Single variant two-point linkage analysis and collapsed haplotype pattern (CHP) two-point linkage analysis were performed. CHP two-point linkage analysis used rare variants (MAF <= 0.05) to create multi-allelic pseudo-markers that correspond to a gene or section of a gene. This gene-based test can improve power in the presence of allelic heterogeneity. Our results found a genome-wide significant linkage on chromosome 1q43-44 (HLOD = 3.5) and 12q23 (HLOD = 3.3) in both sets of linkage analyses. Most interesting was that both signals were driven by a single (different) family. Both regions contain several linked exonic variants including rare variants located in good candidate genes. In conclusion, we have located two significantly linked regions for small posterior fossa that are driven by linked variants in 1 family each and are potentially causal. Further laboratory work is needed to confirm these candidate genes.

2824W

Genetics of age-related cognitive decline and relationship to Alzheimer's and other neurodegenerative diseases. *M.P. Reeve[1,4], M. Alanne-Kinnunen[4], H. Mattson[4], H. Laivuori[4], S. Koskinen[9], T. Paajanen[8], K. Estrada[5], J. Hutz[1], A. Mitchell[1], D. Diogo[7], C. Whelan[6], M. Perola[9], A. Palotie[4,2,10], M. Daly[2,3], S. Ripatti[4].* 1) Eisai, 4 Corporate Drive, Andover, MA 01810; 2) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Richard B. Simches Research Building, 185 Cambridge St, CPZN-6818, Boston, MA 02114; 3) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, P.O.Box 20, FI-00014 University of Helsinki, Finland; 5) Biogen, 225 Binney St, Cambridge, MA 02142; 6) Pfizer, 1 Portland St, Cambridge, MA 02139; 7) Merck, 33 Avenue Louis Pasteur, Boston, MA 02115; 8) Finnish Institute of Occupational Health Helsinki, Uusimaa, Finland; 9) National Institute for Health and Welfare (THL),Mannerheimintie 166, Helsinki, Finland; 10) The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

GWAS studies of Alzheimer's disease (AD) and other neurodegenerative diseases have identified numerous genetic associations, but as yet the biological pathways leading to disease are not well understood, nor is the impact of these variants throughout life.   We aimed to study how genetic polymorphisms associated with AD might be related to general age-related cognitive decline (for instance, whether the APP variant A673T is protective not only for AD, but also general cognition) or whether they speak specifically to the pathology of AD and potentially related dementias. We sought to address these and other questions using a Finnish population-based cohort (Health 2000) of more than 7703 middle-aged adults (mean age 55) who were given a battery of health and cognitive tests in 2000 and then again in 2011 – enabling unique opportunities to examine the genetics of age-related cognitive decline in the general population. We performed genome-wide association on the rate of change of cognitive function as assessed by CERAD and other standard tools, and aligned these data to the individual hits and polygenic risk scores from the largest published AD meta-analysis to date (Lambert *et al.*, 2013). By integrating these data with Finnish Health Registry data, we identified more than 500 individuals who were assessed in the year 2000 or later who received a diagnosis of AD or dementia, in addition to a total of nearly 1400 AD patients who had died before 2014. These data enabled the assessment of which cognitive tests were most correlated to future AD or dementia risk, and permitted the creation of a joint predictor (using proportional hazards model) of AD risk that includes age, specific CERAD results, and polygenic risk scores – providing a substantially improved risk predictor which derives power through an integration of pre-existing genetic risk and current cognitive status. Specific cognitive test results (e.g. word list recall) were highly significant predictors of AD and dementia ($p < 5e-7$) independent of APOE genotype.   Our findings add substantial value to the definition of cohorts at high risk for AD. This approach is particularly promising in Finland, where in addition to rich genetic data, there exists national registries for health care, hospitalization, prescription medications, social care and deaths, and these registries are available for research and cross-linking using unique electronic identifiers to create custom endpoints.

2825T

Gene-gene interaction tests for genetic-imaging data analysis. *W. Peng[1], N. Lin[2], X. Wu[1], M. Xiong[2].* 1) School of Public Health, Bengbu Medical College, Bengbu, Anhui Province, China, Anhui, China; 2) School of Public Health, University of Texas Health Science Center at Houston.

Investigating the gene-gene and gene-environment interaction effects in genetic-imaging data analysis provide important information to achieve deep understanding of the complex genetic structures of neuroimaging. However, due to high dimension of imaging data, very few statistical methods have been developed for testing gene-gene and gene-environment interaction in the genetic-imaging data analysis. A key issue for high dimensional gene-gene and gene-environment interaction analysis is to effectively extract informative internal representation and features from high dimensional genotype and imaging data. To explore correlation information of imaging signal and genetic variants, effectively reduce data dimensions, and overcome critical barriers in advancing the development of novel statistical methods and computational algorithms for gene-gene and gene-environment interaction tests in genetic-imaging data analysis, we proposed a new statistic method referred to as a quadratically regularized functional CCA (QRFCCA) for gene-gene interaction analysis which combines three approaches: (1) quadratically regularized matrix factorization, (2) functional data analysis and (3) canonical correlation analysis (CCA). We extend CCA from association analysis to gene-gene and gene-environment interaction analysis and propose to use QRFCCA as a unified framework for testing gene-gene and gene-environment interaction analysis with next-generation sequencing and imaging data. To evaluate performance, the QRFCCA is applied to the Neurodevelopmental Genomics studies of Complex Phenotypes in which 8, 841,087 rare variants were collected from 854 individuals with DTI images. We identified 879 pair-wise interactions between genes.

2826F

Low-rank structure based brain connectivity GWAS study. *Z. Zhu[1], F. Zhou[1], L. Yang[2], Y. Shan[1], J. Zhang[1], J. Ibrahim[1], H. Zhu[1,3].* 1) Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX.

Introduction: The genetic basis of brain connectivity, which reveals brain structure and brain functions, has not been well studied. It is thought that brain connectivity is affected by only a few seeds, thus yielding a low-rank structure. Existing methods for dealing with brain connectivity data usually ignore the low-rank structure, and will incur biased results for genetic analysis and the high dimension of the original connectivity will bring much burden for statistical computing. **Methods and Results:** In this paper, we propose a new method for doing connectivity GWAS based on spectral clustering to detect the low-rank structure of brain connectivity and also overcome the drawback of the high dimensionality. In the first step, we perform a spectual clustering algorithm to detect the low rank structure of brain connectivity, and thus extracting only a few features for analysis. The second step is to perform multidimensional phenotype GWAS analysis on the features extracted in the first step. In simulation studies, we show that our method will produce more accurate results compared to using the original correlation matrices. Also, our method also outperforms other existing methods in true positive rate. We applied the method to PNC and PING datasets, which contains both typical brain imaging and genetic information. From the results, using different number of features to be selected, we identified SNPs on 16 different genes in PNC dataset, and most of the significant SNPs are on gene MOV10L1 on chromosome 22 and gene CTB.30L5.1 on chromosome 7. For PING data, we identified SNPs from 15 genes that are significantly associated with brain connectivity, with the most on gene RAB39A on chromosome Our results indicates brain connectivity is significantly associated with several genes, but the association might be different in populations with different diseases. **Conclusion:** We proposed an efficient method for detecting genetic association in brain connectivity. Our method can enhance both prediction accuracy and computation efficiency by recovering the low-rank structure of brain connectivity.

2827W

Genetic causes of death in US infants: Findings from the National Center for Health Statistics. *C. Lally[1], C.C. Jones[2], S.F. Cook[3], W. Farwell[2], J. Staropoli[2], W.D. Flanders[1].* 1) Emory University, Atlanta, GA; 2) Biogen, Cambridge, MA; 3) Epidemiology Associates LLC, Chapel Hill, NC.

Background: The discovery of the genetic causes of many diseases and subsequent development of approaches to modify or ameliorate the associated morbidity and mortality of some genetic diseases warrant a re-examination and quantification of the burden of infant mortality from these causes. We examined the relative frequency of the leading genetic causes of infant mortality in the United States. Methods: We quantified infant mortality from the most common genetic causes. This included single gene disorders and chromosomal abnormalities. We compiled a comprehensive list of genetic causes of death and their associated ICD-10 codes from the National Human Genome Institute and linked these disorders to the Orphanet website to identify genetic causes with age of onset in infancy or listed as "any age". Using data from the National Center for Health Statistics we obtained a count of the number of deaths from each of the identified causes of death. Data were queried by ICD-10 codes using the Centers for Disease Control and Prevention, National Center for Health Statistics Underlying Cause of Death 2014 on the CDC WONDER Online Database. Infant deaths following a live birth in 2014 were queried for age 1-364 days by single cause of death based on the identified ICD-10 codes. Results: There were 23,215 infant deaths among 3,988,076 live births in 2014. Among these, 1,120 (4.8%) deaths were from genetic causes of death; the most frequent were: 463 (41.3%) deaths from Edwards syndrome (Trisomy 18), 246 (22%) from Patau syndrome (Trisomy 13), 95 (8.5%) from holoprosencephaly (commonly associated with three copies of chromosome 13), 70 (6.2%) from Down syndrome, and 62 (5.5%) from spinal muscular atrophy (SMA). Edwards syndrome, Patau syndrome, holoprosencephaly, and Down syndrome, all chromosomal abnormalities, are the four leading genetic causes of death. SMA is the leading cause of death from a single gene disorder. Conclusion: Chromosomal abnormalities make up the largest percentage of infant deaths due to genetic causes. When considering deaths from a single-gene disorder, SMA is the leading cause of infant death. This study was funded by Biogen. C.L., S.F.C., and W.D.F. are consultants to and C.C.J., W.F., and J.S. are employees and shareholders of Biogen.

2828T

Rare coding mutations in Alzheimer Disease. *D. Patel[1,2], J. Chung[1,2], X. Zhang[2,3], J. Haines[7], M. Pericak-Vance[8], G. Schellenberg[9], K. Lunetta[3], L. Farrer[1,2,3,4,5,6].* 1) Bioinformatics Graduate Program; 2) Departments of Medicine (Biomedical Genetics); 3) Biostatistics; 4) Neurology; 5) Ophthalmology; 6) Epidemiology, Boston University School of Medicine, Boston, MA; 7) Case Western Reserve University School of Medicine, Cleveland, OH; 8) University of Miami Miller School of Medicine, Miami, FL; 9) University of Pennsylvania, Philadelphia, PA.

Background: Much of the unexplained heritability of Alzheimer disease (AD) may be due to rare variants whose effects are not captured in most GWAS. Because very large samples are needed to observe statistically significant associations with rare variants in genome-wide approaches, we applied a strategy that focused on rare variants occurring only in cases or controls. **Methods:** The Alzheimer's Disease Sequencing Project performed whole-exome sequencing on a sample including non-Hispanic whites (5617 AD cases, 4594 controls). Minor alleles in 110 genes previously associated with AD or dementia were tabulated for rare variants occurring only in AD cases or controls. Top findings were further explored with bioinformatics analyses and protein homology modeling. **Results:** *NOTCH3* SNP rs149307620 had the largest number of rare alleles in AD cases (n=10) and no controls. Multiple *NOTCH3* mutations (not including this variant) have been associated with CADASIL, a diagnostically distinct disorder marked by severe headaches in young adulthood followed by stroke and dementia later in life. A genetic link between AD and *NOTCH3* has not been established, except for an unproven report of a distinct *NOTCH3* mutation shared by several AD-affected members of a single family. 7 subjects with the rs149307620 mutation and available clinical or autopsy data display classic symptoms of AD with progressive loss of memory, and no vascular risk factors or evidence of stroke. The mutation is found in the EGF protein domain near the JAG1-NOTCH3 binding site. Previous studies have shown JAG1-Notch signaling and subsequent hippocampal neurogenesis and astrogenesis to be regulated by BACE1, one of the most promising AD drug targets. Other previously unreported mutations observed only in AD cases include *PSEN1* SNP rs375376095 (n=5) which is 2779 bp from the known AD-associated rs63749824 SNP, and a high-impact splice site *ABCA7* SNP (rs376824416). A novel protective variant in *GLIS3*, found in 4 controls but not in cases, is 3010 bp from rs514716, a common intronic variant previously associated with CSF tau protein level. **Discussion:** We identified several novel rare variants predicted to have high impact on protein structure which may alter AD risk and be useful for future studies aimed at improving our understanding of the basis of AD and developing novel therapeutic targets. We are currently extending this strategy genome-wide and attempting to replicate findings in independent samples.

2829F

Comparison of methods for multivariate gene-based association analysis using common variants for complex disease. *J. Chung[1], GR. Jun[1,6], J. Dupuis[2], LA. Farrer[1,2,3,4,5,6].* 1) Department of Medicine (Biomedical Genetics), Boston University Medical Campus, Boston, MA; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) Department of Neurology, Boston University School of Public Health, Boston, MA; 4) Department of Ophthalmology, Boston University School of Public Health, Boston, MA; 5) Department of Epidemiology, Boston University School of Public Health, Boston, MA; 6) Neurogenetics and Integrated Genomics, Andover Innovative Medicines Institute, Eisai Inc, Andover, MA.

   Background. Complex diseases are usually associated with multiple correlated phenotypes, and the analysis of composite scores or disease status may not fully capture the complexity (or multidimensionality) of a disease. Joint analysis of multiple disease-related phenotypes in genetic tests could potentially increase power to detect SNPs (or genes) associated with a disease. Gene-based tests are designed to detect genes with multiple risk variants with weak association with a univariate trait. **Methods.** In this study, we combined three multivariate association tests (O'Brien method, TATES, and MultiPhen) with two gene-based association tests (GATES and VEGAS), and compared performance (type I error and power) of six multivariate gene-based methods using simulated data. We simulated data (N=2,000) for genetic sequence and correlated phenotypes by varying causal variant proportions and phenotype correlations for various scenarios. We applied these gene-based association methods to a GWAS dataset from the Alzheimer Disease Genetics Consortium containing three neuropathological traits related to Alzheimer disease (AD) including neuritic plaque, neurofibrillary tangles, and cerebral amyloid angiopathy measured in 3,500 autopsied brains. **Results.** Our simulation study showed that the two multivariate association tests (TATES and MultiPhen, but not O'Brien) paired with VEGAS have inflated type I error in all scenarios, while the three multivariate association tests paired with GATES have correct type I error. MultiPhen paired with GATES has higher power than competing methods if the correlations among phenotypes are low ($r^2 < 0.55$). We identified gene-level significant evidence ($P < 2.7 \times 10^{-6}$) in a region on chromosome 2 containing three contiguous genes (*TRAPPC12*, *TRAPPC12-AS1*, and *ADI1*) using O'Brien and VEGAS. The results for these three genes using other approaches (O'Brien with GATES and TATES with GATES) were less significant ($P < 7.0 \times 10^{-5}$). Gene-wide significant associations with these genes were not observed in a univariate gene-based test of each neuropathological trait. **Conclusions**. Our comparison of multivariate gene-based association methods showed noticeable differences among the methods, dependent on the various scenarios. Overall, MultiPhen with GATES performed best for most tested scenarios, but we also suggest O'Brien with VEGAS or TATES with GATES as alternative method because MultiPhen requires raw data (genotypes and phenotypes).

2830W

Relationship between essential tremor and Parkinson's Disease. *A.A. Gosch[1], J.C. Denny[2], M.F. Davis[1].* 1) Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

   Introduction: Essential tremor (ET) is a neurological disorder characterized by symmetrical tremor of the upper limbs without other neurological defects. Parkinson's disease (PD) is also characterized by tremor, with a key differences: ET tremor is mostly seen during action, while PD tremor is more prevalent while at rest. PD is also characterized by problems with walking and balance, while tremor is the primary outcome of ET, without balance problems. Lewy body dementia is an umbrella disorder which covers both Parkinson's disease dementia and dementia with Lewy bodies (DLB). The genetic relationship between these three disorders has been investigated, but remains unclear. Cross-sectional studies analyzing the prevalence of ET in a subset of PD patients suggest a prevalence from 5.9 to 7.1%. The prevalence of ET has been compared in other diseases, like migraines, in large scale datasets but the prevalence of PD development in ET patients in a prospective large scale analysis has yet to be performed. In this study, we compared the prevalence of PD development in ET patients to age matched population controls obtained from the Vanderbilt University Synthetic Derivative (SD) database to better understand the risk of developing PD in patients with a previous diagnosis of ET. Methods: We identified ET patients from the Vanderbilt University Medical Center's Synthetic Derivative database to examine the prevalence of PD in ET patients compared to the prevalence of PD in age and gender-matched controls. We also compared the prevalence of DLB. Criteria for diagnosis of ET for a patient was determined by the presence of an ICD9 or ICD10 billing code in their medical records (333.1 & G25.0, respectively). Patients were then identified to see if they also had an occurrence of PD or DLB in the same fashion using their respective billing codes (PD - 332 & G20; DLB - 331.82 & G31.83). Results: A two-sample Student's T-test for proportion was used to compare the proportion of individuals diagnosed with PD in the ET and control populations (p <0.0001). An odds ratio of 18.34 was calculated from the proportion of individuals diagnosed with PD in our ET dataset compared to the individuals diagnosed with PD in the control population. These data strongly suggest PD is significantly more prevalent in patients with ET than in controls. Additional analyses to investigating the shared genetic risk factors between the two diseases is ongoing.

2831T

Multivariate genome-wide association study for volumes of structural MRI regions of interest measures via a genetic correlation network modular analysis. *J. Liang, X. Zhu, Alzheimer's Disease Neuroimaging Initiative.* Population Quantitative Health Science, Case Western Reserve University, Cleveland, OH.

Challenges in imaging genetic studies are the high dimensionality, multi-modality and high noise data with relative small sample size in conjunction with increased evidence of polygene and pleiotropy. Multivariate and network approaches, which can accommodate correlated traits, are powerful to identify potentially weak complex effects buried in high dimensional datasets and extract independent components or modules in both imaging and genetic data analysis. We calculated the pairwise genetic correlations among 145 baseline structural MRI regions of interest (ROIs) from GWAS summary statistics of volumes of ROIs in ADNI1 participants (230 cognitively normal individuals, 200 Alzheimer's patients, and 410 patients with Mild cognitive impairment). The genetic correlation was used to define a brain imaging genetic correlation network. In this network, the nodes are 145 baseline structural MRI ROIs spanning the entire brain, while edges are genetic correlation estimated by LDSC method reflecting the correlation between genetic influences on ROIs traits. We adapted a weighted gene co-expression network analysis (WGCNA) framework and used the method of topological overlap matrix (TOM) elements in hierarchical clustering to identify module structures. We next performed multivariate GWAS using cross-phenotype association analysis (CPASSOC) program to combine GWAS summary statistics of all ROIs as well as within each of identified module. We observed two variants in genes *SGCZ* and *TOMM40* (region harboring APOE gene) significantly associated with ROIs of volumetric measures of occipital pole and right/left hippocampus (P< 5×10⁻⁸). In the within-module CPASSOC analysis, we observed 10 independent variants in 7 loci (*TOMM40/APOE, APCS, APOC1, ADGRL3, CLSTN2, PSD3, SSBP2*) with P < 5×10⁻⁸, in which three of them (*TOMM40/APOE, APCS, SSBP2*) reached experimental wide significance (P < 2.27×10⁻⁹) after adjusting for multiple tests. Among the identified loci, three genes (*TOMM40/APOE, APCS, APOC1*) have been reported associated with Alzheimer's Disease. Our results suggest that incorporating a brain genetic correlation network and multivariate analysis of GWAS summary statistics of brain structural imaging ROIs improves power to detect genes in imaging GWAS.

2832F

***SORBS2* is associated with extended Alzheimer disease related phenotypes in *PSEN1* mutation carriers in Puerto Rico.** *R. Cheng[1,2], B. Vardarajan[1,2], G. Tosto[1,2], R. Lantigua[1,3], D. Reyes-Dumeyer[1,2], M. Medrano[5], IZ. Jimenez-Velazquez[6], R. Mayeux[1,2,3,4], JH. Lee[1].* 1) Taub Inst, Columbia Univ, New York, NY; 2) Department of Neurology, Columbia University Medical Center, New York, NY; 3) Department of Medicine, Columbia University Medical Center, New York, NY; 4) Department of Psychiatry, Columbia University Medical Center, New York, NY; 5) School of Medicine, Pontificia Universidad Catolica Madre y Maestra, Santiago, Dominican Republic; 6) Department of Internal Medicine, University of Puerto Rico School of Medicine, San Juan, Puerto Rico.

We previously reported that variants in *SORBS2* modified the expression of *PSEN1* in carriers of the G206A founder mutation based on a small set of early onset Alzheimer disease (AD). This gene is located in 4q35.1 where we observed the strongest linkage signal from our earlier study. This SH3-binding domain gene transcribes a brain-specific splice variant known as nARGBP2, and this variant has been reported to influence the integrity of f-actin in the dendritic spines of neurons. Others have reported overexpression of nARGB2 in mice caused aggregation of f-actin bundles in dendritic spines and may influence phenotypic expression, thereby possibly altering age at onset by affecting dendritic spine morphologic and subsequent neurodegenerative processes. Here, we present the findings from our 2-stage genome study. Extending the earlier study, we examined 305 individuals from 45 families in Puerto Rico with at least one G206A mutation carrier to determine whether the association between variants in *SORBS2* and AD related phenotype, namely AD, age at onset of AD (AAO), memory performance. We performed whole genome sequencing (WGS) on a subset of family members and imputed WGS sequence onto the genome wide SNPs in family members who were not sequenced. We first examined whether *SORBS2* was associated with the above phenotypes, and then we examined the SNPs within *SORBS2*. Subsequently, to confirm the findings from the G206A carrier families, we examined the variants in the *SORBS2* in a set of late onset familial AD in Puerto Rican and Dominican families. In G206A carrier families, gene-based analysis revealed that *SORBS2* was significantly associated with the AAO (p=0.0021); however, those of *SORBS2* for global memory or delay recall were not significant. When we examined the cohort with late onset AD (EFIGA), we observed that 15 SNPs within *SORBS2* were associated with AD (p ranging from 0.0056 to 0.05), and they localized to 186.52Mb-186.82Mb. Further, we observed 18 SNPs associated with the AAO (p ranging from 0.0014 to 0.05), and the locations overlapped. Two SNPs were associated with both phenotypes. Interestingly, a number of other 'SH3' family of genes (*SH3BP5L, SH3RF2, SH3RF3, SH3BP4, SH3TC1, SH3PXD2B* and *SH3BGRL2*) were also associated with the AAO. Our analysis of *SORBS2* in Caribbean Hispanics revealed that *SORBS2* is a candidate genetic modifier of the age at onset of AD in both familial EOAD and familial LOAD among our Caribbean Hispanics cohorts.

2833W

Polygenic risk scores applied to UK Biobank data highlight the interplay between behaviour and psychiatric disorders. *P.F. O'Reilly, J. Maxwell.* SGDP Centre, IoPPN, King's College London, London, United Kingdom.

Polygenic risk scores (PRS) are commonly used to infer shared genetic aetiology across phenotypes. Here we construct PRS for autism and schizophrenia and test their relationship with 17 behavioural and mood-related traits in the UK Biobank ($N = 135\,726$). Genetic burden for autism and schizophrenia in the *healthy population* is associated with numerous behavioural and mood-related traits, including positive correlations between autism PRS and educational attainment ($P = 2\text{x}10^{-20}$) and cognition ($P = 1\text{x}10^{-7}$), and schizophrenia PRS and risk-taking ($P = 4\text{x}10^{-26}$). Self-reported risk-taking is associated with UK-based migration in terms of distance-moved ($P = 2e\text{-}67$) and increasing population density ($P = 4e\text{-}73$), and thus part of schizophrenia aetiology may be the genetics of risk-taking, leading to migration, urbanicity or drug-taking – known risk factors for schizophrenia. We also gain insights into the effects of disease onset and medication by contrasting the trends of behaviour observed with increasing genetic load in healthy individuals to that of medicated and non-medicated diagnosed individuals. .

2834T

Severity modifiers in autism spectrum disorder: WGS perspective. *S.P. Smieszek, J.L. Haines, R. Igo.* Department of Population and Quantitative Health Sciences Case Western Reserve University, Cleveland, OH.

Autism Spectrum Disorder (ASD) comprises a complex of neurodevelopmental disorders primarily characterized by deficits in verbal communication, impaired social interaction and repetitive behaviors. The genetic architecture has proved to be complex and encompasses profound clinical heterogeneity, which poses challenges in understanding its pathophysiology. We conducted a large-scale association analysis of the MSSNG whole genome sequencing data to elucidate potential modifiers of ASD severity. Using the additive linear model method (PLINK) we have directly tested the associations between 6,198,166 SNPs (Quality Control: MAF > 0.05, HWE $P < 1 \times 10^{-6}$, Mendelian errors, removal of samples with dis-concordant sex status, twins, samples with unreported relatedness) and Vineland Adaptive Behavior Scores. Interestingly, the top variants direct us to an 850kb region containing 21 variants within 3 genes on Chromosome 2: *LYPD1* a member of the Lynx family of neurotransmitter receptor-binding proteins implicated in anxiety, *NCKAP5* previously implicated in autism (CNV, 2 cases) and *GPR39*, a product of which has been implicated in depression. Some other interesting loci include *CACNA2D2* ($P < 1 \times 10^{-7}$, 16 markers) encoding a subunit of the voltage-dependent calcium channel complex and axon guidance receptor gene, *DCC* ($P < 1 \times 10^{-7}$). Furthermore, to leverage the size of the data we conducted a pathway enrichment analysis of the set of highly significant results ($P < 1 \times 10^{-6}$) using PARIS and DAVID software. The most significant category is tobacco use disorder, with $P < 1 \times 10^{-5}$, with 28% of genes contributing to the significance, followed by the brain development and structural component of myelin sheath pathways. Genes categorized a neurological, developmental and immune-related constitute 65% of all the genes contributing in these pathways. We took variants from contributing genes from significantly overrepresented categories to test how much variability in the VABS scores can be explained by the variants. The cumulative effect of the single top pathway enrichment alone on affection status is 2% ($P = 6.34 \times 10^{-6}$). We detect a region that may be a hallmark of severity in ASD. As genetic predisposition may be different for almost every ASD individual, understanding the common mechanisms for endophenotypes may help elucidate ASD causal mechanisms.

2835F

Diagnostic changes leading to ASDs' prevalence increase altered the disorders' average genetic architecture. *E.M. Wigdor[1,2], W.K. Thompson[3], J.A. Kosmicki[1,2,4], J. Grove[5,6,7], J. Taylor[1,2], T.M. Werge[5,8,9], A.D. Børglum[5-7], P.B. Mortensen[5,7,10], E. Agerbo[5], M.J. Daly[1,2,4], E.B. Robinson[1,2].* 1) Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA; 4) Department of Medicine, Harvard Medical School, Boston, MA, USA; 5) iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; 6) iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark; 7) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 8) Mental Health Centre Sct. Hans, Institute for Biological Psychiatry, Capital Region of Denmark, Roskilde, Denmark; 9) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 10) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark.

The estimated prevalence of Autism Spectrum Disorders (ASD) has significantly increased in the United States over the last several decades, from an estimated 0.01% in 1970 to 1.46% in 2012 (Fombonne, JARID 2005; ADDM CDC, 2012). This increase is predominantly a function of diagnostic changes; the fraction of ASD cases with Intellectual Disability (ID), for example, has simultaneously decreased (Braun et al., PLoS ONE 2015). Because ASD cases are now more heterogenous, this period is often referred to as one of 'diagnostic expansion.' This study aimed to investigate the impact of the diagnostic expansion on ASDs' average genetic architecture, specifically on the rate of *de novo* variants within ASD cases. We conducted a literature review to estimate the change in average ID rate among ASD cases over the last several decades. In parallel, we examined the change in epilepsy rate among all ASD cases born in Denmark between 1981 and 2005. Using sequencing results from the Simons Simplex Collection (SSC) and Boston Autism Consortium (BAC), we then simulated the impact these changes would have on ASDs' average genetic profile. In our review, we found a decrease in average ID rate from 67.8% in studies published prior to 1990 to 32.8% in studies published by the US Centers for Disease Control after 2008. In the Danish National Patient Registry, we found that, within ASD cases diagnosed by age 5, the rate of epilepsy declined from 21.5% in cases born before 1990 to 6.6% in cases born after 2000 (OR=0.26 p=9.84e-09). The same trend was observed in ASD cases diagnosed at older ages (e.g. 5-10, OR=0.33, p=3.75e-11). In the SSC&BAC, both ID (p=0.002) and a history of seizures (p=0.003) were associated with rate of *de novo* constrained protein truncating variants (CPTVs), defined as PTVs in genes intolerant of heterozygous loss of function variation (Lek et al., Nature 2016). In our simulation analysis, the estimated change in ID rate would result in a 17.8% decline in the percent of ASD cases with a CPTV. Similarly, the observed change in epilepsy rate in the Danish data would result in an approximate 11.7% decline in the percent of ASD cases, diagnosed by age 5, with a CPTV. These results suggest that diagnostic trends can substantially impact the average genetic architecture of neuropsychiatric disorders.

2836W

CNVs among Japanese individuals with neuropsychiatric diseases effect dosage sensitivity in ohnologs and genes expressions. *M. Yamasaki[1,2], T. Makino[3], SS. Khor[1], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Biostatistics Section, Department of Data Science, Clinical Science Center, National Center for Global Health and Medicine, Tokyo, Japan; 3) Department of Ecology and Evolutionary Biology, Graduate School of Life Sciences, Tohoku University, Sendai, Japan.

;Copy number variants (CNVs) have been reported to be associated with various diseases, traits and evolution. However, when it comes to clinical applications, several issues still remain. For an example, CNVs often spanned a few mega bases and covered several genes, and this makes it hard to determine which gene should be prioritized for downstream functional experiments if the particular CNV is rare. In this study, dosage sensitivity of genes and expression was applied to individual CNVs in five different neuropsychiatric diseases, in order to overcome this issue. ;A hypothesis of two round whole genome duplication (2R-WGD) was proposed by Susumu Ohno in 1970, and subsequently the hypothesis was confirmed in the past decade. Studies of ohnologs, which existed at age of 2R-WGD, revealed that a pattern of retained ohnologs was not random and that current ohnologs were undergoing gene balance selection. In this study, CNVs of individuals with five different neuropsychiatric diseases were analyzed in order to assess the burden of dosage sensitive ohnologs on these diseases. ;Effects of CNVs on genes expressions were recently reported. Copy number variations were calculated to contribute to 17~99% of genes expressions. Integration of GWAS and eQTL data was widely applied nowadays; however, it is hard to estimate influence of CNVs on expression in whole-genome CNV studies. Oftentimes, both whole-genome data and transcript for targeted sample-set were required to assess effects of CNVs on expression. Therefore, we utilized open-access data from The Genotype-Tissue Expression (GTEx) project in order to assess the burden of whole-genome CNVs on genes expressions. ;Affymetrix Genome-Wide SNP Array 6.0 was used to detect CNVs by PennCNV and CNV Workshop. After quality controls for population stratification, family relationship and CNV detection, 287 narcolepsy, 133 essential hypersomnia, 380 panic disorders, 164 autism, 784 Alzheimer and 1,280 healthy individuals remained for enrichment analysis. ;Overall, significant enrichment of dosage sensitive genes was found across individuals with narcolepsy, essential hypersomnia, panic disorders and autism. Particularly, significant enrichment of dosage sensitive genes in duplications was observed across all diseases except for Alzheimer disease. Significant enrichments of genes with expression sensitivity in brain were observed in subjects with panic disorder and autism. .

2837T

Testing the moderation of quantitative gene by environment interactions in unrelated but dependent individuals. *R. Tahmasbi[1], L. Evans[1], E. Turkheimer[3], M. Keller[1,2].* 1) Institute for Behavioral Genetics (IBG), Boulder, CO; 2) Department of Psychology and Neuroscience, University of Colorado, Boulder, CO; 3) Department of Psychology, University of Virginia, Charlottesville, VA.

The environment can moderate the effect of genes – a phenomenon called gene–environment (GxE) interaction. There are two broad types of GxE modeled in human behavior – qualitative GxE, where the effects of genes differ depending on some environmental moderator, and quantitative GxE, where the additive genetic variance changes as a function of an environmental moderator. Tests of heritability and GxE have traditionally relied on comparing the covariances between twins and close relatives, but recently there has been much interest in testing such models on unrelated individuals measured on genomewide data. However, to date, there has been no ability to test quantitative GxE effects in unrelated individuals; e.g., the ability to do this using the popular GCTA software does not exist because such models require solving nonlinear constraints. Here, we introduce a maximum likelihood approach with parallel constrained optimization to fit such models. We use simulation to estimate the accuracy, power, and type I error rates of our method and to gauge its computational performance, and then apply this method to IQ data measured on 40,172 individuals with whole-genome SNP data from the UK Biobank. We found that the additive genetic variation of IQ tagged by SNPs increases as SES decreases, opposite the direction found by Turkheimer et al. (2003) and several other twin studies conducted in the U.S., but consistent with several studies from Europe and Australia (Tucker-Drob & Bates, 2016).

2838F

Improved prediction of genetic predisposition to psychiatric disorders using genomic feature best linear unbiased prediction models. *P.D. Rohde[1,2,3], D. Demontis[2,3,4], A.D. Børglum[2,3,4], P. Sørensen[1], iPSYCH-Broad Consortium.* 1) Molecular Biology and Genetics, Aarhus University, Tjele, Tjele, Denmark; 2) iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark; 3) iSEQ, Center for Integrative Sequencing, Aarhus University, Aarhus, Denmark; 4) Department of Biomedicine, Aarhus University, Aarhus, Denmark.

Accurate prediction of unobserved phenotypes from observed genotypes is essential for the success in predicting disease risk from genotypes. However, the performance is somewhat limited. Genomic feature best linear unbiased prediction (GFBLUP) models separate the total genomic variance into components capturing the variance by a genomic feature (e.g. GO term) and the remaining genomic variance by differential weighting of the genetic variants within the two groups. Previously we have demonstrated (on pigs and fruit flies) increased predictive ability when the genomic feature is enriched for causal variants. Here we apply the GFBLUP model to a small schizophrenia case-control study to test the promise of this model on psychiatric disorders, and hypothesize that the performance will be increased when applying the model to a larger ADHD case-control study if the genomic feature contains the causal variants. The schizophrenia study consisted of 882 controls and 888 schizophrenia cases genotyped for 520,000 SNPs. The ADHD study contained 25,954 controls and 16,663 ADHD cases with 8,4 million imputed genotypes. The predictive ability for schizophrenia for the null model (all SNPs weighted equally) was low (0.07). Few GO terms did show a tendency of increased predictive ability; e.g. GO:0008645 had a predictive ability of 0.11 (unadjusted t-test p-value = 7.4×10^{-5}), and explained 9% of the genomic variance, and 1.5% of the total phenotypic variance (0.6% for the null model). The improvement in predictive ability for schizophrenia was marginal, however, greater improvement is expected for the larger ADHD data.

2839W

Quantifying the effect of copy-number variants on general intelligence in unselected populations. *G. Huguet[1,2], C. Schramm[1,2,3], E. Douard[1,2], L. Jiang[3], A. Labbe[4], JB. Poline[5], E. Loth[6], R. Toro[7,8], G. Schumann[6], P. Conrod[1,2,6], Z. Pausova[9], C. Greenwood[3,10], T. Paus[11,12], T. Bourgeron[6,7,13], S. Jacquemont[1,2], IMAGEN Consortium.* 1) Université de Montréal, Montreal, QC, Canada; 2) CHU Sainte-Justine, Montreal, QC, Canada; 3) Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada; 4) Département de Sciences de la Décision, HEC, Montreal, QC, Canada; 5) Berkeley University, Berkeley, CA, USA; 6) Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; 7) Institut Pasteur, Paris, France; 8) CNRS URA 2182 "Genes, synapses and cognition", Paris, France; 9) The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 10) McGill University, Montreal, QC, Canada; 11) Rotman Research Institute, University of Toronto, Toronto, ON, Canada; 12) Child Mind Institute, New York, NY, USA; 13) Université Paris Diderot, Sorbonne Paris Cité, Paris, France.

Introduction: Detection of Copy Number Variants (CNVs) are routinely performed in patients with neurodevelopmental disorders (NDs) and "Clinically significant" CNVs, defined as rare and large CNVs contributing to disease, are identified in 10-15% of these patients. Effects of these clinically significant CNVs on cognitive traits have been studied for only a small number of recurrent CNVs. In addition, case-control studies are impossible to perform for > 75% of CNVs that are non-recurrent. As a result, their effects on neurodevelopment are neither characterized nor understood. *Objectives:* To examine the effect of CNVs on measures of Performance and Verbal Intelligence Quotient (PIQ and VIQ) in general populations. To model and predict the effect of CNVs using variables that characterize gene content and non-coding regions involved in CNVs. *Method:* We called CNVs from genotyping data with PennCNV and QuantiSNP on two cohorts drawn for the general population: Imagen (n=1804) and the Saguenay Youth Study (n=968). Rare (<1/1000) CNVs ≥ 50kb were validated by visual inspection. Gene and regulatory content were annotated for all CNVs based on scores of intolerances to mutation, temporal and tissue expression, as well as gene function. We tested several models to select the independent variables most predictive of CNV effect on IQ using information based criteria. *Results:* We identified rare deletion and duplications larger than 250kb in 10% of individuals. Both the size and gene content of rare deletions decrease IQ, eg. deletions ≥250Kb decrease IQ by 6 points (p=2.10^{-3}). We were unable to detect a significant effect of rare duplications on IQ. For estimating the effect of all deletions on IQ, a stepwise linear model procedure converged on a model including mutation intolerance scores (pvalue <10^{-4}). To validate our model which is built mainly on rare non-recurrent deletions, we compared its predictions to empirical measures of IQ loss from the literature for 12 known recurrent CNVs. The Intraclass Coefficient Correlation shows that predictions reliably match known IQ losses (ICC > 80%, p≤2.10^{-3}). *Conclusion:* Our results suggest that the effects of deletions on general intelligence can be reliably modeled and represent a new perspective for the study of non-recurrent CNVs. Results will help clinicians estimate the impact of CNVs on cognition in their patients.

2840T

Genome-wide association study of dental treatment-related fear and anxiety nominates novel genes. *J.R. Shaffer[1], J.M. Chernus[1], C.L. Randall[2], C.D. Wright[2], D.W. McNeil[2], R.J. Crout[2], R.J. Weyant[1], M.L. Marazita[1].* 1) University of Pittsburgh, Pittsburgh, PA; 2) West Virginia University, Morgantown, WV.

Motivation: Fear and anxiety related to professional dental treatment is associated with avoidance of dental healthcare services, which, in turn, contributes to poorer oral health. About 40% of adults in the USA acknowledge some apprehension related to dental treatment, and 10-20% report significant levels of dental fear. Like other forms of fear and anxiety, the etiology of dental fear and anxiety includes a substantial genetic component, although few specific genes have been identified to date. Therefore, we performed a genome-wide association study (GWAS) of dental fear to nominate susceptibility genes. **Methods:** Dental fear and anxiety was measured on 990 adults and adolescents from north central Appalachia using the Dental Fear Survey, which yields a total score and sub-scales pertaining to avoidance, physiological arousal, and dental treatment-specific stimuli. Participants were genotyped on the Illumina Human610-Quad_v1_B array and imputed to the 1000 Genomes Project phase 1 (June 2011) reference. GWAS scans were performed for the dental fear and anxiety total score and each of the three sub-scales using a mixed model approach while adjusting for age, sex, education, household income, and genetic sharing. **Results:** For the dental fear and anxiety total score, the most significantly associated locus was at 10q26.11 (rs12412905; p=5.1E-8), which included variants in and upstream of *SLC18A2*, a vesicular monoamine transporter implicated in multiple neuropsychiatric disorders. This locus was also significantly associated with the avoidance subscale (p=2.0E-8) and showed suggestive evidence for the physiological arousal subscale (p=1.1E-7). Additional significant loci were observed for dental fear subscales, including 12q13.13 (upstream of *MUC1*; p=7.5E-9) for avoidance, and 13q21.1 (intergenic; p=1.1E-8) and 22q13.1 (intronic of *NPTXR*; p=3.0E-8) for physiological arousal. *NPTXR*, a neuronal pentraxin receptor, is notable given that pentraxins are involved in excitatory synapses, and a previous large GWAS of neuroticism implicated a gene that ubiquitinates *NPTXR*. **Conclusion:** We nominate these new gene candidates for future investigation into the biological basis of dental fear and anxiety. Understanding the etiology of dental fear and anxiety may lead to better strategies for breaking the fear-avoidance cycle, thereby improving oral health quality of life. U01-DE018903, R01-DE014899, R03-DE024264, F31-DE023493, T32-GM081741, and U54-GM104942.

2841F

A penalized parametric bootstrap approach for self-contained pathway analysis of gene-environment interaction. *B.J. Coombes[1], J. Biernacka[1,2].* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN.

Understanding the interplay between genes and environments can give us insight into disease etiology. Gene-environment (GE) interaction can be investigated using a Genome Environment Wide Interaction Study (GEWIS) analysis. However, detection of interactions usually requires around four times the sample size to maintain power compared to testing for main effects. Additionally, the multiple testing burden of a GEWIS further reduces power to detect interaction. Thus, researchers have begun studying GE interactions at the gene or pathway. For a gene-level GE interaction test, we can calculate principal components (PCs) for the variants in a gene and test for interaction between the top PCs and the environment. To extend to a self-contained pathway analysis, we can combine the gene-level p-values using Fisher's method to obtain a pathway-level p-value. However, due to correlation of genes within the same pathway, the pathway-level tests can have inflated type I error. Permutation approaches are commonly used in pathway analyses if we are testing for main effects, but for GE interaction, it is typically not possible to construct an exact permutation test. At the gene-level, it has been shown that the parametric bootstrap is a viable alternative to permutation. A parametric bootstrap simulates null test statistics by simulating data from the null model. For a gene-level analysis of interaction, the null model includes only the main effects of the genetic variants in a gene and the environment. While the number of main effects can be large in this setting, the number of main effects at the pathway-level is considerably larger. In order to estimate the large number of main effects, we propose to fit the null model for a pathway using either a ridge or lasso penalty on the genetic main effects. We can then simulate null data from this model in order to generate pathway-level p-values simulated under the null to compare to our observed p-value. We show that this penalized parametric bootstrap approach can correctly maintain type I error for pathway analysis. We apply our method to test whether bipolar disorder is associated with the interaction between BMI and genetic variation in the WNT signaling pathway.

2842W

Identification of novel genetic variants of DSM-5 alcohol use disorder: Genome-wide association study in NESARC-III. *J. Jung, H. Zhang, B. Grant, P. Chou.* Epidemiology and Biometry Branch, NIH/NIAAA, Rockville, MD.

Alcohol use disorder (AUD) is a complex psychiatric disorder that affects about 32% of European Americans (EA) and 22% of African Americans (AA). Moderate to severe DSM-5 AUD and DSM-4 alcohol dependence (AD) are similarly influenced genetically with heritability of 50-65% and genetic correlation of 95%. We analyzed data from the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) to perform a genome-wide association study of alcohol use disorder with moderate to severe symptoms (AUD >=4). EA and AA in NESARC-III samples were genotyped by Affymetrix Axiom Exome Chip 319K and 100K additional customized SNPs. EA samples composed of 2,465 cases without any illicit substance dependence, and 6,599 controls with exposure of alcohol but no illicit substance dependence, and AA samples of 543 cases and 2583 controls were similarly identified. A single SNP-based analysis with common variants (minor allele frequency (MAF)>= 0.01) was performed to test association with moderate to severe AUD using an additive model. Moreover, a gene based analysis with a combined rare and common variants was conducted using Sequence Kernel Association Test after controlling for sex, age, family income, marital status, education, 7 major psychiatric disorders, and 2 population stratification scores. In a gene-based analysis, we identified 3 novel genes (*PPP2R2D, SPX, TRIM13*) in EA and 3 novel genes (*PPIP5K1, TSPAN8, BCAS4*) in AA with p-value <2.5e-6 of a gene level threshold, and *GABRP* was detected in both EA (7.8e-3) and AA (6.4e-4). Based on a single SNP analysis, we identified 7 novel SNPs in *TSPYL5* (rs34070970), *RBM11* (rs7280643), *IL23R* (rs924080), *PLCD1* (rs61755441), *PTPRD* (rs72614092, rs10958948), and *KANK3* (rs11260079), with p-value < 0.01 in both EA and AA. Given that *GABRP* was detected in EA, we replicated the results in AA from a gene-based analysis. With the SNP based approach we further found that among the newly discovered 7 genes in EA and AA, 4 SNPs were associated with either alcohol-related phenotypes or psychiatric disorders. In summary, in our EA and AA samples we were able to confirm previous findings that *GABRP* was associated with alcohol dependence. Our study results identified novel associations which provided evidence and rendered promises for greater understanding of the etiology of AUD.

2843T

Allelic heterogeneity across psychotic disorders and related phenotypes. *T. Polushina[1,2], V. Steen[1,2], O. Andreassen[3], S. Le Hellard[1,2].* 1) NORMENT - K.G.Jebsen Center, Department of Clinical Science, University of Bergen, Bergen, Norway; 2) Dr. Einar Martens Research Group for Biological Psychiatry, Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 3) NORMENT- KG Jebsen Centre, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

   Background Major mental illnesses have been shown to overlap at the clinical and genetic levels. The genetic overlaps have been so far explored at the single genetic variants level, but very few studies have explored how independent variants within a locus could contribute to the genetic overlaps. In our study, we intend to increase the information captured from GWASs by focusing on allelic heterogeneity, i.e. the contribution of several independent markers within one genetic locus, within a trait and across related traits. **Methods** Using summary statistics from GWASs of traits related to mental illnesses: psychotic disorders, cognitive traits and brain volumes, we first selected independent genomic regions associated in each traits after conditional regression (Yang et al.[1]). All the genetic variants in LD with the associated signal were included in the genomic regions. We then first explored the overlaps in the regions within traits and across traits. We also scored each genomic region in each of the traits, using the Brown score for each bin, and explored the overlap in the significant regions. **Results** We observed allelic heterogeneity within and across traits. For instance: 147 genomic regions were associated with independent markers (not in LD) across several traits. We have established a map of genetic overlaps for these clusters across psychiatric disorders and relevant phenotypes (brain volumes, cognitive and personality traits). We have established a pipeline for identification of allelic heterogeneity across different phenotypes. Several of the GWAS included were too limited in power to provide significant hits yet, and will need bigger samples to yield more significant results. **Discussion** We identify allelic heterogeneity across traits, demonstrating that some genetic regions harbor independent associations with related phenotypes. Our approach is complementary to studies that explore genetic overlap at the single marker level. This improves our understanding of the impact of genetic factors in main psychotic disorders and related phenotypes, and could help to direct functional studies later. [1] Yang J. et al. Nat. Genet. 2012.

2844F

Meta-analysis of *de novo* variants from 9246 probands finds that genes previously associated with autism spectrum disorder harbor more *de novo* variants in probands with intellectual disability/developmental delay without autism. *J.A. Kosmicki[1], L. He[2], K.E. Samocha[1], E.B. Robinson[1], J.C. Barrett[2], M.J. Daly[1].* 1) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA; 2) Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, UK.

   Studies of *de novo* variation, which is identified by exome sequencing parent-offspring families, have successfully identified over 100 genes associated with neurodevelopmental disorders, most notably intellectual disability / developmental delay (ID/DD) or autism spectrum disorders (ASDs). Mutations in the genes first discovered in *de novo* studies of ASD were specifically enriched in those ASD individuals with comorbid ID/DD, and these genes were commonly assumed to affect both traits. As such, some criticized these studies for failing to separate ID/DD genes from those genes that confer risk to ASD without ID/DD. Winner's curse suggests the effect size for these genes would be greatest in ASD (i.e., the trait it was first discovered in). Through a meta-analysis of *de novo* variants from 3982 probands ascertained for ASD (892 with comorbid ID/DD) and 5264 probands ascertained for ID/DD (503 with comorbid ASD) we show that, in fact, the reverse is true: previously associated ASD genes are more often mutated in individuals with ID/DD who do not have ASD. To identify disease associated genes, we compared the number of observed *de novo* missense and protein-truncating variants (PTVs) to the number expected from a null mutational model across three sets: ASD+ID/DD, ASD only, and ID/DD only. This yielded a total of 98 significant genes. The contribution of evidence towards each gene's association is overwhelmingly provided by ID/DD compared to ASD (4.9-fold excess; P<9e-58), with only 8 of the 98 genes possessing more *de novo* missense and PTVs in ASD cases than expected by chance; the remaining 90 genes are associated with ID/DD. Of the 8 ASD-associated genes, only 3 genes, *ANK2*, *DSCAM*, and *CHD8*, are significantly more strongly associated with ASD than ID/DD, with the *de novo* variants in *DSCAM* observed solely in ASD probands. *De novo* missense and PTVs in the remaining 5 ASD-associated genes are observed 2.9 times more in 4789 probands with ID/DD who do not have ASD (P<7e-7). This same finding extends to prior published ASD gene lists (1.8-fold excess in ID/DD; P<0.002), indicating that most ASD-associated genes from previous *de novo* studies are more strongly associated with ID/DD. Therefore, researchers should be cautious with using published lists of ASD genes for follow-up functional studies, especially if the goal is to learn about the genetic etiology of core ASD features in the absence of cognitive impairment - namely social and behavioral impairments.

2845W

Smoking and neuroticism: Using Mendelian randomization to investigate causality. *H. Sallis[1,2], G. Davey Smith[2], M. Munafò[1,2].* 1) School of Experimental Psychology, University of Bristol, Bristol, United Kingdom; 2) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom.

Smoking is one of the leading modifiable risk factors for disability, disease and death. There is a well-documented association between smoking and neuroticism, with smokers reporting increased levels of neurotic symptoms. However, much of this data comes from observational studies and we are unable to make any causal inference regarding this relationship. Mendelian randomization (MR) is a method of assessing causality using observational data through the use of genetic instrumental variables for modifiable risk factors. Recent GWAS have identified variants robustly associated with both smoking phenotypes and neuroticism. This enabled us to use a range of MR methods to attempt to unpick this relationship. We use publicly available summary statistics from these studies in addition to data from UK Biobank to investigate whether there appears to be a causal link between smoking and neuroticism. We found evidence of a modest genetic correlation between smoking initiation and neuroticism (r_G = 0.124, p=0.008). However, we found no strong evidence of a causal relationship in either direction when using 2-sample MR or when looking at individual level data in UK Biobank. Within UK Biobank it was possible to stratify on smoking status, which meant that in addition to smoking initiation, we were able to look at the association between smoking heaviness and neuroticism. We found some evidence of a causal relationship from neuroticism to increased smoking heaviness (β=0.198, p=0.014). In conclusion, although much of the observed association between smoking and neuroticism appears to be non-causal, we found evidence of a modest genetic correlation, suggesting some shared genetic aetiology. We also found some evidence of a causal relationship from neuroticism to smoking heaviness, which would lend support to the self-medication hypothesis.

2846T

Genetic validation of bipolar disorder identified by automated phenotyping using electronic health records. *C. Chen[1,2,3,4,5], P. Lee[1,3,4,5], V. Castro[1,6,7], J. Minnier[8], S. Murphy[7,9,10], V. Gainer[7], T. Cai[10,11], P. Sklar[12,13,14], R. Perlis[1,3,4,5,6], J. Smoller[1,3,4,5], International Cohort Collection for Bipolar Disorder Consortium.* 1) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 4) Department of Psychiatry, Massachusetts General Hospital, Boston, MA; 5) Broad Institute of MIT and Harvard, Cambridge, MA; 6) Center for Experimental Drugs and Diagnostics, Massachusetts General Hospital, Boston, MA; 7) Partners Research Information Systems and Computing, Partners HealthCare System, Charlestown, MA; 8) Oregon Health & Sciences University, Portland, OR; 9) Department of Neurology, Massachusetts General Hospital, Boston, MA; 10) Department of Biomedical Informatics, Harvard Medical School, Boston, MA; 11) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 12) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY; 13) Institute for Genomics and Multiscale Biology, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 14) Friedman Brain Institute, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY.

Bipolar disorder (BD) is a heritable mood disorder with 1% lifetime prevalence in general population. To expand research resources for BD genetics, we seek to leverage electronic health records (EHR) database combined with genome-wide data. As part of the International Cohort Collection for Bipolar Disorder (ICCBD), we developed automated phenotyping algorithms that can identify BD patients using codified data and concepts extracted by natural language processing (NLP) from clinical narratives. We identified BD cases and healthy controls in the Partners HealthCare system with the algorithms and genome-wide genotyped these samples. We previously showed that these algorithms have high positive predictive value using a gold standard of in-person structured interviews (Castro et al. Am J Psychiatry, 2015). Here, we attempted to genetically validate the EHR-based BD cases and controls by estimating SNP-based heritability (h^2_g) and genetic correlation (rg) between EHR-based BD and BD cases using traditional ascertainment (diagnosis/interview) in genome-wide analyses by Psychiatric Genomics Consortium (PGC) and ICCBD. We evaluated European ancestry BD cases identified using 4 EHR-based algorithms: an "NLP-based algorithm", and 3 rule-based algorithms using codified EHR with decreasing levels of stringency – "coded-strict", "coded-broad", and "coded-broad based on a single clinical encounter" (coded-SV). We used LD score regression to estimate h^2_g and rg. The analytic sample comprised 862 NLP, 1968 coded-strict, 2581 coded-broad, 408 coded-SV BD cases, and 3952 controls. The estimated h^2_g were 0.24 (p=0.015), 0.09 (p=0.064), 0.13 (p=0.003), 0.00 (p=0.591) for NLP, coded-strict, coded-broad and coded-SV BD, respectively. The h^2_g for all EHR-based cases combined was 0.11 (p=0.006). These h^2_g were lower than those observed by the PGC/ICCBD (0.23, p=3.17E-80, total N=33181). However, the rg between PGC/ICCBD BD and the EHR-based cases were high for NLP (0.66, p=3.69E-5), coded-strict (1.00, p=2.40E-4), and coded-broad (0.74, p=8.11E-7). The rg between EHR-based BDs ranged from 0.90 to 0.98. These results provide the first genetic validation of automated EHR-based phenotyping for BD and suggest that this approach identifies cases that are highly genetically correlated with those ascertained through conventional methods. High throughput phenotyping using the large data resources available in EHRs represents a viable method for accelerating psychiatric genetic research.

2847F

Detecting tissue-specific genetic correlation between complex psychologic disorders using GWAS summaries. *Q. Fan, F. Zhang, W. Wang, A. He, J. Hao, Y. Wen, L. Liu, X. Liang, Y. Du, P. Li, C. Wu, S. Wang, X. Wang, Y. Ning, X. Guo.* Key Laboratory of Trace Elements and Endemic Disea, Xi'an Jiaotong University, Xi'an, Shaan Xi, China.

Genetic correlation analysis of complex diseases can provide novel clues for pathogenetic and therapic studies of complex diseases. Complex diseases generally implicated in various tissues. Incorporating tissue-specificity into genetic correlation analysis has the potential to improve the performance of genetic correlation analysis. In this study, we proposed an analysis framework for tissue-specific genetic correlation analysis utilizing GWAS summary and eQTLs datasets. Our approach was applied to large-scale GWAS summary datasets of schizophrenia (SCZ), bipolar disorder (BIP), autism spectrum disorder (ASD), major depressive disorder (MDD) and attention deficiency disorder (ADD). The eQTLs datasets of ten brain tissues were driven from the Genotype-Tissue Expression database. Linkage disequilibrium score regression (LDsc) was used to detect genetic correlation between the five disorders. Without considering tissue specificity, we observed significant genetic correlation between bipolar disorder, schizophrenia and major depressive disorder. Further tissue-specific analysis detected significant genetic correlation in cerebellum (P-value=6.98×10^{-5}), cortex (P-value=1.50×10^{-3}) nucleus accumbens (P-value=2.00×10^{-2}) and putamen (P-value=4.07×10^{-2}) between schizophrenia and bipolar disorder. Additionally, although correlation analysis without considering tissue-specificity did not detect significant correlation between bipolar disorder and autism spectrum disorder, further tissue-specific analysis observed significant genetic correlation in cerebellar hemisphere (P-value=3.80×10^{-2}). Our method may provide new clues for elucidating the mechanism of genetic correlation among various complex diseases.

2848W

Genetically predicted gene expression in the brain and peripheral tissues associated with PTSD. *L.M. Huckins[1,2], K. Girdhar[1,2], A. Dobbyn[1,2], T. Jovanovic[3], C.M. Nievergelt[4,5], D. Ruderfer[6], G.E. Hoffman[1,2,7], A.P. Wingo[3,8], S.J.H. van Rooij[3], M. Breen[9], L.E. Duncan[10,11,12], A.X. Maihofer[4], P. Roussos[1,2], M. Fromer[1,2], N. Cox[3,13], H.K. Im[3,13], S. Sieberts[14], B. Devlin[15], E.B. Binder[3,16], J.W. Smoller[11,17], D.G. Baker[4,5], R. Yehuda[9,18,19], M.B. Stein[5,20], P. Sklar[1,2], K.J. Ressler[21,22], J.D. Buxbaum[9], E.A. Stahl[1,2], N.P. Daskalakis[9,18,21,22], CommonMind Consortium; PGC PTSD Working Group.* 1) Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York City, NY, USA; 2) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York City, NY, USA; 3) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine , Atlanta, GA , USA; 4) Department of Psychiatry, University of California San Diego, San Diego, CA, 92093; 5) Veterans Affairs San Diego Healthcare System and Veterans Affairs Center of Excellence for Stress and Mental Health, San Diego, CA, 92061; 6) Vanderbilt University Medical Center, Nashville, TN; 7) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, , New York City, NY, USA; 8) Atlanta Veterans Affairs Medical Center, Atlanta, GA, USA; 9) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 10) Department of Psychiatry, Stanford University, Stanford, CA 94305; 11) Broad Institute of MIT and Harvard, Stanley Center for Psychiatric Research, Boston, MA; 12) The Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 13) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA; 14) Systems Biology, Sage Bionetworks, Seattle, WA; 15) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; 16) Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany; 17) Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, and Department of Psychiatry, Massachusetts General Hospital, Boston, MA; 18) Mental Health Patient Care Center, James J. Peters Veterans Affairs Medical Center, Bronx, NY, USA; 19) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York City, NY; 20) Department of Psychiatry and Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA; 21) Department of Psychiatry, McLean Hospital, Belmont, MA; 22) Department of Psychiatry, Harvard University, Cambridge, MA.

Post-traumatic stress disorder (PTSD) is a debilitating psychiatric disorder occurring in individuals exposed to trauma. To date, little is known about the genetic aetiology of the disorder, although the latest GWAS (carried out by the PGC-PTSD working group) demonstrates that genetic heritability is in line with other psychiatric disorders. PTSD development involves multi-systemic dysregulation in multiple brain regions and diverse peripheral tissues. Some peripheral systems are particularly interesting since epidemiologic evidence suggests that PTSD patients commonly have cardiovascular, metabolic and immune dysregulation. Transcriptomic Imputation approaches use machine-learning methods to impute gene expression from large genotype data using curated eQTL reference panels. These offer an exciting opportunity to compare gene associations across neurological and peripheral tissues. Here, we apply CommonMind Consortium (CMC) and Genotype-Tissue Expression (GTEx) derived gene expression prediction models to the PGC-PTSD data (9,245 cases/ 24,285 controls). Models included 12 brain regions, five cardiovascular tissues, 2 endocrine tissues, the tibial nerve, adipose tissue and whole blood. We identified 24 significant gene-tissue associations, of which 5 were in peripheral tissues (adrenal gland, heart atrial appendage, tibial artery, tibial nerve). We stratified analyses according to trauma type (civilian vs. combat trauma), sex, and self-defined ancestry. Our three strongest associations were identified in military cohorts only, which supports the hypothesis that there is substantial genetic heterogeneity between civilian and combat PTSD risk. We used the PsychENCODE neuronal and non-neuronal reference map for two histone marks associated with transcription and open chromatin (3-trimethyl-lysine 4, H3K4me3; H3-acetyl-lysine 27, H3K27ac) to understand patterns of histone modification among our PTSD-associated genes. Preliminary analyses indicate a significant correlation between prediXcan p-values and the presence of histone marks (highest correlation: H3K4me3 mark in neurons, Pearson R=0.87, p=3.99×10^{-5}). We will expand this analysis to include a wider range of histone modification marks. We will further expand these analyses to identify tissue specific gene clusters and enriched pathways across tissues or in specific tissues. Finally, we will use neuroimaging data and physiological cardiovascular data to functionally validate our results.

2849T

Whole genome sequencing in families with bipolar 1 disorder implicates cysteine transport process and synaptic neurotransmission pathway. *A. Parrado[1], T. Lovenberg[2], Q. Li[3].* 1) Discovery Sciences, Janssen Pharmceutical R&D, Spring House, PA; 2) Neuroscience Drug Discovery, Janssen Research & Development, LLC, 3210 Merryfield Row San Diego, CA; 3) Neuroscience Therapeutic Area, Janssen Research & Development, LLC, 1125 Trenton-Harbourton Road, Titusville, NJ.

Our endeavor was to discover novel gene variants that co-segregate or confer statistical association with Bipolar 1 Disorder (BP-I) to nominate as potential drug targets to prevent, diagnose or treat BP-I. We hypothesize that rare variants within and across genes and families collectively contribute to BP-I. In this research study we generated whole-genome sequencing for 32 BP-I patients, 5 Schizophrenia patients and 27 unaffected relatives, among fourteen Bipolar 1 disorder families and 1 Schizophrenia disorder family. We took a two-prong analysis approach assuming allelic and locus heterogeneity in our families. Since BP-I is a highly heritable disorder we hypothesize that rare variants would be shared among affected relatives in either a dominant or recessive mode of inheritance and would not be observed in the remainder of the study subjects and in subjects not diagnosed with Bipolar or Schizophrenia from other studies. Furthermore, to discover novel common variants that confer risk or protective association across families we performed transmission disequilibrium tests (TDT) to access linkage disequilibrium with a disease predisposing-locus. We identified six variants (missense, deletions, and intronic) that co-segregate with disease status in six genes reported in OMIM to have a phenotype-gene relationship with neuronal disorders (i.e. Schizophrenia, Depression, and Amyotrophic Lateral Sclerosis). Although our study was not highly powered for statistical testing, three variants approached a genome-wide significant association level ($P < 1 \times 10^{-6}$) in our TDT analysis. The most promising variants from our co-segregation analysis are in genes involved in transmembrane transport, short term memory, transmission of nerve impulse, and neuropeptide receptor activity. One exonic variant is located in a gene that is part of the G protein-coupled receptor family and plays a role in the synaptic neurotransmission pathway.

2850F

Gene interaction between DRD4 and DAT1 Loci is a ADHD-risk factor in females of Chilean ancestry. *G. Pathak[1], R. Chakraborty[1], F. Rothhammer[2].* 1) Institute of Molecular Medicine, University of North Texas Health Science Center, Fort Worth, TX; 2) Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile.

Background Attention-deficit/hyperactivity disorder (ADHD) is a common neurological disorder characterized by hyperactivity, inattention, and impulsivity. ADHD is mostly diagnosed in children, but it often persists into adulthood. In this study, we investigated the interaction between genes coding for dopamine receptors and transporters - dopamine receptor 4 (DRD4) and dopamine transporter 1 (DAT1) in ADHD affected individuals of Chilean ancestry. **Methods** We analyzed pedigree data, disease status and genotype data of DRD4 and DAT1 in 51 Chilean families consisting of 172 individuals, consisting of 125 cases and 72 controls. For DRD4, we grouped genotypes as Heterozygous-7, 7-allele present, and absence of 7-repeat. For DAT1 loci, we grouped individuals as Homozygous-10 R allele, Heterozygous-10 R allele, and Presence of 10-R allele. The two loci, DRD4 and DAT1 are considered simultaneously, and each combination is tested versus all other genotype grouped together, in a 2x2 contingency table between cases and controls using Fisher's exact tests. Each of these significant combinations of genotype frequencies of DRD4 and DAT1 conditions was further evaluated with kinship coefficient derived from the pedigree data for males and females separately using chi-square correction for allelic association testing. **Results** The genotype frequencies of DRD4 7- R heterozygous allele with DAT1 10-repeat homozygous was found to be significantly higher in female cases (p= 0.0061), and presence of 7-R allele (2/7, 4/7, 7/7) with DAT1 10-R homozygous allele (10/10) was also found to be significantly higher in female cases (p=0.0002). None of the conditions were found to be significant in males. **Conclusion** The corrected chi-square test for allelic associations uses kinship coefficient to account for cryptic relatedness between cases and controls in a family-based study. Together with the higher prevalence of ADHD in males than females, our study observes a strong allelic association between the simultaneous occurrence of both DRD4 and DAT1 genotypes with ADHD pathogenesis in Chilean females than in males. The simultaneous presence of DAT1/10-repeat homozygosity and DRD4 7-repeat heterozygosity and homozygosity indicate a low level of dopamine owing to increased transporter activity and a reduced response of the DRD4 7-repeat receptor to dopamine, provides a plausible explanation in Chilean females with ADHD.

2851W

Proper joint and conditional meta-analysis of sequence data in the presence of missing summary association statistics. *D. Liu.* Department of Public Health Sciences, Penn State University, Hershey, PA.

Meta-analysis of genetic association studies plays a pivotal role in understanding the genetic architecture of complex traits. After significant associations are identified, it is of interest to estimate their joint and conditional effects in order to dissect independent novel associations and pinpoint causal variants. In practice, the summary statistics may only be available in a subset of the contributing studies, which makes it challenging to utilize studies with missing data for conditional analysis. Existing meta-analysis methods either replace missing summary statistics with zero or discard studies with missing data. These naïve approaches can bias genetic effect estimates and lead to seriously inflated type-I or II errors for conditional analysis. To address this widespread issue, we develop a method to borrow strength across participating studies and consistently estimate the joint effects for candidate and known variants. Based on the consistent estimator, we propose a pseudo-score statistic for conditional analysis of single variant and gene-level associations. Through the extensive analysis of simulated and real data, we showed that the new method produces well-calibrated type I errors and is substantially more powerful than existing approaches. We illustrated the effectiveness of the proposed methods in a large-scale meta-analysis of nicotine addiction phenotypes.

2852T

Joint analysis of rare and common variants with the adaptive combination of Bayes factors method. *W. Lin[1,2], W. Chen[1,2,3,4], C. Liu[5], H. Hwu[1,5], S. McCarroll[6,7,8], S. Glatt[9], M. Tsuang[10,11].* 1) Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan; 2) Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan; 3) Institute of Brain and Mind Sciences, College of Medicine, National Taiwan University, Taipei, Taiwan; 4) Genetic Epidemiology Core Laboratory, Division of Genomic Medicine, Research Center for Medical Excellence, National Taiwan University, Taipei, Taiwan; 5) Department of Psychiatry, College of Medicine and National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan; 6) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 8) Department of Genetics, Harvard Medical School, Boston, MA, USA; 9) Departments of Psychiatry and Behavioral Sciences and Neuroscience and Physiology, Medical Genetics Research Center, SUNY Upstate Medical University, Syracuse, New York, USA; 10) Center for Behavioral Genomics, Department of Psychiatry, University of California San Diego, La Jolla, California, USA; 11) Institute for Genomic Medicine, University of California San Diego, La Jolla, California, USA.

Both rare and common variants contribute to the etiology of complex diseases. Some specialized arrays such as PsychChip were designed for the detection of both rare and common variants. There is a need to develop a powerful statistical method for this purpose. To exclude neutral variants from a gene, previous gene-based methods remove variants with larger *P*-values. However, a *P*-value carries no information from the alternative hypothesis and power, which varies with minor allele frequencies (MAFs). Another summary measure, Bayes factor (BF), is defined as the ratio of the probability of the data under the alternative hypothesis and that under the null hypothesis. It quantifies the "relative" evidence in favor of the alternative hypothesis. We show here that BF is superior to *P*-value in removing neutral variants when both rare and common variants are under consideration. We therefore propose an "adaptive combination of Bayes factors" ("ADABF") method that can be directly applied to variants with a wide spectrum of MAFs. Comprehensive simulations show that ADABF is more powerful than single-nucleotide polymorphisms (SNP)-set kernel association tests and burden tests. Moreover, we applied the method to 1,109 case-parent trios from the Schizophrenia Trio Genomic Research in Taiwan, and found that the *MAP2K7* gene was associated with schizophrenia (*P*-value = $1.7*10^{-6}$). Our ADABF method is applicable to analyses of case-parent trios or unrelated subjects (allowing covariate adjustment). The R code to implement this method can be downloaded from http://homepage.ntu.edu.tw/~linwy/ADABF.html.

2853F

Flipping GWAS on its head: A statistical approach to identify genetically distinct disease subphenotypes. *A. Dahl[1], C. Na[2,3], J. Flint[4], N. Zaitlen[1].* 1) Department of Medicine, University of Califronia San Francisco, San Francisco, CA; 2) Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 3) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 4) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA.

GWAS have found many associations by comparing genotypes to pre-defined diseases. But standard GWAS ignore many known complexities of disease, including misdiagnosis, biologically and medically distinct subtypes, and genetic variants that may act differently in different samples. For example, psychiatric diseases, like bipolar disorder and schizophrenia, are often confused; diseases once considered homogeneous, like breast cancer, have genetically distinct subtypes with drastically divergent prognoses and treatments; and many are actively seeking novel subphenotypes in diseases from cancer to major depressive disorder (MDD). These complications all decrease GWAS power, hamper downstream interpretation, and prevent precision clinical treatment. To address these concerns, we develop a statistical approach that reverses GWAS: we use known GWAS hits to partition traits into distinct, latent, and heritable architectures. Our EM algorithm iterates between dividing samples into latent subphenotypes and estimating genetic effects for each subphenotype. The inferred latent factors are then used in a standard interaction test for genetic heterogeneity in the subphenotypes. Our supervised approach identifies phenotypically relevant latent factors without prior knowledge. In particular, it can be used in datasets without direct measurements of subphenotypes, which dramatically eases replication of subphenotype associations. Finally, our multitrait method provides a principled decomposition of disease correlations. The method runs in minutes for tens of thousands of samples and tens of SNPs and traits; can incorporate arbitrary covariates; and can exploit penalization. We show in simulations that our method controls the false positive rate and has power to detect heterogeneity at published GWAS SNPs. By collecting more traits or samples or increasing main genetic and environmental effects, our power approaches an oracle interaction test that uses the true subphenotypes. We apply our method to breast cancer data to recover SNPs with effects known to depend on estrogen receptor status. We then study notoriously-complex psychiatric traits using the CONVERGE MDD dataset, which measured tens of risk factors and related disorders. These precise phenotypes provide a unique opportunity to characterize our inferred subphenotypes post hoc. We then analyze depression in the UK Biobank and explore the detected heterogeneous SNPs by studying their phenotypic segregation in CONVERGE.

2854W

FOLD: A method to optimize power in meta-analysis of genetic association studies with overlapping subjects. *E. Kim[1], H. Buhm[1,2].* 1) Asan Institute for Life Sciences, Seoul, South Korea; 2) Department of Convergence Medicine, University of Ulsan College of Medicine, Seoul, South Korea.

Motivation: In genetic association studies, meta-analyses have been widely used to increase the sample size and achieve sufficient statistical power by aggregating summary statistics from multiple studies. In meta-analyses, participating studies often share the same individuals due to the public availability of control data, accidental recruiting of the same subjects, or utilizing the same controls for meta-analysis. Particularly, a repeated use of the same controls is prevalent in cross-disease meta-analyses where one combines association results of multiple diseases to uncover pleiotropic loci. As such overlapping subjects induce correlations between studies and thus can inflate false positive rate, overlapping subjects are traditionally split in the studies prior to meta-analysis, which requires access to genotype data and is not always possible. Fortunately, recently developed meta-analysis methods can systematically account for overlapping subjects at the summary statistics level. **Results:** We identify and report a phenomenon that these methods for overlapping subjects can yield low power. For instance, in our simulation involving a meta-analysis of five studies that share 20% of individuals, the traditional splitting method achieved 80% power, whereas the power of the other new methods was less than 32%. This low power results from the unaccounted differences between shared and unshared individuals in terms of their contributions towards the final statistics. Here, we propose an optimal summary-statistic-based method termed as FOLD (Fully-powered strategy for OverLapping Data) that increases the power of meta-analysis involving studies with overlapping subjects. In this method, we categorize subjects based on their contributions to the final statistic and then calculate the summary statistic for each category. We analytically show that the FOLD estimator achieves smaller variance than the current methods. Moreover, we apply our method to the cross-disease meta-analysis design using the Psychiatric Genomics Consortium (PGC) data and show that FOLD outperforms other competing methods. **Availability:** Our method is available at http://software.buhmhan.com/FOLD .

2855T

Integrating eQTL data with GWAS summary statistics identifies novel genes and pathways associated with schizophrenia. *C. Wu, W. Pan.* Division of Biostatistics, University of Minnesota, Minneapolis, MN.

The largest genome-wide association study (GWAS) of schizophrenia (SCZ) to date has identified 108 risk loci; however, the biological mechanism underlying these associations is not well understood. Some new gene-based methods, such as transcriptome-wide association study (TWAS) (Gusev et al. 2016, Nature Genetics 48: 245-252), have been proposed to boost statistical power and offer biological insights by imputing gene expression with reference eQTL dataset. We first note that TWAS is simply a weighted Sum test with the weights determined by the reference eQTL data. By noting some well-known shortcomings of the (weighted) Sum test, we propose adopting other more powerful gene-based tests while integrating a reference eQTL dataset to further boost power. Specifically, we applied and compared TWAS and an adaptive association test called aSPU (Pan et al. 2015, Genetics, 197: 1081-1095) with eQTL datasets drawn from brain, blood, and adipose tissues and two SCZ GWAS summary association datasets, denoted SCZ1 and SCZ2, based on 20,000 and 80,000 subjects respectively. 90% of the significant genes identified by the SCZ1 data set were confirmed to be significant by the SCZ2 data set, validating the high reproducibility of the proposed methods. For SCZ2, TWAS and aSPU identified 152 and 242 significant genes, respectively, showcasing much improved statistical power of applying our proposed aSPU test. In the end, we identified 59 novel SCZ genes, of which 41 have not yet been reported. Noting that causal genes of the same disease may share biological function, next we propose a pathway-based association analysis integrating pathway annotations, a reference eQTL dataset, and a GWAS summary association dataset. The basic idea is to impute the gene expression for each gene in a pathway, then adaptively test for association between imputed expression levels of the genes and a GWAS trait. We applied our proposed approach with the KEGG and GO pathways to the eQTL datasets and the SCZ1 and SCZ2. Most of the significant pathways identified by analyzing the SCZ1 were confirmed by the SCZ2. Importantly, we identified 31 novel significant pathways associated with SCZ, such as GO:0007158 (Neuron Cell Adhesion), which could not be uncovered by the gene-based TWAS or aSPU analysis. Our results showcase the power of incorporating reference gene expression data for GWAS. The newly identified genes and pathways provide insights into the mechanism underlying SCZ.

2856F

Identifying highly damaging missense mutations in over 10,000 developmental disorder trios using a regional missense constraint metric. *K.E. Samocha[1,2,3,4], J.A. Kosmicki[2,3,4], K.J. Karczewski[2,3], A.H. O'Donnell-Luria[2,3,5], J. McRae[1], G. Gallone[1], E. Pierce-Hoffman[2,3], J.C. Barrett[1], D.G MacArthur[2,3], B.M. Neale[2,3], M.J. Daly[2,3,4], Deciphering Developmental Disorders.* 1) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Massachusetts General Hospital, Boston, USA; 3) Broad Institute of Harvard and MIT, Cambridge, USA; 4) Harvard Medical School, Boston, USA; 5) Boston Children's Hospital, Boston, USA.

Given the increasing numbers of patients who are undergoing exome or genome sequencing, it is critical to establish tools and methods to interpret the impact of genetic variation. This has specifically arisen in the study of *de novo* (newly arising) variants, which have an important role in the genetic architecture of neurodevelopmental disorders. While the largest excess has been found for *de novo* protein-truncating variants (roughly 2-fold enriched over expectation), there is a significant, but more modest (1.3-fold), enrichment of *de novo* missense variants in cases with a neurodevelopmental disorder. Unfortunately, the ability to interpret missense variation remains particularly challenging since missense variants can have drastically different effects depending on both the precise location and specific amino acid substitution of the variant. Previous tools used to predict the deleteriousness of missense variation have relied primarily on sequence conservation across species (e.g. SIFT), structural features of the protein (e.g. PolyPhen-2), or combinations of these metrics (e.g. CADD). Few metrics have yet to take advantage of the knowledge of constrained genes and regions where natural selection most aggressively removes variation within the human population. In order to better evaluate missense variation, we combined information about the local depletion of missense variation with variant-level metrics to create a novel missense deleteriousness metric named MPC (for **M**issense badness, **P**olyPhen-2, and **C**onstraint). We applied MPC to *de novo* missense variants from over 10,000 trios with a neurodevelopmental disorder and identified a category of *de novo* missense variants with the same impact on risk for developing a neurodevelopmental disorder as truncating variants (rate ratio = 5.8, $p < 10^{-50}$). Furthermore, we showed that MPC vastly outperforms three other metrics that predict missense pathogenicity (PolyPhen-2, CADD, and M-CAP) and provide MPC scores of all potential missense variants in the 18k transcripts studied as a resource for the community.

2857W

Genotype-phenotype study of *OPHN1* and *IL1RAPL2* genes mutations in children with intellectual disability. *Y.M. Khimsuriya[1], N. Kharod[2], J.B. Chauhan[1].* 1) P. G. Department of Genetics, ARIBAS, SPU, Anand, Gujarat, India; 2) H. M. Patel Academic Centre, Karamsad, Anand, Gujarat, India.

   Background: Intellectual disability (ID), a generalized neurodevelopmental disorder defined by significantly underdeveloped intellectual and adaptive skills. The sign and symptoms of ID are mainly behavioral. Genetic analysis of such children helps to understand the genetics aspects of ID. The present study was undertaken to correlate phenotypic and genetic features in two genes *OPHN1* and *IL1RAPL2*, known to cause non-specific X-linked intellectual disability. **Materials & methods:** The PCR-RFLP method was used for genotyping of rs199985543 in *OPHN1* and rs9887672 in *IL1RAPL2* genes in 140 children with ID and 100 normal children of the Indian population. The clinical features of patients were recorded, statistically analyzed and correlated with genotype data. The Multifactor dimensionality reduction (MDR) analysis carried out by GMDR software to predict gene-gene interaction for both selected genes mutation. **Results:** In the present study, genotype frequencies of rs199985543 and rs9887672 were significant and increased disease risk compared to the normal children (χ^2= 15.22 & 14.32; P= 0.005 & 0.0001; OR= 32.56 & 30.56; CI= 1.94-545.08 & 1.821-512.76 respectively). Our study population has 70% males that were most probably due to XLID genes mutations as per reported studies. Interestingly, statistical data of reported clinical features revealed that majority of XLID features were correlated with both the selected mutations. The MDR model also revealed that both mutations have epistasis interactions (p=0.001 and cross-validation consistency= 10/10). **Conclusion:** The study was statistically significant for selected polymorphism in West Indian population but we may not say that these polymorphisms are only responsible for such phenotypes. Our clinical features data demonstrated that the ID affected group have the majority of X-linked phenotypic appearance but due to the limitation of genotyping study we cannot identify the all possible genetic mutations. .

2858T

Global developmental delay: Genetic causes in a group of Mexican patients. *M.L. Arenas-Sordo, LL. Flores-García, EP. Linares-Mendoza.* Genetics, Instituto Nacional de Rehabilitacion "Luis Guillermo Ibarra Ibarra", Av Mexico-Xochimilco 289. Arenal de Guadalupe, 14389 Tlapan, Mexico City, Mexico City, Mexico.

   Introduction. Global developmental delay (GDD) is defined as delay in two or more developmental areas: gross motor, fine motor, language, cognition and social or daily living activities. The delay needs to be more than 2 ST (McDonald et al. 2006). This term is used only in children 5 years old or less. This condition has high prevalence, around 1% to 3%, but it could be higher (Moeschler et al. 2006). The causes of GDD are many but if we grouped them, the most important group is the genetic one (Srour et al. 2006). In some studies is as high as 50%. Development of language and speech is considered by experts as a useful indicator of cognitive abilities and global development of a child (Riou et al. 2009; Spiel et al. 2001; Shevell et al. 2003; Gooch et al. 2015; Busari and Weggelaar 2004). **Methods.** An observational clinic investigation was carried out, retrospective, transversal and descriptive, revising patients' files of those who had gone for the first time to the Genetic Service. From 154 files only 92 complied with the inclusion criteria: To have global developmental delay and complete file. **Results.** Most patients were between the ages of 2 to 4 yrs. The great majority of them (87%) presented affectation of the 5 neurodevelopmental areas. We found 33 different syndromes as a cause of the GDD. We reached the etiological diagnosis in 64 (69%) **Conclusions.** To conclude we can say the patients were studied in an adequate and carefully way. We reached the etiological diagnosis in 69% of the cases, the same percentage mentioned in the literature with the diagnosis tools we used. The patients referred to the Genetic Service are in general, persons with diseases which compromise many cerebral areas. .

2859F

The investigation by WES of inborn errors of metabolism as an underlying cause of idiopathic intellectual disability and/or unspecific congenital malformations in a series of 550 patients. *N. Houcinat[1,2], A. Brunel[1,2], P. Kuenz[1,2], J. Thevenon[1,2], S. Moutton[1,2], Y. Duffour[1,2], P. Callier[1,2], N. Jean[1,2], A.L. Mosca[1,2], F. Tran Mau-Them[1,2], D. Lehalle[1,2], S. El Chehadeh[3], C. Francannet[4], M. Lebrun[5], L. Lambert[6], M. Jacquemont[7], M. Gérard-Blanluet[8], M. Willems[9], J.L. Alessandri[10], T. O'Grady[11], C. Robinet-Thauvin[1,2], L. Faivre[1,2].* 1) Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement, CHU Dijon Bourgogne, Dijon, France; 2) Equipe d'Accueil 4271, Génétique des Anomalies du Développement, Université Bourgogne Franche-Comté, Dijon, France; 3) Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, Hôpital de Hautepierre, Strasbourg, France; 4) Service de Génétique Médicale, CHU de Clermont-Ferrand, Clermont-Ferrand, France; 5) Department of Genetics, Hôpital Nord, CHU Saint Etienne, 42055 Saint-Etienne Cedex 2, France; 6) Service de Génétique, CHU, Nancy, France; 7) Unité de Génétique Médicale, CHU de la Réunion, Saint-Pierre, France; 8) Department of Clinical Genetics, APHP Robert Debré Hospital, Paris, France; 9) Unité de Génétique Clinique, CHI de Créteil, Créteil, France; 10) Pôle Enfants, CHU de la Réunion - Hôpital Félix Guyon, Saint-Denis, France; 11) Department of Paediatrics, Midland Regional Hospital, Mullingar, Co. Westmeath, Ireland.

Inborn errors of metabolism (IEM) comprise heterogeneous and rare genetic diseases with overlapping and/or unspecific phenotype. They are caused by a significant block in a metabolic pathway resulting either in accumulation of substrate or deficiency of product. These disruptions can generally be assessed biochemically through the detection of metabolites in biological fluids allowing the diagnosis. However, biomarkers are not always available especially in new emerging, extremely rare and complex IEM. Moreover, the specificity and sensitivity of some of these biomarkers are not always high. All these considerations make the diagnosis of IEM challenging. Here, we evaluate the contribution of WES strategy combined to reverse clinical and biochemical phenotyping in the diagnosis of complex IEM, in a series of patients of all ages. They presented with unspecific phenotype of developmental delay (DD) and/or congenital malformations and were referred for exome sequencing after negative diagnosis workup. The genetic diagnosis was achieved in 32% of patients (177/550), among them 15 were rare IEM, identified in 19 unrelated patients, including the identification of a novel gene *SLC13A5*. They were first misdiagnosed mainly because of atypical or mild presentations or unspecific symptoms. Moreover, specific and reliable biomarkers in the blood or urines were not available in at least 10 disorders. In this series, despite the absence of clinical or biochemical suspicion of IEM, the diagnosis rate was 3.6%. It represented 11% of all genetic diagnosis. This study highlights the value of WES strategy in the diagnosis of complex IEM which is important for appropriate patient genetic counselling and treatment. Indeed, at least two (GLUT1 and SPR deficiency) of these disorders have specific therapy.

Gene	Disease/metabolic pathway	Clinical manifestations	Nb index cases
MAN2B1	mannosidosis	Deafness,DD	2
SLC13A5	Cerebral citrate transport	DD,seizures (S)	1
SPR	Neurotransmitter deficiency	DD,abnormal movement (AM)	1
GLUT 1	Cerebral glucose transport	DD,S	1
PPT1	ceroid lipofuscinosis (CLN) 1	DD,S	2
CLN3	CLN3	Retinitis pigmentosa	2
ADCK3	coenzyme Q10 deficiency	DD,S	2
SLC6A1	GABA transport	DD,S	1
NGLY1	Protein deglycosylation	DD,S,alacrima	1
DRP1	Mitochondrial/peroxysomal fission	DD,S	1
ST3GAL5	GM3 synthase	DD,S	1
NPC1	Nieman Pick C	fetal hepatosplenomegaly, hydrops	1
PIGN	GPI-anchor biosynthesis	DD,S	1
ALDH18A1	Glu γsemialdehyde synthetase	fetal corpus callosum agenesis, skeletal anomalies	1
SLC30A10	Zinc transport	DD, AM	1

2860W

Significant association at the Duffy blood group locus with mitochondrial copy number. *X. Geng[1], J. Ding[2], E. Schmidt[1], DE. Arking[3], AE. Ashley-Koch[4], K. Barnes[5], M. de Andrade[6], J. Fetterman[7], H. Gui[8], MN. McDonald[9], C. Montgomery[10], JR. O'Connell[11], N. Pankratz[12], JR. Shaw[13], HK. Tiwari[14], LK. Williams[15], J. Wilson[16], T. Blackwell[1], D. Levy[17], G. Abecasis[1], C. Liu[17] On behalf of the TOPMed mitochondrial DNA working group.* 1) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI; 2) Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Duke Molecular Physiology Institute, Duke University, Durham, NC; 5) Department of Medicine, John Hopkins University School of Medicine, Baltimore, MD; 6) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 7) Department of Medicine and Whitaker Cardiovascular Institute, Boston University School of Medicine, Boston, MA; 8) University of Hong Kong, Henry Ford Health System, Hong Kong; 9) Division of Pulmonary, Allergy, and Critical Care Medicine, School of Medicine, University of Alabama, Birmingham, AL; 10) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 11) Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 12) Division of Laboratory Medicine & Pathology, University of Minnesota, Minneapolis, MN; 13) Jackson Heart Study, Jackson State University, Jackson, MS; 14) Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 15) Department of Internal Medicine, Henry Ford Health System, Detroit, MI; Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI; 16) Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, UK; 17) Population Sciences Branch and Framingham Heart Study, National Heart, Lung and Blood Institute (NHLBI), Bethesda, MD.

Mitochondrial DNA (mtDNA) copy number is a cellular phenotype which can be estimated directly from whole-genome sequence data. Here, we investigate mtDNA copy number variation in diverse populations using whole-genome sequence data from Phase 1 of the NHLBI Trans-Omics for Precision Medicine (TOPMed) program, which has sequenced blood-derived DNA samples for 18,526 individuals. We have estimated mtDNA copy number from short-read sequence data in 13,648 individuals (8 studies) by comparing the average sequence coverage of mitochondrial and autosomal genomes. The estimates in 13,648 participants average 148.9 mitochondrial genomes per cell, with 80% of the samples between 89.0 and 224.3. We observe differences in mtDNA copy number distribution between studies as a result of different strategies for extracting and processing DNA. We see a modest decrease in mtDNA copy number due to sex and age effects. On average, males have 5.1 fewer mitochondrial genomes per cell than females; for each additional 10 years of age, individuals have on average 4.3 fewer mitochondrial genomes per cell. An apparent strong positive association is found between mtDNA copy number and the Duffy null blood group variant on Chromosome 1 (*rs2814778*) (effect size: 0.321 standard deviation per allele, $p=1.1\times10^{-35}$). This signal is also found in male-only and female-only association studies ($p=1.0\times10^{-20}$ and 4.3×10^{-20} respectively). This result is driven by the high frequency of the Duffy null allele, which causes ethnic neutropenia in persons of African Ancestry. Indeed, reduced neutrophil count alone predicts mitochondrial copy number more accurately than the Duffy genotype. In 1,000 Genomes sequence data, the non-reference allele at this locus is present at 0.6% in 1,006 samples of European ancestry and 96% in 1,322 samples of African and African American ancestry. This finding highlights the value of incorporating ethnic diversity included in the TOPMed sequencing program.

2861T

Targeted sequence analysis of human mitochondrial DNA using an IDT xGen® Lockdown® probe panel. *R. Lopez, N. Downey.* Integrated DNA Technologies, Coralville, IA, USA.

Mitochondria play an essential role in producing energy for cells. In humans, 80% of the ~17 kb mitochondrial genome encodes RNA, in contrast to the nuclear genome where <2% codes for proteins and up to 80% codes for RNAs that feed into the complex systems of a cell . Hence, if the smaller mtDNA chromosome incorporates a random mutation in mitochondrial DNA (mtDNA) it is much more likely to have direct functional consequences compared to mutations in the nuclear genome. Human mtDNA is also maternally inherited, thus tracking mutations in the mtDNA hypervariable regions over generations can be used in population genetic studies. In addition, cells in tissues that have higher energy demands (i.e., need more ATP) also have higher copy numbers of mtDNA than cells with lower energy requirements which lends itself to coverage depth challenges when analysis is done in conjunction with nuclear genomic DNA.   To facilitate studies of the mitochondrial genome and its potential disease contributions, IDT has developed the xGen® Human mtDNA Research Panel, which consists of 138 synthesized xGen Lockdown® Probes that provide reliable coverage for the entire mitochondrial genome. The ability to detect single nucleotide polymorphisms (SNPs) that lead to functional or metabolic disorders was measured using a Leber Optic Atrophy cell line, which contains a G>A transition in the NADH dehydrogenase subunit 4 gene (Coriell NA10744). Three generations of CEPH/Utah pedigree 1463 were used to demonstrate inheritance of variants in the hypervariable region of the mitochondrial genome. IDT xGen panels are modular, so the xGen Human mtDNA Research Panel may be used as a spike-in to other larger panels (e.g., xGen Exome Research Panel or AML Cancer Panel) to evaluate mitochondrial mutations and/or copy number for cancer research. Here, the mitochondrial panel was used as a spike-in with a custom xGen Lockdown panel of 508 probes and the larger Exome Research Panel across a broad dynamic range of mtDNA copies to assess the relative coverage. Together these data demonstrate the utility of the xGen Human mtDNA Research Panel for disease and cancer research   .

2862F

Association study for common and rare genetic variation contributing to exfoliation syndrome. *R.P. Igo[1], J.N. Cooke Bailey[1], J.H. Kang[2], P. Kraft[3], R. Allingham[4], M. Houser[5], J. Fingert[6], R. Ritch[7], A. Sit[8], R. Lee[9], M.A. Pericak-Vance[10], W. Scott[10], L.R. Pasquale[11], J.L. Haines[1], J.L. Wiggs[11].* 1) Population and Quantitative Health Sciences, Case Western Reserve Univ, Cleveland, OH; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Program in Genetic Epidemiology and Statistical Genetics, T.H. Chan Harvard School of Public Health, Boston, MA; 4) Ophthalmology, Duke University Medical Center, Durham, NC; 5) Medicine, Duke University Medical Center, Durham, NC; 6) Ophthalmology, University of Iowa, College of Medicine, Iowa City, IA; 7) Einhorn Clinical Research Center, Department of Ophthalmology, New York Eye and Ear Infirmary of Mt. Sinai, New York, NY; 8) Ophthalmology, Mayo Clinic, Rochester, MN; 9) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami FL; 10) Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 11) Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA.

Exfoliation syndrome (XFS), characterized by proteinaceous deposits within the eye, is a major risk factor for glaucoma, and therefore an important cause of blindness. *LOXL1*, a well-characterized risk locus, accounts for much of the heritable variation for XFS. Two missense variants in *LOXL1* are strongly associated with XFS, but the risk allele of each is reversed in some populations, indicating complex genetic architecture, and no causal variant for XFS has yet been identified. Six other genetic loci have been implicated in XFS through genomewide association studies (GWAS), but additional causal loci for XFS have not yet been found. To search for additional genetic variation contributing to XFS, we conducted a GWAS and gene-based tests of rare variation on a sample of 1,118 XFS cases and 3,661 controls genotyped for the Illumina OmniExpressExome chip and imputed to the Haplotype Resource Consortium panel. A GWAS on 7,626,807 common (minor allele frequency (MAF) ≥ 0.01) single-nucleotide polymorphisms (SNPs) was carried out by Firth bias-corrected logistic regression using allele dosage as a predictor, adjusting for sex and six principal components. We found overwhelming evidence for association at *LOXL1* (index SNP rs8027022 in intron 1, OR = 0.25 for the A allele, p = 4.6E-134) and strong association at a novel locus on chromosome 17q21 (index SNP rs739499 in intron 2 of *SAMD14*, OR = 1.31 for the G allele, p = 7.2E-08) near the type 1 collagen subunit a1 (*COL1A1*) gene. We found suggestive evidence for association at the known XFS risk locus *CACNA1A* (rs78771862 in intron 2, OR = 2.68, p = 2.49E-05). Gene-based tests for association of rare genetic variation (MAF < 0.01), using SKAT-O, revealed highly significant signals for *LOXL1* (p = 2.3E-10 for all rare variants; p = 1.5E-09 for exonic variants only) and *LOXL1-AS1* (p = 5.1E-09 for all variants). The rare alleles of all seven exonic SNPs in *LOXL1* were enriched in XFS controls; two missense polymorphisms, rs201884374 (G120S) and rs138183635 (A160P), are predicted to be deleterious by SIFT and PolyPhen. However, gene-based test results were nonsignificant (p > 0.05) when previously described *LOXL1* common SNPs G153D and L141R were included in the SKAT-O model. Overall, these analyses add new SNPs, within and outside *LOXL1*, to the growing list of potential causal variants for XFS, and reveals additional complexity of how *LOXL1* variation contributes to the disease. Grant support: NIH/NEI R01 EY020928.

2863W

Whole exome sequencing reveals candidate variants for elevated intraocular pressure in the Beaver Dam Eye Study. *W. Li[1], F. Chen[1], A.P. Klein[1,2,3], B.E.K. Klein[4], R. Klein[4], K.E. Lee[4], P. Duggal[1].* 1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 2) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland, USA; 3) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA; 4) Department of Opthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin, USA.

Nearly 2% of people in the US > 40 years are affected with glaucoma that can lead to vision loss/blindness. Elevated intraocular pressure (IOP) is a key risk factor for glaucoma. To assess the genetic underpinnings of IOP, we performed whole exome sequencing on 135 participants of the Beaver Dam Eye Study (BDES) with extreme baseline IOP or refraction measurements. Single-nucleotide variants (SNVs) and insertions/deletions (INDELs) were called across all 135 samples by GATK Unified Genotyper. All variants were annotated by ANNOVAR, with gene and functional annotations from RefGene and allele frequencies from the 1000 Genome Project. We excluded variants if (1) they failed quality filters based on the variant quality score lod-odds (VQSLOD) derived from selected quality metrics per GATK's best practice, or (2) they were in segmental duplication regions based on UCSC genomic annotations. A total of 309,066 variants with a mean transition transversion ratio (Ti/Tv) of 2.54 were obtained. Principal component analysis (PCA) confirmed the European descent of all individuals and identity-by-descent analysis (IBD) identified 95 first-degree and 102 second-degree relatives pairs, as expected. We focused our analysis on premature terminate variants (PTV) defined as splicing or nonsense SNVs and frameshift INDELs based on RefGene annotations. A total of 30 unrelated individuals with IOP > 21 mmHg or IOP/glaucoma treatment were defined as cases and 11 unrelated individuals with IOP < 15 mmHg were defined as controls. We filtered for PTV variants: (1) in autosomal chromosomes, (2) with allele frequencies in the CEU population of 1000 Genome Project < 1%, and (3) uniquely seen in cases but not in controls. A splicing variant (12:81671194) of the PTPRF interacting protein alpha 2 (*PPFIA2*, [MIM# 603143]) gene was uniquely seen in 7 out of 30 (23%) unrelated high IOP individuals and in 1 relative with high IOP. A frameshift INDEL (6:30995054) of the mucin 22 (*MUC22*, [MIM# 613917]) gene was uniquely seen in 4 out of 30 (13%) unrelated high IOP individuals and in 2 relatives with high IOP. Previously we also identified the prostaglandin E receptor 1(*PTGER1*, [MIM# 176802]) gene associated with elevated IOP. To validate our findings, we selected 23 variants from these 3 genes and genotyped them in 936 samples with IOP measurements including 316 high IOP and 254 low IOP individuals. Identification of genes associated with IOP may lead to better and targeted therapeutics.

2864T

Individualized glaucoma risk evaluation using the genomic profile. *X. Gao, F. Fan.* Dept Ophthalmology & Visual Sci, Univ Illinois at Chicago, Chicago, IL.

Glaucoma is a leading cause of irreversible blindness worldwide and has a substantial impact on the quality of life of millions of Americans. Primary open angle glaucoma (POAG) is the most common form of this disease and has a strong genetic component. POAG is usually asymptomatic early in the disease pathogenesis and can be far advanced before patients notice an extensive visual field loss. Therefore, early detection of this disease is important to prevent vision loss. Genomic profiles in combination with recent advances in statistical learning may provide a solution to identify individuals at increased risk for developing POAG. In recent years, genetic risk scores (GRS) have been used in risk prediction for many diseases, such as age-related macular degeneration, type 2 diabetes and coronary artery disease. Currently, few GRS studies have been reported for glaucoma. Furthermore, Hispanics/Latinos have historically been understudied and genetic studies in these admixed populations remain a challenge. Using genomics data available from the Mexican American Glaucoma Genetic Study (MAGGS), we carry out a precision-medicine study. In this study, we construct GRS for POAG including its endophenotypes and build risk prediction models using advanced statistical learning approaches, such as LASSO regression, artificial neural network and Xgboost. We evaluate their risk-classification performance using cross-validation and receiver operating characteristic analyses in these Hispanic/Latino subjects. Since the sample size of any single genetic study can be limited, we further conducted this investigation using large-scale whole-genome simulations (n=100,000). Our results demonstrate that increasing both the sample size of the training dataset and the number of genetic markers in the risk prediction models can improve the prediction accuracy.

2865F

Family based association tests of myopia reveal a potentially hidden association signal upstream of two *GABA* receptor genes. *C.D. Middlebrooks[1], C.L. Simpson[1,2], A.M. Musolf[1], L. Portas[3], F. Murgia[3], E. Ciner[4], D. Stambolian[5], J.E. Bailey-Wilson[1].* 1) National Human Genome Research Institute, Baltimore, MD; 2) Department of Genetics, Genomics and Informatics, Univ of Tennessee Health Science Center, Memphis, TN; 3) Institute of Population Genetics, CNR, Li Punti, Sassari, Italy; 4) Salus University, Elkins Park, PA; 5) Ophthalmology-Stellar Chance Lab, University of Pennsylvania, Philadelphia, PA.

   Myopia is an eye condition in which the light entering the eye does not focus on the retina resulting in distant objects appearing out of focus. Within recent years, the incidence and prevalence of myopia have increased in most populations and has reached epidemic proportions in several Asian countries. We have performed a family based association study using Exome Chip genotyping (Illumina Human Exome v1.1 array plus 24,263 custom SNPs) in five family cohorts for a total of 1,718 subjects in 261 families. These cohorts include Amish, Ashkenazi Jewish, African American, Caucasian and Chinese American families who have multiple individuals affected with myopia. Individuals in the families were defined as myopic if their average refractive error was <= -1 Diopter (D) and were considered unaffected if their average refractive error was > 0.0 D. Children were considered unaffected as follows: MSE>=+2D (ages 6-10) or MSE>=+1.5D (ages 11-20). After quality control, there were ~127,000 polymorphic SNPs available for analysis. Both gene-level and single-variant association analyses were performed using Family Based Association Test (FBAT) software. This resulted in a significant signal in a novel region upstream of two gamma-Aminobutryric Acid (*GABA*) receptor genes (*GABRA6*; *GABRB2*). GABA is a neurotransmitter that has previously been implicated in refractive development. The associated SNP, rs1373602, is not found in the Genotype-Tissue Expression (*GTEx*) project, but a nearby SNP, rs62381591, has been identified as an expression quantitative trait locus for the *GABRA6* gene. As the significant variant is common (~48% across populations), we wondered why the larger, population-based associations studies of Myopia have not found a signal in this regions. Upon further analysis, we learned that this variant is not in high Linkage disequilibrium (LD) with any other variants in our dataset (highest r² was ~0.002) and is indicated as triallelic in the 1000 genomes dataset (although it was biallelic in our smaller dataset).   Hence, this triallelic SNP may be filtered out before GWAS and there may not be another SNP that tags this region. We plan to follow-up this analysis by collaborating with groups that have performed genome-wide genotyping studies of Myopia to determine if this signal was missed due to the aforementioned reasons.

2866W

Examination of a rare risk variant in complement factor H for age-related macular degeneration in the Amish. *A.R. Waksmunski[1,2], K. Miskimen[3], Y.E. Song[3], R. Laux[3], D. Fuzzell[3], S. Fuzzell[3], L.D. Adams[4], L. Caywood[4], M. Prough[4], W.K. Scott[4], D. Stambolian[5], M.A. Pericak-Vance[4], J.L. Haines[2,3].* 1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, U.S.A; 2) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, U.S.A; 3) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, U.S.A; 4) Hussman Institute for Human Genomics, University of Miami, Miami, FL, U.S.A; 5) Ophthalmology and Genetics, University of Pennsylvania, Philadelphia, PA, U.S.A.

   The Amish constitute a genetically isolated and socially segregated population located in North America. The present-day Amish population arose from a few hundred founders, who immigrated to North America 300 years ago, resulting in a population bottleneck that has been sustained since that time. Members of this faith community practice endogamy and adhere to a strict, uniform lifestyle. Consequently, they are substantially more genetically and environmentally homogeneous compared to the general Caucasian population. Previous work identified a rare missense mutation for age-related macular degeneration (AMD) in complement factor H (rs570523689, *CFH* P503A) in the Amish. This variant was initially found in 19 Amish individuals from Ohio and Indiana. We performed custom TaqMan genotyping assays and identified an additional 33 carriers of this variant in these Amish populations. This variant was not identified in a cohort of 495 Amish individuals from Lancaster County, Pennsylvania, which may indicate that this variant segregated in particular Amish sub-isolates in the United States. Nearly all of the newly identified carriers are heterozygous for the variant and are the children, siblings, or close relatives of the original 19 *CFH* P503A carriers. Using data from the Anabaptist Genealogy Database, the carriers are connected through a 10 generation pedigree that dates back to the 1700s. The *CFH* P503A carriers appear to have a younger age of first AMD diagnosis compared to elderly non-carrier Amish individuals with AMD. The average age of first AMD diagnosis in the non-carriers is 85, and the average age of first AMD diagnosis in the carriers is 81. Studies are underway to calculate genetic risk scores for the carriers and closely related non-carriers to determine if the carriers have a higher genetic burden for AMD. Functional studies are ongoing to elucidate the consequences of this variant in individuals with AMD.

2867T

Genetic risk score is associated with vertical cup-disc ratio and improves prediction of primary open angle glaucoma in Latinos. *D.R. Nannini, H. Kim, F. Fan, X. Gao, Mexican American Glaucoma Genetic Study Group.* Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL.

Previous genome-wide association studies have identified multiple genetic variants associated with vertical cup-disc ratio (VCDR). Genetic risk scores (GRS) examine the joint genetic effect of individual variants on a trait by aggregating these separate genetic variants into a single measure. The purpose of this study is to construct a GRS for VCDR using previously published single nucleotide polymorphisms (SNPs) and to evaluate whether the GRS is associated with VCDR and whether the GRS increases the discriminatory ability for primary open angle glaucoma (POAG) in a Latino population. We performed a population based-association study using 4,018 Latino study participants. We constructed unweighted and weighted GRS using 68 previously reported VCDR SNPs. Linear and logistic regression analyses were performed to examine the association between both GRS with VCDR and POAG. To evaluate the discriminatory ability of the GRS for POAG, receiver operating characteristic (ROC) analyses were conducted. The GRS were significantly associated with VCDR ($P < 0.0001$), after adjusting for age, gender, central corneal thickness, intraocular pressure, and education. Compared to traditional risk factors for VCDR, the weighted GRS explained an additional 2.75% of the variation in VCDR. Moreover, compared to the lowest quintile of the GRS, study participants in the highest quintile experienced significantly higher odds of POAG. Addition of the weighted GRS improved the discriminatory ability for POAG during ROC analyses ($P = 0.03$). To our knowledge, we are the first to report the association between GRS and VCDR, and the improvement in discriminatory ability of POAG, in a Latino population.

2868F

Grouped association analysis for very rare variants using Fisher's Exact Test and external controls. *A. Kwong[1], D. Stambolian[2], E.Y. Chew[3], A. Swaroop[4], G. Abecasis[1].* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania Medical School, Philadelphia, PA, United States; 3) Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD, United States; 4) Neurobiology-Neurodegeneration and Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, MD, United States.

Developments in next-generation sequencing technology have led to an exponential increase in the number of very rare variants available for analysis in genetic association studies. Very rare variants with large effect sizes can provide specific clues about disease mechanisms, but identifying these variants is challenging. One major challenge is the requirement of very large sample sizes to achieve power for rare variant association testing, even for burden tests. We evaluate the potential for rare variant association studies to use large external control sets, such as variant lists and genotype counts from the Exome Aggregation Consortium (ExAC), to increase power. Our approach compares counts of very rare variants (allele frequency < 0.1%) in study cases and study controls to those in an external control sets using gene-level comparisons, incorporating variant list filters that consider a range of annotation information, including allele frequency, predicted function, and CADD score. Each individual comparison is based on Fisher's Exact Test for carrier status for rare variants vs. case-control status of samples. We tested this approach on our Age-Related Macular Degeneration (AMD) whole-genome sequencing case-control study with 2,394 cases and 2,393 controls, containing 46,946,619 variants in all. For this analysis, we focused on genes within known associated loci containing very rare frameshift, essential splice, or stop gain variants, totaling 652 variants in 392 genes. We discovered associations for genes with rare variants known to be associated with AMD (*CFH*, *CFI*), along with results suggesting additional potentially associated genes. In conclusion, our approach provides a way to increase the power of very rare variant association analysis by incorporating external controls.

2869W

The genetic architecture of the AVSD risk in Down syndrome: Results from chromosome 21 genome sequencing. *X. Blanc[1], F.A. Santoni[1], C. Borel[1], P. Makrythanasis[1], R. Sailani[1], M. Guipponi[1], S. Nikolaev[1], F. Bezrukov[1], E. Falconnet[1], P. Ribaux[1], U. Radhakrishna[1], F. Bajolle[2], D. Laux[2], J. Okoronkwo[2], M.A. Alcántara-Ortigoza[3], J. Rubens-Figueroa[3], B. Estandía-Ortega[3], A. González-del Angel[3], S. Vicari[4], K. Rabionet[5,6,7], R. Touraine[8], S. Kitsiou[9], H. Fryssira[9], G. Merla[10], L.A. Pérez-Jurado[6,11], J.M. Delabar[12], D. Bonnet[2], S.E. Antonarakis[1,13].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Geneva 1211, Switzerland; 2) M3C-Necker, Pediatric Cardiology, University Paris Descartes, 75015 Paris, France; 3) Department of Human Genetics, National Institute of Pediatrics, Mexico City, Mexico; 4) Department of Neuroscience, Children's hospital Bambino Gesù, 00165 Rome, Italy; 5) Centre for Genomic Regulation, Biomedical Research Park (PRBB), E-08003 Barcelona, Spain; 6) Hospital del Mar, Research Institute (CRG), E-08003 Barcelona, Spain; 7) CIBER in Epidemiology and Public Health (CIBERESP), E-08003 Barcelona, Spain; 8) CHU de Saint-Etienne, Hôpital Nord, 42055 Saint-Etienne Cedex 2, France; 9) Department of Medical Genetics, University of Athens, 11527 Athens, Greece; 10) Medical Genetics Service, IRCCS, 71013 San Giovanni Rotondo, Italy; 11) Department of Experimental and Health Sciences, Universitat Pompeu Fabra, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), E-8003 Barcelona, Spain; 12) Unité de Biologie Fonctionnelle et Adaptative (BFA), CNRS EAC 4413, Université Paris Diderot, Sorbonne, Paris Cité, 75205 Paris, France; 13) iGE3 Institute of Genetics and Genomics of Geneva, Geneva 1211, Switzerland.

Down syndrome (trisomy 21; T21) patients are often born with congenital heart defects (CHD), including atrioventricular septal defect (AVSD). The incidence of AVSD is 43% in CHD T21 patients (17% in whole T21 population) and very rare in the general population. The genetic architecture of CHD has been previously studied with SNP arrays; with this approach, our group identified two SNPs and three CNVs regions significantly associated with the risk of CHD (Sailani et al., Genome Research, 2013). Here we use chromosome 21-targeted deep sequencing (>120x) to perform a case-control association study in a T21 population to identify SNPs, structural variants and CNVs associated with the AVSD in T21. The study cohort includes 102 Down syndrome patients with AVSD and 145 Down syndrome patients without detectable cardiac defect as controls. A total of 220671 SNPs and 21756 short indels has been called through our in-house pipeline implementing BWA, GATK and Annovar. The association study has been performed with Fisher's exact test and logistic regression following different genetic models. No variant has reached strict 5% Bonferroni level significance; however, the analysis revealed several prominent loci such as rs57685080 (p=1.2x10^{-6} on allelic model), a 3 bp deletion 2.9 kb upstream APP, or rs79882024 (p=3.5x10^{-6} on allelic model, p=7.9x10^{-6} on additive model, p=7.8x10^{-6} on dominant model), a SNP located within an intron of *PCP4*. Epistasis analysis using logistic regression showed an interaction between rs2824960 (within an intron of MIR548XHG) and rs9975251 (50.1 kb upstream MRPS6) with p=2.1x10^{-8}, albeit not reaching 5% Bonferroni level. We have also performed CNVs detection with CNVkit and Nexus and association analysis with CNVRuler. This study identified 4 CNVs, one spanning over both exonic and intronic regions of *FAM207A* (3.9 kb, FDR = 0.029) and others over intronic regions of *PCBP3* (0.87 kb, FDR = 0.029) and *CLDN14* (9.6 kb, FDR = 0.029). This study shows that common variants on chromosome 21 may contribute to the developmental phenotype of AVSD and provide insights into the understanding of the genetic mechanisms underlying the development of congenital heart defects. A larger sample size may better characterize the genetic architecture of the chromosome 21 contribution to the AVSD.

2870T

Tracing the dark matter: Prevalence of copy number variants across Mendelian disorders. *R. Truty, S. Lincoln, M. Kennemer, J. Paul, E. Gafni, S. Fay, R. Nussbaum, S. Aradhya.* Invitae, San Francisco, CA.

The prevalence of exonic copy number variants (CNVs) in the human genome is not well understood. We applied a next-generation sequencing method to simultaneously detect sequence variants and intragenic CNVs in a large population undergoing clinical testing for neurological, pediatric, hereditary cancer, or cardiac disorders. From testing more than 76,000 unrelated individuals for subsets of 1002 genes (the equivalent of 2.2 million single-gene tests), we identified 1307 clinically reportable CNVs in 221 genes, including 830 deletions and 477 duplications. While CNVs were observed in only 1.7% of patients, they were pathogenic in 9.3% of patients with a positive finding. Most deletions (93%) were pathogenic while only 46% of duplications were. Only 17% of CNVs included an entire gene and, in several instances, these likely represented larger cytogenetic events encompassing several neighboring genes. The frequency of intragenic CNVs among pathogenic variants ranged from 4.8% in cardiac disorders to 32.2% in neurological disorders. CNVs were highly prevalent in neuropathies, muscular dystrophies, neurofibromatosis, epilepsy, familial hypercholesterolemia, and other conditions with loss-of-function (LOF) mutational mechanisms, but were rare in Noonan syndrome, cardiomyopathies, and others which mostly result from gain-of-function variants. We also evaluated genes unrelated to the presenting clinical phenotype in >76000 individuals to estimate the baseline population prevalence of non-morbid intragenic CNVs. Up to 600 genes were examined per individual, representing a total of 13 million single-gene analyses. We observed 2676 high-quality CNVs, with an average of one CNV for every 4534 genes sequenced. The majority of CNVs (64%) were duplications. Most (82%) CNVs were heterozygous variants in recessive genes and/or in genes in which disease-causing variants do not show a LOF mechanism. We identified commonly polymorphic CNVs (e.g., at *NPHP1* and *NIPA1*) but most were novel, rare events. Our data demonstrate that universal exon-level CNV analysis is valuable in diagnostic testing, particularly in pediatrics and neurology, because these events explain a disproportionately high frequency of pathogenic variants. This study describes one of the large clinical datasets on small intragenic CNVs and provides a deeper view into their prevalence in human genes.

2871F

Iterating from discovery to epidemiological consequence through disease mechanism. *J. Brown, G. Unlu, E. Gamazon, E. Knapik, N. Cox.* Vanderbilt University, Nashville, TN.

In early studies of PrediXcan in BioVU, we observed an association of *GRIK5* with many different eye phenotypes (glaucoma, cataract, vitreous body abnormalities, retinal detachment, etc.), and subsequent studies of zebrafish knockout of the ortholog confirmed the importance of normal expression of this gene for normal eye development in the zebrafish. More detailed studies have uncovered vascular biology as the primary mechanism underlying these effects. The *GRIK5* ortholog is most highly expressed in 3-day old embryos in the developing eye, ear, and brain, and we observe fewer blood vessels in these tissues as well as leaky blood vessels in both knock-out and knock-down models. Motivated by these observations, we are making a systematic study of the relationship of vascular phenotypes and eye disease in both Bio-VU, the biobank at Vanderbilt University with more than 240,000 samples with DNA, and the entire Synthetic Derivative, the clinical data warehouse with the deidentified and continuously updated image of the electronic health records (EHR) in more than 2.7 million subjects receiving healthcare at Vanderbilt University Medical Center. If vascular biology is the primary genetic mechanism giving rise to the increased risk of eye disease in individuals with reduced expression of *GRIK5*, we hypothesized that eye diseases would be found at increased frequency in individuals reported to have congenital anomalies of the vascular system. We found that subjects with congenital anomalies of the vascular system have a 2-fold increased risk of glaucoma, cataract, retinal detachment, all types of macular degeneration, hypotony, disorders of the vitreous body, scleritis and episcleritis and visual field defects of the eye, and a 3-fold increased risk of diabetic retinopathy. Subjects with congenital anomalies of the vascular system also have a 3-fold increased risk of congenital anomalies of the eye. Many eye phenotypes show a smaller increased risk, not quite 2-fold, including disorders of globe, retinal disorders, disorders of refraction and accommodation, optic atrophy, pain/swelling/discharge of eye, and cancer or neoplasm of the eye. These studies illustrate the utility of iterating between human and model systems as well as the advantages of being able to consider validation studies at the level of phenome relationships in a large clinical data warehouse.

2872W

Genome-wide association analyses in large-scale multi-ancestry cohorts: Statistical challenges and opportunities. *C. DeBoever[1,2], C.R. Gignoux[3], S. Buyske[4], C.D. Bustamante[1,2], M.A. Rivas[1].* 1) Department of Biomedical Data Science, Stanford University; 2) Department of Genetics, Stanford University; 3) Department of Biostatistics and Center for Personalized Medicine, University of Colorado Anschutz Medical Campus; 4) Department of Statistics and Biostatistics, Rutgers University.

Large-scale sequencing projects and biobanks with phenotype data such as the NIH Genome Sequencing Program, which plans to sequence 200-250k whole genomes from a variety of ancestries, present new opportunities to identify genetic associations using larger samples sizes and more diverse cohorts than previous GWAS. However, the size and design of these initiatives also pose new statistical challenges for GWAS. Here, we investigated the utility of diverse GWAS cohorts by developing a Bayesian mixture model to identify genes that are enriched for rare loss-of-function (LoF) variants in subjects with African (AFR) and admixed American (AMR) ancestry relative to Europeans (NFE) in the gnomAD database. We find that 3,950 genes and 3,553 genes are enriched for LoF variants in AFR and AMR subjects, respectively, of which 564 and 659 are associated with traits through GWAS. For instance, *IL2RA* (type 1 diabetes, Crohn's, rheumatoid arthritis, multiple sclerosis, inflammatory bowel disease) LoF alleles are present at frequency 0.003% in NFE but are ~2.5 fold enriched in AFR and AMR and are observed at frequencies 0.007% and 0.006% respectively. Similarly, *CDKN2A* (type 2 diabetes, myocardial infarction, coronary artery disease, endometriosis, glaucoma) LoF alleles are at frequency 0.016% in NFE but are observed twice as frequently (0.033%) in AMR. The differing frequencies of LoF alleles demonstrates the utility of diverse GWAS cohorts for the purpose of pursuing knockout carriers and identifying gene-disease associations. While including subjects from diverse ancestries in GWAS allows for testing a broader range of genes and variants, it also poses challenges for statistical analyses. To explore this, we simulated hundreds of thousands of genomes from different ancestries using msprime, a coalescent simulator that models human demographic history, and examined how the power to detect associations changes under different study designs. We also examined the effect of using unascertained controls on the power to detect rare variant associations and found that both disease prevalence and the case/control ratio can strongly affect statistical power. In particular, we find that the power to detect rare protective variants is more sensitive to the case/control ratio than the power to detect rare risk variants. This work demonstrates the value of utilizing subjects from diverse ancestries in GWAS and suggests guidelines for effectively analyzing such datasets.

2873T

Genetic factors that modulate the relationship between education and Alzheimer's disease. *R.A. Bhatta[1], S.-Y. Chou[2,3], Y. Zhao[1], L. Qu[1], A. Kuzma[1], G. Schellenberg[1], L.-S. Wang[1].* 1) Penn Neurodegeneration Genomics Center, Dept of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 2) Dept of Economics, Lehigh University, Bethlehem, PA; 3) National Bureau of Economic Research, Cambridge, MA.

Neither genetics nor environmental factors alone fully explains observed phenotypic variation in Alzheimer's disease (AD), suggesting a possible role for gene-environment interactions in the etiology of AD. The purpose of this study is to examine one environmental factor implicated to have a protective effect against AD: education. Specifically, using known genetic risk/protective factors for AD, we consider how gene-education interactions relate to AD risk. Using phenotype data from National Alzheimer's Disease Coordinating Center (NACC) and HRC imputation data from Alzheimer's Disease Genetics Consortium (ADGC), we employ a regression approach utilizing linear probability models. All models control for sex, race/ethnicity, APOE, APOE interacted with sex, age at first visit, genotyping cohort, and study recruitment site. Standard errors are clustered by recruitment site, allowing for within-site error correlation. In the first step of our analysis, we interact the top genome-wide significant AD single nucleotide polymorphisms (SNPs) (see Lambert et al., 2013) with indicator variables for education level. The main outcome variable, the likelihood of AD, is regressed on these interaction terms and all control variables. Any statistically significant SNP-education interactions are used for localized analysis in the second step, in which we incorporate all SNPs located within close proximity of this "top" significant SNP. Results from the first step identify a nominally significant, negative interaction between education and rs2718058 (7:37841534), a SNP located near the *NME8* gene. The negative coefficient estimates indicate that higher educated groups possessing a greater number of minor alleles are less likely to develop AD when compared to the least educated group, i.e. the effect of education is amplified by the minor (in this case, protective) allele. Upon closer examination of neighboring SNPs, located within ~300Kbp of this top SNP, we discover both positive and negative interactions with education. These localized results suggest the possibility of at least three haplotypes conferring interaction effects with education level. Further analyses will exploit advanced econometric techniques as well as Mendelian Randomization for causal inference on the relationship between education and AD.

2874F

Vitamin D deficiency: Analysis shows season and dietary vitamin D intake influence the effect of *GC*, *CYP2R1*, *DHCR7* and *CYP24A1* genes on vitamin D levels. *K.E. Hatchell, C.D. Engelman.* Population Health Sciences, University of Wisconsin-Madison, School of Medicine and Public Health, Madison, WI.

Statement of Purpose: Vitamin D deficiency is determined both environmentally (i.e. diet and sun exposure) and genetically, and is associated with numerous adverse health effects. Genome-wide meta-analyses have shown that *GC*, *CYP2R1*, *DHCR7/NASYN1* and *CYP24A1* play key roles in the biological mechanism of vitamin D deficiency. We sought to determine the additive effect of the most influential SNP in each of these genes on 25-hydroxyvitamin D (25(OH)D) levels in the blood while stratifying by season and dietary vitamin D intake. Methods Used: Participants, ages 30-74, were selected from the Survey of the Health of Wisconsin. We calculated a weighted (by beta-coefficients from a linear regression model) polygenic risk score (PRS) using the most significant SNP in *GC*, *CYP2R1*, *DHCR7/NASYN1* and *CYP24A1*. Data were analyzed using linear regression controlling for sex, age, season of blood draw, sun exposed and inherited (unexposed) skin color measured by spectrophotometry, vitamin D intake from food and supplements, waist circumference, and genetic ancestry. Two additional models tested the interaction between the PRS and season of blood draw or quartile of dietary vitamin D intake. Summary of Results: The PRS ranged from 0 to 13.3. Preliminary results (N=200) for subjects of European descent, show effect modification of the association between the PRS and 25(OH)D blood levels by both season of blood draw and quartile of dietary vitamin D intake. When stratified by season, the effect of each one unit increase in the PRS on lowering 25(OH)D blood levels is larger in summer (b=-1.1 ng/ml; p<.0001) compared to winter (b=-0.44 ng/ml; p=.29; PRS X season interaction p=0.25). When stratified by quartile of dietary vitamin D intake, the effect of each one unit increase in the PRS on lowering 25(OH)D blood levels is larger as dietary vitamin D intake increases: b=-0.38 (p=.35), -0.65 (p=.18), -0.82 (p=0.07) and -1.27 (p=.006) ng/ml for the 1st, 2nd, 3rd and 4th quartiles of intake, respectively (PRS X dietary intake interaction p=0.03). These findings replicate and expand the findings of our previous study in an independent cohort and have important implications for both study design and precision medicine.

2875W

Evidence of *ZKSCANS, SULT2A1, TRIM4* and *BCL2L11* for serum dehydroepiandrosterone sulfate (DHEAS) levels: Replication from the Long Life Family Study (LLFS). *P. An[1], B. Thyagarajan[2], J.H. Lee[3], M.K. Wojczynski[1], J. Sanders[4], P.A. Lenzini[1], S.J. Lin[1], A. Yashin[5], K. Christensen[6], T. Perls[7], A.B. Newman[8], M.A. Province[1].* 1) Department of Genetics Division of Statistical Genomics, Washington University School of Medicine, St Louis, MO; 2) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN; 3) Sergievsky Center / Taub Institute, Columbia University Medical Center, New York, NY; 4) Massachusetts General Hospital Department of Medicine, Boston, MA; 5) Duke University Population Research Institute, Durham, NC; 6) Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, Odense, Denmark; 7) Division of Geriatrics, Department of Medicine, Boston University Medical Center, Boston, MA; 8) University of Pittsburgh, Graduate School of Public Health, Department of Epidemiology, Center for Aging and Population Health, Pittsburgh, PA.

Dehydroepiandrosterone sulfate (DHEAS) is an adrenal androgen and the major precursor of sex steroids. Levels of DHEAS decline substantially over the life course. DHEAS has a modest but significant genetic component (h^2=15% in the Long Life Family Study, LLFS). A recent meta-analysis of seven GWA studies in 14,846 subjects of Caucasian origin revealed eight common loci (*BCL2L11, ARPC1A, ZKSCAN5, TRIM4, HHEX, CYP2C9, BMF, SULT2A1*) associated with DHEAS levels. To validate these loci, we assessed eight reported SNPs and conducted association tests among 3,686 participants of Caucasian origin in the LLFS. Further, we assessed their pleiotropic effects on healthy aging index in the LLFS. DHEAS levels were natural log-transformed to approximate normality, and adjusted for age, sex, field centers, and principal components for ancestry. A linear mixed effects model, assuming additive genetic effects, was used for association tests; a kinship model was used to correct for random effects of relatedness among family members. In this analysis, $p<0.006$ (0.05/8) was considered significant for replication. Four of the eight SNPs were replicated. They included *ZKSCAN5* (rs11761528, p=1.649e-5, minor allele frequency, MAF=0.08), *SULT2A1* (rs2637125, p=6.549e-5, MAF=0.17), *TRIM4* (rs17277546, p=2.375e-5, MAF=0.05), and *BCL2L11* (rs6738028, p=1.404e-4, MAF=0.43). These SNPs may have regulatory signals and long-range interactions according to our GWAS3D, GSEA, TSEA, CSEA, HaploReg and RegulomeDB queries. Each accounted for less than 1% phenotypic variation in DHEAS levels. Interestingly, *SULT2A1*-rs2637125 ($p<0.01$) was also associated with healthy aging index (p=0.003, β=-0.2) among women in the LLFS. The *SULT2A1* encodes sulfotransferase that converts DHEA to DHEAS, and is expressed in the adrenal cortex and liver. The remaining four SNPs (*ARPC1A*-rs740160, *HHEX*-rs2497306, *CYP2C9*-rs2185570, *BMF*-rs7181230) were not replicated in this cohort. Of the eight known candidate loci associated with DHEAS, we replicated *ZKSCAN5, TRIM4, BCL2L11* and *SULT2A1* in the LLFS, along with initial evidence supporting roles of *SULT2A1* in aging mechanisms.

2876T

Gene by environment interaction in human longevity as observed in Danish birth cohorts from 1905 to 1915. *Q. Tan.* Epidemiology, University of Southern Denmark, Odense, Denmark.

The frequencies of variants of major candidate genes (APOE and FOXO3A) have been shown to vary in long-lived individuals of different Danish birth cohorts, e.g. with significantly decreased frequencies in the recent birth cohorts for the e4e4 genotype and also for the minor allele of rs7762395 of FOXO3A gene. We collected genotype and survival information on two birth cohorts, 1905 (1424 subjects) and 1915 (1105 subjects), aged over 95 years at in-take. Since genotype frequency in a population cannot be expected to change over a very short period of 10 years (1905 to 1915), we assume that the reported change in genotype frequencies could reflect cohort-specific genetic risk on mortality resulted from gene-environment interaction. Based on the same observations and recently updated mortality information in the two birth cohorts, we conduct a survival analysis introducing cohort- and sex-specific population survivals from population statistics and specifying and estimating genotype-specific relative risks in cohort 1905 and change in relative risk from 1905 to 1915, with consideration of unobserved frailty. We estimated a trend of risk reduction for APOE4 allele carriers (-0.119, 95% CI: -0.355 to 0.119) while the risk for APOE2 allele carriers remained constant (-0.007, 95% CI: -0.200 to 0.206), in nonagenarians and centenarians. The risk of death for carriying the rs479744 minor allele of FAXO3A showed a trend of increase (0.106, 95% CI: -0.049 to 0.292) while that for rs7762395 was unchanged (-0.038, 95% CI: -0.204 to 0.156). Our novel analysis provided evidence to the varying risks of major candidate genes on mortality at advanced ages due to interaction with the changing environment.

2877F

Genetics of the human microbiome and implications in obesity associated measures. *C.T. Finnicum[1], S. Doornweerd[2,3,4], C.V. Dolan[3,5], J.M. Luningham[6], J.J. Beck[1], C. Davis[1], G. Willemsen[3,5], E.A. Ehli[1], D.I. Boomsma[1,3,5], R.G. Ijzerman[2,3,4], G.E. Davies[1,5], E.J.C. de Geus[3,5].* 1) Avera Institute for Human Genetics, Avera McKennan Hospital and University Health Center, Sioux Falls, SD; 2) Department of Internal Medicine, Vrije Universiteit Medical Center, Amsterdam; 3) Amsterdam Public Health Research Institute, Vrije Universiteit medical center, Amsterdam; 4) Amsterdam Cardiovascular Sciences, Vrije Universiteit Medical Center, Amsterdam; 5) Department of Biological Psychology, Behavioral and Movement Sciences, Vrije Universiteit, Amsterdam; 6) . Department of Psychology, University of Notre Dame, Notre Dame, IN.

　Background: The microbial organisms residing within the gut microbiota of humans have been shown to play a pivotal role in human health and disease development. One area in particular that has garnered a lot of interest is the study of the gut microbiota in relation to the development of obesity. Although it is becoming clear that there is a relationship between these two factors, more work is needed to parse out the intricacies of this relationship. **Methods:** In order to explore the relationship between BMI and gut microbiota composition, two separate genetically informative study designs have been formed from participants of the Netherlands Twin Register (NTR). The first group of individuals (N=50) was selected from a large population (N=11,495) for which BMI and polygenic risk score scores for BMI were available. This allows for the use of a four-corner design where the study participants are selected from the top and bottom 25% of the BMI distribution, and the top and bottom 20% of the distribution of BMI polygenic risk scores. The second design that will be used to study the relationship between BMI and the gut microbiota consists of monozygotic (MZ) twin pairs (N=30, 15 pairs total) discordant for BMI (d = > 2 kg/m^2). Fecal samples were collected for all individuals and subjected to DNA extraction. Microbiota composition was explored through the use of 16s rRNA sequencing. **Results:** Our results highlight a negative association between BMI and alpha diversity of the gut microbiota. The low genetic risk / high BMI group of individuals had a lower gut microbiota alpha diversity when compared to the other three groups. Within the discordant MZ design the gut microbiota of the heavier twins had a lower average inverse Simpson value relative to the mean value of the lean twin counterparts; however, this difference failed to reach significance. There were 7 operational taxonomic units (OTUs) associated with either a leaner or heavier phenotype, with enrichment for OTUs classified to the *Ruminococcaceae* and *Oxalobacteraceae* taxonomic families. **Conclusion:** Our study presents evidence of a negative relationship between both body fat and BMI with the alpha diversity of the gut microbiota. In addition to these findings, a number of OTUs were found to be significantly associated with host BMI. These findings may highlight separate sub-types of obesity, one driven by genetic factors, the other more heavily influenced by environmental factors.

2878W

Leukocyte Telomere Length (LTL) as a marker of biological aging in Iranian healthy adult population: Report on assay establishment and recent finding. *F. Larti[1], S. Dehdahsi[1], F. Larti[2], A. Rad[3], R. Jazayeri[1,4], M. Mohammad-Zadeh[3], K. Kahrizi[1], H. Najmabadi[1].* 1) Genetics research center, University of social welfare and rehabilitation sciences, Tehran, Iran; 2) Cardiology Department, Imam Khomeini Complex Hospital, Tehran University of Medical Sciences, Tehran, Iran; 3) Cellular and Molecular Research Center, Sabzevar University of Medical Sciences, Sabzevar, Iran; 4) Department of Biochemistery, Genetic and Nutrition, Faculty of Medicine, Alborz University of Medical Sciences, Karaj, Iran.

　Leukocyte telomere length (LTL) is a marker of biological aging that can be affected by several factors including ethnicity. In population-based human disease studies, especially age-related diseases, knowing the actual body age is very important in interpretation of the results. To investigate the normal variation of LTL in Iranian population, and establish the assay for LTL measurement, we aimed to study relative LTL in a cohort of 1000 healthy adult individuals (>25 years) including eight different Iranian ethnicities, by qPCR method. At the phase of pilot study, 72 healthy individuals (29-77 years), 46 male and 28 female from Persian ethnicity were selected. All subjects were confirmed healthy by detailed clinical and paraclinical tests. They were not related and all lived at least ten years in their current environment. A number of 173 subjects (19-85 years) were also selected from Genetics Research Center DNA bank with unknown health status to examine the reliability of the method. All samples quantified against a 5 point standard curve using a mixture of 5 genomic DNA with different chronological ages as a reference DNA. K562 genomic DNA was included in each plate as a quality control and to avoid more than 5% coefficient of variation between different plates. Analysis was done based on relative standard curve using standards to determine the relative quantity of telomere for each sample. The ratio of quantity of telomere to single copy gene (36B4) for each sample was calculated and normalized by this ratio for K562 genomic DNA from the same plate. Data analysis showed that telomere length decreased as age increased. By classifying the 72 subject, in 10-year age groups, a reverse correlation was found between age and relative LTL (R = -0.886), that is fitted in a linear regression model (R^2= 0.730, P-value= 0.019). Reverse correlation (R = -0.882) and linear regression (R^2= 0.734, and P-value= 0.009) were also found for all studied subjects (245 samples), in 10-year age groups. Due to small sample size, analysis in 5-year age groups and without making groups, showed a weak reverse correlation (R=-0.197) between LTL and age. Data presented here, elucidate that variation in LTL may be significantly different in every decade of life, but not in every age of life. Analysis of raw data related to all subjects including other Iranian ethnicities is ongoing.

2879T

Old before our time: Biological ageing in an ethnically diverse cohort of preschool children. *K.N. Ly, C.G. Walker, S.D. Berry.* University of Auckland, Auckland, New Zealand.

Background: Substantial and accumulating evidence exists to link the early life environment with long term health outcomes. However, the precise biological mechanisms that explain how such early life events influence adult disease risk are poorly understood. One proposed mechanism for the mediation of exposure events on later disease risk is via accelerated biological ageing, for which telomere length is understood to be a useful and valid biomarker. Telomeres are specialised DNA structures located at the ends of chromosomes that function to maintain genomic stability. Telomeres naturally shorten with age, and accelerated telomere shortening is observed in individuals experiencing stressful or adverse environments. **Methods:** This study utilised data from the *Growing Up in New Zealand* longitudinal study. Participants provided saliva samples for DNA extraction when they were 4½ years of age (n=4587). DNA was extracted and relative telomere length (RTL) was measured using quantitative PCR (n=4388). Child ethnicity was determined by mother report at 4½ years of age. Linear regression was used to compare mean differences in RTL associated with gender and ethnicity. For analysis, child ethnicity was defined based on parent report and included European, Māori, Pacific, Asian, and Other. **Results:** There was a significant effect of both gender and ethnicity for RTL. Females had longer telomeres than males ($P=5.8\times10^{-7}$) and Pacific people had longer telomeres than all other ethnic groups (4.9×10^{-26}). **Conclusion:** Variation in telomere length is evident in preschool children and differs with gender and ethnicity. Further research is required to determine the basis of these differences, including how exposure to stressors throughout pregnancy and postnatal life impact biological age.

2880F

Admixture mapping of asthma in individuals of mixed African ancestry reveals a novel association on chromosome 6q23.2. *M. Daya[1], N.M. Rafaels[1], S. Chavan[1], M. Boorgula[1], T.M. Brunetti[1], A. Shetti[1], C.R. Gignoux[1], G. Wojcik[2], R. Johnston[3], E.G. Burchard[4], L. Caraballo[5], G.M. Dunston[6], M.U. Faruque[7], E.E. Kenny[8], J.M. Knight-Madden[9], D.A. Meyers[10], E.R. Bleecker[10], C. Ober[11], C. Rotimi[12], J.G. Wilson[13], L.K. Williams[14], H. Watson[15], Z.S. Qin[3], M.A. Taub[16], T.H. Beaty[17], I. Ruczinski[16], R.A. Mathias[17], K. Barnes[1], CAAPA consortium.* 1) Colorado Center for Personalized Medicine, University of Colorado, AURORA, CO; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; 3) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA; 4) Department of Bioengineering and Therapeutic Sciences , University of California, San Francisco, CA, USA; 5) Institute for Immunological Research, Universidad de Cartagena, Cartagena, Colombia; 6) Department of Microbiology, Howard University College of Medicine, Washington, DC, USA; 7) National Human Genome Center, Howard University College of Medicine, Washington, DC, USA; 8) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) Caribbean Institute for Health Research, The University of the West Indies, Kingston, Jamaica; 10) Department of Medicine, University of Arizona College of Medicine, Tucson, AZ, USA; 11) Department of Human Genetics, University of Chicago, Chicago, IL, USA; 12) Center for Research on Genomics & Global Health, National Institutes of Health, Bethesda, MD, USA; 13) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA; 14) Center for Health Policy & Health Services Research, Henry Ford Health System, Detroit, MI, USA; 15) Faculty of Medical Sciences, The University of the West Indies, Queen Elizabeth Hospital, Bridgetown, St. Michael, Barbados; 16) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore, MD, USA; 17) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, USA.

RATIONALE: Asthma is a complex disease with striking disparities across racial and ethnic groups, which may be partly attributable to genetic factors. One of the main goals of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) is to discover genes conferring risk to asthma in populations of African descent. We therefore performed admixture mapping of asthma in 6,686 African American and African Caribbean subjects genotyped on a variety of GWAS platforms plus the African Diaspora Power Chip (ADPC). **METHODS:** RFMix was used to infer local ancestry. The 1000 Genomes Project CEU and YRI populations were specified as reference populations. A total of 273,738 phased SNPs that passed quality control filters and overlapped in both CAAPA and reference population data sets were used as input for inferring local ancestry. A total of 15,824 local ancestry segments were estimated. Linear mixed models were used to test for association between the number of copies of African ancestry at each local ancestry segment (0/1/2) and asthma case-control status. The models incorporated a covariance matrix to account for relatedness between all individuals, as well as population structure to account for global ancestry. The originating CAAPA study data set was included as fixed effect covariate. **RESULTS:** An admixture mapping peak was identified on chromosome 6q23.2, with increased African ancestry associated with increased risk of asthma. One segment of local ancestry within this peak, between hg19 base pair positions 134,149,974-134,300,365, crossed the multiple testing threshold of p < 1.9e-4. Genes falling in this segment include the *Transcription Factor 21 (TCF21)* and *TATA-Box Binding Protein Like 1 (TBPL1)* genes. A previous study showed the *TCF21* gene as differentially expressed in bronchial biopsies of asthmatics compared to controls. The SNP with the most significant asthma association in CAAPA in this region, rs111966851, has a minor allele frequency of 0.308 and 0.006 in Africans and Europeans, respectively. **CONCLUSIONS:** We identified a novel asthma association signal on chromosome 6q23.2 that warrants further investigation.

2881W

Methods to estimate heritability of complex traits under a variety of complex genetic architectures. *L. Evans[1], R. Tahmasbi[1], S. Vrieze[1], G. Abecasis[2], S. Das[2], D. Bjelland[1], T. deCandia[1], M. Goddard[3], B. Neale[4], J. Yang[5], P. Visscher[5], M. Keller[1,6], Haplotype Reference Consortium.* 1) Institute for Behavioral Genetics, University of Colorado, Boulder, CO; 2) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109; 3) Faculty of Veterinary and Agricultural Science, University of Melbourne, Parkville, Victoria, Australia; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 5) Institute for Molecular Bioscience and the Queensland Brain Institute, University of Queensland, Brisbane, 4072, Queensland, Australia; 6) Department of Psychology and Neuroscience, University of Colorado, Boulder, CO.

Statement of Purpose: Complex trait genetic architecture has important implications for disease genetics as well as providing fundamental understanding of human evolutionary genetics. Narrow-sense heritability (h^2) is a central parameter for this, and a number of methods—GREML, LD score regression, LDAK, treelets, etc.— estimate heritability from SNPs (h^2_{SNP}) using genome-wide markers in unrelated individuals. Recent reports have shown that such estimates can be biased or unbiased depending on the assumed genetic architecture. We compared these methods to understand their performance across different genetic architectures, levels of stratification, and marker type (common SNPs from arrays vs. sequence data) using simulated phenotypes in real whole-genome sequence data.   Methods Used: We used a subset of ~21K whole genome sequences from the Haplotype Reference Consortium to generate phenotypes (h^2=0.5) from 1K causal variants (CVs) chosen randomly from different minor allele frequency (MAF) categories from common to very rare. We simulated phenotypes using a variety assumed allelic effects (betas), including betas coupled to minor allele frequency or not, and betas a function of LD weighting from LDAK. We then used either the whole genome sequence data or positions found on commercial SNP arrays to estimate h^2 and examine genomic architecture in mixed effects models using genetic relationship matrices (GRMs) estimated from different methods across levels of stratification, specifically single component GREML and LDAK, and MAF- or LD-& MAF-stratified GREML.   Summary of Results: We confirm a recent report that the assumed genetic architecture can lead to biases in different methods when using dense whole genome sequence or imputed variant data. However, we demonstrate that model fit (log-likelihood) can be increased even when the estimate is strongly biased, and can therefore lead to incorrect heritability estimates. We also found that MAF- & LD-stratified GREML (i.e., GREML-LDMS) is robust to most misspecifications of the true relationship between MAF and allelic effect sizes, and is robust to many simulated relationships between LD weights and effect sizes. Thus, careful consideration of the robustness of methods to model misspecification must be taken when apply these methods to real data, and the use of an increase in log-likelihood can lead to biased heritability estimates.

2882T

Robust inference of population structure from next-generation sequencing data with systematic differences in sequencing. *Y.J. Hu[1], G. Satten[2], P. Liao[1].* 1) Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 30322, USA; 2) Centers for Disease Control and Prevention, Atlanta, GA, 30333, USA.

Inferring population structure is important for both population genetics and genetic epidemiology. Principal components analysis (PCA) has been effective in ascertaining population structure with array genotype data but can be difficult to use with sequencing data, especially when low depth leads to uncertainty in called genotypes. Because PCA is sensitive to differences in variability, PCA using sequencing data can result in components that correspond to differences in sequencing quality (read depth and error rate), rather than differences in population structure. We demonstrate that even existing methods for PCA specifically designed for sequencing data can still yield biased conclusions when used with data having sequencing properties that are systematically different across different groups of samples (i.e., sequencing groups). This situation can arise in population genetics when combining sequencing data from different studies, or in genetic epidemiology when using historical controls such as samples from the 1000 Genomes Project. To allow inference on population structure using PCA in these situations, we provide an approach that is based on using sequencing reads directly without calling genotypes. Our approach is to adjust the data from different sequencing groups to have the same read depth and error rate so that PCA does not generate spurious components representing sequencing quality. To accomplish this, we have developed a subsampling procedure to match the depth distributions in different sequencing groups, and a read-flipping procedure to match the error rates. We average over subsamples and read flips to minimize loss of information. We demonstrate the utility of our approach using two datasets from 1000 Genomes, and further evaluate it using simulation studies.

2883F

Model-based multiple variants test considering causal status. *J. Joo[1], F. Hornozdiari[2], E. Eskin[3].* 1) Department of Computer Science Engineering, Dongguk University-Seoul, South Korea; 2) Epidemiology Dept, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 3) Computer Science Dept, University of California, Los Angeles, Los Angeles, CA, USA.

Over the past decade, GWAS have successfully identified many variants associated with diseases and complex traits. The standard GWAS examine one variant at a time to identify causal variants. However, this univariate approach is often underpowered and usually explains only a small fraction of the genetic variation of complex traits. Several studies have demonstrated that multiple causal variants may exist in a region. For those regions, the standard GWAS may be inappropriate due to its low statistical power. Alternatively, an approach considering multiple causal variants simultaneously may increase statistical power by aggregating the effects of causal variants in a region. Unfortunately, we do not know in advance which variants are causal and there are too many possible causal status. Recently, there are many progresses in fine mapping approaches that try to identify causal variants in a region. Causal variants are the variants that are responsible for the association signal at a locus. However, at each locus, there are often tens to hundreds of variants tightly linked to the reported associated single nucleotide polymorphism, therefore, the linkage disequilibrium hinders the identification of causal variants at risk loci. CAVIAR is one of the recent fine mapping approach that estimates the probability of each variant being causal allowing an arbitrary number of causal variants by jointly modeling the association statistics at all variants in a region accounting for linkage disequilibrium. We propose a new model-based method referred to as MARS (Model-Based Association test Reflecting causal Status) that performs association test between variants in a region and a trait to increase statistical power of the test. We extend the likelihood model of CAVIAR to explicitly incorporate linkgae disequilibrium utilizing MVN distribution conditional on causal status of variants. Applied to both simulated and real datasets, MARS shows significant improvement on the statistical power compared to some of the representative previous multiple variants testing methods.

2884W

Caring without sharing: Genome-wide association and mapping on cohorts fragmented across institutional silos. *A. Pourshafeie[1], C. Bustamante[2,3], S. Prabhu[2,3].* 1) Physics Department, Stanford University, Stanford, CA; 2) Department of genetics, Stanford University, Stanford, CA; 3) Biomedical Data Sciences, Stanford University, Stanford, CA.

We present robust statistical inference on datasets fragmented across multiple institutions as a powerful paradigm for the future of gene mapping and discovery. Though the recent explosion in genomic and phenotypic data from next-gen biomolecular and biometric technologies, respectively, is leading to breakthroughs in our understanding of human disease - as participant data becomes more multifactorial and private, large cohort datasets can be expensive, unsafe and time consuming to generate or maintain at a single centralized facility. Today, "meta-analysis" techniques offer a work-around by combining summary statistics across multiple studies. However, there are a few limitations: (a) the shared summary statistics might be inadequate for some types of inference (e.g. estimating effects from shared two-tailed p-values), (b) subtle differences in models, assumptions and QC can introduce bias (Greco et al. 2013), and (c) parameter estimates can be unreliable when the sample size (N) is small compared to the number of covariates (k). Here, we explore algorithms for inference that allow participant data to remain fragmented across institutional silos. We show that the most popular genetic mapping model - generalized regressions - can easily be performed in such a setting. We compare the performance of three types of algorithms: four stochastic sub-gradient methods with large batch sizes, one stochastic quasi-Newton method and an Alternating Direction Method of Multipliers (ADMM) (Boyd et al. 2010) approach with the estimates from summary statistics and centralized algorithms on simulated datasets. We characterize these algorithms under a variety of conditions - type of predictor (genotype dosage, continuous, categorical), number of silos, sample sizes per silo, etc. Our results show that (a) the derived estimates are identical to simple regression on a centralized dataset (b) ADMM estimates reach the centralized estimates the fastest out of all explored methods and (c) Reliable odds ratio estimates are usually achieved in <1,000 iterations (network communications between the silos). Through simulations, we show that we can recapitulate the findings of seminal GWAS efforts like the WTCCC. Our proof of concept unequivocally demonstrates that multi-institutional mega-analyses can be performed, at scale, in a decentralized setting. A working draft and relevant experiments can be found https://github.com/apoursh/Decentralized_reg.

2885T

Some novel methods of detecting gene-drug interactions. *M. Rao[1], L. Zhang[1], Q. Wang[2], M. Medvedovic[1].* 1) Environmental Hlth, Univ Cincinnati, Cincinnati, OH; 2) Millennium Pharmaceuticals, Boston, MA.

The data that led to current research is the RNA-sequence data from 60 drugs and controls put forward by Mount Sinai Medical center. Each drug is administered to a certain number of cell lines and RNA sequences for 23,895 genes are recorded from each cell line. RNA-sequences are also recorded from a certain number of cell lines none of which received any drug (controls). <u>Goals</u> Compare each drug with the control. Identify drugs significantly different from the control for each gene. Identify blocks of genes and drugs with the property that in each block the drugs are significantly different from the control under each of the genes in the block. <u>Methods</u> Implement an ANOVA procedure a la Dunnett (1991) for comparing drugs with the control for each gene. Following the Dunnett procedure or its variations, the given data will be summarized by a data matrix Y whose rows are indexed by genes and columns by drugs. The entry in the i-th row and j-th column is equal to one if the j-th drug is significantly different from the control under the i-th gene. A bi-clustering method (2017) is invoked on the 0-1 matrix Y. The goal is to find maximal sub-matrices of Y in each of which every entry is equal to one. This is tantamount to identifying drugs significantly different from the control under each of the genes identified in the sub-matrix. <u>Results, Interpretation, and Illustration</u> We discuss extensively what each such sub-matrix (bicluster) is giving us. Identify the pathways of the genes identified by each bicluster discovered. Identify the medical conditions the drugs are meant for. Check with Drug x Medical Conditions and Drug x Gene networks. Several biclusters are identified with sizes 85x3, 14x4, 32x3, 5x5, etc, The bicluster of size 14x4 covers four drugs (Aprepitant; Pazopanib; Sunitinib+Estradiol; Sunitinib+Prednisilone) and fourteen genes (A1BG-AS1; ABCG2; BCHE; C10orf10; C6orf52; CLDN14; COL19A1; FGFBP3; PBK; PDC; PRG4; UBAP1L; UBE2C; XKR6). Drugbank also provides information about drugs and the genes they target. This information is contrasted with the information emanating from the bicluster analysis. <u>References</u> C.W. Dunnett and A.C. Tamhane (1991). Step-down multiple tests for comparing treatments with a control in unbalanced one-way layouts. *Statistics in Medicine*, 10, 939-947. Adetayo Kasim, Ziv Shkedy, Sebastian Kaiser, Sepp Hochreiter, and Willem Taloen (2017). Applied bi-clustering methods for big and high-dimensional data using R. CRC Press, New York.

2886F

Use low-depth and high-depth whole genome sequencing data to predict 36 blood groups. *Y. Sun[1], J. Huang[2].* 1) Department of mathematics & statistics, Boston university, Boston, MA; 2) Harvard Medical School, Boston, MA.

Background: Knowing one's blood type is important for both scientific and medical purpose. As genotyping becomes more common and cost-effective than phenotyping, there is a desire to derive human blood type from genetic data. Recently, whole genome sequencing data form the 1000 genomes' project has been reported to predict as many as 36 blood types, as a proof of principle. **Methods**: Here, we use high-depth whole genome sequencing data from 4,141 individuals enrolled in the Trans-Omics for Precision Medicine program (TopMed), together with the same 2,504 individuals with low-depth sequencing from the 1000 genomes project (1000GP), to predict each person's 36 blood types. We next compare the performance of these two datasets. The genotype data was pre-phased using EAGLE software on the genome-wide scale. The cDNA reference sequences for all known red blood cells and platelet genes listed in the Blood Group Gene Mutation Database were aligned to the human reference genome. **Results**: For TopMed and 1000GP respectively, we discovered a total of 1,671 and 1,330 missense variants in the coding regions of 36 blood group systems. Among them, 443 and 271 are mapped to known blood group polymorphisms, respectively. For 97 individuals in the TopMed dataset also having serological ABO blood types, the concordance between genetically predicated and phenotypically measured blood type is 99.0%. **Conclusion**: We demonstrated empirically the high accuracy of using genetic data to predict blood types. We further show that high-depth whole genome sequencing yields higher resolution of rare genetic variants and haplotypes that are associated with human blood types. More comprehensive evaluation of the performance of genetically predicated blood types would contribute to transfusion medicine and therefore precision medicine.

2887W

Modeling the interactions between coding and non-coding RNA by kernel machines in binary phenotypes. *S. Yang, F. Shao, W. Duan, Y. Zhao, F. Chen.* Nanjing Medical University, Nanjing, China.

The etiology of complex disease is likely associated to the genetic factors, along with the interactions between different level data, e.g. microRNA(miR-NA)-genes. The relations between miRNA and traits are mediated by gene expression. Most studies of the complex phenotype are on binary traits. The traditional chi-square test requires a large number of degrees of freedom to test the main effects and all the corresponding interactions under a parametric assumption, and hence suffers from low power. Consequently, we extend a powerful joint test for testing the effect of genes and their interactions on the binary outcomes. Considering the similarity between the variance component test and the kernel machine function, we construct a score test to test the garrote kernel function. The procedure of the parameter estimation involves iterations because of the logistic regression framework. The use of kernel functions is promising, because it integrates the information of the weight of genes by the prediction numbers. By type I error simulations, it is clear that the type I error of our method is controlled, rather than that of the traditional main-effect method disperses in some scenarios. The increase of differences between the null model and the whole model possibly leads to the increase of power of our method. Also, our model is hardly sensitivity to the different model assumption. We apply the method to the breast cancer data from The Cancer Genome Atlas.

2888T

High frequency of the MEFV c.1437C>G, p.F479L allele among Druze FMF patients. *V. Adir, O. Sadeh, A. Peleg.* Institute of Human Genetics, Carmel Medical Center, Haifa, Israel and the Ruth and Bruce Rappaport Faculty of Medicine, Technion, Israel Institute of Technology, Haifa, Israel.

Introduction: Familial Mediterranean fever (FMF) is a hereditary auto inflammatory disease characterized by brief recurrent episodes of synovitis, usually with accompanying fever. FMF disease is caused by mutations in the MEFV gene and is common among Mediterranean populations. Although FMF is considered to be an autosomal recessive disease, some proportion of heterozygotes also express the clinical symptoms. . **Results:** 328 samples of Druze, Muslims Arabs, Christian Arabs and Jewish patients with clinical symptoms suspected for FMF were tested. The mutations frequencies were: V726A (31%), E148Q (23%), M694V (21%), M680I (6%), M694I (6%), F479L (5.4%) P369S (3.2%), K695R (1.6%), A744S (1.6%), I692del (0%) and R761H (0%). It has been shown that some rare MEFV mutations tend to be over represented in particular ethnic groups. The reported general population c.1437C>G, (p.F479L) mutation frequency is lower than 1%. In our cohort, a frequency of 5.4% was found. Further analysis showed c.1437C>G, p.F479L mutation was limited to Arabs and Druze ethnicity samples. From a total 139 FMF mutated alleles among Arabs and Druze, eleven alleles were found to be F479L (7.9%) and eight of them were found in Druze patients from the same village. **Conclusion:** Our results show that the F479L allele is very frequent among Druze patients from one village in northern Israel. We recommend including this allele in the mutation panel for FMF testing of Druze patients.

2889F

Family-based rare variant association study of familial myopia in Caucasian families. *D. Lewis[1], C.L. Simpson[2], A.M. Musolf[1], K. Long[1,3], L. Portas[1,4], F. Murgia[1,4], E. Ciner[5], D. Stambolian[6], J.E Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, Baltimore, MD; 2) Genetics, Genomics and Informatics, University of Tennessee Health Sciences Center, Memphis, TN; 3) University of Texas at El Paso, El Paso, TX,; 4) Institute of Population Genetics, CNR, Li Punti, Sassari, Italy; 5) Salus University, Elkins Park, PA; 6) Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Myopia is a common refractive error (RF) which affects at least a third of most populations. Individuals with high myopia are vulnerable to ocular complications later in life. Consequently, great efforts have been undertaken to identify and understand the mechanisms underlying the development and progression of myopia. Genome-wide association studies and linkage studies have identified loci influencing the risk of developing myopia but, few causal variants have been identified with the majority of them being common (minor allele frequency > 0.05). Therefore, this study aims to identify regions associated with rare variants that increase myopia risk using dense Exome Chip genotyping (Illumina Human Exome v1.1 array plus 24,263 custom SNPs) data from 75 myopic Caucasian families from the Penn Family Myopia Study. Myopia was defined based on mean spherical equivalent in Diopters (D): affected (≤-1D), unaffected (≥0 D) or unknown (<0D, >-1 D). Children were considered unaffected as follows: MSE>=+2D (ages 6-10) or MSE>=+1.5D (ages 11-20). We used the rare-variant transmission disequilibrium test (RV-TDT) to perform gene-based association tests with rare variants in 25 Caucasian parent-child trios (independent completely-genotyped trios from the families above). After quality control, 27,121 SNPs were analyzed and RV-TDT analysis identified a suggestively associated locus on chromosome 2.q31.1for a non-synonymous coding SNP rs34564141 in *LRP2* after correcting for multiple testing (*P value* = 2.65x10[-3]). *LRP2* has been implicated in a study of a rare form of severe myopia in patients who had a mutation encoding a receptor *LRP2* in the retina. Greatly enlarged eyes have been observed in mice lacking this *LRP2* gene. We are extending these analyses to include 350 individuals from additional myopic family cohorts, including Ashkenazi Jewish and African American families, and will present these results.

2890W

Exploring the effect of minor allele frequency on the inflation of type I error rates for GWA studies of family data with non-normally distributed traits. *J.A. Sabourin, H. Sung, A.J.M. Sorant, A.F. Wilson.* CSGB/NHGRI/NIH, Baltimore, MD.

When the assumptions of an analysis method, such as the normality of the error distribution, are violated, it is often assumed that the type I error rate will be inflated. We previously investigated how the type I error rate in genetic association studies of unrelated subjects was affected by the minor allele frequency (MAF) of the SNV and the degree of deviation from normality. We observed that the deviation from normality only inflated type I error rates for rare SNVs and that transformation of the trait to reduce the degree of non-normality brought the type I error rate closer to the nominal rate. In this study, we extend this work to family data. Simulations were based on the Genetic Analysis Workshop 19 family data which is comprised of whole genome sequencing (8.4million SNVs) of 975 subjects from 20 extended families. Two types of "null" traits were considered. In the first, there was no genetic contribution; all the variability was explained by independent errors. In the second, the null trait had no detectable single SNV genetic effects, but a polygenic effect, whose structure is defined by the estimated kinship matrix, contributed to the phenotypic variability, with the remaining variability explained by independent errors. Several different proportions of the two components were considered. For each null model the normal distribution assumed by the analysis and a gamma distribution (with shape and scale parameters of 3 and 20, respectively) were considered for the independent errors. In addition, null traits based on the gamma distribution were analyzed after transforming with the $\log_{10}$ and rank-based inverse normal transformations. Two hundred replicates of each trait were analyzed. Samples were analyzed with EMMAX with structure assumed to follow the kinship matrix estimated by EMMAX on a pruned set of SNVs. For common SNVs, MAF ≥ 0.05, results showed no inflation of type I error rates in any traits investigated. For rare SNVs, MAF < 0.05, an inflation in type I error for non-normally distributed traits was observed. This inflation increased with both decreasing MAF and increasing degree of non-normality, although this effect was lessened as the proportion of the polygenic component increased. As in the previous study, reducing the non-normality of the trait by appropriate transformation reduces the inflation of type I error rates for rare SNVs.

2891T

To ERV is human: A phenotype-wide scan linking polymorphic human endogenous retrovirus-K insertions to tissue-specific gene expression and complex diseases. *A.D. Wallace[1], G. Wendt[2], AJ. De Smith[3], C. Metayer[1], J. Wiemels[3], LF. Barcellos[1], SS. Francis[2,3].* 1) Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) University of Nevada Reno, Reno, NV; 3) University of California San Francisco, San Francisco, CA.

Approximately 8% of the human genome is comprised of ancient endogenized retroviruses (ERVs) that integrated into germ cells and were subsequently preserved. They are established regulators of gene expression and thus serve as excellent candidates for causal variants contributing to the heritability of complex phenotypes. Until recently it was assumed that the locations of ERV insertions in the genome were fixed across individuals, thus their potential contribution to phenotypic variation was largely overlooked. Recent studies have shown that the most recent ERV integration, HERV-K, is polymorphic in some genomic locations and is actively expressed in a range of conditions from cancer to neurologic diseases to HIV infection. Using an in-house computational pipeline and whole genome sequencing data from the 1000 Genomes population (n=2,504), we identified 172 polymorphic HERV-K insertion sites. We sought to identify adjacent SNPs associated with these HERV-K insertions to detect potential relationships between HERV-K and phenotypes via established SNP annotations. For 129 sites with estimated prevalence ranging from 20-80%, no HERV-K insertion associated SNPs (hiSNPs) were detected. The mean distance of these 129 sites from recombination hotspots compared to HERV-K with hiSNPs or randomly selected genomic sites suggest that the majority of polymorphic HERV-K insertions are in recombination cold spots (p<0.005) and cannot be tagged with SNPs. Pooled hiSNPs for the 48 polymorphic HERV-K sites that exhibited strong SNP associations were statistically enriched (p<1.0E-16) for eQTLs identified by the GTEx Project Consortium across 44 human tissues. When we restricted to the hiSNP sets identified from the 30 polymorphic HERV-K insertion sites in Europeans, we observed enrichment (p<0.05) of hiSNPs as eQTLs in at least one tissue type for 21 of the 30 insertion sites by Fisher's exact test. Further, a subset of hiSNPs for 15 of the 30 European HERV-K insertions are annotated in the EMBL-EBI GWAS Catalog and are associated with >100 phenotypes in total. Experimental factor ontology enrichment analyses suggest that polymorphic HERV-K insertions may contribute to neurologic and immunologic disease phenotypes, including traits related to intracranial volume (FDR 2.00E-09) and autoimmune diseases (FDR 1.80E-09). For these and other phenotypes identified via this method, polymorphic HERV-K insertions may play an etiologic role and further study is warranted.

2892F

Robust, accurate, and efficient pedigree reconstruction and pedigree-aware distant relatedness detection in 120 rhesus macaques (*Macaca mulatta*) from the Tulane National Primate Research Center using dense whole genome sequence data. *L.E. Petty[1], M. Raveendran[3], R.A. Harris[3], H.M. Kubisch[2], D.M. Muzny[3], R.A. Gibbs[3], R. Bohm[2], J. Rogers[2], J.E. Below[1].* 1) The Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 2) Tulane National Primate Research Center, Tulane University, Covington, LA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Genetic relatedness is informative for many genetic analyses including phasing, imputation, detecting evidence of natural selection, demographic comparisons, estimating heritability, and maximizing power and minimizing type 1 error in trait analysis. We have developed some of the most popular and powerful tools for pedigree reconstruction and distant relatedness detection in humans. However today, whole genome and whole exome sequencing (WGS, WES) are increasingly being applied in cohorts of nonhuman species, in which the complexity of population genetics is exacerbated by incomplete information and complex mating patterns. To fully characterize genetic relatedness in 120 Indian-origin rhesus macaques (*Macaca mulatta*) from the Tulane National Primate Research Center colony, we performed whole genome sequencing at >35x average coverage using 2x150 reads generated on the Illumina HiSeqX platform. Reads were mapped to the rhesus reference genome (NCBI Mmul_8.0) using BWA-mem and SNPs called using GATK with standard quality filtering. The applicability and accuracy of tools developed to estimate segmental sharing (GERMLINE, PLINK, IBDseq) and relatedness (ERSA, PLINK, KING), as well as reconstruct pedigrees (PRIMUS) and identify pedigree-aware distant relatedness (PADRE) was evaluated using colony records. Thirty-two sire groups were identified by first- and second-degree relationships, and all possible pedigrees were reconstructed and validated using 43 first-,14 second-, and 14 third-degree relatives within PRIMUS. Using PADRE, pedigrees were connected via pedigree-aware founder relationships by calculating the composite likelihood across all relationship probabilities between and within each pedigree. Predicted relationships were validated against colony records, demonstrating the robustness of these approaches for accurate assessments of varying degrees of kinship and precise pedigree reconstruction in nonhuman primate research populations, and as a natural extension, to wild populations with unknown pedigree structure. Nonhuman primates are central to many aspects of biology and medicine as crucial models for study of infectious disease, neurobiology, and psychobiology. Detecting cryptic relatedness and validating pedigrees are essential for reliable genetic studies. Determination of kinship among nonhuman primates increases our ability to investigate fundamental aspects of primate population biology and disease associations in research populations.

2893W

Genome-wide scan of pulmonary phenotypes on local ancestry in African Americans reveals novel genes interacting with smoking. *A. Ziyatdinov[1], M.H. Cho[2,3], M. Parker[2], T.H. Beaty[4], P. Kraft[1], H. Aschard[1,5].* 1) Departmentof Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA; 4) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 5) Centre de Bioinformatique, Biostatistique et Biologie Integrative (C3BI), Institut Pasteur, Paris, France.

   Abstract: It is now well established that the proportion of African ancestry is associated with lower pulmonary function for a given age and height [1]. At the same time, smoking is the strongest environmental risk factor for reduced pulmonary function. However, the interplay between the two has been seldom explored due to current methodological challenges in interaction analysis as well as limited sample sizes of current studies. This work aims to identify local ancestry-by-smoking interactions in the largest genetic cohort of pulmonary phenotypes in 3,300 African-American individuals from the (Chronic Obstructive Pulmonary Disease) COPDGene study. We performed a systematic screen for local ancestry-by-smoking interaction leveraging multiple pulmonary phenotypes available in COPDGene: the COPD (chronic obstructive pulmonary disease) status and five pulmonary phenotypes, namely FEV1 (force expiratory volume in one second), FVC (force vital capacity), FEV1/FVC ratio, percent emphysema and TLC (total lung capacity as measured by CT scan). We developed novel methods for ancestry-by-smoking interaction that account for confounding due to linkage disequilibrium differences between African and European populations and dependence among association tests using local ancestries. These methods include univariate and joint analysis of multiple phenotypes using principal component analysis; gene-set analysis; and grouping predictors in a risk score. We identified local ancestry segments showing interaction effects with current smoking status. These segments contained several candidate genes, all of which were associated with multiple phenotypes: (1) the *HS6ST3* gene, the nominal p-values between 1×10^{-4} and 1×10^{-3}; (1) the *PARVA* gene, p-values between 3×10^{-5} and 5×10^{-4}; (3) the *HNRNPD* gene, p-values 5×10^{-5} and 3×10^{-4}. Previous studies identified and association between *PARVA* methylation and smoking status [2]. Out findings confirm the validity of the approach in using estimated local ancestry information for detection of interaction effects between genes and exposures. Future results will likely identify novel loci, helping unravel the genetic architecture of COPD. **References**: [1] Kumar, R., *et al*. (2010). Genetic Ancestry in Lung-Function Predictions. New England Journal of Medicine, 363(4), 321–330. [2] Wan, Emily S., et al. (2015). Smoking-associated site-specific differential methylation in buccal mucosa in the COPDGene study. AJRCMB, 53(2), 246-254.

2894T

Genetic analyses for antiepileptic drug-induced cutaneous adverse reaction in a HK population. *J. Ding[1], M. Kowk[2], H. Gui[1], C. Tang[1], P. Sham[1], S. Cherny[1], P. Kwan[3].* 1) Department of Psychiatry, University of Hong Kong, Hong Kong; 2) Department of Medicine & Therapeutics, the Chinese University of Hong Kong , Hong Kong SAR, China; 3) Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia.

   Given the widespread use of anti-epileptic drugs (AEDs) in clinical treatments, the appropriateness and safety of this group of medications have posed a serious issue. Previous findings have revealed AEDs may be associated with life-threatening cutaneous adverse drug reactions (cADRs), with evidence pointing to a possible genetic basis, where HLA-B*15:02 allele has been consistently reported (Zhang et al., 2011). Our study, adopting the whole-genome sequencing method, focuses on identifying rare variants underlying AEDs-induced cADRs and establishing a multi-faceted genetic predisposition model for predicting the prognosis of AEDs. 80 AEDs patients who had experienced severe cADRs (SJS/TEN) and 72 matched AEDs-tolerant control pairs were recruited. Both groups of patients were assayed on Illumina HiSeq x Ten platform. Additional 316 sequenced population controls from collaborative data were included to increase the statistical power. Our results showed that rare variants in gene ADAMTS5, ADAMTS8, UROC1 and SPINK6 were marginally significant signals for cADRs. Although strong signals have not been detected, these genes were previously found playing key roles in the immune system and skin-related conditions.

2895F

Bayesian hierarchical modeling of genic sub-region intolerance. *T.J. Hayeck[1], N. Stong[1], B. Copeland[1], D. Goldstein[1], A. Allen[1,2].* 1) Institute for Genomic Medicine, Columbia University, New York City, NY; 2) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA.

Understanding the local distribution of intolerance within genes will be key in correctly deciphering relationships between genetic variation and disease. Genic intolerance scores have been developed to identify genes subject to purifying selection in the human lineage. These scores have come to play an important role in the interpretation of clinically identified variants, but given they are computed at the gene-level, they are unable to prioritize variation within genes. Frequently, pathogenic mutations can cluster in particular sub-regions of a disease gene, while other benign mutations may occur in other parts of the same gene. Previous methods have had success describing regional intolerance using marginal measures of departure from the amount of common putatively functional variation given the total genic variation. The earlier approaches however have limited utility because many of the subregions of genes carry too little information to provide reliable intolerance scores. We address this limitation and expand on the basic framework by developing a hierarchical Bayesian model framework for analysis of genomic variation, simultaneously modeling genes and their sub-regions (domain or exon). The Bayesian hierarchical framework allows for direct inference on the joint posterior distribution of gene level and sub-region level effects and allows for identification of genes with high levels of variation in regional intolerance. Using standing human variation from 123,136 exome sequences in Genome Aggregation Database (gnomAD), we demonstrate classification of intolerant sub-regions in key genes such as MECP2, VHL, and SCN1A. We further assess our ability to identify regions of the genome harboring pathogenic disease mutations through comparison of our regional intolerance scores and the presence of disease mutations from ClinVar and HgMD mutation databases. Initial results demonstrate significant association between sub-region intolerance scores and the genic sub-regions where pathogenic variants have been reported in ClinVar and HgMD (p=2.39e-15). These results suggest our framework can provide improved classification of intolerant regions and improve diagnostic interpretation of both known and novel variation.

2896W

Mixed-model adjustments for tests of epistasis reduce confounding by other loci. *N. Patel[1], W. Bush[1], A. Fish[2], J. Capra[2].* 1) Department of Population and Quantitative Health Sciences, Institute of Computational Biology, Case Western Reserve University, Cleveland, OH; 2) Department of Biological Sciences, Vanderbilt Genomics Institute, Vanderbilt University, 22525 West End Avenue, Nashville, TN.

Recent studies have identified numerous potential examples of epistasis among cis-regulatory variants influencing gene expression traits. However, these tests are subject to different forms of confounding than single-variant association tests - most notably, that multi-locus genotype combinations can tag the effect of nearby cis-eQTLs. As a result, many statistically significant interaction models for eQTL SNPs are more parsimoniously explained by the presence of another nearby SNP, and post-hoc conditional analyses are necessary to identify and characterize these findings. In this study, we develop and evaluate a mixed linear model approach, to adjust for genetic relatedness, in order to correct for this potential confounding. In a prior study, we identified 1,119 cis-regulatory interactions in a discovery set of 210 individuals with array-based gene expression and dense genotyping from lymphoblastoid cell lines, following rigorous false discovery corrections. Even though many of these interactions replicated in whole blood samples from the Genotype Tissue Expression (GTEx) project, all were found to be likely false positives in post-hoc tests for confounding. We re-analyzed these potential interactions using models including a random effect adjustment for the other SNP within the cis-regulatory region. This adjustment dramatically reduced the amount of confounding; likelihood ratio tests for the effect of the interaction terms were non-significant for all but six models. These results illustrate that mixed-model adjustments are a powerful approach for addressing confounding in SNP-SNP interaction analyses. .

2897T

The usage of local ancestry to Inform eQTL mapping in African Americans. *Y. Zhong[1], E. Gamazon[2], M. Perera[1].* 1) Department of Pharmacology, Northwestern University, Chicago, IL; 2) Division of Genetic Medicine, Vanderbilt University, Nashville, TN.

Expression quantitative trait loci (eQTL) are genetic variants significantly associated with gene expression. In order to control for population stratification, principal components (PCs) representing the global population structure are commonly used in eQTL mapping. However, this method may be inadequate for admixed populations such as African Americans in which the proportion of ancestry varies widely within each chromosome. Given the paucity of eQTL information in African Americans, it is critical to develop methods that adequately adjust for admixture and thus correctly identify eQTLs in this population. Thus, we investigated whether adjusting for local ancestry (LA) is more effective in identifying true eQTL associations than adjusting for PCs alone in African Americans. Here we developed a highly efficient algorithm to calculate millions of associations between SNPs and gene expression while adjusting for LA. We evaluated this method in an African American (AA) lymphoblastoid cell line (LCL) dataset. We first performed simulations in which we used actual genotype data and simulated gene expression. In simulations, LA adjustment and PC adjustment controlled for population stratification equally well. However, both methods had lower power to uncover associations with the correct effect size for high *Fst* eQTLs than for low *Fst* eQTLs. We also conducted genome-wide eQTL mapping using both LA and PC adjustment and found 221 eQTLs (FDR < 0.05) shared by both methods. For these shared eQTLs, local ancestry adjustment gave more significant p-values than PC adjustment, indicating that the local ancestry adjustment is more powerful to uncover true eQTLs. PC adjustment and LA adjustment identified 50 and 70 unique eQTLs respectively. Interestingly, several eQTLs unique to PC adjustment had highly significant p-values but were not significant in LA adjustment, and this pattern was not observed in eQTLs unique to LA adjustment. Further validation in an independent LCL dataset showed these discordant eQTLs were false positives due to the confounding of local ancestry to gene expression. Overall our analysis demonstrates that the LA adjustment outperforms PC adjustment in removing spurious associations in eQTL mapping.   .

2898F

Evidence for a major gene for myopia risk in Han Chinese-American families at 10q26. *J.E. Bailey-Wilson[1], A.M. Musolf[1], C.L. Simpson[1,2], B.A. Moiz[1], K.A. Long[1], D.D. Lewis[1], C.D. Middlebrooks[1], L. Portas[3], F. Murgia[3], E. Ciner[4], D. Stambolian[5].* 1) Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD; 2) Dept of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN; 3) Institute of Population Genetics, CNR, Li Punti, Sassari, Italy; 4) Salus University, Elkins Park, PA; 5) Dept of Ophthalmology, University of Pennsylvania, Philadelphia, PA.

Myopia is caused by an overgrowth of the eye which causes light to focus in front of the retina, leading to blurry vision. We studied 34 Han Chinese-descent families from Pennsylvania who were ascertained with an apparent dominant inheritance of myopia to search for linkage between genomic variants (Illumina Human Exome v1.1 array plus 24,263 custom SNPs) and the disease. Affection status was based on mean spherical equivalent (MSE) in Diopters (D): affected (≤ -1D), unaffected (≥0 D) or unknown (<0D, > -1 D). Children were considered unaffected as follows: MSE≥+2D (ages 6-10) or MSE≥+1.5D (ages 11-20). Three types of parametric linkage analyses were performed: single variant two-point, multipoint, and collapsed haplotype pattern variant linkage (CHP). CHP uses multiple rare variants to create a multi-allelic pseudomarker that corresponds to short haplotypes in a genomic region. The resulting pseudomarkers improve information content for a gene-based linkage test. This increases power to detect situations where multiple families harbor different rare mutations in the same causal gene. Two-point linkage analysis is then performed using CHP markers.   CHP linkage analysis identified a genome-wide significant locus at 10q26.13, centered on *TACC2* (HLOD=3.73). This pseudomarker consisted of several rare exonic SNPs from the *TACC2* gene. CHP analysis also found 6 more suggestive linkage signals at other genes in 10q24.2-26.2. Single variant two-point identified 34 suggestively linked SNVs on 10q24-26, while multipoint identified evidence of suggestive linkage in 10q26.11-13. Many of the suggestive linkages in both analyses were located in *HTRA1*, a known age-related macular degeneration gene. Three of the families with strongest linkage to this region had rare intronic variants in HTRA1 on their strongest linked haplotype. Several other promising candidate genes, such as *BAG3* and *DOCK1*, are also present in the region. Multipoint analysis also identified a highly suggestive linkage at 9q33.1. This region includes *TLR4*, a gene known to interact with alpha-crystallin in the retina. Targeted sequencing for both regions is planned to elucidate the causal variants.

2899W

Platelet-derived growth factor genes, maternal binge drinking and obstructive heart defects. *M.A. Cleves[1], X. Tang[1], J.K. Eberhart[2], M. Li[3], S. MacLeod[1], W.N. Nembhard[4], C.A. Hoobs[4], The National Birth Defects Prevention Study.* 1) Section of Biostatistics, Department of Pediatrics, College of Medicine, University of Arkansas for Medical Sciences; 2) Department of Molecular and Cell and Developmental Biology, Institute for Cellular and Molecular Biology and Institute for Neuroscience, University of Texas, Austin, Texas,; 3) Department of Epidemiology and Biostatistics, School of Public Health, Indiana University at Bloomington, Bloomington, Indiana; 4) Section of Birth Defects Research, Department of Pediatrics, College of Medicine, University of Arkansas for Medical Sciences.

   Background: More than 85% of congenital heart defects (CHDs) are thought to result from a complex interaction between maternal exposures, genetic susceptibilities, and epigenetic phenomena. Right- and left-sided obstructive heart defects (OHDs) are the most lethal subtypes of CHDs. Platelet-derived growth factor receptor alpha gene (*Pdgfra*) has been shown in a zebrafish model to interact with ethanol during development leading to craniofacial structural malformations. Previous studies have reported an association between alcohol exposure during the periconceptional period and the risk of CHDs. **Objective:** The goal of our study was to investigate the interaction between variants in the platelet-derived growth factor (*PDGF*) genes and periconceptional alcohol exposure on the risk of OHDs. **Methods:** A genome-wide association study was carried out using buccal cell samples from 806 case-parental triads and 995 control-mother dyads enrolled in Arkansas, California, George, Iowa and Texas sites of the National Birth Defects Prevention Study (NBDPS) with birth dates between 1997 and 2011. DNA samples were genotyped for more than 4.3 million single nucleotide polymorphisms (SNPs) by use of the Illumina Infinium Omni5Exom-4 chip. Among 4.3 million SNPs, 102 SNPs in the *PDGF* family of genes (*PDGFRA*, *PDGFRB* and *PDGFB*) were selected for the current study. Data on maternal alcohol exposure was self-reported via a maternal telephone interview conducted by NBDPS. The exposure window was defined as one month prior to conception to three month after conception. A family-based hybrid log-linear model was used to estimate the maternal and fetal genetic relative risk of each allelic variant in three *PDGF* genes. **Results:** The maternal CT and TT genotypes of rs869978 variant in the *PDGFRA* gene were associated with a 1.51 (95% confidence interval (CI): 1.06 to 2.15) and 2.28 (95% CI: 1.12 to 4.62) increased risk of OHDs compared to the CC genotype, respectively, among women who binge drink during the periconceptional period. In contrast, fetal genotypes of four SNPs (rs2291591, rs2228230, rs1547904, and rs869978) in the PDGFRA gene were associated with a decreased risk of OHDs among women who binge drink. **Conclusion:** Our findings suggest that variants in the maternal *PDGFRA* gene may interact with binge drinking, resulting in an increased risk of obstructive heart defects, whereas variants in the infant *PDGFRA* gene may protect against obstructive heart defects among binge drinkers.

2900T

Iranome: A human genome variation database of eight major ethnic groups that live in Iran and neighboring countries in the Middle East. *M.R. Akbari[1,2,3], Z. Fattahi[4], M. Beheshtian[4], M. Mohseni[4], H. Poustchi[5], E. Sellars[1], H. Nezhadi[6], A. Amini[7], S. Arzhangi[4], K. Jalalvand[4], P. Jamali[8], B. Davarnia[4], P. Nikuei[9], M. Oladnabi[4], A. Mohammadzadeh[4], E. Zohrehvand[4], E. Shamsi-Gooshki[10,11], S. Börno[12], B. Timmermann[12], R. Najafipour[13], H.R. Khorram Khorshid[4], K. Kahrizi[4], H. Najmabadi[4].* 1) Women's College Research Institute, Univ Toronto, Toronto, ON, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 4) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, 19857 Tehran, Iran; 5) Liver and Pancreatobiliary Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran; 6) Department of computer Science, University of Toronto, Toronto, Ontario, Canada; 7) Information Technology Office, University of Social Welfare and Rehabilitation Sciences, 19857 Tehran, Iran; 8) Shahrood Genetic Counseling Center, Welfare Office, 36156 Semnan, Iran; 9) Molecular Medicine Research Center, Hormozgan University of Medical Sciences, Bandarabbas, Iran; 10) Medical Ethics and History of Medicine Research Center, Tehran University of Medical Sciences, Tehran, Iran; 11) Department of Medical Ethics, Faculty of Medicine, Tehran University of Medical Sciences, Tehran, Iran; 12) Max-Planck-Institute for Molecular Genetics, 14195 Berlin, Germany; 13) Cellular and Molecular Research Centre, Qazvin University of Medical Sciences, Qazvin, Iran.

   Access to clinical genetic testing has been growing continuously around the globe since the introduction of the next generation sequencing technology to the field of genetics about a decade ago. Widespread access to genetic testing will have a remarkable impact on realizing the vision of precision medicine to improve the prevention, diagnosis and treatment of human disorders, many of which have genetic etiology. However, many ethnic groups are not represented in current human genome variation databases. The benefits of precision medicine may not be realized for these groups if we do not address this gap. With this in mind, we established the Iranome database (www.iranome.com) by performing whole exome sequencing on 800 individuals from eight major ethnic groups that live in Iran. The groups include 100 healthy individuals from each of Arabs, Azeris (Turk), Balochs, Kurds, Lurs, Persians, Persian Gulf Islanders and Turkmen ethnic groups, which represent over 80 million Iranians and to some degree half a billion individuals who live in the Middle East. These ethnic groups are among the most underrepresented populations in currently available human genome variation databases. The mean depth of coverage of the exons of the human genome based on CCDS Release15 was 82x with 95% and 91% coverage at 10x and 20x or more, respectively. Principle component analysis indicates that except the Iranian Baloch and Persian Gulf Islander populations, which form their own clusters, the rest of the populations are genetically very close to each other. We identified 1,368,380 variants, which passed filter based on depth of coverage and several other variant calling quality metrics. Of those, 196,776 are insertions or deletions. Also, 44.4% of the variants are singleton and 21.9% of them have a minor allele frequency (MAF) of higher than 1%. Loss of function (LoF) variants constitute 1.2% (15,892 variants) of the database, of which 14.5% are located at OMIM genes with a reported associated phenotype. We found 422,095 novel variants, which were not reported in the Genome Aggregation or 1000 Genome databases, that represent 30.1% of the entire detected variants. As expected, the majority of these novel variants are singleton (65.1%) and only 7.5% and 3.3% of them have a MAF of higher than 1% and 5% respectively. This shows the importance of representing diverse populations in the human genome variation databases to create a more complete picture of human genetic variations.

2901F

Genome-to-genome analysis of host-pathogen interactions in human tuberculosis. *N. Chaturvedi[1,4], S. Gagneux[2], M. Coscolla[2], D. Brites[2], D. Portevin[3], J. Fellay[1,4].* 1) School of life sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland; 2) Tuberculosis Research Unit, Swiss Tropical and Public Health Institute, University of Basel, Switzerland; 3) Clinical Immunology Unit, Swiss Tropical and Public Health Institute, University of Basel, Switzerland; 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

Tuberculosis (TB), resulting from infection with *Mycobacterium tuberculosis* (MTB), remains the leading cause of human death in developing countries. It is an ancient disease and MTB has been associated with its human host for a very long time, likely leading to some degree of host-pathogen co-evolution. However, formal evidence of co-evolution and its contributions towards host-pathogen interactions in human TB is lacking. In this work, we used paired human and MTB genomic information from infected patients to assess the potential influence of host genetic variation on MTB genomic diversity. Paired human genome-wide genotyping and MTB whole-genome sequencing were obtained for 81 HIV-infected patients in the Swiss HIV Cohort Study (SHCS) who also developed active TB disease. Human DNA was obtained from whole blood and bacterial DNA was extracted from mycobacteria isolated from sputum. We generated binary variables for 1,812 non-conserved MTB amino acids in 4,000 MTB coding genes. Out of these 1,812 binary variables, 16 were located in known human T cell epitopes. We ran separate single-marker logistic regressions for each dependent outcome and applied Bonferroni correction ($p < 1.7 \times 10^{-12}$) to control for multiple testing. Our sample set was heterogeneous in terms of patient ethnicity and MTB phylogenetic lineage, which has the potential to result in human and bacterial stratification. To avoid spurious associations due to these systematic differences, we added the first five host principal components and the first two pathogen phylogenetic principal components as covariates in all models. This pilot study is statistically powered to detect very strong associations (e.g. odds ratio of >20 for associations between human and pathogen variants). We did not identify any significant associations between specific binary amino acid variables in the MTB genome and human SNPs, which could be due to the large number (~ 27 billion) of tests performed in parallel and a small sample size. The smallest association p-value was 8.9×10^{-5}, between host SNP rs138060913 (gene TJP2, chromosome 9 position 71850428) and amino acid position 252 in the MTB gene Rv1872c. This is the first evaluation of host genetic restriction on MTB genomic diversity, and larger sample sizes are needed to get a clearer picture of the interactions between the human and MTB genomes.

2902W

Genotype imputation performance using an African-American population. *L. Franco[1,3], C.I. Vergara[2], C. Thio[2], A.V. Valencia-Duarte[3], D. Thomas[2], P. Duggal[4].* 1) University of Antioquia, National Faculty of Public Health, Medellin, Antioquia, Colombia; 2) Johns Hopkins University, School of Medicine, Baltimore, MD.,USA; 3) Universidad Pontificia Bolivariana, School of Medicine, Medellín, Antioquia, Colombia; 4) Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD.,USA.

Genotype imputation is used to estimate unobserved genotypes from genotypic data, to increase genome coverage and power for genomewide arrays. Imputation has been successful for European ancestry populations. In comparison, African descent populations have been limited on available reference panels but there are enhanced African populations in 1000 genomes (1000G) and The Consortium On Asthma Among African-Ancestry Populations In The Americas (CAAPA). We aimed to compare the performance of these reference panels to impute variation in 447 African Americans (AA) genotyped for 774,792 SNPs using the Illumina OmniQuad array. Eaglev2.3 was used for haplotype phasing and imputation was performed with Minimac3 on the Michigan Imputation Server. The haplotypes of 2,504 individuals (1000G) and 883 (CAAPA) were used as reference. R^2 was used for imputation quality, and accuracy was estimated by a "masked analysis" with 25,000 SNPs (MAF:0.01-0.5). We evaluated 17 individuals with both GWAS and 30x WGS for genotype concordance. 1000G imputed 1.56x more variants than CAAPA irrespective of quality. Using an $R^2>0.5$ threshold, 1000G imputed 1.4x more variants. 35% and 9.1% of them were unique for 1000G and CAAPA, respectively. From the 16,814,797 overlapping variants between the two panels, 99% had concordant genotype calls, but the quality was higher for 1000G (R^2:0.7-0.9 vs R^2:0.2-0.6). For masked SNPs the concordance between the actual genotypes and the imputed data was similar for both panels (97% in 1000G, 96% in CAAPA), but more accurate for 1000G. Using a single chromosome (chr22) as an example, the concordance between sequenced and imputed data was similar (98.9% of 213,472 variants in 1000G, 99.1% of 190,005 in CAAPA). Functional annotation using ANNOVAR identified 8.7% and 3.6% of variants for 1000G and CAAPA, respectively. Both panels increased the number of variants from a GWAS array for African American individuals with high quality and accuracy although 1000G was higher. Despite the larger reference size of 1000G, a sensitivity analysis restricting to the African populations (N=661), still showed increased quality and accuracy for 1000G. This likely reflects the greater contribution of African ancestry in 1000G as compared to the admixed individuals in CAAPA. The large number of variants that are uniquely imputed with each panel make them complementary but suggests that even more primary African sample references will improve imputation.

2903T

Data-driven genetic encoding (DAGE) allows flexible identification of novel main effects and SNP-SNP interactions. *M.A. Hall[1], A.M. Lucas[2], Y. Bradford[2], J.H. Moore[1], M.D. Ritchie[2].* 1) Institute for Biomedical Sciences, University of Pennsylvania, Philadelphia, PA; 2) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA.

Certain assumptions are made about a SNP's biological action when choosing a genetic encoding. For each encoding, risk incurred by one copy of the alternate allele in relation to two copies varies: the heterozygous genotype (HET) is coded to bear 0%, 50%, and 100% the risk of homozygous alternate (HA) for recessive (REC), additive (ADD), and dominant (DOM) encodings, respectively. Yet, HET may yield any portion of the risk of HA. Further, SNPs across the genome are unlikely to demonstrate the same genetic action as one another. Choosing one encoding lacks flexibility, while running every encoding raises the multiple testing burden. We present a novel, flexible alternative: the data-driven genetic encoding (DAGE). Here, a HET value is assigned, based on individual SNP action in the dataset. DAGE outperformed traditional encodings in power tests for main effect and genetic interactions in a comprehensive combination of simulated genetic models. Further, DAGE maintained a low false positive rate, while the additive and dominant encodings demonstrated significant inflation. We applied DAGE to data from the Electronic Medical Records and Genomics (eMERGE) Network using age-related macular degeneration, age-related cataract, glaucoma, type 2 diabetes, hypothyroidism, and resistant hypertension and assessed these traits with ADD, REC, and DOM encodings as well. Across each trait, DAGE categorized SNP action consistent with p-values for ADD/DOM/REC (e.g., SNPs categorized by DAGE as having dominant action were found to have low significance for REC, high significance for DOM, and moderate significance for ADD). Further, DAGE identified numerous Bonferroni significant main effect and interaction signals not identified by traditional encodings. For example, DAGE found an interaction between SNPs rs4695885 and rs7787286 (Bonferroni-adjusted p: 0.018) predictive of age-related cataract that no other encoding identified. SNP rs4695885 was categorized as additive and rs7787286 as recessive using DAGE. This is the first time these intronic SNPs have been implicated in age-related cataract, perhaps due to the limitation of traditional encodings. Our results demonstrate the impact of choosing a genetic encoding and the need for an individualized method for assigning SNP action. We offer DAGE as a flexible and robust method to identify novel main effect and SNP-SNP interactions in GWAS data.

2904F

The genetic architecture of 25-hydroxyvitamin D. *X. Jiang[1], P. O'Reilly[2], H. Aschard[3], Y. Hsu[4], B. Richards[5], T. Spector[6], V. Vazquez[6], J. Dupuis[7], T. Wang[8], E. Hypponen[9], P. Kraft[1], D. Kiel[10] on behalf of the SUNLIGHT consortium.* 1) Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Social Genetic & Developmental Psychiatry, King's College London, Institute of Psychiatry, UK; 3) Department of Genomes and Genetics Center of Bioinformatics, Biostatistics and Integrative Biology, Institute Pasteur; 4) Institute for Aging Research, 1200 Center Street Boston, 02131, MA, USA; 5) Department of Medicine, Human Genetics, Epidemiology and Biostatistics, Quebec, Canada; 6) The Department of Twin Research & Genetic Epidemiology, King's College London, UK; 7) Department of Biostatistics, Boston University School of Public Health, 02118, MA, USA; 8) Division of Cardiovascular Medicine; Physician-in-Chief, Vanderbilt Heart and Vascular Institute; 9) Center for Population Health Research, Sansom Institute for Health Research, University of South Australia, Australia; 10) Institute for Aging Research, Hebrew Senior life,Department of Medicine Beth Israel Deaconess Medical Center and Harvard Medical School.

Vitamin D is a steroid hormone precursor that has established associations with a range of human traits and diseases. Despite relatively high heritability estimates from family studies, GWAS of serum 25-hydroxyvitamin D concentrations from the SUNLIGHT Consortium have identified only four loci (*GC, NADSYN1/DHCR7, CYP2R1, CYP24A1*) harboring genome-wide significant variants. To better understand the genetic architecture underlying serum 25-hydroxyvitamin D, as well as to test for interactions between dietary vitamin D intake and genetic factors, we conducted the largest GWAS meta-analysis to date on this important vitamin. We expanded our previous GWAS dramatically, and undertook a large, multicenter, genome-wide association study. This single stage discovery meta-analysis consisted of 79,366 samples of European descent drawn from 31 studies, five times larger than our previously published GWAS. We also performed a genome-wide screen leveraging the possibility that genetic effects may depend on dietary vitamin D intake. We performed linkage disequilibrium score regression analysis to estimate the SNP-heritability of serum 25-hydroxyvitamin D concentrations. We additionally assembled summary statistics from GWAS of 39 traits performed in individuals of European descent, and calculated the pairwise genetic correlations between 25-hydroxyvitamin D and each trait. We also conducted cell-type-specific analysis to identify the enrichment of tissues in vitamin D. Our larger GWAS yielded only two additional loci harboring genome-wide significant variants (P=4.7X10^{-9} at rs8018720 in *SEC23A*, and P=1.9X10^{-14} at rs10745742 in *AMDHD1*). The overall heritability of 25-hydroxyvitamin D serum concentrations attributable to GWAS SNPs was 7.54%. Statistically significant loci accounted for 38% of this total (2.84% out of 7.54%). Further investigation identified signal enrichment in immune hematopoietic tissues, and clustering with autoimmune diseases in cell-type-specific analysis. Genome-wide genotype-by-dietary vitamin D interaction analyses identified no significant interactions. These results suggest circulating 25-hydroxyvitamin D has an oligogenic genetic architecture and is largely influenced by a small number of loci. Efforts to ensure adequate vitamin D concentrations and to understand the known variability in responses to vitamin D supplementation may be able to focus on the limited number of genetic determinants that we have identified.

2905W

Improved genotype imputation in disease-relevant regions with inclusion of patient sequence data: Lessons from cystic fibrosis. *N. Panjwani[1,14], B. Xiao[1,14], L. Xu[2], J. Gong[1], K. Keenan[3], F. Lin[1], G. He[1], Z. Baskurt[1], L. Zhang[4], S. Kim[5], M. Esmaeili[1], S. Blackman[6], H. Corvol[7,8], M. Drumm[9,10], M. Knowles[11], G. Cutting[6,12], J.M. Rommens[1,13], L. Sun[4,5], L.J. Strug[1,2,4].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Statistics, University of Toronto, Toronto, ON, Canada; 5) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 6) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 7) Pediatric Pulmonary Department, Hospital Trousseau, Assistance Publique-Hôpitaux de Paris (AP-HP), Institut National de la Santé et la Recherche Médicale (INSERM), U938, Paris, France; 8) Pierre et Marie Curie University–Paris 6, Paris, France; 9) Department of Pediatrics, Case Western Reserve University, Cleveland, Ohio, USA; 10) Department of Genetics, Case Western Reserve University, Cleveland, Ohio, USA; 11) Cystic Fibrosis Pulmonary Research and Treatment Center, University of North Carolina, Chapel Hill, North Carolina, USA;; 12) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 13) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 14) These authors contributed equally to this work.

Genotype imputation improves fine mapping and enables meta-analysis of cohorts genotyped on different platforms. Traditional imputation uses whole genome sequence (WGS) reference panels e.g. 1000 Genomes Project (1KGP, n=2,504). However, imputation with the 1KGP failed to impute variants in the cystic fibrosis (CF) transmembrane conductance regulator (*CFTR*) for the International CF Gene Modifier Consortium, genotyped on the Illumina Human660W-Quad BeadChip (n=1,995). *CFTR* displays significant allelic heterogeneity and is associated with CF co-morbidities of complex inheritance such as intestinal obstruction. Using larger reference panels such as the Haplotype Reference Consortium (HRC; n=32,470), or alternatives that incorporate in-sample WGS may be preferable. We compare imputation in *CFTR* using the HRC to a composite reference combining the 1KGP with WGS from 101 CF patients; the latter enriching the reference with study-specific haplotypes. The 1KGP, HRC and composite reference provide, respectively, 1,438, 2,164 and 2,439 biallelic variants for imputation in *CFTR*. The composite reference results in more variants imputed; greater concordance between imputed and known CF-causing mutations, e.g. W1282X with 98.3% vs. 96.6% in HRC; and imputation of the most common CF mutation (p.Phe508del) missed by HRC which does not support indels, a consequence of low coverage sequencing. Notably, the composite reference displayed greater ability to detect association between *CFTR* and intestinal obstruction over 1KGP or HRC panels. Results suggest that traditional imputation can omit the most disease-relevant genotypes when there is allelic heterogeneity at causal loci, but incorporating WGS on a subset of the study population can improve imputation and causal variant identification.

2906T

Assessing pleiotropy and mediation in loci associated with chronic obstructive pulmonary disease. *M.M. Parker[1], S.M. Lutz[2], B.D. Hobbs[1,3], R. Busch[1,3], M.L. McDonald[4], P.J. Castaldi[1,5], E.K. Silverman[1,3], M.H. Cho[1,3].* 1) Channing Laboratory, Brigham and Womens Hospital/HMS, Boston, MA; 2) Department of Biostatistics and Informatics, University of Colorado, Anschutz Medical Campus, Denver, CO, USA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 4) Division of Pulmonary, Allergy, and Critical Care Medicine, University of Alabama, School of Medicine, Birmingham AL, USA; 5) Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA.

Rationale: COPD is a complex, heterogeneous disease. Large GWAS studies have identified several variants associated with case/control status. Many of these loci also show significant associations with other COPD-related phenotypes (e.g. emphysema). We aimed to distinguish if these loci act directly on a set of COPD-related traits or if the association is predominately due to a mediator (e.g. cigarette smoking). **Methods:** We tested the association of 22 COPD loci identified from a recent case-control GWAS with 12 COPD-related phenotypes in 5,873 non-Hispanic whites from the COPDGene study. SNPs associated with >1 phenotype at $p < 10^{-4}$ were tested for pleiotropy via the pleiotropy R package. In cases of significant pleiotropy, we performed mediation analysis (using the R package mediation) to test if SNPs had a direct or mediated effect on the outcome of interest. Causal models assumed FEV_1/FVC or cigarette smoking as the mediator. **Results:** Of the known COPD loci, 5 were associated with > 1 outcome at $p < 10^{-4}$ (nearest genes: *RARB*, *RIN3*, *CHRNA5*, *AGER*, *HHIP*). Tests for pleiotropy between genotype and associated outcomes were significant for all loci ($p < 0.05$). The *RARB* and *RIN3* SNPs had a significant mediated effect on emphysema through FEV_1/FVC; however there was not a significant direct effect. In contrast, we found evidence for both direct and mediated effects for *AGER*, *CHRNA5*, and *HHIP*. At the *CHRNA5* locus, we confirmed an effect on lung function partially mediated through cigarette smoking (p direct effect = 0.01), though the effects of SNP on emphysema and emphysema distribution were entirely mediated through FEV_1/FVC (p direct effect = 0.19, 0.07, respectively). The SNP in *AGER* was associated with FEV_1/FVC, emphysema, and total lung capacity. The association with emphysema and total lung capacity was partially mediated through FEV_1/FVC, but there was also a direct effect of genotype on both phenotypes (p direct effect = < 0.001, p=0.01, respectively). Lastly, the *HHIP* locus was associated with FEV_1/FVC, emphysema, and emphysema distribution. Effects were partially mediated by FEV_1/FVC but there was also a direct effect of SNP on both emphysema phenotypes (p direct effect = 0.02, <0.001). **Conclusions:** We used causal modeling to identify 3 SNPs with direct associations to multiple outcomes (*AGER*, *CHRNA5*, *HHIP*), suggesting these SNPs may act directly on these COPD-related phenotypes and that this pleiotropic relationship is not solely due to mediators.

2907F

Genetic determinants of urinary biomarkers in the UK Biobank. *D. Zanetti[1], S. Gustafsson[2], E. Ingelsson[1].* 1) Department of Medicine, Division of Cardiovascular Medicine, Stanford University, School of Medicine, CA; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Sweden.

Background: Urine biomarkers, such as sodium, potassium, creatinine and microalbumin, are strongly associated with several common diseases including diabetes mellitus, cardiovascular disease and chronic kidney disease. Knowledge about genetic determinants of these biomarkers can shed light on pathophysiological mechanisms underlying disease development. Methods: We performed four genome-wide association studies (GWAS) of urinary sodium, potassium, creatinine and microalbumin in up to 116,860 individuals of European ancestry from the UK Biobank, a large population-based cohort study of ~500,000 individuals aged 40-69 years recruited at 21 centers across UK in 2006-2010. Further, we explored tissue and pathway enrichment using DEPICT in an expanded set of independent variants ($P<10-5$). Results: In these preliminary analyses of the first release of data from the UK Biobank, we found 7 genome-wide significant ($P<5e-8$) independent loci for sodium (4), potassium (1), creatinine (1) and microalbumin (1). We confirmed the well-known association with microalbuminura near *CUBN* (P=4.41e-12), and one of the associations with sodium was near *GCKR*, a well-known diabetes locus that also has been associated with chronic kidney disease in prior studies (P=6.43e-10). The other associations were novel in terms of urinary biomarkers and kidney function, but signals for urinary sodium and creatinine mapped in loci previously associated with breast cancer (*MLLT10/DNAJC1*) and coffee intake (*CYP1A1/CYP1A2*), respectively. Tissue enrichment analyses in DEPICT highlighted parietal brain lobe as significantly enriched among associations with urinary sodium (FDR<0.05). We detected some degree of overlap across associations between the four urinary biomarkers. Specifically, two genome-wide significant loci for urinary sodium and one for potassium showed a strong association also with urinary creatinine ($P<10-5$). Conclusion: We report seven genome-wide significant associations with urinary sodium, potassium, creatinine and microalbumin in 116,860 individuals from the UK Biobank, confirming several known associations and providing new insights into the genetic basis of these traits, and their connection to chronic diseases.

2908W

Whole genome sequencing association analysis of red blood cell traits in a multi-ethnic population from the Trans-Omics for Precision Medicine (TOPMed) Project. *X. Zheng[1], Y. Hu[2], D. Jain[1], C.A. Laurie[1], S.M. Gogarten[1], M. Chen[3,8], J.R. O'Connell[4], J.P. Lewis[4], L.M. Raffield[5], A. Correa[6], L.A. Cupples[3,7], K.M. Rice[1], A.D. Johnson[3,8], C.C. Laurie[1], A.P. Reiner[2,9], the TOPMed Hematology and Hemostasis Working Group.* 1) Department of Biostatistics, University of Washington at Seattle, Seattle, WA; 2) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) NHLBI's the Framingham Heart Study, Framingham, MA; 4) School of Medicine, University of Maryland, Baltimore, MD; 5) Department of Genetics, University of North Carolina, Chapel Hill, NC; 6) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 7) Department of Biostatistics, Boston University, Boston, MA; 8) NHLBI, Population Sciences Branch, Bethesda, MD; 9) Department of Epidemiology, University of Washington, Seattle, WA.

Red blood cell (RBC) measurements are polygenic traits, and genome-wide association studies (GWAS), exome chip and sequencing analyses have identified hundreds of associated genetic variants in European-, Asian-, African- and Hispanic-ancestry populations. However, additional loci remain undiscovered and causal variant(s) at each locus have not been well characterized. We performed whole genome sequencing (WGS) analyses for hemoglobin (HGB), hematocrit (HCT), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), RBC count and red cell distribution width (RDW) in a multi-ethnic population from phase 1 of the NHLBI TOPMed Project. 7,490 Amish-, European- and African-ancestry individuals from the Old Order Amish Study, the Framingham Heart Study and the Jackson Heart Study, respectively, were pooled for single variant tests using inverse-normal transformed residuals as outcomes, with adjustment for age, sex, study, relatedness, population structure, and residual heteroscedasticity. We identified nine putatively novel loci reaching genome-wide significance ($P<5E-8$), including two each for HGB (*4q22*-rs10030052 and *ADAMTS20*-rs563075310, MAF=0.001), HCT (*4q22*-rs10030052 and chrX: 53479481, MAF=0.001), MCH (*1p11*-rs114421285 and *2q14.1*-rs150467776, MAF=0.013 and 0.005, respectively) and MCHC (*ELOVL6*-rs111995643 and *PHACTR*-rs141007575, MAF=0.014 and 0.001, respectively), and one each for RBC count (*ITPR3*-rs555173304, MAF=0.004) and RDW (chr22: 51068271, MAF=0.001). The novel variants showed disparate allele frequencies across European and African ancestral populations, and will require validation in additional independent samples. We also confirmed seven previously reported loci for associations with multiple RBC traits. The lead variant for six (*HFE, BYSL, MYB, TMPRSS6, CD36* and *G6PD*) was the same as the reported index SNP or in strong LD (r²>0.8). The lead variants at *ITFG3/LUC7L* for MCH, MCHC, MCV and RBC were in strong LD with the reported index SNP while lead variants for HGB and RDW were not. At the *HBB* locus, the strongest association signal was rs34598529, with novel evidence of associations with MCH, MCV, RBC and RDW ($P≤9E-9$). This variant lies in a putative erythroid regulatory element and has been identified in African American beta thalassemia patients. In summary, our initial results suggest identification of novel rare RBC variants from WGS analyses of a multi-ethnic population.

2909T

Large-scale inference in population cohorts. *M.A. Rivas.* Biomedical Data Science, Stanford University, Stanford, CA.

Whole genome sequencing studies applied to large population cohorts, biobanks, and case-control studies with extensive phenotyping raise novel analytic challenges. The need to consider many variants at a locus or group of genes simultaneously and the potential to study many correlated phenotypes with shared genetic architecture provide opportunities for discovery and inference that are not addressed by the traditional one variant-one phenotype association study. Here we introduce an online resource, Global Biobank Engine (GBE, biobankengine.stanford.edu), along with novel methodology for estimating correlation, scale, and location of genetic effects across a group of genetic variants, phenotypes, and studies implemented in GBE using the probabilistic programming language, STAN. We consider two components: 1) Bayesian model comparison, and 2) an Empirical Bayes mixture model, which can scale to millions of samples. After a signal is detected by model comparison, the mixture model estimates the underlying mixture of neutral and functional variants and thereby improves estimation of the effect size profile of any particular variant across a broad range of traits. Notably, this helps to assess whether protective alleles exist (location) without severe side-effects (correlation) and whether differences in strength of effects (scale) exist among various annotation classes of genetic variants such as mildly deleterious missense variants and highly disruptive protein-truncating variants. Taken together, the inference generated by the application of MRP to large-scale human genetic data should improve the generation of effective therapeutic hypotheses. Here, we present the Global Biobank Engine and its application to the first genetic data release by UK Biobank from 150,000 participants, pinpointing protective alleles in: IL33 conferring protection to asthma (posterior odds ratio=.59, [.49-.72]), IFIH1 to hypothyroidism (number of cases=7539, posterior odds ratio=.75 [.64-.88]), and TYK2 to psoriasis, rheumatoid arthritis, hypothyroidism, and sarcoidosis (posterior odds ratio range .5-.8); and estimation of genetic parameters including genetic correlation among various measurable phenotypes and diseases including Parkinson's disease and severe depression (rho_g = .63 [0.53-0.74]).

2910F

Population pharmacokinetics of sulindac and genetic polymorphisms of *FMO3* and *AOX1* in women with preterm labor. *J. Yee[1], K. Lee[2], J. Chung[3], J. Sung[1], S. Park[1], Y. Kim[4], H. Gwak[1].* 1) College of Pharmacy & Division of Life and Pharmaceutical Sciences, Ewha Womans University, Seoul, South Korea; 2) College of Pharmacy, Chungbuk National University, Cheongju, South Korea; 3) College of Pharmacy, Sungkyunkwan University, Suwon, South Korea; 4) School of Medicine, Ewha Womans University, Seoul, South Korea.

This prospective study was aimed to establish a population pharmacokinetic model of sulindac and its active metabolite and to evaluate the effects of genetic polymorphisms in sulindac-related metabolizing enzyme genes including *FMO3* and *AOX1* on the population pharmacokinetics of sulindac in pregnant women with preterm labor. Sixty-eight patients diagnosed with preterm labor between 16 and 37 weeks' gestation were enrolled and data of 58 patients were analyzed. Plasma samples were collected at 1.5, 4, and 10 hr after the first oral administration of sulindac. Plasma concentration of sulindac and its active metabolite (sulindac sulfide) were determined using HPLC. To evaluate the effects of metabolizing gene polymorphisms on the pharmacokinetics of sulindac, 10 single nucleotide polymorphisms (SNPs) were analyzed by Snapshot and TaqMan genotyping assays. Pharmacokinetic analysis was performed with NONMEM 7.3. The mean maternal age and gestational age at dosing were 32.5 ± 4.4 (range, 20-41) years old and 27.4 ± 4.4 (range, 16.4-33.4) weeks, respectively. Of the patients, 38 (65.5%) were over 26 weeks of gestational age at dosing and the mean body weight was 60.2 ± 9.7 (range, 45.0-92.0). In population pharmacokinetic analysis, one depot compartment model of sulindac with absorption lag time best described the data. The metabolism of sulindac and sulindac sulfide was described using Michaelis-Menten kinetics. In stepwise modeling, gestational age (divided into groups of ≥ 26 weeks and <26 weeks) impacted volume of distribution and *FMO3* rs2266782 was shown by the Michaelis constant to affect conversion of sulindac sulfide to sulindac in the final model. The present study details the pharmacokinetic characteristics of sulindac and its metabolites in pregnant woman and verifies that genetic polymorphisms of *FMO3* and *AOX1* affect the pharmacokinetics of sulindac in women who undergo preterm labor. It is expected that the results could help clinicians predict the efficacies of sulindac in the development of individualized treatment plans for patients who undergo preterm labor.

2911W

Genome wide meta-analysis for dental caries in childhood. *S. Haworth[1], D. Shungin[2], J. van der Tas[3], S. Vucic[3], C. Medina Gomez[4], V. Yakimov[5], J. Shaffer[6], M.K. Lee[6], M. Standl[7], C. Wang[8], K. Bønnelykke[9], J. Waage[9], L.E. Jesson[9], P. Nørrisgaard[9], R. Joro[10], I. Seppälä[11], T. Lehtimäki[11], T. Lakka[10], H. Bisgaard[9], C. Pennell[8], J. Kühnisch[7], M. Marazita[6], F. Geller[5], F. Rivadeneira[4], E. Wolvius[3], N.X. West[12], S.J. Thomas[12], P.W. Franks[13,14,15], N. Timpson[1], I. Johansson[2].* 1) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 2) Department of Odontology, Umeå University, Umeå, Sweden; 3) Department of Oral and Maxillofacial Surgery, Special Dental Care and Orthodontics, Erasmus Medical Center, Rotterdam, The Netherlands; 4) The Generation R Study Group, Erasmus Medical Center, Rotterdam, The Netherlands; 5) Statens Serum Institut, Copenhagen, Denmark; 6) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA, USA; 7) Institute of Epidemiology I, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 8) Division of Obstetrics and Gynaecology, The University of Western Australia, Perth, Australia; 9) Copenhagen Prospective Studies on Asthma in Childhood, University of Copenhagen, Copenhagen, Denmark; 10) Physical Activity and Nutrition in Children Study group, University of Eastern Finland, Joensuu, Finland; 11) Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 12) School of Oral and Dental Sciences, University of Bristol, Bristol, UK; 13) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden; 14) Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; 15) Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA.

   Background: Prior studies suggest dental caries traits in childhood are heritable, but there has been no large-scale consortium genome-wide association study (GWAS) to date. **Participants and methods**: We performed GWAS for dental caries status on a case:control basis. The participants were children (aged 2.5-18.0 years) of European ancestry. Phenotype definitions were created for the presence or absence of treated or untreated dental caries, stratified by primary and permanent dentition. All studies tested genotype dosage (imputed to Haplotype Reference Consortium or 1000 Genomes phase 1 version 3 panels) whilst accounting for population stratification. Fixed–effects meta-analysis was performed weighted by standard error. Analysis included 17,037 individuals / 6,922 cases for primary teeth; 13,353 individuals / 5,875 cases for permanent teeth. Heritability was estimated using univariate linkage disequilibrium score regression. **Results:** There was modest evidence for association with caries status at single variants including rs1594318-C for primary teeth (intronic *ALLC*, OR 0.85, EAF 0.60, p 4.13e-8, effects consistent in all but 2 of the 22 sex-stratified results files) and rs7738851-A (intronic, *NEDD4*, OR 1.28, EAF 0.85, p 1.63e-8, effects consistent in all but 2 of the 14 files) for permanent teeth. Consortium-wide estimated heritability of caries was low (h^2 of 1% [95% CI: 0%:7%] and 6% [95% CI 0%:13%] for primary and permanent dentitions, respectively). **Discussion:** This study was designed to identify common genetic variants with modest effects which are consistent across different populations. We did not identify convincing evidence for single variant association in the primary dentition under these assumptions. Phenotypic heterogeneity between cohorts may have contributed; these findings could also reflect complexity not captured by our study design, such as genetic effects which are conditional on environmental exposure. Prior studies have nominated fluoride exposure and tooth age as potential modifiers of genetic effects. In contrast to previous estimates from single cohorts, we observed low consortium-wide heritability estimates, highlighting the importance of non-genetic factors for the development of dental caries in children. The nomination of *NEDD4* as a novel candidate is notable given that neural crest cells participate in enamel and dentine formation. *NEDD4* transcripts are reported to regulate cranial neural crest signalling pathways.

2912T

Leveraging whole genome sequence data to improve imputation and increase power in GWAS of diverse populations. *C. Quick[1], C. Fuchsberger[1,2], M. Boehnke[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 2) Center for Biomedicine, European Academy of Bolzano/Bozen (EURAC), affiliated with the University of Lübeck, Bolzano, Italy.

   Imputation reference panels are instrumental in genome-wide association studies (GWAS), providing increased genomic coverage and statistical power from genotyping arrays. Current reference panels, e.g. from the Haplotype Reference Consortium (HRC), include tens of thousands of largely European (EU) individuals. These panels provide high imputation quality in many EU populations, but lower quality in non-EU, admixed, and isolate populations, particularly for rare variants. Capturing genetic variation across diverse populations is crucial to detect population differences in genetic risk and pharmacogenetic factors, maximize gene discovery, and identify causal variants through trans-ethnic fine-mapping. In diverse populations, multi-ethnic or population-matched reference panels can increase imputation quality. Here, we consider an approach in which a subset of participants is sequenced and the rest are array-genotyped and imputed using an augmented reference panel comprised of the sequenced participants and individuals from an external panel. We identify cost-effective approaches to increase imputation quality and maximize statistical power for GWAS in African American (AA), Latin American (LA), Sardinian, and Finnish populations. We used sequence data from 3412 AA and 2068 LA individuals in the TOPMed WGS project and from 2995 Finns and 3445 Sardinians in the HRC to compare imputation quality across multiple reference panels and genotyping arrays. For each population, we selected 2K individuals to use in reference panels and imputed the rest using Illumina Core, OmniExpress, Omni2.5, and Omni5 array SNPs. We assessed imputation quality by comparing true and imputed genotypes for each array using reference panels comprised of 0, 100, …, 2K individuals from the target population and ~29K individuals from the HRC, and used Monte Carlo methods to estimate power as a function of sequenced and imputed sample size and genotyping array. Augmented reference panels substantially increased imputation quality for populations with high genetic distance from the HRC EU panel. For example, adding 2K population-matched individuals to the HRC panel increased coverage for MAF 0.05-1% variants 3.7-fold for AA vs. 1.1-fold for Finns. Our results suggest that sequencing a fraction of participants typically maximizes power given fixed resources in a single study. More broadly, constructing diverse reference panels will be critical to facilitate GWAS in diverse populations.

2913F

GWAS data analysis with non-local prior based Bayesian iterative variable selection-regression. *N. Sanyal, C.-H. Chen.* Radiology, University of California, San Diego, San Diego, CA.

We focus on multiple SNP analysis of genome wide association study (GWAS) data. Because of the ultra high dimension of GWAS data, most multiple SNP analysis approaches, both in the frequentist and in the Bayesian domain, effectively perform some sort of SNP selection by means of shrinkage of the effect sizes. Most commonly used frequentist shrinkage methods such as LASSO and elastic net often produce large number of false positive SNPs. On the other hand, though taking the Bayesian approach helps to reduce the number of false positives, high dimensional GWAS data analysis using existing Bayesian methods such as for example piMASS is computationally quite prohibitive. We present a novel Bayesian GWAS data analysis method that purports to ameliorate the above issues by harmoniously combining two features - dimension reduction as yielded by the use of stepwise selection approach and parsimonious selection as yielded by the use of recently proposed non-local priors. Our proposed method performs SNP selection in an iterative manner, followed by prediction based on the selected SNPs. Iterations proceed guided by a few tuning parameters that specify our method algorithmically, and within each iteration, SNPs are selected by using non-local prior based variable selection. Our method can be applied to both GWAS genotype level data and GWAS summary statistics. In order to illustrate the usefulness and flexibility of our method, we have performed several simulated data analyses using both independent and real genotyped SNPs and both binary and continuous phenotypes. Compared to several existing frequentist and Bayesian methods, our method has shown overall superior performance in terms of number of false positives, estimation error in the effect sizes, and mean squared error. Finally, several real GWAS data analyses demonstrate the accuracy of our method by comparing the results with existing established knowledge.

2914W

Genome-wide association study of HIV-1 subtype C in Botswana population. *A.K. Shevchenko[1], S.V. Malov[1], D.V. Zhernakova[1,2], N. Cherkasov[1], A. Svitin[1], W. Xie[3,4], V. Novitsky[3,4], M. Essex[3,4], S.J. O'Brien[1].* 1) Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russian Federation; 2) University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands; 3) Harvard T.H. Chan School of Public Health AIDS Initiative, Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; 4) Botswana Harvard AIDS Institute, Gaborone, Botswana.

HIV-infection is a deadly chronic disease spreading across all continents. More than 36 000 000 people currently live with HIV and more than 1 000 000 die due to AIDS (UNAIDS, 2015). Despite efforts and resources directed to the fight against the HIV epidemics for more than three decades, still currently there exists no effective vaccine or cure. According to UNAIDS, Botswana falls into top 3 countries with the highest HIV prevalence in the world with adult aged 15 to 49 HIV prevalence rate of 22.2% (UNAIDS, 2015). Moreover, HIV-1 subtype C (widespread in Botswana), which accounts for over half of HIV-1 infections worldwide, is understudied, because most studies focus on other subtypes. Genome-wide association studies (GWAS) allow to identify common variants associated with the trait in question. In order to efficiently search for the genetic associations we have previously developed GWATCH web platform (Svitin et al, 2014). The broad goal of the Botswana GWAS project is to identify genetic determinants of susceptibility and resistance to infection by HIV-1 subtype C among people severely affected by HIV/AIDS in Botswana. By conducting GWAS analysis on HIV-1C case/control dataset consisting of 955 Tswana people (combined from two partly overlapping datasets of 809 microarray and 362 WGS samples), we found several gene regions slightly below significance level. One of these loci was successfully replicated in the longitudinal study on CD4 and viral load (VL) trajectories in a total of 556 treatment-naive HIV-1C infected individuals in Botswana, which was the first GWAS for HIV-1C acquisition and control of viral replication and the first GWAS in this population (Xie at al, 2016). Despite different types of clinical data in the compared cohorts (HIV infection & AIDS progression), this is an interesting and noteworthy replication. Currently we are working on the replication of our findings in other available datasets. 1. Svitin A. et al. (2014). GWATCH: a web platform for automated gene association discovery analysis. GigaScience. 2. Xie W. et al. 2016. Genome-Wide Analyses Reveal Gene Influence on HIV Disease progression and HIV-1C Acquisition in Southern Africa. AIDS Research and Human Retroviruses.

2915T

Genome-wide association study identifies novel susceptibility loci for tanning ability in Japanese population: From ToMMo cohort study. *K. Shido[1], K. Kojima[3], A. Hozawa[2,3], S. Ogishima[3], N. Minegishi[2,3], Y. Kawai[3], G. Tamiya[3], K. Tanno[4,5], K. Yamasaki[1], S. Aiba[1], Y. Suzuki[2,3], M. Nagasaki[3].* 1) Department of Dermatology, Tohoku University Graduate School, Sendai, Miyagi prefecture, Japan; 2) Graduate School of Medicine, Tohoku University, Sendai, Japan; 3) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 4) Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, Yahaba, Iwate, Japan; 5) School of Medicine, Iwate Medical University, Morioka, Japan.

Genome-wide association studies (GWAS) in European populations identified genes such as MC1R, MATP, OCA2, HER2, SLC24A4, and IRF4 as candidate genes involved in melanogenesis and associated with hair, eye, and skin color. In this study, we tried to identify genes relating to skin types in a Japanese population. We employed GWAS using data from the Tohoku Medical Megabank Organization (ToMMo) cohort study. Japanese skin type I-III were classified based on an individual's susceptibility to sunburn and ability to tan: type I, always burns, never tans; type II, moderately burns, moderately tans; type III, never burns, always tans. Single nucleotide polymorphisms (SNPs) were tested for association with three types of Japanese skin type among 9,966 Japanese who participated in ToMMo cohort study in Miyagi and Iwate prefectures. As a result, we identified two candidate genes at genome-wide significance of less than 5 ×10-8; chr15:28228553 C>T (rs74653330) at 15q12-13 (Oculocutaneous albinism II (OCA2) gene) and chr9:16808172 C>G (rs10122901) at 9p22.3-22.2 (Basonuclin 2 (BNC2) gene) in the population of Miyagi. The association between the nonsynonymous SNP in the OCA2 gene and normal skin pigmentation of East Asian and Japanese has already been reported in the previous studies. We confirmed that the SNP of OCA2 gene had a strong relation to Japanese skin type, and the SNP located in a non-coding region of BNC2 gene was also closely linked with Japanese skin type in the population of Miyagi as well as Iwate prefecture. Furthermore, we additionally found the following three loci in the combined analysis of Miyagi and Iwate prefecture: KITLG rs11104947 on chromosome 12, RAB32 rs11104947 on chromosome 6, and DBNDD1 rs138160245 on chromosome 16. The recent functional analysis showed these genes are involved in the process of melanogenesis. Thus our data from ToMMo cohort study confirmed the involvement of OCA2, BNC2, and KITLG and added two new gene loci RAB32 and DBNDD1 in human skin types in real life.

2916F

Comparison of power of summary based methods for identifying expression-trait associations. *Y. Veturi, M.D. Ritchie.* Biomedical and Translational Informatics, Geisinger Health System, State College, PA.

Transcriptome-wide association studies (TWAS) have recently been employed as one way to draw upon the advantages of both GWAS and gene expression studies to identify genes associated with complex traits. Unlike standard GWAS, TWAS only require summary level data and offer improved statistical power. Two popular TWAS methods include either (a) imputing the *cis* genetic component of gene expression from smaller sized studies into much larger effective sample sizes afforded by GWAS or (b) using summary-based Mendelian randomization (SMR). Although these methods have been effective at detecting functional variants, it remains unclear how extensive variability in the genetic architecture of complex traits and diseases impacts TWAS results. Our goal was to investigate the different scenarios under which these methods yielded enough power to detect significant expression-trait associations. In this study, we conducted extensive simulations on 6000 randomly chosen, unrelated Caucasian males from Geisinger's MyCode population to compare the power to detect *cis* expression-trait associations (within 500 KB of a gene) using the above-described approaches. To test TWAS across varying genetic backgrounds we simulated gene expression and phenotypes with different numbers of quantitative trait loci per gene and gene expression/ genomic heritability using genetic models that differentiate the effect of direct causality from that of pleiotropy. For each gene, on a training set of 1000 individuals, we either (a) imputed gene expression and estimated corresponding weights using five different methods: LASSO, elastic net, Bayesian LASSO, Bayesian spike-slab, and best linear unbiased predictor or (b) obtained the top performing *cis*-eQTL. We then sampled with replacement 50,000, 150,000, and 300,000 individuals respectively from the testing set of the remaining 5000 individuals and conducted GWAS on each set. Subsequently, we integrated the GWAS summary statistics derived from the testing set with the weights (or eQTLs) derived from the training set to identify expression-trait associations using (a) imputed gene expression and (b) SMR. Finally, we examined the power to detect functionally relevant genes using the different approaches under the considered simulation scenarios. We observed substantial differences in power between the methods under varying genetic architectures. All the reported results are based on simulated data from the MyCode population.

2917W

Improving imputation by maximizing power. *Y. Wu[1], E. Eskin[1,2], S. Sankararaman[1].* 1) Computer Science Department, UCLA, los angeles, CA; 2) Department of Human Genetics, UCLA, los angeles, CA.

GWAS estimates the correlations between disease status and collected genetic variants. After estimating the correlations, we perform a statistical test to indicate if each of the estimated correlation is statistically significant. If the absolute value of the association statistics is smaller than the threshold we set, we reject the null hypothesis. In GWAS, imputation is used to aid the interpretation of a GWAS by predicting the association statistics at untyped variants. People perform imputation in two sets of ways. One way is that to impute the genotypes directly at the untyped variants, and then perform a statistical test. The other sets of ways is to utilize summary statistics and impute the association statistics directly. Having the predicted statistics of the untyped variants, people indicate significance if two ways. One way is to adjust the rejection threshold for the untyped SNPs, basing on the correlation between the untyped and typed variants, to control the family-wise error rate. The other way is to set the same reject threshold for all association statistics of the typed and untyped SNPs. In this paper, we closely examine the different methods of imputation and the significant indicating methods. We found that while imputation increasing power, previous methods to indicate significance loose power. Thus we propose our method, where we impute untagged SNPs using a causal model, which assumes a subset of SNPs in the locus being causal, and then we nd a set of thresholds to maximize the power of tagged SNPs. In an association test, by rejecting null hypothesis for SNPs at different threshold, we can control the false positive rate while maximizing the power.

2918T

Genome-wide analysis of age-related macular degeneration progression. *Q. Yan[1], Y. Ding[2], Y. Liu[1,2], T. Sun[1,2], L. Fritsche[3], T. Clemons[4], R. Ratnapriya[5], M. Klein[6], R. Cook[7], Y. Liu[8], L. Wei[8], R. Fan[9], G. Abecasis[3], A. Swaroop[5], E. Chew[10], D. Weeks[2,11], W. Chen[1,2,11].* 1) Division of Pulmonary Medicine, Allergy and Immunology, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 4) The Emmes Corporation, Rockville, MD; 5) Neurobiology Neurodegeneration and Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, MD; 6) Casey Eye Institute, Oregon Health & Science University, Portland, Oregon; 7) Department of Statistics and Actuarial Science, University of Waterloo, Canada; 8) State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, Guangdong Province, China; 9) Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center, Washington, DC; 10) Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD; 11) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, PA.

Age-related macular degeneration (AMD) is a heritable eye disease and the leading cause of blindness in elderly population in the United States. Family-based and population-based genetic studies have successfully identified multiple disease-susceptibility genes for AMD. However, most studies to date have focused on AMD status (disease or normal) while the genetic influences on AMD progression are largely unexplored. In this study, we performed a genome-wide bivariate time-to-event analysis to test for association of time-to-advanced-AMD (either choroidal neovascularization [CNV] or geographic atrophy [GA]) with ~9 million variants (either from exome chip or imputed) on 2,721 Caucasians from a large multi-center randomized clinical trial, the Age-Related Eye Disease Study. We used a robust Cox proportional hazards model to appropriately account for between-eye correlation when analyzing the progression time in the two eyes of each participant. We identified four susceptibility loci associated with AMD progression with $P < 5 \times 10^{-8}$. All of them were previously reported to be associated with AMD risk, near *ARMS2-HTRA1* ($P = 8.1 \times 10^{-43}$), *CFH* ($P = 3.5 \times 10^{-37}$), *C2-CFB-SKIV2L* ($P = 8.1 \times 10^{-10}$), and *C3* ($P = 1.2 \times 10^{-9}$). In a secondary analysis, limited to the 34 known AMD risk variants, *LIPC* and *CTRB2-CTRB1* were also found to be associated with AMD progression ($P < 0.0015$). Finally, our analysis replicated the finding that rs142450006 near *MMP9* was specifically associated with progression to choroidal neovascularization ($P = 0.0006$) but not with geographic atrophy. In this study, we performed the first genome-wide association study for AMD progression. Our findings expand our understanding of AMD genetics in a clinically-relevant manner by assaying genetic effects on AMD progression.

2919F

Logolas: A tool for visualizing enrichment of genetic signature profiles. *K. Dey, M. Stephens.* University of Chicago, chicago, IL.

Logo plots are popular in genomics for identifying transcription factor (TF) binding motifs. Popular logo plotting tools, such as seqLogo and motifStack, use a library of 4 symbols - A, C, G and T corresponding to 4 bases - to represent the TF binding motifs. While it is adequate for viewing TF binding signature profile, the limited size of the library of symbols makes it hard to extend the applicability of logo plots to other areas of genetics and beyond. We propose a new R package called Logolas, which has a much wider library of logos comprising of all alphabets from A to Z, numbers from 0 to 9 and other symbols like dashes, arrows etc. Most importantly, Logolas allows the user to easily combine the logo representations of these alphanumeric characters to create string logos. This enables the user to represent almost anything that can be written using alphanumerics, say methylation profile, histone names, protein amino acids, names of metagenomic/ecological taxa and also mutations (symbols like C->T, T->A etc). Additionally, Logolas provides a simple user interface where an user can build from scratch a logo for a symbol of her own choice and add it to the library. Besides the entropy based representation of enrichment of symbols commonly used in TF binding motifs, Logolas also presents an alternative representation aimed at highlighting both the enrichment and depletion of symbols, compared to the background, in each position of the signature pattern of a motif. We show a number of applications of Logolas to demonstrate how it can be used for visualization for representing TF binding motifs, protein amino acid motifs, mutational profiles, with additional applications in phylogenetics and histone modifications. The package can be installed from Bioconductor (https://bioconductor.org/packages/release/bioc/html/Logolas.html). .

2920W

A semi-supervised method for predicting functional consequences of genome-wide coding and noncoding variants. *Z. He, I. Ionita-Laza.* Biostatistics, Columbia University, 722 W 168th Street.

The tremendous progress in massively parallel sequencing technologies enables investigators to efficiently obtain genetic information down to single base resolution on a genome-wide scale. However understanding the functional effect of genetic variants is much more challenging especially for variants in noncoding regions. Projects such as Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics provide a rich set of functional genomics annotations that can be used to predict functional effects of variants. Most of the methods integrating diffferent functional annotations are unsupervised, particularly due to a lack of experimentally validated functional variants. However as more experimental data across different tissues and cell types becomes available it becomes feasible to employ semi-supervised methods for functional prediction. We propose here such an approach based on manifold regularization to utilize both experimentally confirmed regulatory variants and billions of unlabeled variants, with the goal to improve the accuracy of functional predictions. Through extensive simulations, we show that the proposed method significantly improve prediction compared to existing unsupervised methods. We apply the proposed method to several experimental datasets to predict functional consequences of genetic variants in different cell types, tissues and disease phenotypes and show substantially improved power relative to existing unsupervised approaches.

2921T

Genome-to-genome analysis: Correcting for population stratification in joint association analysis of host and pathogen genomes (G2G) reduces false positive and negative results. *O. Naret, N. Chaturvedi, C. Hammer, J. Fellay.* School of Life Sciences, EPFL, Lausanne, Switzerland.

Background: Genome-wide association studies of host and pathogen genomic variation, which we here call genome-to-genome (G2G) analysis, can identify sites of genomic conflicts between host and pathogens. Due to stratification of host and pathogen populations, G2G analysis can suffer from inflated type I (false positive) and type II (false negative) error rates. We demonstrated through simulation studies that correcting for host and pathogen stratifications reduces spurious signals and increases power. We confirmed the simulation by showing comparable results in a G2G analysis of paired human and HIV genomes. **Method:** To generate human SNPs and pathogen variants, we used the Balding-Nichols model. Additionally, to simulate the causal association between a SNP and a pathogen variant, we used a logistic model. To test associations, we used logistic regression with pathogen variants as dependent variables and host SNPs as independent variables. To correct for host stratification, we included the top five principal components of the SNPs data as covariates. To correct for pathogen stratification, we included phylogenetic principal components as covariates. We then compared the association p-values trend without correction, with correction on host side, with correction on pathogen side and with correction on both sides. **Results and discussion:** In the simulation, we observed a better control of false positive signals after adjustment for host and pathogen covariates. The median association p-values were $p_{no\ correction} = 5.7*10^{-3}$, $p_{host} = 1.9*10^{-1}$, $p_{pathogen} = 3.4*10^{-1}$ and $p_{both} = 4.9*10^{-1}$. For true associations, we observed an overall improvement in the strength of the signal after adjustment for host and pathogen covariates. The median association p-values were $p_{no\ correction} = 1.1*10^{-1}$, $p_{host} = 3.6*10^{-4}$, $p_{pathogen} = 4.1*10^{-5}$ and $p_{both} = 8.3*10^{-7}$. We then confirmed what we observed in the simulation on the HIV data. For the likely false positive association between *rs4913471* on chromosome 12 and amino acid 67 of the protease, the p-values were $p_{no\ correction} = 1.5*10^{-16}$, $p_{host} = 3.2*10^{-7}$, $p_{pathogen} = 4.5*10^{-5}$ and $p_{both} = 3*10^{-4}$. For the known association between *rs9266628* in the HLA region and amino acid 135 of the reverse transcriptase, the p-values were $p_{no\ correction} = 5.2*10^{-18}$, $p_{host} = 1.12*10^{-18}$, $p_{pathogen} = 5.9*10^{-19}$ and $p_{both} = 3.7*10^{-20}$. Our results demonstrated that correcting for stratification on human and pathogen sides offers a better control of type I and II errors.

2922F

Mapping genetic organization and disease liabilities of human cortical surface with summary statistics of vertexwise genome-wide association studies. *C. Fan[1,2], N. Schork[3].* 1) Cognitive Science, University of California, San Diego, La Jolla, CA; 2) Center for Multimodal Imaging and Genetics, University of California, San Diego, CA; 3) J. Craig Venter Institute, La Jolla, CA.

Defining genetic distances from one trait to another has been a critical tool for understanding human diseases and complex traits. This is particularly important for imaging genetics, where the phenotypes in interest are inherently high dimension, composing thousands of measures simultaneously extracted from the same individual. Although genetic correlations were oftentimes used for this context, the logistic hurdles of imaging phenotypes make either moment based summary statistic methods or mixed effects model based methods difficult to apply. Here we proposed two summary statistic based methods to map the genetic topologies of human cortical surface. The first genetic distance metric we defined is weighted Euclidean distance on summary statistics from vertexwise genome-wide association study (vGWAS). This enables us to identify homogeneous clusters from measurements based on the same group of individuals. The second metric is weighted summary statistic based Mendelian randomization (weighted SMR) that gauges the magnitude of associations from brain regions to complex traits. Our simulations indicate our methods identify the genetic topologies well even when signal to noise ratio is low. We applied our methods empirically to imaging genetic cohorts (n=2039) and demonstrated that our methods can provide biological insight on the formation of human brain. In addition to homogeneous modules of cortical surface regions, we reveal the topological maps varying according to the complex traits in interest, including schizophrenia, education attainment, and Alzheimer's disease.

2923W

Gene-based pleiotropic analysis of multiple survival traits via functional regressions with applications to eye diseases. *R. Fan[1], C. Chiu[2], T. Sang[1], R.J. Cook[3], Y. Ding[4], W. Chen[4,5,6], D.E. Weeks[4,5], M.L.T. Lee[7], E.Y. Chew[8], M.M. Xiong[9].* 1) Department of Biostatistics, Bioinformatics, and Biomathematics, GEORGETOWN UNIVERSITY MEDICAL CENTER, Washingon, DC; 2) Biostatistics and Bioinformatics Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (NIH), Bethesda, MD 20892; 3) Department of Statistics and Actuarial Science, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1; 4) Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital, of Pittsburgh at The University of Pittsburgh, Pittsburgh, PA 15224; 5) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Crabtree Hall, Room A303 130 DeSoto Street, Pittsburgh, PA 15261; 6) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Crabtree Hall, Room A303 130 DeSoto Street, Pittsburgh, PA 15261; 7) Departments of Epidemiology and Biostatistics, University of Maryland, College Park, MD 20740; 8) Division of Epidemiology and Clinical Applications, National Eye Institute, NIH, Bethesda, MD 20892; 9) Human Genetics Center, University of Texas – Houston, P.O. Box 20334, TX 77225.

One way to analyze a set of phenotypic traits caused by pleiotropy is to analyze the traits one by one. This approach may lead to low power since it ignores the extra information obtained by combining multiple traits in one unified analysis. In the literature, there are no gene-based statistical methods for a joint analysis of multiple survival traits. We build on our previous work involving functional Cox regression for censored univariate time to event data. Specifically we develop methods for a combined association analysis between a pleiotropic gene (rather than a single SNP) and multiple survival traits through use of multivariate mixed effect Cox proportional hazard model using functional regressions. The methods are evaluated and refined by extensive simulation studies. An application to a motivating study of AMD progression is provided for illustration where we analyze jointly the times to vision loss in left and right eyes.

2924T

Exploring genetic associations using self-reported phenotypes in genes for good. *A. Pandit[1], E.M. Schmidt[1], G.J.M. Zajac[1], W. Li[1], K.K. Brieger[1], G. Datta[2], S.I. Vrieze[2], G.R. Abecasis[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI,, USA; 2) Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA.

Genes for Good is a genetics study that engages 10,000s of individuals through social media to help better understand the genetic basis of human health and disease. As an incentive to contribute, participants receive ancestry results and the option to download their raw, uninterpreted genetic data, including imputed genotypes. All phenotypes are self-reported through online surveys in the Genes for Good Facebook app. To ensure a sufficient amount of phenotypic data per participant, participants must complete a minimum number of surveys in order to receive a spit kit in the mail and provide a sample for genotyping. As of June 2017, >45,000 individuals from all 50 U.S. states have enrolled. While collecting phenotypes via social media is convenient for both parties, the reliability of self-reported measures is hotly debated. To assess the accuracy of our data, we evaluate our ability to replicate known associations. When analyzing an initial freeze of 4,286 European individuals, we were able to reproduce several known genetic association signals, for example associations between body mass index and *FTO* ($p=2.9 \times 10^{-8}$) and between high LDL cholesterol (binary) and *APOE* ($p=5.2 \times 10^{-10}$). Population-level associations between health traits, such as diabetes and obesity, were also replicated.The convenience of collecting phenotypes online allows us room to explore interesting behavioral outcomes, for instance the number of meals or snacks that one eats per day and whether someone prefers to eat alone. When analyzing the number of meals per day in our relatively small sample size, we noted multiple hits approaching genome-wide significance on *CASP1*, a protein-coding gene that plays an integral role in cell apoptosis and has been linked to appetite regulation and body weight ($p=4.7 \times 10^{-7}$). As we continue to explore these relationships in our growing population, we improve our ability to uncover novel genetic associations in diverse behavioral and health-related traits.

2925F

The 1M Africa genotype array: A powerful tool for medical genetic research globally. *T. Carstensen[1,3], A. Mittal[2,3], M.H. Woehrmann[2,3], J. Gollub[2,3].* 1) Wellcome Trust Sanger Institute, Cambridge, Cambridgeshire, United Kingdom; 2) Thermo Fisher Scientific, Santa Clara, CA, USA; 3) AGR investigators.

Large-scale genome-wide association studies in Africa are essential to better understand genetic determinants of complex traits in this region. The high levels of genetic diversity across Africa, and the high proportion of population specific variation are unlikely to be captured well by existing arrays that are based primarily on European populations, or a small subset of relatively homogenous populations across western Africa. Development of representative and cost-effective genotype arrays is key to facilitating medical genetics research across the continent. Here, we present the development and design of an Africa specific array, a cost-effective array, with high accuracy across African populations and globally. African 1M genotype array has been developed based on the African Genome Resource, a resource of low coverage sequence data (4x, 6x and 8x) encompassing 5000 individuals and 10 African populations, including populations from the 1000 Genomes Project Phase 3. Markers were selected using the Axiom™ genotyping array technology. We used a hybrid imputation and tagging approach to identify the most efficient set of 1 million tag SNPs across populations, to maximise accuracy of capture of common variation across all populations studied. We calculated coverage in each population as the proportion of common variants imputed with >0.80 accuracy (r2) imputed with the AGR reference panel using 5-fold cross validation, also excluding each target population from the reference panel in turn. We additionally included pre-selected markers important in the context of disease research and pharmacogenomics in global populations. The Africa 1M genotype array provided >90% coverage of common variants (MAF>=5%) across all African, European and Asian populations, including the most diverse indigenous population groups from Africa (the Nama), providing strong evidence of the utility of the array across the region, and in multi-ethnic populations globally. The African 1M array developed using the African Genome Resource has the potential to provide a cost-effective and powerful tool for large-scale GWAS in African and multi-ethnic populations worldwide.

2926W

Clonal hematopoiesis: Genetic and phenotypic associations. *C. Tian, P. Fontanillas, J. Tung, D. Hinds.* Research, 23andme, Mountain View, CA.

Age-related clonal hematopoiesis is a common condition that is associated with an increase in hematologic cancer and in all-cause mortality. Studies have shown that 10% of individuals over the age of 65 years have recurrent somatic mutations and large-scale somatic events. We analyzed SNP array data with custom probes for 13 previously characterized single-nucleotide somatic mutations (*DNMT3A* r882c/r882h, *IDH1* r132h/r132l, *IDH2* r140q, *JAK2* v617f, *KRAS* g12d/g13d, *NRAS* g13r, *NRAS* q61r, *SF3B1* k700e, *TP53* r248q and *U2AF1* s34p) in ~800k 23andMe consented research participants and we further used the data to detect large mosaic structure alterations. The presence of the 13 somatic single-nucleotide mutations was detectable in 1.15% of the study population, and it increased from 0.7% in young individuals (<= 45 years old) to 2.0% in older individuals (> 65 years old). Large mosaic structural alterations were detected in 1.5% of the study population. To understand risk factors that contributed to having a detectable somatic mutation, we performed a genome-wide association scan for each of the 13 single-nucleotide somatic mutations. As previously described, *JAK2* (rs12349785, P=4.0e-95, OR=0.436), *TERT* (rs2736100, P=1.4e-42, OR=1.722), *TET2* (rs199741557, P=6.0e-14, OR=0.497), *SH2B3* (rs653178, P=5.4e-9, OR=0.795) and *ATM* (rs11212666, P=1.3e-9, OR=1.27) were significantly associated with *JAK2* v617f carrier status. We identified novel associations with *IFT80* (rs201009932, P=1.2e-15, OR=3.14), and with two regions that were both in DNase hotspots (rs7803075, DNase hotspot in leukemia tissue, P=1.7e-13,OR=0.735; and rs621940, 9q34.13, GFI1B-[]--GTF3C5, P=1.7e-10, OR=1.376). *TERT* is also significantly associated with having *DNMT3A* r882c/r882h mutations. Variants in 15q26.3 region (rs76685039, P=6.3e-16, OR=1.3) are found to be significantly associated with having *KRAS* g13d somatic mutations. A region in 6q21 indexed by rs1832777 (P=2.7e-11, OR=1.5), previously reported to be associated with red blood traits, is significantly associated with having the *SF3B1* k700e mutations. These data indicate that the germline variants predispose individuals to clonal hematopoiesis, which may later cause an overt neoplasm. The presence of somatic mutations was also analyzed for association with self-reported phenotypes. Somatic mutation carriers reported higher instances of having ever taken chemotherapy meds (P=6.1e-15, OR=1.81) and being current tobacco users (P=2e-4, OR=1.2).

2927T

Summary statistic GWAS joint analyses across 50+ traits. *H. Aschard[1,2], C. Lasry[1], P. Lechat[1], H. Menager[1], B.J. Vilhjalmsson[2,3], V. Guillemot[1].* 1) C3BI, Institut Pasteur, Paris, France; 2) Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, USA; 3) Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark.

Genome-wide association studies (GWAS) have been performed on hundreds of traits and diseases, identifying thousands of significant genetic associations. Besides genome-wide associated variants these studies carry large amount of information that can be used to further deciphering the genetic architecture of human phenotypes. In particular, the joint analysis of multiple diseases and traits has received limited attention, even though multivariate analyses have the potential to improve the detection of new genetic variants. Here we propose JASS (Joint Analysis of Summary Statistics), a computationally efficient framework for the joint analysis of multiple phenotypes based on GWAS summary statistics. JASS allows for the application of several multi-trait statistical tests, including standard omnibus approach and hypothesis driven weighted sum of individual phenotypes statistics. Our framework also solves several practical and methodological issues which have been overlooked in previous studies. We demonstrate how these issues can result in bias when testing association between single nucleotide polymorphisms (SNPs) and multiple phenotypes, illustrate the resulting increased false positives rate in real data, and developed solutions to avoid such pitfalls. We applied JASS for the joint analysis of publicly available GWAS summary statistics of more than 50 phenotypes altogether, or using sub-groups of phenotypes based on shared genetics or shared environmental variances. We detected dozens of genome-wide significant variants missed by univariate phenotype screenings. The newly identified variants include strong candidate genes. For example the joint analysis of 5 psychiatric diseases identified *PCDHA1*, a gene with a potential role on cell-cell connections in the brain, while the joint analysis of 5 auto-immune diseases, identified *PLCL1*, a gene known to play a role in allergy (while our auto-immune phenotype set did not include allergy-related phenotypes). Comparison between joint tests highlighted higher power of the omnibus approach on average, while weighted sum of individual phenotype statistics show better performances only in a few specific scenarios. Finally, we present a publicly available implementation of our framework (jass.pasteur.fr/static/), which includes all 50 publicly available GWAS curated by our group, thus allowing investigators to assess association with any subset of phenotypes of interest for a specific disease.

2928F

Polymorphisms in the *HSF2, LRRC6, MEIG1* and *PTIP* genes correlate with sperm motility. *S. Rajender[1], N. Gupta[1], S.N. Sankhwar[2].* 1) Endocrinology Division, Central Drug Research Institute, Lucknow, UP, India; 2) Dept. of Urology, King George's Medical University, Lucknow, India.

Abnormality in the genes involved in the process of sperm production or maturation may cause morphological and functional changes in spermatozoa, which would be reflected in the form of poor semen parameters or loss of fertility. Implication of sex chromosomal genes (X or Y) in male infertility has been well studied, particularly the association of Y-deletions with loss of fertility. However, alterations in the autosomal genes could also be a contributing factor in the etiology of male infertility. We have genotyped 24 functionally important SNPs from 24 infertility strong autosomal candidate genes in 500 infertile men. The genotype data for these polymorphisms were segregated into two groups according to the dominant model of genetic analysis (11 and 12 plus 22). Sperm count and motility were compared between the alternate genotypes at each test locus. We observed a significant change in sperm motility between the alternate genotypes for *LRRC6* rs200321595 C>G polymorphism (P = 0.012; CC: 35.98 ± 18.06, CG+GG: 45.47 ± 17.01)*, MEIG1* rs150031795 G>A polymorphism (P = 0.040; GG: 36.18 ± 18.04, AA: 47.57 ± 20.04)*, HSF2* rs143986686 *T>C* polymorphism (P = 0.0.021; TT: 36.18 ± 18.06, CC: 47.95 ± 18.32) and *PTIP* rs61752013 polymorphism (P = 0.013; CC: 35.97 ± 17.79, CT: 46.70 ± 19.21). Sperm count between the above comparisons did not differ significantly. Multi-dimensionality reduction analysis revealed that rs61752013 (*PTIP*) and rs200321595 (*LRRC6*) were the best interacting SNPs for sperm motility analysis. We conclude that polymorphisms in HSF2 (rs143986686), LRRC6 (rs200321595), MEIG1 (rs150031795) and PTIP (rs61752013) correlate with sperm motility and hence infertility risk. .

2929W

Transformation of summary statistics from linear mixed model association on all-or-none traits to odds ratio. *L.R. Lloyd-Jones[1], J. Yang[1,2], P.M. Visscher[1,2].* 1) Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia; 2) Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia.

Genome-wide association studies (GWAS) have identified thousands of loci that are robustly associated with complex diseases. GWAS are often hindered by population stratification and cryptic relatedness, which can induce spurious associations if not properly accounted for. Principal component correction in combination with logistic regression is a popular method for GWAS of disease traits when population structure is of concern. The use of linear mixed model (LMM) methodology for GWAS is becoming more prevalent due to its ability to control for population structure and cryptic relatedness. The odds ratio (OR) is a common measure of the association of a disease with an exposure (e.g. a genetic variant) and is readably available from logistic regression. However, when the LMM is applied to all-or-none traits it provides estimates of genetic effects on the observed 0-1 scale, a different scale to that in logistic regression. This limits the comparability of results across studies, for example in a meta-analysis, and makes the interpretation of the magnitude of an effect from an LMM GWAS difficult. In this study, we derive transformations from the effects generated under the LMM to the OR that only rely on summary statistics. To test the proposed transformations, we used real genotypes from two large (60,000 and 150,000 individuals) publicly available data sets to simulate 50 all-or-none phenotypes for each of five scenarios that differed in underlying model, disease prevalence and heritability. We measured the performance of each transformation by estimating the slope and adjusted R^2 from regression of the OR estimates from the LMM (from 10,000 loci for each scenario) on the simulated truth. For each scenario, the slope and adjusted R^2 were close to unity indicating a very high concordance between the transformed OR and the truth. Furthermore, we applied these transformations to GWAS results for type 2 diabetes generated from 108,042 individuals in the UK Biobank. We observed very high concordance between the transformed OR from the LMM and those estimated from logistic regression, reinforcing the reliability of the derived transformations. These transformations improve the comparability of results from prospective and already performed LMM GWAS by providing a common comparative scale for the genetic effects. We have developed an R Shiny application implementing the methodology, which is available at http://cnsgenomics.com/shiny/LMOR/. .

2930T

Assessing the causal impact of smoking and drinking on human health outcomes: Using 239 novel genetic associations for smoking and drinking addictions and a robust ensemble method for causal inference. *F. Chen, Y. Jiang, DJ. Liu.* Penn State College of Medicine, Hershey, PA.

Cigarette smoking and alcohol consumption are modifiable risk factors for human health outcomes. While generally believed to be harmful, their causal impacts on many human health outcomes remain elusive. Conventional correlation analyses of smoking/drinking traits with diseases are subject to the influence of confounders and reverse causation, so cannot be used to assess causal impact. Mendelian Randomization (MR) is a powerful approach to identify the causal relations between risk factors and diseases. But the success of MR relies critically on the availability of smoking/drinking associated genetic variants that can be used as instrument variables. Despite the known heritability of smoking/drinking addictions, very few loci were consistently implicated in previous studies, which limits the applicability of MR. Recently, the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) project identified 239 novel independent loci associated with five traits of smoking/drinking: cigarettes per day, smoking initiation, smoking cessation, drinking per week and drinker vs. nondrinker, which enabled unprecedentedly powerful MR study. Taking advantage of these newly identified loci for smoking/drinking and publically available summary association statistics, we performed MR analysis on a myriad of human health outcomes, ranging from metabolic traits, cardiovascular diseases to cancer. We synthesized the results from multiple approaches using a novel ensemble method, including traditional multiple instrument MR, egger regression and two-stage regression. We identified several robust pairs of causal relations for smoking/drinking. For example, we noted that tobacco consumptions causes an increased risk for coronary artery disease (CAD). An increased consumption of 5 cigarettes per day will cause a 1.4 fold increase in the odds of developing CAD (95% CI: 1.16-1.72, p<0.01). However, smoking showed an interesting protective effect on gout, for which smoking reduces the OR by 10% (95% CI for OR: 0.08- 0.16, p<0.01). The protective effect is supported by the biological connections between nicotine in-take and urate crystals. Together, these results provide robust novel insights on the causal relations between smoking/drinking and human health outcomes. It also provides support for previous inconclusive causal inference. The method and results will be very useful for predicting disease risk and guiding clinical intervention.

2931F

Simulating autosomal genotypes with realistic linkage disequilibrium and a spiked in genetic effect. *M. Shi, D.M. Umbach, A.S. Wise, C.R. Weinberg.* Biostatistics and Computational Biology Branch, NIEHS, Res Triangle Park, NC.

To evaluate statistical methods one needs to be able to simulate realistic genome-wide genotypes. We here describe a method that can simulate autosome-wide single-nucleotide polymorphism (SNP) data with realistic linkage disequilibrium (LD) and with spiked in, user-specified, multi-SNP causal effects. Our construction uses existing genome-wide association (GWAS) data from unrelated case-parent triads, augmented by including a hypothetical complement triad for each triad (same parents but with a hypothetical offspring who carries the non-transmitted parental alleles). We assign offspring qualitative or quantitative traits probabilistically through a specified risk model and show that our approach destroys the risk signals from the original data. Our method can simulate genetically homogeneous or stratified populations and can simulate cross-sectional studies, case-parent studies, studies of quantitative traits, or studies based on outcome-dependent sampling. We show that allele frequencies and LD structure in the original GWAS sample are preserved in the simulated data. We have implemented our method in a R package "TriadSim".

2932W

X wide association analysis identifies a novel *FRMPD4* locus associated with the differential sex risk for multiple sclerosis. *Y. Zhou, J. Charlesworth, I. van der Mei, B. Taylor, ANZgene Consortium.* Menzies Institute for Medical Researchi, Hobart, Tasmania, Australia.

Multiple sclerosis is characterised by a significant sex imbalance with an increasing female to male ratio of 2.3-3.5:1 for relapse onset MS. Characterising the role of the X chromosome in MS will enable a unique biological perspective on MS aetiology. However, the vast majority of GWAS studies have either not assessed or under assessed the X chromosome, as most of the analysis pipelines are designed for autosomal loci. The availability of a recently developed software (XWAS) that is specially designed for X chromosome analysis has allowed us to assess the role genetic variation of the X chromosome in MS onset. We therefore conducted an X-wide association analysis using the ANZgene MS GWAS data (2482 MS cases and 4304 controls) as a discovery sample involving 6,518 SNPs and replicated using Australian MS exome array data (583 MS cases and 410 controls). We identified a significant association in *FRMPD4* (rs6641026) that showed different effects by sex. In the discovery stage, the MS risk OR for males was 2.06 (P=1.84×10^{-6}), while the OR for females was 0.94 (P=0.35). In the validation stage, the OR for males was 5.09 (P=0.03) while the OR for females was 0.83 (P=0.38). The combined test for differentiated effect size between sexes was significant after multiple testing (P=3.9×10^{-6}). This gene *FRMPD4 functions* as a positive regulator of dendritic spine morphogenesis and density, and is required for the maintenance of excitatory synaptic transmission. The finding of an association with MS risk only in males may suggest a significant parent of origin effect or suggest a gene dosage effect is important and may be protective in females. This gene thus has high biological plausibility for involvement in MS risk. The differential effect by sexes requires more research to elucidate the mechanisms.

2933T

Genome-wide association study identifies novel genetic loci in the Major Histocompatibility Complex (MHC) associated with reduction in *Clostridium difficile* Infection (CDI) recurrence in patients treated with bezlotoxumab. *J. Shen[1], D.V. Mehrotra[1], M.B. Dorr[1], Z. Zeng[1], J. Li[2], M.H. Wilcox[3], R. Mogg[1], R.L. Blanchard[1], P.M. Shaw[1].* 1) Merck & Co., Inc., Kenilworth, NJ; 2) BGI-Shenzhen, Shenzhen, China; 3) University of Leeds, Leeds, UK.

Bezlotoxumab (BEZ) and actoxumab (ACT) are monoclonal antibodies against *C. difficile* toxins B and A, respectively. Patients receiving a single infusion of BEZ alone or with ACT in the MODIFY I/II trials showed a reduction in the rate of CDI recurrence (rCDI) over a 12-week period compared with a placebo (PBO) infusion. To identify genetic variants that associate with the treatment effect on rCDI, we performed a pharmacogenetics genome-wide association study (GWAS) of pooled data from the MODIFY I/II trials (191 with rCDI and 513 without rCDI). Genetic data were generated on a commercial Affymetrix Axiom array platform. Genotype imputation was performed using the 1000 Genomes Phase 3 reference data and Impute2 software after genetic quality control. Classic human leukocyte antigen (HLA) alleles at *HLA-A*, *-B*, *-C*, *-DPB1*, *-DQA1*, *-DQB1* and *-DRB1* were imputed using HIBAG, and HLA association analyses were conducted. We identified a common intergenic SNP associated with treatment effect on rCDI in the MHC of chromosome 6 located between the HCP5 and MICB genes (rs2516513, MAF = 0.24, P = 3.04E-08, per allele odds ratio [OR] in treated patients = 0.31 and 95% CI: 0.18 to 0.48). Imputation of classic HLA alleles identified two class II HLA alleles in high linkage disequilibrium (LD r^2=0.98) associated with the treatment effect on rCDI (*HLA-DRB1*07:01*, P = 1.93E-05, per allele OR in treatment arms = 0.19 and 95% CI: 0.06 to 0.44 and *HLA-DQA1*02:01*, P = 5.18E-05, per allele OR in treatment arms = 0.21 and 95% CI: 0.08 to 0.46). Carriers of the T allele of rs2516513 and *HLA-DRB1*07:01* allele were associated with statistically significant reduction in rCDI in BEZ and Bez+ACT-treated patients compared with PBO-treated patients (rCDI reduction = 21.5% and 32.3%; one-sided Fisher's exact test P = 1.10E-05 and 7.18E-06, respectively). Conditional regression analyses showed that the signals from rs2516513 and *HLA-DRB1*07:01* were relatively independent although the majority of the association was explained by the SNP. The identified genetic associations within the MHC support a potential role of immune mechanisms in the rCDI response to BEZ and may provide new opportunities for predicting patients more likely to respond to treatment. Further work is required to validate these associations in an independent validation study.

2934F

A *CREBRF* missense mutation substantially affects height in Samoans. *S.L. Rosenthal[1], R.L. Minster[1], N.L. Hawley[2], G. Sun[3], O.D. Buhule[4], H. Cheng[3], T. Naseri[6], M.S. Reupena[5], R. Deka[3], D.E. Weeks[1,4], S.T. McGarvey[7,8].* 1) Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA; 2) Department of Epidemiology (Chronic Disease), Yale University School of Public Health, New Haven, Connecticut, USA; 3) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA; 4) Department of Statistical Methods and Actuarial Science, Makerere University, Kampala, Uganda; 5) Bureau of Statistics, Government of Samoa, Apia, Samoa; 6) Ministry of Health, Government of Samoa, Apia, Samoa; 7) International Health Institute, Department of Epidemiology, Brown University School of Public Health, Providence, Rhode Island, USA; 8) Department of Anthropology, Brown University, Providence, Rhode Island, USA.

The Samoan population is a genetically isolated population ideal for genetic studies as evidenced by the recent identification of a missense mutation (Arg457Gln) in *CREBRF* that significantly affects body mass index (BMI). This SNP, rs373863828 (G>A), is virtually absent in non-Pacific Islander populations (minor allele frequency, MAF < 0.0001), despite being common in Samoans (MAF=0.259). In prior research, we observed that *CREBRF*, and rs373863828 specifically, is also associated with other body size phenotypes—waist circumference, hip circumference, and percent body fat, although not waist–hip ratio. We were concerned that perhaps the increase in BMI might, in part, be due to a decrease in height, which is in the denominator of the equation for BMI. We tested the hypothesis that the minor A allele of rs373863828, which is associated with increased BMI, was associated with decreased height using a cross-sectional sample of Samoan individuals (n=3,068) recruited from 33 villages across both islands of the Independent State of Samoa. We used linear regression to calculate the residuals with age and sex as fixed effects and tested the association of this missense mutation with height residuals adjusting for relatedness using an empirical kinship matrix. Our results show the A allele was associated with an increase, not a decrease, in height (p=1.94 x 10^{-5}, effect size=0.65 cm per copy of the A allele). We repeated our analysis in a replication sample of 2,191 Samoan individuals collected in 1990–91 and 2002–03 from Independent Samoa and American Samoa. When calculating residuals for the heights in the replication sample, we included the additional fixed effects of polity and study cohort in our regression model. Results from the replication study supported the findings in our discovery group (p=4.42 x 10^{-5}, 0.85 cm per copy of the A allele). These results suggest that rs373863828 strongly affects height as well as BMI, and alters height at an effect size among the largest seen in common variants affecting height.

2935W

Investigation of post-colonial demographic structure and the implications for association analyses. *K.A. Rand[1], J.M. Granka[1], D. Garrigan[1], J. Byrnes[1], E.L. Hong[1], C.A. Ball[1], K.G. Chahine[2].* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Lehi, UT.

Recent work has highlighted the need to evaluate the impact of fine-scale population structure on association analyses. Despite thousands of high-powered genome-wide association studies (GWAS), only about one-fifth of all studies include individuals of non-European descent. It is understood that GWAS conducted in European populations are not always generalizable to non-European populations, and that continental population structure is an important factor to consider. However, it remains unclear whether fine-scale population structure, particularly more recent within-continental structure, contributes to bias in the interpretation of GWAS results. From our database of over four-million AncestryDNA customers, we visualize regional fine-scale structure, within continental ethnicity groups, in a subset of over one million AncestryDNA customers who have consented to research. In most cases, such regional United States population structure, identified to relate to post-colonial demography during the last several hundred years, is undetectable in a principal component analysis (PCA). However, we still observe regional allele frequency variation at the single-variant level. To explore whether these regional allele frequency differences play a role in association analyses, we examine the frequency distribution of previously-identified variants in the GWAS and ClinVar catalogs. To account for the effects of continental population structure, we examine regional sub-population frequencies of variant subsets that have been identified in each specific continental population (i.e., SNPs identified in a GWAS of European ancestry for European-descent US subpopulations). While observed frequency differences occur across only some regional groups, in some cases these differences suggest residual confounding due to within-continental population structure. It is unclear whether the observed differences can be attributed to true differences in disease risk, or whether they are solely an artifact of generalizing GWAS results to other, unstudied populations. Our results show that, even when considering very recent time scales, small frequency differences can impact GWAS conclusions and interpretations. This work highlights the fundamental need to understand the generalizability of GWAS results through the lens of both continental, and within-continental, population structure, particularly in large-scale analyses.

2936T

Comparison of PC-based and LME-based population structure adjustment using GWAS and WES markers. *Y. Chen[1], GM. Peloso[1], AL. DeStefano[1,2], S. Seshadri[2], J. Dupuis[1].* 1) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 2) Department of Neurology, Boston University School of Medicine, Boston, MA.

Whole exome sequencing (WES) captures only ~1% of the genome that is in the coding regions of genes. This provides a lower coverage of the entire genome compared with genotyping array-based data. Population structure can cause inflated type-I error and spurious association in genome-wide association studies (GWAS), a problem that is also inherent with WES data. Two established approaches to adjust for population structure are principal components (PCs) and linear mixed effect (LME) models. Due to the lower number of variants available in WES, the performance of these two approaches in WES data may be inadequate. We compare the PCs and model performance of the two approaches using GWAS and WES variants by simulation using 1000 Genome data. We consider the scenarios where both European and African Ancestry individuals are included, and a more subtle population structure scenario with only European Ancestry samples. We find that the differences in GWAS-computed PCs and WES-computed PCs are due to the number of markers included in the calculation, with GWAS-computed PCs having a smaller variance. Association results are similar when adjusting for GWAS-computed PCs or WES-computed PCs, which extends to LME using a genetic relationship matrix computed from GWAS or WES markers. While the approach based on PCs performs better for variants not confounded by population structure, the LME-based approach can achieve higher power for variants confounded by population structure. Finally we explored the above models in a real data set from the Framingham Heart Study (FHS). .

2937F

Pharmacological insights from genetic mapping of the plasma proteome. *J.C. Maranville[1], B.B. Sun[2], J.E. Peters[2,3], R.M. Plenge[1], C.S. Fox[1], J. Danesh[2,3,4,5], A.G. Day-Williams[1], A.S. Butterworth[2,5], H. Runz[1].* 1) Genetics and Pharmacogenomics, MRL, Merck & Co., Inc., Kenilworth, NJ; 2) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) British Heart Foundation Cambridge Centre of Excellence, Division of Cardiovascular Medicine, Addenbrooke's Hospital, Cambridge, UK; 4) Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 5) NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

Genetic perturbations of biological processes can be used to inform the development of drugs with similar mechanisms of action. To identify genetic variants that influence circulating protein levels, we mapped protein quantitative trait loci (pQTL) for 2,994 plasma proteins using 3,301 healthy donors from the INTERVAL study. We found pQTL for 1,478 proteins, including 244 known drug targets. These proteins were targeted by drugs at all stages of development, including; 119 in preclinical studies, 76 in clinical trials, and 49 approved compounds. We overlapped pQTL with publically available GWAS data to provide insight into the potential phenotypic consequences of perturbing drug targets. In some cases genetics appears to support proposed indications. For example, we found that indications match associated traits for 7 of the 15 proteins with cis-acting pQTL that overlap GWAS signals. For other drugs targets, pQTL-GWAS overlaps do not support indications and/or point to safety concerns. We also identified novel disease-protein connections that could guide future drug development. For example, we found 77 novel candidate drug targets that are not currently targeted by any reported drug but which have pQTL that overlap disease GWAS signals. These analyses highlight the potential to use pQTL to inform drug development efforts.

2938W

Measuring the rate and heritability of aging using machine learning methods. *J. Ding[1], E.D. Sun[1], J.Y. Zhao[1], G.R. Abecasis[2], F. Cucca[3], I.G. Goldberg[1], D. Schlessinger[1].* 1) Laboratory of Genetics and Genomics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan Ann Arbor, MI; 3) Institute of Genetic and Biomedical Research, National Research Council, Monserrato, Cagliari, Italy.

It is widely thought that individuals age at different rates. A method that measures physiological age independently of chronological age could therefore be a first step to understand relevant mechanisms; but searches for individual biochemical markers of physiological age have had limited success. In this study, we assessed the extent to which an individual's physiological age could be determined as a composite score inferred from a broad range of biochemical and physiological data. Data were collected in population studies in Sardinia ("SardiNIA") and Tuscany ("InCHIANTI"). We used machine learning strategies on data for ~6,000 Sardinian participants, who ranged in age from 12 to 81. The best predictive models were determined from multiple combinations of dimensionality reduction, classification, and regression algorithms. They reached very strong correlations (R > 0.9) between predicted and actual ages, and showed relative stability in successive visits of the same individuals (R>0.5). Similar results were also seen for participants in the InCHIANTI study. We then defined an Effective Rate of Aging (ERA) for each participant, a continuous trait measured as the ratio of an individual's predicted age to his/her chronological age. The inference that individuals have a characteristic rate of aging is supported by findings that in the SardiNIA cohort, the inferred values of ERA showed genetic heritability of 40%. This has been sufficient to initiate genome-wide association studies that identify genetic variants influencing the rate of aging in genes that affect telomere length and metabolic activity.

2939T

Potentially causal rare variants identified using whole genome sequencing of distant relatives from multiplex families with oral clefts. *F. Begum[1], A. Bureau[2], I. Ruczinski[3], J. Hetmenski[1], M. Parker[4], H. Albacha-Hejazi[5], AF. Scott[6], JC. Murray[7], ML. Marazita[8], JE. Bailey-Wilson[9], TH. Beaty[1], MA. Taub[3].* 1) Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 2) Médecine Sociale et Préventive, Université Laval, Québec, QC G1V 0A6, Canada; 3) Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 4) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; 5) Hejazi Clinic, Clinic, Riyadh, Saudia Arabia 11461; 6) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21224; 7) Department of Pediatrics, School of Medicine, University of Iowa, Iowa City, Iowa 52242; 8) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15219; 9) Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore Maryland 21121.

Many genetic studies have contributed to the long list of genes recognized as influencing risk to non-syndromic oral clefts (NSOC). Majority of the identified common risk loci are not directly causal and collectively can only account for a small portion of heritability. A few sequencing studies have identified some potentially causal rare variants for NSOC. Our main goal is to identify single nucleotide variants (SNVs) as potentially causal for NSOC using whole genome sequence data. We sequenced 170 distantly related members of 58 multiplex cleft families with four different ethnic backgrounds: European-American, Central-American, Middle Eastern and Asian. We restricted our analysis to coding non-synonymous SNVs in autosomes with minor allele frequency reported as less than 1% in the 1000 Genomes project, ExAC and ESP samples. We performed further filtering using CADD, SIFT and PolyPhen to target potentially causal variants. After quality control and variant filtering steps, we included 151 distantly related subjects with NSOC from 53 families in the final analysis. We calculated the exact probability of all affected individuals in a family sharing an SNV, provided one of the affected family members carried it. We also calculated the sharing probability across families irrespective of ethnic groups and corrected for multiple testing. In our analysis of 151affected family members, ten potentially causal variants achieved statistical significance after adjusting for multiple testing. The most significant SNV in *PERM1* (*p*=2.4e-9) was shared by 6 Filipino families followed by an SNV in *LOC100129697* (*p*=6.9e-8), which was shared by 7 families (4 Asian and 3 Syrian). All individuals (totaling 13) in 3 Filipino families shared 2 SNVs in *RNF39* and 2 in *TRIM31*. One SNV in *COL11A2* was shared by 4 Filipino families. One SNV in *SESN2* and one SNV in *ZFYVE26* were shared by 3 and 4 Filipino families, respectively. One SNV in *ZNF574* was shared by 2 Filipino and 1 Chinese families. We identified 10 novel SNVs in 8 protein coding genes using an exact probability method for families from four ethnic groups. Some of these families show enrichment of potentially causal deleterious variants. However, these sharing probabilities (and p-values) may be underestimated if the variants are more frequent in these populations than in the reference databases used. We plan to further assess the role of each of these variants and perform functional analyses where feasible.

2940F

Gene-based association testing of dichotomous traits with generalized linear mixed models using extended pedigrees. *C. Chiu[1], Y. Jiang[2], Q. Yan[3], W. Chen[3], M. Gorin[4], Y. Conley[5], M. Lakhal-Chaieb[6], R. Cook[7], C. Amos[8], A. Wilson[9], J. Bailey-Wilson[9], F. McMahon[10], A. Vazquez[11], A. Yuan[12], X. Zhong[12], M. Xiong[13], D. Weeks[1,14], R. F[12].* 1) Biostatistics and Bioinformatics Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh; 3) Division of Pulmonary Medicine, Allergy and Immunology, Children's Hospital of Pittsburgh at The University of Pittsburgh; 4) Department of Ophthalmology, David Geffen School of Medicine, Stein Eye Institute, University of California Los Angeles; 5) Department of Health Promotion and Development, University of Pittsburgh; 6) Department de Mathematiques et de Statistique, Universite Laval; 7) , Department of Statistics and Actuarial Science, University of Waterloo; 8) Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth; 9) Computational and Statistical Genomics Branch National Human Genome Research Institute, NIH; 10) Human Genetics Branch and Genetic Basis of Mood and Anxiety Disorders Section, National Institute of Mental Health, NIH; 11) Department of Epidemiology and Biostatistics, Michigan State University; 12) Department of Biostatistics, Bioinformatics, and Biomathematics, Georgetown University Medical Center; 13) Human Genetics Center, University of Texas - Houston; 14) Departments of Human Genetics and Biostatistics, Graduate School of Public Health, University of Pittsburgh.

In recent years, fixed effect functional and additive regression models have been developed for unrelated samples to test for gene-based association between a quantitative/dichotomous/survival trait and multiple genetic variants in a gene region. The functional regression models reduce high dimensionality of next-generation sequencing data to draw useful information. In major gene association analysis, these models have similar or higher power than sequence kernel association tests (SKAT), its optimal unified tests (SKAT-O), and a combined sum test of rare and common variant effect (SKAT-C). Here we extend this approach to accommodate family-based data using additive generalized linear mixed models (GLMM) and generalized functional linear mixed models (GFLMM) to test for association between a dichotomous trait and genetic variants in a defined chromosome region. The GLMM and GFLMM model the effect of a major gene as a fixed mean, the polygenic contributions as a random variation, and the correlation of pedigree members by kinship coefficients. Then likelihood ratio test (LRT) statistics are used to test if the fixed mean is zero to perform gene-based association tests. Simulation results indicate that the LRT statistics accurately control the type I error rates for a pedigree dataset of moderate sample size (i.e., 456 individuals in 50 pedigrees). We illustrate the behavior of these statistics by applying them to an age-related macular degeneration (AMD) family data set.

2941W

TRUFFLE: Tests of undetermined relationships between founders - fast, light and efficient. *A. Dimitromanolakis[1,3], L. Sun[1,2], A.D. Paterson[4,2].* 1) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Division of Biostatistics and Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada; 4) The Hospital for Sick Children Research Institute, Toronto, Ontario, Canada.

As studies generate GWAS data on large (>0.5M) numbers of subjects, estimating relatedness and co-ancestry among pairs of individuals is important for family-based analyses, but a computationally intensive challenge. Maximum likelihood methods, such as PLINK and PREST are not very accurate in estimating distant relatedness and have bias in mixed ethnicities. More recently, GERMLINE, ERSA and fastIBD estimate recent shared ancestry between pairs of individuals by identifying small shared chromosomal segments using dense genetic data. However, they typically require phasing which is computationally intensive, and can introduce errors. We develop a new method and software (TRUFFLE) that enables the simple and accurate identification of IBD1 and 2 segments, calculation of total IBD1/2 probabilities and provide graphical reporting of distribution of shared segments across pairs of individuals of mixed ethnicities. By skipping the haplotype phasing step and, instead, relying on a simpler region based model, we avoid the influence of haplotype phasing errors and simplify the analysis pipeline. In addition this makes our method much more computationally efficient. TRUFFLE has high power to distinguish 5th to 6th degree relatives from unrelated pairs and the IBD1/2 estimates are as accurate as MLE for close relatives. In addition the method is robust to population structure. We illustrate our approach using phase 3 1000 Genomes data from 2504 individuals. Despite the large number of pairs to be examined, the running time of TRUFFLE was only 30 minutes, using 16 cores of a 2Ghz Xeon processor and memory usage was very low at < 1 GB. There is little influence of TRUFFLE to the effects of the extreme population stratification in this dataset, especially when compared to estimates from PLINK and KING. When examining 2nd-cousin relationships, PLINK reported 687,084 pairs with kinship coefficient > 0.0156, many of which are likely false. While KING was more robust to population structure, it still reported 57,839 such pairs. In contrast, TRUFFLE estimated only 142 such pairs. The TRUFFLE estimated kinship is in high agreement with BEAGLE IBD procedure, but without the very significant computational burden of BEAGLE and the necessity of haplotype phasing.

2942T

A fast algorithm for Bayesian multi-locus model in genome-wide association studies. *W. Duan[1,2,3], Y. Zhao[1,2,3], Y. Wei[1,2,3], F. Chen[1,2,3].* 1) Department of Biostatistics, Nanjing Medical University, Nanjing, Jiangsu, China; 2) The Key Laboratory of Modern Toxicology of Ministry of Education, Nanjing Medical University, Nanjing, Jiangsu, China; 3) Key Laboratory of Biomedical Big Data, Nanjing Medical University, Nanjing, Jiangsu, China.

Purpose: Many multi-locus shrinkage models have been proposed to identify single nucleotide polymorphisms (SNPs) associated with complex traits under a Bayesian framework. However, most use Markov Chain Monte Carlo (MCMC) algorithm, which are time consuming and challenging to apply to data from genome-wide association studies (GWAS). Therefore, an accelerating strategy is needed. **Methods:** Here we propose a variational inference of Bayesian lasso with adaptive prior (BAL-VI) for detecting underlying susceptible variants by fitting all SNPs in GWAS simultaneously. BAL-VI approximates the intractable joint posterior distribution of model with a factorized form including a tractable distribution and minimizes the Kullback-Liebler (KL) distance between this factorized form and full posterior distribution. We use extensive simulations and real data analysis to appreciate the accuracy, time using, and other properties of this model. MCMC version of Bayesian lasso (BL) and BAL are compared with BAL-VI. **Results:** BAL-MCMC and BAL-VI give nearly equal effect estimators for causal variants and much more accurate results than BL. BAL-VI obtains the highest power (TPR) with a well-controlled FPR among all methods by 95% credible interval (Table 1). Compared with BAL-MCMC, the narrow credible interval of BAL-VI makes it easy to falsely include neutral variants. In real data analysis, a data from GWAS of lung cancer susceptibility in China is used. BAL-VI presents similar association signal with BAL-MCMC, but requires less time (Table 2). Sensitive analysis of hyper-parameters on shrinkage factors indicate that BAL-VI under given hyper-parameters performs well in variable selection and effect estimation. Table 1. Simulation: False positive rate (FPR) and true positive rate (TPR) in variable selection with 50 replications (h^2=0.2; n=1,000; p=7,349)

Method	FPR	TPR
BL	9.62E-4	0.3
BAL-MCMC	4.90E-5	0.528
BAL-VI	1.07E-3	0.913

Table 2. Computation time for application in simulation trials and GWAS[a] Time for GWAS.

Data dimensionality	Computation time, hours (SD)		
	BL	BAL-MCMC	BAL-VI
n=500; p=1,000	0.21(0.02)	0.21(0.02)	0.005(0.001)
n=1,000; p=7,349	2.92(0.33)	2.83(0.32)	0.020(0.006)
n=3,435; p=569,615[a]	—[b]	~120	~13

[b]BL model was not applied to GWAS data. **Conclusions:** BAL-VI is a fast and accurate model and can be applied to analyze all SNPs simultaneously in GWAS.

2943F

The SUPERBABY PROJECT: Genetic determinants of the favorable NICU course in premature newborns. *K.M. Gnona[1,2,4], WCL. Stewart[2,5,6], P. White[3,5], M. Klebanoff[1,5], LD. Nelin[1,5], IA. Buhimschi[1,5].* 1) Center for Perinatal Research; 2) Battelle Center for Mathematical Medicine; 3) Institute of Genomics Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio; 4) Department of Biophysics; 5) Department of Pediatrics; 6) Department of Statistics,The Ohio State University, Columbus, OH.

OBJECTIVE: Premature infants are at increased risk for complications [e.g. bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), intraventricular hemorrhage (IVH), and retinopathy of prematurity (ROP)], and many of these conditions can lead to long-term disability. However, among premature infants there is remarkable variation in the presence and severity of complications. We hypothesize that the accumulation of rare variants influences the risk for neonatal complications, which may also explain the clinically noted gender-ancestry effect. **METHODS:** We sequenced the exomes of 209 extensively phenotyped premature newborns (born between 26-31 completed weeks of gestation) who were cared for in the same time period and NICU setting. Ninety infants did not develop severe complications (termed "superbabies", SB), whereas 119 newborns developed at least one severe complication (IVH gr≥3 n=9, NEC gr≥2 n=6, ROP gr≥2 n=27, BPD n=106) (termed "controls", CRL). For each gene, we performed a rank-sum test on the total count of rare variants, and we used a false discovery rate (FDR) approach to identify genes (and pathways) of interest. **RESULTS:** Across 24,121 genes, the distribution of p-values (p<7x10-5) was not uniform, implying that SB and CRL are genetically different. To confirm control of type 1 error, we used real data to compute the distribution of p-values in the absence of association, and found that we could not reject the null hypothesis of "no association" (p=0.87). Furthermore, we used an FDR of 20% to identify genes with excess accumulation of rare variants in either SB or CRL infants. Among these, we highlight *HSP90AA1* (q-value≈0.08), which is in the *HSP90* protein-folding pathway and has known anti-oxidant and anti- protetoxic function. Lastly, in contrasting the exomes of CRL Caucasian males (n=52) with the exomes of SB African-American females (n=14) we confirmed the non-uniform distribution of p-values (p=1.8x10-14) suggesting that the gender-ancestry interaction effect could in fact be strong. **CONCLUSIONS:** With respect to rare variants, the genomic architecture of SB infants is different from CRLs, and this difference appears to influence risk for neonatal complications. We propose that exposure of the premature newborn lung to increased oxygen concentrations carries heightened risk for complications for carriers of *HSP90* and/or DNA repair pathway mutations.

2944W

Integration statistics suggest gene expression in the exocrine pancreas may contribute to intestinal obstruction in cystic fibrosis. *J. Gong[1], L. Zhang[2], Z. Baskurt[1], N. Panjwani[1], B. Xiao[1], G. He[1], S. Kim[1,3], F. Lin[1], M. Esmaeili[1], K. Keenan[1], H. Corvel[4,5], S. Blackman[6], G.R. Cutting[6,7], M.L. Drumm[8,9], M.R. Knowles[10], J.M. Rommens[1,11], L. Sun[2,3], L.J. Strug[1,3,12].* 1) Program in Genetics and Genome Biology, Research Institute, the Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Statistical Science, University of Toronto, ON, Canada; 3) Division of Biostatistics, Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 4) Pediatric Pulmonary Department, Hospital Trousseau, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France; 5) Sorbonne Universités, UPMC Univ Paris 06, INSERM, Centre de Recherche Saint-Antoine (CRSA), Paris, France; 6) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 7) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 8) Department of Pediatrics, Case Western Reserve University, Cleveland, Ohio, USA; 9) Department of Genetics, Case Western Reserve University, Cleveland, Ohio,USA; 10) Cystic Fibrosis Pulmonary Research and Treatment Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; 11) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 12) The Centre for Applied Genomics SickKids Research Institute, Toronto, ON, Canada.

Cystic Fibrosis (CF) is caused by dysfunction of the CF transmembrane conductance regulator (CFTR). *CFTR* does not fully explain the variation in disease severity across multiple affected organs such as the lungs, pancreas, and intestine. The international CF gene modifier consortium conducted a genome-wide association study (GWAS) of intestinal obstruction at birth (MI) in 6,770 individuals with CF. Development of a novel statistic that integrates gene expression enabled insights into the mechanism by which gene modifiers influence the presence of MI in CF. Three loci with genome-wide significant variants just 5' of *SLC26A9, SLC6A14* and *ATP12A*, and suggestive *PRSS1* variants (that were independently replicated)*,* were identified. The 5' location of the associated variants combined with the identification of *PRSS1*, a pancreas-specific enzyme, suggested that the loci may impact MI through gene regulation in the pancreas. To test this hypothesis, expression quantitative trait loci (eQTLs) p-values were downloaded from the genotype tissue expression project (GTEx), and a test statistic was defined to assess the evidence for significant GWAS and gene-expression variants colocalizing. The statistic is a standardized difference in summed GWAS Wald association statistics between variants that overlap GTEx eQTLs versus variants that are not eQTLs at a given locus. The corresponding p-value is obtained by permutation, accounting for correlation between variants. To assess the most probable contributing gene and tissue, a similar statistic contrasted standardized average GWAS test statistics for tissue- and gene-specific eQTLs, with a companion R script displaying gene-by-tissue colocalization heatmaps. Generalizations using a unifying regression framework are underway, providing scalable analytic solutions for more complex data. Applying this method to the CF MI loci demonstrated that the associated variants colocalized with eQTLs for *SLC26A9, SLC6A14* and *ATP12A* in the pancreas, and showed significant tissue-specific enrichment in the pancreas over all other tissues investigated, eg. small intestine, transverse colon, stomach and esophagus. Results suggest that *SLC26A9, SLC6A14 and ATP12A* are the most likely contributing genes, and may impact MI by varying gene expression in the CF pancreas. The proposed methodology is applicable to any trait where summary association statistics are available to integrate with functional data to test mechanistic hypotheses.

2945T

Bayesian methods for genetic associations and causal inference yield potential biological insight for genetics of gene regulation. *B. Jo[1], C. Brown[2], B. Engelhardt[3].* 1) Department of Quantitative and Computational Biology, Princeton University, Princeton, NJ, USA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA; 3) Department of Computer Science and Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA.

Expression quantitative trait loci, or eQTLs, are enriched for polymorphisms that have been found to be associated with disease risk. Meanwhile, Mendelian randomization (MR) uses the naturally randomized genotype as instrumental variables to conduct statistical tests of causality, such as among gene expression levels. With the recent advances in genomic sequencing technologies, large-scale projects such as the Genotype Tissue Expression (GTEx) project version v6p, including of 449 individuals with RNA-sequencing data across 44 tissue types, allow us to identify both local and distal eQTLs with sufficient power, and to use MR to test for causal, potentially mechanistic associations among gene expression levels. In this study, we develop a Bayesian method for MR, and we compare different methods for causal inference to study the mechanisms of distal expression QTLs and genome-wide association studies. A number of methods that have similar goals exist, including Mendelian randomization, mediation analysis (e.g., PrediXcan), and co-localization analysis (e.g., RTC or coloc). In this work, we combine the association Bayes Factor with the co-localization analysis to yield a novel set of tissue-specific Bayesian causal genetic networks, as well as a biological interpretation of the expected frequency of local and distal eQTLs encoded in terms of priors. We then compare the frequentist causal inference Wald statistic to the Bayesian causal analysis. Therefore, this work extends the current Bayesian causal inference methods, and then provides a novel, rigorous comparison of frequentist and Bayesian methods in genetic association and causal inference studies in a large-scale, multi-tissue expression dataset. Finally, we also provide a potential biological insight for genetics of gene regulation derived from this application.

2946F

Human knockouts in the Ashkenazi Jewish population. *A. Kleinman, O. Sazonova, S. Pitts, V. Vacic, R. Gentleman.* 23andMe, Inc., Mountain View, CA.

The Ashkenazi Jews are a founder population that went through a population bottleneck event approximately 700 years ago. As a result, the population features both pervasive IBD sharing and a number of deleterious variants fixed at relatively high frequency. This means that by only sequencing a relatively modest number of Ashkenazim, we expect to be able to build a imputation panel that captures a relatively large amount of the haplotypic diversity within the population, and that we should be able to impute a number of deleterious variants well. We started with a set of 40,000 genotyped consented research participants whose genomes we classified as > 85% Ashkenazi. To aid imputation, we selected samples for sequencing according to a computationally efficient greedy algorithm that seeks to maximize the total amount of population genome that is in IBD with one or more of the sequenced individuals. This resulted in a set of 481 individuals who covered an average of ~3040 cM (89%) of the autosome of each non-sequenced individual. By comparison, we computed that a random selection algorithm would have to choose more than 510 individuals to achieve the same coverage. We sequenced the chosen individuals at a median ~30x coverage and annotated the variants using VEP/LOFTEE, which showed a total of 5,291 autosomal loss-of-function SNPs and indels in 3,714 genes. We removed variants which we observed only once, leaving us with 16.3M SNPs and 4.7M indels, and then imputed these variants into 48,916 additional genotyped and prephased Ashkenazim. The imputed variants included 4,151 loss-of-function variants. There are 1,768 genes in which we observed at least one loss-of-function variant, and 394 in which we observed more than one. The genes HLA-A, HLA-B, HLA-DQA1, HLA-DRB1 and HLA-DRB5 each had more than 5 loss-of-function variants, and together accounted for 59 of the total. There are 2178 imputed individuals (4.4%) who are homozygous loss-of-function in at least one of 664 genes. We restricted to variants with accurate imputation and calculated transmission probabilities for these LOFs in 1736 trios. In comparison to the expectation under Mendelian inheritance, we observed a deficit of 46 double transmissions of the minor allele per 1,000 transmissions from a pair of heterozygous parents (p=0.062), with 61 of 300 possible double transmissions corresponding to a double-transmission probability of 20.3%, suggesting that homozygous loss-of-functions are counter-selected.

2947W

A genetic variants simulation program to simulate high order epistatic interactions for family-based studies. *Q. Li, J. Bailey-Wilson.* Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD.

In recent years, with the increased popularity of sequence data, the research community has paid more attention to family-based studies of complex traits. Therefore, an efficient simulation program is in demand to accommodate the current variety of genetic variant data and complex disease mechanisms. Most of the existing simulation software rely on gene-dropping methods and can be inefficient to generate high order gene-gene interaction effects for extended pedigrees. Previously, we developed an R package, trio, which incorporated a haplotype-based simulation of case-parent trios when the disease risk depends on higher-order epistatic interactions or gene-environment interactions. In this work, we extend that algorithm to simulate genetic variants, including rare variants, in extended pedigrees. By conditioning on the haplotypes of the set of initially ascertained, affected children, we can calculate the haplotype frequencies as well as the probability of affection status of the parents and other relatives. We allow various pedigree configurations and gene-based epistatic interactions which may involve different causal rare variants in the same genes within, or across pedigrees. We compared the validity and the efficiency of our simulation program with other existing simulation software, SIMLA, SIMLINK and SIMULATE.

2948T

Modeling functional enrichment improves polygenic prediction accuracy in UK Biobank and 23andMe data sets. *C. Marquez-Luna[1], S. Gazal[2,3], P. Loh[2,3], N. Furlotte[4], A. Auton[4], A. Price[1,2,3], 23andMe Research Team.* 1) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) 23andMe Inc., Mountain View, CA.

Recent methods for polygenic prediction place priors on unobserved causal effect sizes and shrink observed effect sizes to obtain posterior means, accounting for statistical noise as well as linkage disequilibrium (LD) between variants (e.g. LDpred-inf; Vilhjalmsson et al. 2015 AJHG). Here, we extend LDpred-inf to incorporate functionally informed priors using the baseline-LD model trained on a set of 31 independent diseases and complex traits (average N=84,686; Gazal et al. bioRxiv). This model jointly infers enrichment of functional elements (coding, conserved, histone marks, enhancers, etc.) as well as LD-dependent causal effect size architectures using stratified LD score regression (Finucane et al. 2015 Nat Genet). Our polygenic prediction method specifies a prior distribution of causal effect sizes based on the expected per-SNP heritability under the baseline-LD model and computes posterior means accordingly, again accounting for statistical noise as well as LD between variants. We applied the method to predict height in the UK Biobank. We used association statistics from N=113,660 UK-ancestry samples as training data (correcting for 10 PCs) and N=8,441 samples of other European ancestries as validation data, to minimize confounding due to uncorrected population stratification or cryptic relatedness. We analyzed 2.1 million genotyped and imputed SNPs with minor allele frequencies greater than 1% that passed quality control filters. As expected, LDpred-inf obtained higher prediction accuracy (r^2=0.239) compared to standard pruning/thresholding methods (r^2=0.229). Incorporating priors from the baseline-LD model produced an additional 10% increase in prediction accuracy (r^2=0.263; P<10^{-16} for improvement). We repeated the analysis using European training data from 23andMe customers who consented to participate in research (N=429,528), and observed moderately higher prediction accuracies (r^2=0.280-0.301). The moderate improvement despite the much larger sample size is consistent with the observed 30% higher heritability in UK Biobank, which is similar to previous work (Ge et al. 2017 PLoS Genet). Overall, our results show that modeling functional enrichment substantially improves polygenic prediction accuracy, bringing polygenic prediction of complex traits closer to clinical utility.

2949F

Total serum IgE whole genome sequence association analysis in families from Barbados. *A. Shetty[1], M. Daya[1], N. Rafaels[1], MP. Boorgula[1], S. Chavan[1], C. Foster[2], P. Maul[3], T. Maul[3], H. Watson[4], I. Ruczinski[5], TH. Beaty[6], RA. Mathias[2], KC. Barnes[1].* 1) Center for Personalized Medicine and Biomedical Informatics, School of Medicine, University of Colorado, Anschutz Medical Campus, Denver, CO; 2) Johns Hopkins Asthma & Allergy Center, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Genetics and Epidemiology of Asthma in Barbados, The University of the West Indies, Barbados; 4) School of Clinical Medicine and Research, Queen Elizabeth Hospital, The University of the West Indies, Barbados; 5) Department of Biostatistics, School of Public Health, Johns Hopkins University, Baltimore, MD; 6) Department of Epidemiology, School of Public Health, Johns Hopkins University, Baltimore, MD.

RATIONALE: Asthma is a complex disease with striking disparities between ethnic and racial groups, Asthmatics of African descent have more severe asthma, poorer response to therapy, and higher total serum IgE (tIgE) levels compared to asthmatics of European ancestry. tIgE has been shown to play an important role as both a marker and mediator in allergic disease. Genetics of asthma research has been conducted in the Barbados Asthma Genetics Study (BAGS) for more than two decades focused on a population of African descent. As part of the NHLBI's Trans-Omics for Precision Medicine (TOPMed) Program, high coverage whole genome sequencing (WGS; 39.4 average depth) was performed on samples collected from 189 families (mean pedigree size of 5.3 individuals, ranging from 1 to 48 individuals). This dataset provides us with the opportunity to investigate genetic variants that underpin tIgE levels in the BAGS population. **METHODS**: We used TOPMed freeze 3a data of 967 BAGS samples that passed quality control filters to test for association of tIgE levels. The dependent variable, tIgE, was logarithmically (natural log) transformed to normalize its distribution and diminish the effect of outliers. tIgE association analysis was then performed using linear mixed models implemented in the GENESIS R-package. Age, sex, asthma case-control status, tIgE batch and principal components associated with tIgE were included as fixed effect covariates. A kinship matrix was included as random effect to account for relatedness between subjects. **RESULTS**: Previously, the NHLBI-supported EVE Consortium performed a genome-wide meta-analysis of tIgE in 4292 ethnically/racially diverse (i.e. African American, Latino, and European American) asthma cases and controls. We investigated the 10 most significant associations from EVE, and found that SNPs rs10944017 (6:85300836), rs9469220 (6:32658310), rs7751374 (6:107981856), rs10124954 (9:82836415) and rs16977747 (17:71699790) were similarly associated in BAGS, with effect directions consistent with EVE. However, the p-values from BAGS (p=0.01-0.04) were marginal, which may be explained by environmental differences and/or the much smaller sample size of BAGS. In future, we will be performing burden tests, fine-mapping and pathway analysis.

2950W

Sum ranking, simple but powerful method for detecting pleiotropic loci. *G.V. Roshchupkin[1,2,3], H.H.H. Adams[2,3], R.A. Poot[4], M.A. Ikram[2,3,5], W.J. Niessen[1,3,6], H. Tiemeier[2,7,8].* 1) Department of Medical Informatics, Erasmus MC, Rotterdam, Netherlands; 2) Department of Epidemiology, Erasmus MC, Netherlands; 3) Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, Netherlands; 4) Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands; 5) Department of Neurology, Erasmus MC, Rotterdam, Netherlands; 6) Faculty of Applied Sciences, Delft University of Technology, Delft, Netherlands; 7) Department of Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, the Netherlands; 8) Department of Psychiatry, Erasmus MC, Rotterdam, the Netherlands.

Genetic pleiotropy is when a single gene influences more than one trait. Genome-wide associations studies (GWAS) have identified many variants that each affects multiple trails, suggesting that pleiotropy may be common phenomena for many human complex phenotypes. However, systematic detection of such variants based on meta-analysis summary data is challenging and current methods do not satisfy pleiotropy requirements for the null hypothesis test. Here we propose a simple and feasible statistical test for single loci pleiotropy detection. This test is based just on GWAS summary p-values, which convert to individual rank per every variant, and has elegant combinatorics interpretation. Compare to classical methods for combining summary statistics, our approach is robust to extreme values. The higher rank "overwrites" the effect of small one. Therefore, the null hypothesis holds true if some SNP has very significant association with only one trait, but not with another. The null hypothesis will be rejected only if the rank of association is below some threshold for all trails. This property is an exact requirement for pleiotropy. Using this method, we confirmed the known pleiotropic loci (from *TOMM40, TMEM161B, AKAP6, TNRC6B genes*) as well as identified new variants *(from PEF1, ITIH1, CADM2, C3orf38, POU3F2, KNOP1, CDH7 genes)* between body mass index and cognition, with pleiotropy p-values $< 5 \times 10^{-8}$. Moreover, because the method is aimed to detect a single variant we were able to identify 28 independent pleiotropic loci between seven subcortical brain structures (putamen, hippocampus, amygdala, thalamus, caudate, accumbens, pallidum) and schizophrenia, while absent of significant LD score genetic correlation. They are located in such genes as *MEI1* for hippocampus, *DCC* for putamen, *CACNA1I* for thalamus.

2951T

Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *H.K. Finucane[1,2,6], Y.A. Reshef[3], V. Anttila[4,5,6], K. Slowikowski[7], A. Gusev[2], A. Byrnes[4,5,6], S. Gazal[2], P.R. Loh[2], G. Genovese[5,6], A. Saunders[8], E. Macosko[8], S. Pollack[2], J.R.B. Perry[9], S. Raychaudhuri[6,10], S. McCarroll[5,6,8], B.M. Neale[4,5,6], A.L. Price[2], The Brainstorm Consortium.* 1) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Department of Computer Science, Harvard University, Cambridge, MA; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 5) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 6) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 7) Bioinformatics and Integrative Genomics, Harvard University, Cambridge, MA; 8) Department of Genetics, Harvard Medical School, Boston, MA; 9) Medical Research Council Epidemiology Unit, Cambridge Biomedical Campus, Cambridge, UK; 10) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

There are many diseases whose causal tissues or cell types are unknown. Identifying these tissues and cell types is critical for exploring gene regulatory mechanisms that may contribute to disease. Different types of data characterizing tissue- and cell-type-specific activity have been analyzed together with GWAS data to identify disease-relevant tissues and cell types: histone marks, DHS, eQTLs, and gene expression. Of these, gene expression data--without genotypes or eQTLs--has the advantage of being available in the widest range of tissues and cell types. We analyze gene expression data from several sources, together with genome-wide association study (GWAS) summary statistics for 48 diseases and traits with an average sample size of 86,850, to identify disease-relevant tissues and cell types. We develop and apply a new method, LDSC-Specifically Expressed Genes (LDSC-SEG), that uses stratified LD score regression to test whether disease heritability is enriched in regions surrounding genes with the highest specific expression in a given tissue. We show via simulations that LDSC-SEG is well-powered to identify true enrichments for polygenic traits while avoiding genomic confounding, in contrast to other methods that use information about tissue-specific expression. We then analyze two gene expression data sets containing a wide range of tissues to infer system-level enrichments. We identify immune enrichments for immunology diseases, brain enrichments for psychiatric diseases, liver enrichments for lipid traits, and other enrichments consistent with known biology. We also analyze chromatin data from the Roadmap Epigenomics project across the same set of diseases and traits, and conclude that gene expression and chromatin provide complementary information. We then analyze system-specific gene expression data sets that allow us to achieve higher resolution. We detect an enrichment of neurons over other brain cell types for several brain-related traits, enrichment of inhibitory neurons over excitatory neurons for bipolar disorder, and enrichments in the striatum for migraine and in the cortex for schizophrenia, depressive symptoms, bipolar disorder, and BMI. We identify enrichments of alpha beta T cells for asthma and eczema, B cells for primary biliary cirrhosis, and myeloid cells for lupus and Alzheimer's disease. Our results demonstrate that our polygenic approach is a powerful way to leverage gene expression data for interpreting GWAS signal.

2952F

Multivariate generalized linear model for genetic pleiotropy. *D.J. Schaid[1], X. Tong[2], A. Batzler[1], J.P. Sinnwell[1], J. Qing[2], J.M. Biernacka[1].* 1) Health Sciences Research, Mayo Clinic, Rochester, MN; 2) School of Statistics, Beijing Normal University, Beijing, China.

When a single gene influences more than one trait, known as pleiotropy, it is important to detect pleiotropy to improve the biological understanding of a gene. Yet, most current multivariate methods to evaluate pleiotropy test the null hypothesis that none of the traits are associated with a variant; departures from the null could be driven by just one associated trait. A formal test of pleiotropy should assume a null hypothesis that one or fewer traits are associated with a genetic variant. We have developed statistical methods to analyze pleiotropy for analysis of binary, ordinal, or quantitative traits, or a mixture of these types of traits, based on generalized linear models and estimating equations. Our framework provides a sequential approach to test the null hypothesis that $k+1$ traits are associated, given that the null of k traits are associated was rejected. This provides a method to determine the number of traits associated with a genetic variant, as well as which traits, while accounting for correlations among the traits. By simulations, we illustrate the Type-I error rate and power of our new methods, describe how they are influenced by sample size, the number of traits, and the trait correlations, and apply the new methods to a genome-wide association study of multivariate traits. Our new approach provides a quantitative assessment of pleiotropy, enhancing current analytic practice.

2953W

Quantification of MAF-dependent architectures in 14 UK Biobank traits reveals strength of genome-wide negative selection. *A. Schoech[1,2], P. Loh[1,2], D. Jordan[3], S. Gazal[1,2], L. O'Connor[1,2], D. Balick[2,4], P. Palamara[1,2], H. Finucane[1,2,5], S. Sunyaev[2,4], A. Price[1,2].* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Icahn School of Medicine at Mount Sinai, New York City, NY; 4) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA.

Understanding the role of rare variants is important in elucidating the genetic basis of human diseases and complex traits. It has been hypothesized that negative selection may cause rare variants to have larger per-allele effect sizes than common variants. Here, we develop a method to estimate the minor allele frequency (MAF) dependence of SNP effect sizes. We employ a model in which per-allele effect sizes have variance proportional to $[p(1-p)]^\alpha$, where p is the MAF and negative values of α imply larger effect sizes for rare variants. We estimate α by maximizing its profile likelihood in a linear mixed model framework. Simulations based on imputed genotypes from the UK Biobank show that our method obtains unbiased estimates of α under a wide range of genetic architectures, including LD-dependent SNP effects (Gazal et al. bioRxiv). We estimated α for each of 14 UK Biobank diseases and complex traits, including height, BMI, hypertension and asthma, using imputed genotypes from British ancestry individuals (N=114K). All 14 traits produced negative α estimates with 9 significantly negative, implying larger rare variant effect sizes. The inferred best-fit distribution of true α values across traits has mean –0.38 and standard deviation 0.06, with statistically significant heterogeneity across traits. Despite the larger rare variant effect sizes, we show that for most traits analyzed, rare variants (MAF < 1%) explain <10% of total SNP-heritability. We interpret our results using the Eyre-Walker model (2010 PNAS), which quantifies how frequency-dependent architectures depend on the joint distribution of SNP effects on trait and fitness. By performing forward simulations using SLiM 2 (Haller and Messer 2017 MBE) under a realistic European demographic model, we find that an average genome-wide selection coefficient of -10^{-4} or stronger is necessary to explain the α values that we inferred, under plausible levels of coupling between trait and fitness effects. In conclusion, our results show that rare variants have increased effect sizes on human diseases and complex traits due to significant genome-wide negative selection, but this increase is not large enough for rare variants to explain a substantial proportion of total SNP-heritability.

2954T

Local genetic correlation gives insights into the shared genetic architecture of complex traits. *H. Shi[1], N. Mancuso[2], S. Spendlove[4], B. Pasaniuc[2,3].* 1) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA; 2) Department of Pathology and Laboratory Medicine, UCLA, Los Angeles, CA; 3) Department of Human Genetics, UCLA, Los Angeles, CA; 4) Department of Biology, Brigham Young University, Provo, UT.

Although genetic correlations between complex traits provide valuable insights into epidemiological and etiological studies, a precise quantification of which genomic regions contribute to the genome-wide genetic correlation is currently lacking. Here, we introduce ρ-HESS, a technique to quantify the correlation between pairs of traits due to genetic variation at a small region in the genome. Like cross-trait LD Score regression (Bulik-Sullivan et al., Nat Genet 2015), our approach only requires GWAS summary-level data, however, it makes no distributional assumption on the causal variant effects sizes while accounting for linkage disequilibrium (LD) and overlapping GWAS samples. We evaluated our method through extensive simulations, and showed that ρ-HESS is approximately unbiased and more consistent than cross-trait LD Score regression. We then analyzed summary association data from 11 large-scale GWAS consortia across 36 complex traits (630 pairs) spanning 1703 LD-independent loci, and identified 25 genomic regions that contribute significantly to the genetic correlation among these traits. As an example, our analysis indicated the region chr2:21-23M, showing significant local genetic correlation for HDL and TG, $r_{g,local}$ = -0.94 (95% CI [-1.00,-0.65]). We note that this region harbors the APOA gene, known to associate with lipid phenotypes (e.g. HDL, LDL, TG; GLGC, Nat Genet 2013). Notably, we also find 6 genomic regions that contribute to the genetic correlation of 10 pairs of traits that show negligible genome-wide correlation, including the HBS1L region (chr6:134-136M), showing significant local genetic correlation for PLT (number of platelets) and MCH (mean cell hemoglobin), $r_{g,local}$ = 1.00 (95% CI [0.72,1.00]). Interestingly, this region has been previously identified to associate with blood phenotypes (e.g. PLT, RBC, MCH; van der Harst et al, Nature 2012). These results further showcase the power of local genetic correlation analyses. Finally, we report the distribution of local genetic correlations across the genome for 55 pairs of traits that show putative causal relationships. As an example, our bi-directional analyses showed that causal regions for BMI consistently increase triglycerides (TG), $r_{g,local,BMI}$ = 0.47 (95% CI [0.37, 0.57]), but not vice versa $r_{g,local,TG}$ = 0.02 (95% CI [-0.14, 0.10]), providing further evidence that BMI causally increases triglycerides (Mancuso et al., AJHG 2017; Pickrell et al., Nat Genet 2016).

2955F

MEGA analysis of alcohol consumption in diverse populations: The Population Architecture using Genomics and Epidemiology (PAGE) Study.

K.L. Young[1], L. Park[2], I. Cheng[3], Y. Li[3], S. Bien[4], M. Graff[1], H. Highland[1], R. Tao[5], L. LeMarchand[6], C. Schurmann[7], D. Stram[3], S.P. David[8], U. Peters[4], K.E. North[1], T. Matise[9], S. Buyske[9] on behalf of the PAGE Study. 1) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 2) University of Southern California, Los Angeles, CA; 3) Cancer Prevention Institute of California, Fremont, CA; 4) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN; 6) University of Hawaii Cancer Center, Honolulu, HI; 7) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 8) Division of Primary Care and Population Health Stanford University School of Medicine, Palo Alto, CA; 9) Rutgers University, New Brunswick, NJ.

Harmful use of alcohol results in 3.3 million deaths annually and 139 million disability-adjusted life years worldwide, and is a risk factor for obesity, hypertension, and stroke. According to the NIAAA, while some minority groups (African Americans and Hispanics) are more likely to abstain from alcohol, those who do drink have higher rates of alcohol consumption and binge drinking. In addition, most genetic studies of alcohol have focused on dependence phenotypes. In populations of European and East Asian ancestry, genes related to alcohol metabolism, including *ADH* and *ALDH*, have been identified, and *AUTS2* and *SLC6A1* have been associated with alcohol consumption. Studies of diverse populations are necessary to provide insight into population-specific genetic variability to guide precision screening and intervention efforts for those at greatest risk of alcohol misuse. We performed the first discovery study for alcohol traits using the Multi-Ethnic Genotyping Array (MEGA), a custom array designed to capture genetic variants in diverse populations to allow for imputation to low minor allele frequencies across ancestries, as well as identification of clinically relevant variants and deep genotyping of previously identified GWAS regions associated with complex diseases. MEGA genotype data imputed to 1KGP3 in African American, Hispanic, Asian, Native Hawaiian, and Native American PAGE participants were used to identify genetic variants influencing alcohol consumption. We compared moderate vs. more than moderate drinkers over the age of 21 (N=30,264) in study-specific pooled single variant additive regression models controlling for sex, age, ancestry, top 10 global ancestry principal components, and study center. Five known variants reached genome-wide significance (p<5E-8), one near *ADH1B* and four near *ALDH2*. An additional 12 novel variants reached suggestive significance (p<1E-6), most of which were low frequency (effect allele frequency [EAF] <5% in the 1000 Genomes EUR reference population), but common (EAF ≥ 5%) in PAGE. Three risk variants were common in African descent groups (OR [95%CI]=1.65[1.37-1.99] near *MIR548AB/MIR548A3*, 1.80[1.44-2.25] near *PIGZ/MFI2,* and 1.40[1.23-1.60] near *PER2/NDUFA4*). One risk variant near *MIR548AU* (OR [95% CI]=1.43[1.25-1.64]) was common in all PAGE populations except Africans. Future work will focus on analyses of other alcohol consumption phenotypes, as well as fine-mapping known alcohol-related loci typed on the MEGA chip.

2956W

Association detection between ordinal trait and rare variants based on adaptive combination of p-values. *Y. Zhou, M. Wang.* Heilongjiang University, Harbin, China.

Next generation sequencing technology not only presents a new method for the detection of human genomic structural variation, but also provides a large number of genetic data of rare variants for us. Currently, how to detect association between human complex diseases and rare variants using genetical data has attracted extensive attention. In the field of medicine, many of people's health and disease conditions are measured by ordinal response variables, namely, the trait value reflects the development stage or severity of a certain condition. However, most existing methods to test for association between rare variants and complex diseases are designed to deal with dichotomous or quantitative traits. Association analysis methods of ordinal traits are relatively fewer, and considering ordinal traits as dichotomous and quantitative traits will inevitably lose some valuable information in the original data. Therefore, in this paper, we extend an existing method of adaptive combination of P-values (ADA) and propose a new method of association analysis for ordinal trait based on it (called OR-ADA) to test for possible association between ordinal trait and rare variants. In our method, we establish a cumulative logistic regression model, in which the regression coefficients are estimated by the NR algorithm and the likelihood ratio test is used to test the association. Through a large number of simulation studies and an example, we demonstrate the performance of the new method and compare it with several methods. The analysis results show that the OR-ADA strategy is robust to the signs of effects of causal variants and more powerful under many scenarios.

2957T

A large-scale genome-wide enrichment analysis identifies new trait-associated genes, pathways and tissues across 31 human phenotypes. *X. Zhu[1], M. Stephens[1,2].* 1) Department of Statistics, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL.

Genome-wide association studies (GWAS) aim to identify genetic factors that are associated with complex traits. However, individual genetic variants have small effects, making them hard to identify. In addition, lists of individual variant associations give limited biological insights. "Enrichment analyses" can address these problems by focussing on *a set of related genes* (e.g. genes in a biological pathway), instead of individual genetic variants. Here we develop an efficient enrichment analysis method that jointly models GWAS summary statistics at millions of common variants, and use it to analyse 3,913 biological pathways and 113 tissue-based gene sets in 31 human phenotypes. Our results highlight several relevant pathways and tissues for complex traits that were not identified in previous analyses of the same data. For example, the *endochondral ossification pathway* is enriched for associations with adult height, and *liver-related* genes are enriched for Alzheimer's disease. A key feature of our method is that inferred gene set enrichment automatically informs new trait-associated genes. For example, enrichment in *lipid transport* genes suggests strong evidence for association between *MTTP* and low-density lipoprotein cholesterol levels, whereas conventional analyses of the same data found no significant variant near this gene.

2958F

VikNGS: A C++ Variant Integration Kit for Next Generation Sequencing across research studies for robust rare and common variant association analysis. *Z. Baskurt[1], S. Mastromatteo[2], J. Gong[1], LJ. Strug[3,4].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Computer Science, University of Toronto, ON, Canada; 3) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, ON, Canada; 4) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada.

The use of next generation sequencing (NGS) in association studies of complex traits is on the rise. Large sample sizes and consortium efforts will be necessary to increase power. However, integration of NGS from different research groups is a challenge because uncertainty in genotype calls differs across experimental designs. For example, different sequencing platforms, parameters (eg. read depth), and variant calling algorithms, which can all impact genotype uncertainty differentially, can result in spurious association findings. Previously we developed a robust score statistic (RVS) and a workflow (Derkach, 2014) and an R package (Gong, 2016, ASHG) for case-control association analysis of common and rare variants when cases and controls were not sequenced as part of the same experimental design, but the aligned reads in BAM format were available for both groups. The RVS for case control studies has been useful for several research groups (eg. Luo et al. Nat Gen 2017). However, analysis of large sample sizes is not sufficiently enabled by the R package, and more functionality is needed. Here we introduce *Variant Integration Kit for NGS* (vikNGS). VikNGS provides a fast and computationally efficient C++ implementation of the RVS workflow for rare and common variant analysis in case-control studies, as well as generalizations for quantitative trait analysis and the integration of NGS from an arbitrary number of cohorts (k). The VikNGS provides functionality for covariate adjustment, user-defined specification for rare variant collapsing, and power estimation for study planning. VikNGS requires a multi-sample VCF file and a tab-separated file containing individual-level information as input. We derive RVS generalizations for integrating sequence from k-cohorts, and for quantitative trait analysis. Simulation studies demonstrate that the generalizations enable integration of sequence across cohorts with different error models and sequencing depth, retaining the expected Type I error and comparable power to analyses with the true genotypes. We provide a benchmark for computing time between the C++ and R implementations, with the C++ implementation demonstrating significant speed advantages while enabling computation with larger data sets. Large samples sizes will be required to identify rare variation contributing to disease, and vikNGS will enable mega-analysis from large consortia despite differences in sequencing design.

2959W

GLMM-seq: Gene-based detection of allele-specific expression by RNA sequencing. *J. Fan[1], J. Hu[1], M. Reilly[2], R. Xiao[1], M. Li[1].* 1) Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA; 2) Cardiology Division, Department of Medicine, Columbia University Medical Center, New York, NY.

Allele-specific gene expression (ASE) analysis, an alternative and complementary approach to eQTL analysis, is a powerful tool for identifying variation in gene expression. ASE quantifies the relative expression of two alleles in a diploid individual, and the imbalance of expression of the two alleles may explain phenotypic variation and disease pathophysiology. To understand the role of ASE in complex diseases, we can exploit this imbalance by RNA sequencing (RNA-seq), which provides allele-specific read counts distinguished by heterozygous sites. Existing ASE detection methods using RNA-seq data, e.g. MBASED and GeneiASE, report evidence of ASE in single samples, and for each sample, the ASE is quantified for each SNP and a gene-level measure of ASE is obtained by integrating the ASE effects across SNPs in the same gene. However, evidence of ASE is often shared across individuals. It is desirable to have an ASE detection method that can simultaneously model both multi-SNP and multi-individual information. To this end, we developed GLMM-seq, a generalized linear mixed-effects model, which allows the modeling of the RNA-seq read counts across multiple individuals and multiple SNPs in the same gene and the adjustment of covariates. This method is able to detect gene-level ASE for RNA-seq samples under one condition, and can also be modified to detect differential ASE between two conditions (e.g., diseased vs. healthy controls). To evaluate the performance of GLMM-seq when haplotype phase information is known or unknown, we conducted extensive simulations and found that GLMM-seq performs consistently well and outperforms MBASED and GeniASE, especially when sample size is small, read depth is low, and the degree of expression imbalance is low to moderate. We will further apply GLMM-seq to RNA-seq data generated in the Genetics of Evoked Response to Niacin and Endotoxemia Study. Results from this analysis will provide novel candidates for modulation of innate immune responses in humans.

2960T

Reverse regression enables disease only case-control association studies for burden tests. *J. Tom[1], A. Wuster[1,2], D. Chang[1,2], K. Mukhyala[1], J. Vogel[1], J. Reeder[1], J. Hunkapiller[2], M. Brauer[1], T. Behrens[2], W. Forrest[1], T. Bhangale[1,2].* 1) Bioinformatics and Computational Biology, Genentech, South San Francisco, CA; 2) Human Genetics, Genentech, South San Francisco, CA.

Acquiring an adequate number of sequenced controls is non-trivial due to high cost and the difficulty in locating ancestry-matched disease free samples. Using samples with known diseases as controls is possible but problematic, as associations cannot definitively be attributed to the cases or disease of interest. Reverse regression in the context of pleiotropy uses genotype as the dependent variable and disease indicators as the independent variables. For case-control analysis we first performed a genome-wide association study with our disease of interest compared against all other diseases (acting as controls). Then, in order to ensure that associations were due to the disease of interest, we implemented reverse regression (with genotype for common variants or burden score for rare variants as the dependent variable), covariates (such as age, sex, and eigenvectors), and disease indicators relative to the disease of interest followed by model selection. For a given variant, the subset of disease indicators selected determined which disease the variant was associated with. For common variants, we tested reverse binomial regression followed by model selection on a 30x whole genome sequenced dataset jointly genotyped using GATK best practices (n=6,585). In this dataset, Asthma (n=3,225) was the disease of interest, and Rheumatoid Arthritis (RA, n=2,185) and Age-Related Macular Degeneration (AMD, n=1,175) samples acted as controls. We found that an indicator for AMD best explained the known AMD loci CFH and ARMS2-HTRA1 and that the known Asthma locus TSLP was attributed to Asthma. For rare variants, associations were screened using reverse logistic regression with burden score as the dependent variable. This was tested in a 30x whole exome sequenced jointly genotyped dataset (n=15,885) where RA (n=2,299) was the disease of interest and samples from eight diseases were used as controls. Of the 17,957 genes tested with rare variants, we found evidence 926 of these were associated with disease(s) other than RA. As examples, the genes C3, ARMS2, and CFH were attributed to AMD and the gene MUC5AC was attributed to IPF. Simulations were used to assess power.

2961F

Estimating cell-type-specific DNA methylation effects in the presence of cellular heterogeneity. *Y. Feng[1], L. Liang[1,2].* 1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.

DNA methylation is an epigenetic modification that controls cell lineage and regulates gene expression. Signatures of DNA methylation differ across tissues and cell types, and cell composition can largely confound the association between phenotype and methylation when samples consist a mixture of cell populations (e.g. whole blood). Many statistical methods have been developed to adjust for this potential bias. More importantly, examining cell-type-specific DNA methylation effects can help identify the causal cell type(s) to follow up and gain insight into the underlying biology. However, purified cell types are usually not available in large scale studies due to impediment cost. In this work, we proposed a method to estimate cell-specific methylation-phenotype associations from unsorted whole tissue data where cell type proportions are also available. We used a framework that combines Monte Carlo EM algorithm and Metropolis-Hastings sampler to recreate the unobserved cell-specific methylation and to estimate its effect on phenotypes. Through simulations, we demonstrated that the method can successfully identify the true effects under various parameter settings, even when the causal cell type is relatively rare. Application to an Illumina HumanMethylation27 dataset showed that cell-specific methylation pattern decomposed from whole blood using the algorithm matches the directly measured methylation status in purified cell types. The method can be readily applied to existing EWAS datasets and is free of bias due to cell type distribution.

2962W

Generalizing genetic risk scores from Europeans to Hispanics/Latinos. *T. Sofer[1], K. Grinde[1], Q. Qi[2], T. Thornton[1], S. Liu[3,4], A. Shadyab[5], K.H. Chan[6], A. Reiner[7].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Epidemiology and Population Health, Albert Einstein College of Medicine, New York; 3) Epidemiology, Brown University School of Public Health, Rhode Island; 4) Medicine, Alpert School of Medicine, Brown University, Rhode Island; 5) 5Division of Epidemiology, Department of Family Medicine and Public Health, University of California, San Diego School of Medicine, California; 6) 6Laboratory of Molecular Epidemiology and Nutrition, Department of Epidemiology, Brown University, Rhode Island; 7) Department of Epidemiology, University of Washington, Washington.

Introduction: Genetic risk scores (GRSs) are constructed by taking the simple or the weighted sum of all trait-increasing alleles, the latter with weights equal to the estimated effect sizes of the variants. GRSs are widely used in studies involving trait or risk prediction and Mendelian randomization. Single nucleotide polymorphisms (SNPs) used in the GRSs and their weights are often based on results from genome-wide association studies (GWASs), most of which were conducted in populations of European ancestry (EA). However, optimal GRSs may differ between Hispanic/Latino and EA populations due to differences in linkage disequilibrium (LD) structure, gene-environmental interactions, and allelic heterogeneity. **Methods:** We investigated methods for constructing GRSs in samples of Hispanics/Latinos by combining information from results of GWAS in large EA populations with results from GWAS in a medium-sized study of Hispanics/Latinos. We studied approaches for selection of both SNPs and weights in constructing the GRSs. We built GRSs for a few anthropometric, blood cell count, and blood pressure (BP) traits using publically available summary statistics from GWAS in EA studies and the Hispanic Community Health Study/Study of Latinos HCHS/SOL, (n=12,784). We evaluated the performance of these GRSs in an independent study of 3,582 Hispanic women from the Women's Health Initiative. **Results:** For anthropometric traits (height, BMI, waist circumference, waist-to-hip ratio, and hip circumference), selecting SNPs based on the EA GWAS results or on the meta-analysis of the EA GWAS and the HCHS/SOL GWAS (dubbed Meta) performed equally well. Using SNP weights calculated in META was slightly better. For BP traits (Systolic-, Diastolic-, mean arterial-, and pulse-pressure), selection of SNPs using EA GWAS performed poorly, unless SNP weights were derived from the effect sizes estimated in the HCHS/SOL. When selecting SNPs based on Meta, GRSs were more robust to weight selection. For blood count traits (platelet, white blood cells, and hemoglobin counts) GRSs, using the effect sizes from the HCHS/SOL as weights was beneficial for all approaches of SNP selection. **Conclusion:** Variant selection for a GRS based on the EA GWAS performed well, as long as the SNP weights were derived from the Hispanics/Latinos study (i.e., for blood pressure and blood count traits), or in the meta-analysis of the EA study and the Hispanics/Latinos study (i.e., for anthropometric traits).

2963T

Imputation of exome array variants to the Haplotype Reference Consortium (HRC). *S. Bomotti[1], A.P. Klein[1,2,3], C. Vergara[4], C. Valencia[1], B.E.K. Klein[5], K.E. Lee[5], R. Klein[5], P. Duggal[1].* 1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD; 3) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD; 4) Division of Infectious Diseases, Johns Hopkins University School of Medicine, Baltimore, MD; 5) Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Exome arrays are a cost-effective alternative to exome sequencing for interrogating less frequent variants. To assess imputation quality and to determine if genome-wide coverage of the exome array can be improved, we imputed 1,871 European-Americans from the Beaver Dam Eye Study who had been genotyped on the Illumina exome array (95,376 variants) to the Haplotype Reference Consortium reference panel using SHAPEIT2 and Minimac3 on the Michigan Imputation Server. Two common metrics were used to assess imputation quality: empirical R^2 (EmpR^2) and imputation R^2 (ImpR^2). EmpR^2 is the masked squared correlation between imputed dosages and actual discrete allelic dosages obtained from the exome array. ImpR^2 is the squared correlation between imputed genotypes and true, unobserved genotypes. Using an ImpR^2 of 0.6, the imputed dataset included 1,024,985 variants. Imputation quality increased with increasing minor allele frequency (MAF). Common variants (MAF > 0.05) and rare variants (MAF < 0.01) had a mean ImpR^2 of 0.99 and 0.92, respectively. Mean EmpR^2 was lower for common variants (mean EmpR^2 =0.66) compared to rare variants (mean EmpR^2= 0.79). The imputed geographic coverage was 1.8 variants per 5 kb. While coverage was dense in regions originally targeted by the exome array, regions like the short arm of chromosome 22 with no exome array variants were not imputed. Overall, the total number of high quality variants increased 10-fold. These metrics consistently suggest the exome array performed well in spite of its low marker density when imputed to the HRC, but they show opposing trends with MAF. The increasing ImpR^2 with increasing MAF is expected. The high ImpR^2 among rare variants in particular may reflect the strong number of less frequent and rare variants on the exome array. The decreasing EmpR^2 with increasing MAF suggests that reference alleles of rare variants are easier to correctly impute by chance alone, which may indicate a weakness of this metric. Moreover, this metric treats exome array genotypes as a gold standard, which may not be true. While both metrics are informative, the ImpR^2 appears more useful. These results indicate imputation of exome array variants could enhance our ability to detect associations with a larger number of variants, including common variants, in the exome. However, these imputed variants do not provide complete genome-wide coverage and should not be considered as a comprehensive representation of the genome.

2964F

Fast permutation tests and related methods for association between rare variants and binary outcomes. *A. Sondhi, K.M. Rice.* Department of Biostatistics, University of Washington, Seattle, WA.

In large scale genetic association studies, a primary aim is to test for association between genetic variants and a disease outcome. The variants of interest are often rare, and appear with low frequency among subjects. In this situation, statistical tests based on standard asymptotic results do not adequately control the Type I error rate, especially if the case:control ratio is unbalanced. In this paper, we propose the use of permutation and approximate unconditional tests for testing association with rare variants. We use novel analytical calculations to efficiently approximate the true Type I error rate under common study designs, and in numerical studies show that the proposed classes of tests significantly improve upon standard testing methods. We also illustrate our methods in data from a recent case-control study, for genetic causes of a severe side-effect of a common drug treatment.

2965W

Epistasis detection for human complex diseases in structured populations. *K. Van Steen[1,2,3], K. Chaichoompu[1], F. Abegaz[1].* 1) GIGA-R Medical Genomics - BIO3, University of Liege, Liege, Liege, Belgium; 2) WELBIO, University of Liège, Liège, Belgium; 3) Department of Human Genetics, University of Leuven, Leuven, Belgium.

Epistasis, or interaction between genes, has been identified as a component of complex phenotypes in a number of studies. When explaining multifactorial trait variation in humans, gene-gene interactions should not be ignored and potentially complex population-dependent modes of inheritance need to be assumed. Even simple genetic diseases may be complex. For example, Mendelian disorders such as Hirschsprung's disease and cystic fibrosis are documented examples of epistasis where modifier genes have been identified to affect phenotypic differences. In general, epistasis studies help to identify novel drug targets and biomarkers relevant to the underlying mechanisms of disease that are not captured by single locus analysis. Moreover, there have been various studies to find pharmacogenetic evidence of epistatic interactions underlying drug resistance, for example, in malaria[1], epilepsy[32] and influenza[3]. Despite the fact that more and more researchers explore epistasis or genome-wide association interaction (GWAI) studies in an attempt to discover more of the hidden or missing heritability of complex traits, there are still several important challenges and considerations to bear in mind. This study aims to investigate the effect of population substructures and admixture on epistasis detection and to develop and apply remedial measures to confounding by shared genetic ancestry in epistasis analyses. Both real-life data and synthetic data will be used for illustration. Starting point is the versatile epistasis analysis tool Model-Based Multifactor Dimensionality Reduction (MB-MDR)[4]. It is non-parametric in that it does not make any assumptions about the epistasis inheritance model and has several advantages over classic regression-based detection tools. **References** 1. Duraisingh, M.T. & Refour, P. Multiple drug resistance genes in malaria -- from epistasis to epidemiology. *Mol Microbiol* **57**, 874-877 (2005). 2. Kim, M.K.*, et al.* Evidence for epistatic interactions in antiepileptic drug resistance. *J Hum Genet* **56**, 71-76 (2011). 3. Wilson, B.A., Garud, N.R., Feder, A.F., Assaf, Z.J. & Pennings, P.S. The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. *Mol Ecol* **25**, 42-66 (2016). 4. Van Lishout, F., Gadaleta, F., Moore, J.H., Wehenkel, L. & Van Steen, K. gammaMAXT: a fast multiple-testing correction algorithm. *BioData Min* **8**, 36 (2015).

2966T

Leveraging polygenic functional enrichment to improve GWAS power. *G. Kichaev[1], G. Bhatia[2], S. Gazal[2], PR. Loh[2], K. Burch[1], M. Kumar[3], A. Scoech[2], B. Pasaniuc[1,3], AL. Price[2,4].* 1) Inter-departamental Program in Bioinformatics, University of California, Los Angeles, Los Angeles CA, 90024 USA; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health. Boston MA, 02115 USA; 3) Department of Human Genetics, University of California, Los Angeles, Los Angeles CA, 90024 USA; 4) Department of Epidemiology, Harvard T.H. Chan School of Public Health. Boston MA, 02115 USA.

Functional genomics data has the potential to increase GWAS power by identifying SNPs that have a higher prior probability of association. In this work, we introduce a powerful association method that discretizes SNPs based on the extent to which they tag functional annotations that are enriched for heritability. Roughly, we stratify SNPs into bins of expected chi-squared statistic from stratified LD score regression (Finucane et al. 2015 Nat Genet) and then apply p-value weights that correspond to the expected fraction of rejections within a bin. Our method operates directly on summary statistics and outputs well-calibrated p-values -- in contrast to Bayesian approaches (Pickrell 2014 AJHG). We assessed false-positive rate and power using simulations with UK Biobank genotypes (N=113K). We used functional enrichment from the baselineLD model (Gazal et al. biorxiv) to inform the genetic architecture of the simulated traits. We compared our methodology to three FDR-controlling procedures that incorporate prior information: stratified FDR (Sun et al. 2006 Genetic Epi), Grouped Benjamani-Hochberg (Hu et al. 2010 JASA) and Independent Hypothesis Weighting (Ignatiadis et al. Nat Methods 2016). We find that our method correctly controls false positives, whereas the other three methods had false-positive rates at null loci 3-6 times greater than expected. At causal loci, our method attains a 10-30% increase in the number of independent loci, depending on the polygenicity of the simulated trait. We applied our method to 49 diseases and complex traits for which summary statistics were available in large sample size (N=26K-330K). We identified 218 new GWAS risk loci at genome-wide significance (P=5 x 10[-8]). For five of these traits, we were able to obtain independent GWAS summary statistics for replication; the index variants at 55 novel risk loci had a highly significant replication slope (0.58, P = 2x10[-16]), with 49 of 55 loci replicating at P<0.05 . In conclusion, we have shown that leveraging polygenic functional enrichment using our method robustly increases GWAS power. .

2967F

Development of an evidence based sequence variant interpretation tool based upon ACMG and AMP variant interpretation consensus guidelines. *F. Suer[1,2], O. Birsoy[1,2], S. Ghimbovshi[1,2], J. Carr[1,2], J. Liao[1,2], A. Birch[1,2], J. Barrows[1,2], M. Delio[1,2], R. Kueffner[1,2], R. Chen[1,2], R. Kornreich[1,2], L. Edelmann[1,2].* 1) Icahn School of Medicine At Mount Sinai, Department of Genetics and Genomic Sciences, Mount Sinai Genetic Testing Laboratory New York, NY,10029; 2) Sema4, a Mount Sinai venture, Stamford, CT, 06902.

   Background Successful clinical reporting of genetic sequencing results relies on the accuracy and consistency of variant interpretation. Standards and guidelines for the interpretation of sequence variants have been published by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP). Our goal was to improve our variant interpretation process for inherited disorders by applying the published practice guidelines, enriching with gene specific rules and defined laboratory policies to develop an evidence–based semi-quantitative scoring tool. **Results** The scoring tool, developed for clinical use, guides the variant curator to enter applicable evidence for specific variants using a quantitative score for each criterion predefined in the tool. As a result, appropriate contextualized evidence is captured in a more robust and semi-automated fashion to enable the establishment of formalized algorithmic reasoning. Once all the evidence is entered into the tool, a pathogenicity score for the variant is generated. Final review and variant score approval is required. A subset of our variant list database ~1000 variants in 229 genes was classified using this tool. Variant curation was conducted by multiple curators to measure inter-user differences. Variants were curated in parallel with and without the scoring system based on ACMG evidence levels. For the vast majority of variants (~94%), the classification obtained using ACMG guidelines matched the scoring tool results. Approximately 3% of the discordant results were due to applying gene specific rules noted in our laboratory policy. The remaining discordance is within acceptable limits specified in guidelines. **Conclusions** Published guidelines provide a consensus and outline for interpreting sequence variation for clinical sequencing. Our laboratory developed tool creates more consistency between curators by robust implementation of guidelines. It also allows for gene-specific and laboratory specific standard policies to be put in place. Furthermore, this tool facilitates recuration of variants as new evidence accumulates in the literature due to the additive nature of the scoring system.

2968W

POLARIS: Polygenic LD-Adjusted Risk Score approach for analysis of GWAS data. *E. Baker[1], K.M. Schmidt[2], R. Sims[1], P. Holmans[1], M.C. O'Donovan[1], J. Williams[1], V. Escott-Price[1].* 1) Institute of Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK; 2) School of Mathematics, Cardiff University, Cardiff, UK.

   Polygenic risk scores (PRS) provide a method for combining information from individual SNPs into a single measure of risk allele burden. In its most widely used form, PRS have been applied to genome wide SNP data where they can capture a useful fraction of genetic liability to polygenic traits. PRS assume independence between SNPs, and thus linkage disequilibrium (LD) pruning is required to use this approach. We propose the adaptation of PRS with an additional component of adjustment for LD; this informs the analysis with previously reported effect sizes of a SNP's association to disease, and accounts for LD between SNPs. We call this method POLARIS: POlygenic Linkage disequilibrium-Adjusted RIsk Score. POLARIS identifies the LD structure of SNPs in the individual genotype data using spectral decomposition of the SNP correlation matrix and adjusts the effect sizes from the summary statistic data used in the risk score. POLARIS scores are calculated per subject and the overall association to disease is determined using logistic regression on the adjusted polygenic risk score and additional population covariates. LDpred is an alternative method which also adjusts effect sizes for LD structure. LDpred uses a Bayesian approach to estimate the posterior mean effect sizes from the summary statistic data and the LD information of the SNPs. We used simulations to compare the Type I Error of POLARIS and of standard PRS to assess the LD adjustment. The power between POLARIS and PRS is also compared in real data. Since the general concept of β adjustment in LDpred is similar to POLARIS, the β adjustments of POLARIS and LDpred are compared in both simulated and real data. We observed that POLARIS successfully adjusts for LD between SNPs since the inflated Type I Error observed using standard PRS in the presence of LD is reduced. POLARIS also has higher power than standard PRS, particularly when higher (p>0.1) inclusion thresholds are used in marker selection. LDpred, like POLARIS, does not adjust the β values for independent SNPs. In our simulations, LDpred had some sensitivity to the sample size of the data used to estimate LD; larger sample sizes result in far less pronounced β adjustments. In comparison, the β adjustment in POLARIS is largely unaffected by the sample size of the data used to estimate LD.

2969T

Sequential fine-mapping from summary statistics in meta-analyses of genome-wide association studies. *C. Benner[1,2], C. Medina-Gomez[3,4], A.S. Havulinna[1,5], F. Rivadeneira[3,4], M.A. Ikram[4,6,7], T. Lehtimäki[8], O.T. Raitakari[9,10], M.R. Järvelin[11,12,13,14], V. Salomaa[5], S. Ripatti[1,2,15], M. Pirinen[1,2,16].* 1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Department of Public Health, University of Helsinki, Helsinki, Finland; 3) Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands; 4) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 5) National Institute for Health and Welfare, Helsinki, Finland; 6) Department of Radiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 7) Department of Neurology, Erasmus University Medical Center, Rotterdam, The Netherlands; 8) Department of Clinical Chemistry, Fimlab Laboratories and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland; 9) Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 10) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 11) Biocenter Oulu, University of Oulu, Oulu, Finland; 12) Center for Life Course Health Research and Northern Finland Cohort Center, Faculty of Medicine, University of Oulu, Oulu, Finland; 13) Unit of Primary Care, Oulu University Hospital, Oulu, Finland; 14) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; 15) Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 16) Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland.

Recently, genetics research has seen a surge of statistical approaches for fine-mapping of causal variants that work on summary statistics from Genome-Wide Association Studies (GWAS). All these approaches require information about the Linkage Disequilibrium (LD) between the variants. Unfortunately, for fine-mapping from summary statistics of large GWAS meta-analyses, publicly available reference panels are too inaccurate to provide the LD. GWAS consortia are therefore conducting fine-mapping 1) by meta-analyzing conditional analysis results from participating studies requiring multiple rounds of time-consuming coordination and repeated analysis efforts, 2) by using meta-analyzed summary statistics under the assumption of a single causal variant in the genomic region even when multiple causal variants may be present, or 3) by using genotype data from one study to obtain LD, which may be inaccurate. We introduce a new idea of sequential fine-mapping that we have implemented in the FINEMAP software. Sequential fine-mapping is a message-passing procedure where the fine-mapping output from the preceding study serves as an input for the subsequent study. This way sequential fine-mapping 1) circumvents the need of sharing LD, 2) requires only minimal communication between studies and no repeated analysis efforts while it still 3) enables fine-mapping down to any allele frequency by using summary statistics locally together with LD from the original GWAS data. We demonstrate with simulations over 10 GWAS regions and a meta-analysis of body mass index combining the UK biobank with six cohorts from Finland and The Netherlands that 4) sequential fine-mapping is as accurate as fine-mapping from meta-analyzed summary statistics and pooled LD. As an important practical result for ongoing GWAS meta-analyses we show that sequential fine-mapping can be more accurate than fine-mapping from meta-analyzed summary statistics assuming a single causal variant or using one study to obtain LD. In a meta-analysis of the association between the *APOE* locus and LDL-C, sequential fine-mapping identified two well-known missense variants (rs7412 and rs429358), a hepatic control region polymorphism and two novel SNPs. The same 5-SNP configuration was verified by meta-analyzing conditional analysis results whereas fine-mapping the whole locus jointly using LD from the largest study missed rs429358 and the approach of assuming a single causal variant highlighted only rs7412.

2970F

A meta-analysis strategy based on the SPA test to combine multiple PheWAS studies. *R. Dey[1,2], S. Lee[1,2].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

With phenome-wide association studies (PheWASs) gaining increasing attention in the genomics literature for unraveling novel genetic variants associated with clinical phenotypes, methods to combine multiple PheWAS results for improved power is of great interest. Because PheWAS can test 1000s of binary phenotypes, and most of them have unbalanced (case:control = 1:10) or often extremely unbalanced (case:control = 1:600) case-control ratios, very few methods can provide well-calibrated tests for associations. For example, traditional p value-based meta-analysis often results in very conservative or anti-conservative type I error rates in such unbalanced scenarios. We propose a meta-analysis strategy to combine multiple PheWAS results using the saddlepoint approximation-based score test method (SPA). Our method involves sharing an approximation of the distribution of the score test statistic from each study using cubic Hermite splines. It can adjust for covariates and control type I error rates even when the case-control ratios in the individual studies are extremely unbalanced. We compare our proposed method with traditional p value-based meta-analysis strategies, and demonstrate the superior performance of our proposed method using numerical simulations and real data applications.

2971W

Admixture mapping: Controlling for multiple testing and spurious associations in the presence of population structure. *K. Grinde, S.R. Browning.* Department of Biostatistics, University of Washington, Seattle, WA.

Admixture mapping in populations with mixed ancestry is a powerful approach for identifying genetic variants associated with complex traits. Genome-wide admixture mapping scans are being employed more and more frequently in recent years, due in part to technological advances and growing international effort to increase the diversity of genetic studies. However, many open questions remain about appropriate implementation of admixture mapping studies, including how best to control for multiple testing and/or spurious associations in the presence of heterogeneous global ancestry proportions (i.e., population structure). In this study, we extend an existing theoretical framework[1] to characterize the pattern of admixture linkage disequilibrium in admixed populations with arbitrary population structure. This theoretical framework motivates our derivation of admixture mapping significance thresholds that control family-wise error rate in the presence of population structure, as well as our proposal of approaches to control for spurious associations induced by population structure. We validate our proposed significance thresholds and population-structure-adjustment approaches via simulation studies and application to genotype data from 8,421 unrelated African American women from the Women's Health Initiative SNP Health Association Resource (WHI SHARe) study. Our theoretical work and data analyses demonstrate that population structure in admixed populations induces long-range correlation in local ancestry, which can in turn cause spurious associations if admixture mapping models do not appropriately account for population structure. We highlight the critical importance of adjusting for global ancestry in admixture mapping studies to control for spurious associations caused by this long-range admixture linkage disequilibrium. Furthermore, we show that particularly careful consideration is needed in data analyses to appropriately adjust for heterogeneous global ancestry and avoid spurious associations. Our findings hold many practical implications for researchers regarding best practices in admixture mapping studies in admixed populations with population structure.
1. Siegmund, David, and Benjamin Yakir. *The Statistics of Gene Mapping.* Springer, 2007.

2972T

Incorporating multiple functional annotations to infer trait-relevant tissues in genome-wide association studies. *X. Hao[1,2], P. Zeng[1,2], X. Zhou[1,2].* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA.

Genome-wide association studies (GWASs) have identified many disease associated loci, the majority of which have unknown biological functions. Understanding the biological mechanism underlying trait associations requires the identification of trait-relevant tissues, as the effects of genetic variants often act through a tissue-specific fashion. Recent computational and experimental studies have produced a rich category of SNP annotations, with which trait-relevant tissues can be inferred. However, existing methods for trait-relevant tissue inference can only make use of one annotation, and examining one annotation at a time is suboptimal as this approach fails to account for other annotations that can contain complementary information. Here, we develop a simple extension of the commonly used linear mixed model to jointly model multiple annotations with GWAS traits to facilitate the identification of trait-relevant tissues. We also develop an accompanying algorithm based on the widely used generalized estimating equation (GEE) framework for parameter inference. GEE allows us to perform inference using summary statistics and naturally accounts for genetic correlation due to linkage disequilibrium. With simulations, we show that, compared with using a single annotation alone, using multiple annotations jointly can improve power for identifying the correct trait-relevant tissues. Because of a close link between our method and the commonly used SNP set test SKAT, we are also able to exploit parameter estimates from the inferred trait-relevant tissues to construct more powerful SNP set tests in new association studies. We apply our method for an in-depth analysis of 43 traits from 26 GWASs using tissue-specific annotations in 105 tissue types derived from ENCODE and Roadmap. Our results reveal new patterns of relationship among traits in terms of their tissue relevance footprint, pinpoint important annotations that are particularly informative of trait-tissue relationship, and illustrate how parameter estimates from the inferred trait-relevant tissues can lead to more powerful gene set tests in the Wellcome trust case control consortium study.

2973F

Increasing the power of meta-analysis of genome-wide association studies to detect heterogeneous effects. *C.H. Lee[1], E. Eskin[2,3], B. Han[1].* 1) Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Songpa-gu, Seoul 138-736, Korea; 2) Department of Computer Science, University of California, Los Angeles, CA 90095, USA; 3) Department of Human Genetics, University of California, Los Angeles, CA 90095, USA.

Motivation: Meta-analysis is essential to combine the results of genome-wide association studies (GWASs). Recent large-scale meta-analyses have combined studies of different ethnicities, environments, and even studies of different related phenotypes. These differences between studies can manifest as effect size heterogeneity. We previously developed a modified random effects model (RE2) that can achieve higher power to detect heterogeneous effects than the commonly used fixed effects model (FE). However, RE2 has some limitations. First, RE2 cannot perform meta-analysis of correlated statistics, which are found in recent research designs. For example, in cross-disease meta-analyses, it is common that some controls are used in more than one studies, which can cause correlations of statistics. The second limitation of RE2 is that the identified variants often overlap with those found by FE. **Results**: Here, we propose RE2C, which increases the power of RE2 in two ways. First, we generalized the likelihood model to account for correlations of statistics to achieve optimal power, using an optimization technique based on spectral decomposition for efficient parameter estimation. Our optimization technique utilizes the properties of restricted maximum likelihood function (Patterson and Thompson, 1971) to reduce the number of unknown parameters of the likelihood function into one, which resulted in a dramatic reduction in computation time. Second, we designed a novel statistic to focus on the heterogeneous effects that FE cannot detect, thereby, increasing the power to identify new associations. We developed an efficient and accurate p-value approximation procedure using analytical decomposition of the statistic. In simulations, RE2C achieved a dramatic increase in power compared with the decoupling approach (71% vs. 21%) when the statistics were correlated. Even when the statistics are uncorrelated, RE2C achieves a modest increase in power. When we applied RE2C to the cross-disease study data of Parkinson's disease (PD) and Alzheimer's disease (AD) by Moskvina et al. (2013), we observed considerable improvements in association p-values relative to the competing approaches at the pleiotropic loci with heterogeneous effects. RE2C is highly efficient and can meta-analyze one hundred GWASs in an hour. **Availability**: The software is freely available at http://software.buhmhan.com/RE2C.

2974W

Integrative analysis of GWAS summary statistics and imputed gene expression in 44 tissues deciphers genetic architecture for many complex traits. *M. Li, Q. Lu, Y. Hu, H. Zhao.* Yale University, New Haven, CT.

Genome-wide association studies (GWAS) have been successful in identifying DNA variants associated with complex traits. Enrichment of these variants in regulatory regions in the genome suggests the importance of genetically predisposed RNA transcription in disease etiology. In order to understand the architecture of gene expression's mediation effects in complex traits, we proposed **G**ene-based **LD S**core **R**egression (GLDSR) to estimate gene-mediated heritability using GWAS summary statistics and imputed gene expression. The method relies on the idea that linkage disequilibrium (LD) between eQTL and other SNPs contribute to the significance of test statistics in GWAS. Jointly modeling SNP-level test statistics and LD provides an estimate of gene-mediated heritability, while distinguishing it from effects of other biological mechanisms. We first evaluated the performance of GLDSR through simulation, which showed that the method provided unbiased estimation of gene-mediated heritability in a range of settings. We then applied our method to summary statistics of 68 GWAS traits (N = 5.1 million) with gene expression imputed in 44 tissues using GTEx data. Our results demonstrate that GLDSR provides robust estimation even when the sample size of GWAS was relatively small. We found that the overall mean proportion of heritability mediated by gene expression across tissues and traits was 0.075 (sd = 0.006), with 0.178 (sd = 0.013) in the most relevant tissue across traits. By comparing the estimation across 44 tissues, we found evidence suggesting that the expression-mediated SNP effect was cross-tissue in most of the psychiatric traits, but tissue-specific in most of the immune and metabolic traits. We further explored the method on multiple immune diseases including Crohn's disease, eczema, rheumatoid arthritis, systemic lupus erythematosus and ulcerative colitis. Blood tissue has the highest gene-mediated heritability estimation for all the diseases, suggesting our method is capable of identifying the most relevant tissue to disease.

2975T

A hierarchical clustering method for joint analysis of multiple pheno-types. *X. Liang, Q. Sha, Y. Rho, S. Zhang.* Mathematical Sciences, Michigan Technological University, Houghton, MI.

Genome-wide association study (GWAS) has become a very effective research tool to identify genetic variants underlying various complex diseases. In spite of the success of GWAS in identifying thousands of reproducible associations between genetic variants and complex disease, in general, the association between genetic variants and each phenotype is usually weak. It is increasingly recognized that joint analysis of multiple phenotypes can be potentially more powerful than the univariate analysis, and shed new light on underlying biological mechanisms of complex diseases. In this paper, we develop a hierarchical clustering method (HCM) for joint analysis of multiple phenotypes in association studies. The proposed method applies a dimension reduction technique by selecting a representative phenotype for each cluster of phenotypes. Then, we can use existing methods to test the association between genetic variants and the representative phenotypes rather than the individual phenotypes. We perform extensive simulation studies to compare the powers of Multivariate Analysis of Variance (MANOVA), joint model of Multiple Phenotypes (MultiPhen), and Trait-based Association Test that uses Extended Simes procedure (TATES) using HCM with those of without using HCM. Our simulation studies show that using HCM is more powerful than without using HCM in most scenarios.

2976F

Testing for goodness rather than lack of fit of a X-Chromosomal SNP to the Hardy-Weinberg Model. *S. Wellek[1], A. Ziegler[2].* 1) Dep. of Biostatistics, CIMH Mannheim/Univ. of Heidelberg, Mannheim, Germany; 2) Prof. Dr. Andreas Ziegler, Moenring 2, D-23560 Luebeck.

The problem of checking the genotype distribution obtainend for some diallelic marker for compatibility with the Hardy-Weinberg equilibrium (HWE) condition arises also at loci on the X chromosome. In that case the genotypes which can be discriminated in terms of the marker depend on the sex of the individual under study: for females, the data generated through determining the genotype are trinomial as in the case of an autosomal locus, whereas for males a binomial proportion is observed. Like in genetic association studies with autosomal SNPs, interest is typically in establishing approximate compatibility of the observed genotype frequencies with HWE which requires to replace traditional methods tailored for detecting lack of fit to the model with an equivalence testing procedure to be derived by treating approximate compatibility with the model as the alternative hypothesis. The test constructed here is based on an upper confidence bound to an easy to interpret combined measure of distance between true and HWE conforming genotype distributions in female and male subjects. A particular focus is on the asymptotic distribution of the test statistic under null alternatives which is not of the usual Gaussian form. A closed sample size formula is provided and shown to behave very satisfactorily in terms of the approximation error.

2977W

Robust genetic prediction of complex traits with latent Dirichlet process regression models. *X. Zhou, P. Zeng.* Department of Biostatistics, University of Michigan, ANN ARBOR, MI.

There has been a growing interest in using genotype data to perform genetic prediction of complex traits. Accurate genetic prediction can facilitate genomic selection in animal and plant breeding programs, and can aid in the development of personalized medicine in humans. Because most complex traits have a polygenic architecture and are each influenced by many genetic variants with small effects, accurate genetic prediction requires the development of polygenic methods that can model all genetic variants jointly. Many recently developed polygenic methods make parametric modeling assumptions on the effect size distribution and different polygenic methods differ in such effect size assumption. However, depending on how well the effect size distribution assumption matches the unknown truth, existing polygenic methods can perform well for some traits but poorly for others. To enable robust phenotype prediction performance across a range of phenotypes, we develop a novel polygenic model with a flexible assumption on the effect size distribution. We refer to our model as the latent Dirichlet Process Regression (DPR). DPR relies on the Dirichlet process to assign a prior on the effect size distribution itself, is non-parametric in nature, and is capable of inferring the effect size distribution from the data at hand. Because of the flexible modeling assumption, DPR is able to adapt to a broad spectrum of genetic architectures and achieves robust predictive performance for a variety of complex traits. We compare the predictive performance of DPR with several commonly used polygenic methods in simulations. We further illustrate the benefits of DPR by applying it to predict gene expressions using cis-SNPs, to conduct PrediXcan based gene set test, to perform genomic selection of four traits in two species, and to predict eight complex traits in a human cohort.

2978T

Estimating higher-order heritability components in GWAS data from 133,515 individuals. *S.R. McCurdy[1], L.F. Barcellos[1,2], O. Zuk[3].* 1) California Institute for Quantitative Biosciences, UC Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, UC Berkeley, Berkeley, CA; 3) Department of Statistics, Hebrew University of Jerusalem, Jerusalem, Israel.

Most genome-wide association studies (GWAS) to date have focused on estimating the additive contributions of genetic variation to the phenotype of interest. Recent work on estimating the dominance contribution to the heritability of multiple phenotypic traits suggests that dominance frequently contributes little to the genetic variation (Zhu et al., 2015). However, higher-order dominant and epistatic contributions to the total genetic variation remain largely unexamined, in part because estimating them reliably is statistically challenging and requires very large sample sizes. Characterization of higher-order dominant and epistatic effects is important for at least two reasons: first, the characterization is a necessary step towards understanding the total genetic variation of a phenotypic trait, the broad-sense heritability; second, understanding the relative importance of the different orders of complex genetic interactions contributing to a specific trait could guide further study on specific interactions. We test and apply new methods for estimating the heritability of complex traits and diseases for general genetic architectures that include higher-order dominant and epistatic effects. We explore these methods through simulation studies of phenotypes with both dense and sparse genetic architectures, and study the sample sizes and other population parameters required to estimate the higher-order contributions reliably. We also estimate higher-order dominant and epistatic effects for dozens of human complex traits, analyzing separately phenotypes collected by the electronic Medical Records and Genomics (eMERGE) Network and the Resource for Genetic Epidemiology Research on Aging (GERA) Cohort. The eMERGE Network collected data on 41 phenotypes, including case-control disease status and continuous phenotypes such as body mass index (BMI), and single nucleotide polymorphism (SNP) genotypes imputed from the 1000 Genomes Project, for 55,029 participants. The GERA Cohort collected data on 22 case-control phenotypes and continuous phenotypes such as BMI, and collected SNP genotypes on high-density Affymetrix Axiom arrays, for 78,486 participants. These cohorts are ideal for estimating these effects due to their size and diversity.

2979F

Using relationships inferred from electronic health records to conduct genetic studies. *F. Polubriaginof[1], R. Vanguri[1], K. Quinnies[1,2], G. Belbin[3], A. Yahi[1], H. Salmasian[1,4], T. Lorberbaum[1,5], V. Nwankwo[1], L. Li[3], M. Shervey[3], P. Glowe[3], I. Ionita-Laza[6], M. Simmerling[7], G. Hripcsak[1,8], S. Bakken[1,9], D. Goldstein[2], K. Kiryluk[10], E. Kenny[3], J. Dudley[3], D. Vawdrey[1,4], N. Tatonetti[1,2,10,11].* 1) Biomedical Informatics, Columbia University, New York, NY; 2) Institute for Genomic Medicine, Columbia University, New York, NY; 3) Department of Genetics and Genomic Sciences, Mount Sinai Health System, New York, NY; 4) Value Institute, NewYork-Presbyterian Hospital, New York, NY; 5) Department of Physiology and Cellular Biophysics, Columbia University, New York, NY; 6) Mailman School of Public Health, Columbia University, New York, NY; 7) Department of Medicine, Weill Cornell Medicine, Cornell University, New York, NY; 8) Medical Informatics Services, NewYork-Presbyterian Hospital, New York, NY; 9) School of Nursing, Columbia University, New York, NY; 10) Department of Medicine, Columbia University, New York, NY; 11) Department of Systems Biology, Columbia University, New York, NY.

Study of heritability of disease is essential to precision medicine efforts, allowing to determine the fraction of phenotypic variance attributable to genetic factors. However, estimation of disease heritability traditionally requires prospective family studies, with laborious patient recruitment and phenotype ascertainment, limiting sample sizes and therefore, power. Electronic Health Records (EHR) are now broadly implemented, capturing health information of thousands of patients daily, representing a novel resource for studying heritability of traits that are not usually explored. Using emergency contact data provided by millions of patients at three large academic medical centers, Columbia, Cornell and Mount Sinai, we used Relationship Inference From the EHR (RIFTEHR) to infer 7.4 million familial relationships. We validated these relationships using both clinical and genetic data sources. The relationships are consistent with clinical data and genetically-derived relatedness. We then computed disease heritability using only EHR data for 500 distinct clinical traits. To compute heritability using observational data, we need to account for ascertainment biases that vary from family to family and therefore can produce unstable heritability estimates. We developed a procedure called SOLARStrap, using repeated subsampling to produce heritability estimates that are robust to this bias. To validate the accuracy and robustness of this method, we used simulations of quantitative and qualitative traits. We found that both SOLAR and SOLARStrap produce accurate heritability estimates in the presence of random missingness; however, SOLARStrap produces accurate estimates in the presence of ascertainment bias. Overall, heritability estimates are correlated across sites (Columbia r=0.35, p=1.32e-05, Cornell r=0.48, p=8.20e-10 and Mount Sinai r=0.36, p=5.48e-03 with other sites) and consistent with the published literature (r=0.45, p=9.11e-03). Using EHR for heritability studies can increase the sample size of traits previously studied, such as height (h_o^2=0.80, CI=0.74-0.86) but it can also clarify traits that are poorly studied. We computed heritability estimates for 500 traits, only 33 of which had been previously studied as part of the latest meta-analysis or identified by our literature review. We present the first systematic study of disease heritability using EHR data. These analyses provide a novel validation of the use of EHRs for genetics and disease research.

2980W

Detecting heritable phenotypes without a model: Fast permutation testing for heritability and set-tests. *R. Schweiger[1], E. Fisher[2], O. Weissbrod[3], E. Rahmani[1], L. Shenhav[2], M. Müller-Nurasyid[4,5,6], S. Kunze[7,8], C. Gieger[7,8], M. Waldenberger[6,7,8], S. Rosset[2], E. Halperin[9,10].* 1) Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel; 2) School of Mathematical Sciences, Department of Statistics, Tel Aviv University, Tel Aviv, Israel; 3) Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 7610001, Israel; 4) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 5) Department of Medicine I, Ludwig-Maximilians-Universität, Munich, Germany; 6) DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany; 7) Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany; 8) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, 85764 Neuherberg, Germany; 9) Department of Computer Science, University of California, Los Angeles, CA, USA; 10) Department of Anesthesiology and Perioperative Medicine, University of California, Los Angeles, CA, USA.

Testing for an association between a set of genetic markers and a phenotype is a fundamental question in genetic studies. Set-tests attempt to increase power in association studies by aggregating weak individual effects of the tested set of markers. When the entire genome is considered, linear mixed model (LMM) set-tests are equivalent to testing for the heritability of a phenotype, defined as the proportion of phenotypic variance explained by genetics. When evaluating the heritability of a phenotype, it is important to accurately measure the statistical significance of obtaining the estimate under the hypothesis of no true phenotypic heritability. The standard approaches for this task, as implemented in commonly used software such as GCTA, strongly rely on the LMM parametric model, and are not robust to deviations from it: studying the p-values for heritability estimates of methylation profiles of 431,366 CpG sites from a cohort of 1,799 individuals, we show that due to model mis-specification, p-values may be extremely inflated, frequently by many (e.g., 15) orders of magnitude, resulting in more than 45% possible false positives. In contrast, permutation testing is a popular nonparametric alternative, which does not require the assumption of a parametric form of the distribution of the statistic, and that it does not rely on asymptotic assumptions. Indeed, we show that permutation p-values are significantly larger than their parametric counterparts for the studied sites. Thus, we conclude that many non-heritable sites are falsely significantly heritable according to the parametric test. Permutation testing, however, is often computationally prohibitive. Here, we propose an efficient method to perform permutation testing for heritability. Our approach can rapidly identify heritable sites out of millions of sites acquired via high-throughput technologies, does not suffer from model mis-specification and is highly efficient, obtaining speedups of up to seven orders of magnitude.

2981T

Estimating effect-size distributions using summary level statistics from genomewide association studies and projecting trajectories of future discoveries for 32 complex traits. *Y. Zhang[1], G. Qi[1], N. Chatterjee[1,2].* 1) Department of Biostatistics, Johns Hopkins University, Baltimore, MD; 2) Department of Oncology, Johns Hopkins University, Baltimore, MD.

   <u>Background</u> Estimation of heritability based on SNP-arrays has been much focus of research in modern GWAS. While such estimates of heritability provides an understanding of the limit of GWAS to explain trait variation, further understanding of effect-size distribution, i.e. how the heritability is distributed across many GWAS markers, is critical for understanding how fast one can approach the limit as a function of sample size and also for developing more efficient approach for analysis of existing GWAS. <u>Method</u> We describe a novel likelihood based approach for analyzing summary-level data to estimate effect-size distribution of underlying susceptibility SNPs. We show that under a mixture-normal model for the effects of the genetic variants in a multivariate model, the distribution of summary-statistics can be approximated by an alternative mixture normal distributions with variance parameters depending on the sample size, LD-score and the unknown numbers of causal variants with different types of effects the individual GWAS SNPs tag, mixed according to a multinomial probability distribution of the last. The representation allows estimating parameters of the underlying effect-size distribution based on an expectation maximization (EM) algorithm. <u>Results</u> We apply the proposed methods on publicly available summary level results across 19 continuous traits and 13 diseases. Our analyses show that fairly parsimonious models for effect-size distribution, involving four or less parameters, can adequately describe distribution of p-values observed in some of the largest GWAS across almost all traits. The estimated values of the model parameters provide insights into heritability, degree of polygenicity and presence of distinct clusters of effects. Among continuous traits, we predict that, even in optimistic scenario, the sample size needed to identify SNPs that will explain 80% of GWAS heritability could be as large as 300K-500K for some of the early growth traits, between 1-2 millions for some anthropometric and cholesterol traits and into multiple millions for the remaining traits. Among disease traits, the corresponding sample size requirement could be 200K-400K for inflammatory bowel diseases, close to one million for most common adult onset chronic diseases, and between 1-2 millions for the psychiatric diseases. <u>Conclusion</u> We provide most comprehensive analysis of effect-size distribution in GWAS and accordingly provide projections for future discoveries.

2982F

Combining sequence data from multiple studies: Impact of analysis strategies on rare variant association results. *Z. Chen, M. Boehnke, C. Fuchsberger.* Biostatistics, University of Michigan, Ann Arbor, MI.

   Although technological advances in high-throughput DNA sequencing have dramatically increased the abundance of genome and exome sequence data, individual studies often have limited power to detect association with rare and low-frequency variants. A common strategy to increase power for association is to aggregate data or results from multiple studies using joint or meta-analysis. For studying rare variants in sequence data, jointly calling and analyzing all samples together is the ideal "gold standard" strategy but is often difficult to implement due to privacy restrictions and computational burden. Here, we explore alternative strategies by comparing the gold standard against single-cohort variant calling and subsequent meta-analysis of single-cohort callsets in terms of: (1) variant detection sensitivity, (2) genotype accuracy, and (3) association power. We address these issues as a function of sequencing coverage and different software pipelines. To do so, we analyzed deep-coverage (~82X) whole exome (WES) and low-coverage (~5X) whole genome (WGS) sequence data on 2,250 European individuals from the Genetics of Type 2 Diabetes (GoT2D) study applying the broadly used GATK and GotCloud pipelines at default settings. We analyzed all 2,250 individuals jointly and then separately within 5 non-overlapping cohorts of 193 to 890 individuals determined by geographic region. We then carried out gene-based and single-variant association tests of type 2 diabetes (T2D) status for the joint and single-cohort callsets and combined the single-cohort results using fixed-effects meta-analysis. We show for deep-coverage sequencing data: (1) 93% of the detected rare single-nucleotide variants (SNVs) are found in both the joint callset and the union of single-cohort callsets; (2) non-reference concordance with a set of highly accurate genotypes is ≥99% for rare SNVs found in both joint and union callsets; (3) meta-analysis have similar power to joint analysis for gene-based tests (Least-square regression (LSR) slope: 0.938-1.03) but are noticeably less powerful than joint analysis for rare SNVs in single-variant tests (LSR slope: 0.837). We propose single-cohort analysis strategies as a viable alternative to joint analysis strategy for deep-coverage sequencing data with some limitations for gene-based testing: there is no information on whether rare SNVs detected in only one cohort are missing, reference, or filtered out in other cohorts.

2983W

Pervasive pleiotropy in the human genome revealed by a novel quantitative analysis of summary association statistics. *D.M. Jordan, M. Verbanck, R. Do.* Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mt. Sinai, New York, NY.

It is well known that a single genetic variant may have effects on multiple traits, a phenomenon known as pleiotropy. The extent and strength of pleiotropy in the human genome has far-reaching implications for the design of future genetic studies, and for our understanding of the genetic architecture of human traits. Attempts have been made to catalog multi-trait causation for all available human traits and variants, but these efforts fall short of being true systematic measurements of pleiotropy. This is largely due to the use of arbitrary and highly correlated lists of traits which overrepresent well-studied diseases and biological processes. To address this problem, we developed a novel statistical method to transform an arbitrary list of traits into a decorrelated list of independent traits, representing the true set of biological traits underlying the measured traits. Using this method, we developed statistics to quantify both the number of independent traits each variant affects and the total magnitude of effect each variant has. We used summary statistics from published genome-wide association studies of human quantitative traits, diseases, and metabolites, spanning a total of >700 traits in >5,000,000 individuals, to assign a two-dimensional pleiotropy score and an associated P-value to each of >1,000,000 genetic variants. As validation, these scores were shown to be correlated with existing, less widely applicable measures of pleiotropy and functional effect derived from gene expression. Overall, we find that approximately 27% of variants are significantly pleiotropic and 26% have significant overall functional impact, with an overlap of 19% of variants both affecting multiple traits and having a high functional impact. This represents a highly significant excess of pleiotropy, meaning that the average genetic variant affects many more traits than expected by chance. In addition, we identify hundreds of individual loci that exhibit extreme levels of pleiotropy or biological function, passing the threshold for genome-wide significance (P < 5x10^{-8}). These effectively represent genetic associations for pleiotropy and/or overall biological function. A subset of these loci have not been reported for any individual trait, and may represent previously unknown biology. In summary, our results suggest that pleiotropy is pervasive in the human genome, and evaluation of this phenomenon is crucial to understanding the biology of human traits and diseases.

2984T

GWAS genes whose expression is implicated by Mendelian randomization are highly connected in tissue-specific regulatory circuits. *E. Porcu[1,2], D. Marbach[2,3], A. Reymond[1], Z. Kutalik[2,4], eQTLGen Consortium.* 1) Center of Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne Switzerland; 3) Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; 4) Institute of Social and Preventive Medicine, CHUV and University of Lausanne, Switzerland.

Genome-Wide association studies (GWAS) have identified thousands of variants associated with complex traits but both the causative genes and pathways often remain unknown. Most of these variants fall into regulatory regions and overlap with expression Quantitative Trait Loci (eQTLs), indicating their potential involvement in regulation of expression. We propose an advanced summary statistics-based Mendelian Randomization (MR) approach that uses simultaneously multiple SNPs and gene expression traits as instruments and exposures, respectively. Having access to eQTLs from multiple tissues, i.e. a large-scale unpublished peripheral blood study (n=14,117) and the GTEx consortium, enabled us to compare the resulting causal genes for a dozen of complex traits across tissues. We highlight the ambiguity on how different tissues point to different genes; for example our approach, in line with literature evidence, identifies *SORT1* expression to be causal for LDL levels specifically in liver. Furthermore, we found several examples of causally associated genes in regions with no genome-wide significant GWAS SNPs. To translate our findings into deeper understanding of biological processes, we evaluated the over-representation of our causally associated genes in biological pathways. Our findings go far beyond state-of-the-art pathway analyses in several aspects: (i) We find enrichment in trait-relevant pathways that were missed by previous studies (e.g., phospholipid metabolic process (GO:0006644) and lipid biosynthetic process (GO:0008610) for HDL); (ii) Some traits, not enriched for classical pathways, exhibit significant connectivity enrichment for tissue-specific regulatory circuits (e.g.,systolic and pulse pressure- associated genes disturb regulatory modules in endothelial cells); (iii) We find stronger connectivity enrichment when using our causally associated gene-set rather than the genes scored by physical proximity to GWAS hits (e.g., genes that are causally associated with BMI show increased connectivity in regulatory circuits of hormonal glands). Our results demonstrate the importance of integrating expression data from various tissues when trying to interpret GWAS results. With the increasing number of samples and tissues, our MR approach should become a standard tool in the interpretation of GWAS results. It will yield a far deeper insight into biological mechanisms underlying complex trait, as well as new avenues of therapeutic intervention.

2985F

Widespread pleiotropy confounds causal relationships between complex traits and diseases inferred from Mendelian randomization. *M. Verbanck[1], C.Y. Chen[2,3,4], B. Neale[2,3,4], R. Do[1].* 1) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, MA.

Mendelian randomization (MR) is a commonly-used method to infer causality between exposures and complex diseases. A fundamental assumption of MR is that the single nucleotide variant (SNV) used as the instrumental variable must not have pleiotropy. Pleiotropy occurs when the SNV has a direct or indirect effect on the outcome that is not mediated through the exposure. Specifically, pleiotropy can confound MR tests, leading to biased estimates and false positive causal relationships. However, the extent to which pleiotropy alters MR results between complex traits and diseases is currently unknown. We developed a method called the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) to both detect and correct for pleiotropy in multi-instrument summary-level MR testing. MR-PRESSO is a unified framework comprised of three components: 1) global detection of pleiotropy, 2) correction of pleiotropy via outlier removal and 3) testing of the distortion of the causal estimate before and after correction for pleiotropy. Through extensive simulations, we showed that our method can reliably detect and correct for pleiotropy, and thoroughly outperforms existing approaches including MR-Egger which was underpowered and the Q test which showed a severely inflated type-I error. An extensive evaluation of pleiotropy was conducted in 4,250 MR analyses from pairwise comparisons of >80 complex traits and diseases from summary level genome-wide association data. First, using a stringent Bonferroni cut-off ($P<1.17\times10^{-5}$), MR-PRESSO detected pleiotropy in 22 % (N = 922) of the 4,250 tests whereas MR-Egger was severely underpowered having detected none. Moreover, when restricting to significant causal relationships, MR-PRESSO detected pleiotropy at a higher rate of 41% (N = 78) of the total number of tests (N = 191). However, using the outlier detection test, MR-PRESSO was able to correct pleiotropy in approximately 47% (N = 437) of the pleiotropic relationships. Finally, at a nominal however commonly-used significance threshold of 5%, pleiotropy is responsible for up to 10% of false positive causal relationships in MR tests and significantly distorts the causal estimate by a 2.8-fold on average. In conclusion, our results demonstrate that pleiotropy is widespread and pervasive, and must be properly corrected for in order to maintain validity of MR.

2986W

Allele specific information in Mendelian randomization. *X. Wang[1], M. Li[2], N. Zhang[3], D. Small[3].* 1) Applied Mathematics and Computational Science, School of Art and Science, University of Pennsylvania , Philadelphia, PA; 2) Department of Biostatistics and Epidemiology, School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA.

Mendelian randomization (MR) is a statistical framework for quantifying the causal effect of genetic-modifiable exposure to phenotypes, in which genetic variants are used as instrumental variables (IV). The "modifiable exposure" examined by MR studies have traditionally encompassed blood pressure, obesity, smoking, alcohol intake etc. With the rapid accumulation of high throughput sequencing technologies, the exposure of interest can now encompass gene expression through the use of expression quantitative trail loci (eQTLs) as the IV. As exemplified by recent transcriptome-wide association studies (TWAS), MR offers a powerful framework for identifying and quantifying the downstream causal effects of gene expression. Sequencing technologies allow quantification of allele-specific gene expression, where the expression level of both alleles at heterozygous loci are measured for each subject. Current methods for Mendelian randomization do not take advantage of allele-specific measurements. For example, two-stage least squares (TSLS), the classical MR method, uses only total expression. We present a new framework that makes use of allele specific expression to estimate the causal effect of gene expression on quantitative phenotypes. We call this new framework allele specific Mendelian randomization (ASMR). Simulations show that ASMR gives unbiased estimate of causal effect size, and, in many cases, improves estimation precision over TSLS. By utilizing allele-specific information, ASMR reduces estimation variance and improves detection sensitivity in cases where the eQTL strength varies across subjects. We illustrate the new framework on the problem of quantifying downstream effects of lncRNA expression on mRNA expression of protein coding genes on data from the GEUVADIS consortium.

2987T

A highly adaptive microbiome-based survival analysis method. *H. Koh[1], A.E. Livanos[2], M.J. Blaser[3], H. Li[1].* 1) Department of Population Health, New York University School of Medicine, New York, NY; 2) Department of Medicine, Columbia University Medical Center, New York, NY; 3) Department of Medicine, New York University School of Medicine, New York, NY.

There has been increasing interest in discovering microbial taxa that are associated with human health and disease, gathering momentum through the advances in next-generation sequencing technologies. Investigators have also increasingly employed prospective study designs to survey survival (i.e., time-to-event) outcomes, but current item-by-item statistical methods have limitations due to the unknown true nature of the state. Thus, here, we propose a new adaptive microbiome-based association test for survival outcomes, namely, optimal microbiome-based survival analysis (OMiSA). OMiSA approximates to the most powerful association test in two domains: 1) microbiome-based survival analysis using linear and non-linear bases of OTUs (MiSALN) which weighs rare, mid-abundant, and abundant OTUs, respectively, and 2) microbiome-based kernel association test for survival traits (MiRKAT-S) which incorporates different distance metrics (e.g., unique fraction (UniFrac) distance and Bray-Curtis dissimilarity), respectively. We illustrate that OMiSA powerfully discovers microbial taxa whether their underlying associated lineages are rare or abundant and phylogenetically related or not. OMiSA is a semi-parametric method based on a variance-component score test and a re-sampling method; hence, it is free from any distributional assumption on the effect of microbial composition and advantageous to robustly control type I error rates. Our extensive simulations demonstrate the highly robust performance of OMiSA. We also present the use of OMiSA with real data applications.

2988F

Lower frequency of genetic mosaicism observed on the X chromosomes of males relative to the X chromosomes of females. *M.J. Machiela[1], W. Zhou[1,2], N.D. Freedman[1], D. Albanes[1], S.I. Berndt[1], S.M. Gapstur[3], M. Dean[1], M. Yeager[1,2], S.J. Chanock[1] on behalf of the Genetic Mosaicism Consortium.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research Inc., Bethesda, MD; 3) Epidemiology Research Program, American Cancer Society, Atlanta, GA.

Genetic mosaicism is the presence of clonal populations of cells harboring post-zygotic mutations that are not present in germline DNA. Mosaic mutations range in size from single point mutations to large structural events that involve an entire chromosome. Population-based studies have detected age-related genetic mosaicism in blood and buccal derived DNA in apparently healthy individuals. Recent evidence suggests select mosaic mutations may predispose to some forms of cancer and chronic disease. Large structural (> 2 Mb in size) mosaicism of the autosomes affects approximately (1%) of adults. Mosaicism of the sex chromosomes appears more frequently, with mosaic Y loss affecting over 15% of elderly men and mosaicism of the female X chromosome observed at 4 times the base pair adjusted rate of events observed in autosomes. Interestingly, in comparison to the autosomes, mosaicism of the female X chromosome more frequently includes the entire chromosome, with a high proportion of mosaic events involving the inactivated X chromosome. To our knowledge, the frequency of mosaicism on the male X chromosome has yet to be estimated. Herein, we investigate the frequency of genetic mosaicism of the male X chromosome in a population of 42,127 males with high genotyping completion rates (≥ 0.98%). We systematically scanned for deviations in log R ratio (LRR) data extracted from commercially available genotyping arrays of blood or buccal derived DNA. We detected 4 (0.01%) men with evidence for large structural (> 2 Mb) chromosome X mosaicism. Furthermore, 4 (0.01%) additional men were detected with constitutional Klinefelter syndrome (XXY). Of the detected instances of male X mosaicism, all 4 events were mosaic gains and not mosaic losses. We observed 3 of the 4 mosaic gains encompassing the entire X chromosome and 1 mosaic gain affecting only the q arm of chromosome X. The lower frequency of X mosaicism in males (0.01%) relative to females (0.25%) plausibly reflects the importance of some or all of the approximately 800 genes residing on the X chromosome; the majority of which are not directly related to sex determination. Whereas in females the loss or gain of an inactivated copy of the X chromosome may be well tolerated, we hypothesize that the loss or gain of the X chromosome in males is substantially less common because such events are not well tolerated, particularly if the mosaic event involves loss of a substantial proportion of the male X chromosome.

2989W

Identifying the clinical impact of loss-of-function intolerant genes using SKAT-O PheWAS. *R. Sivley[1], W. Bush[2], A. Capra[3,1].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 3) Biological Sciences, Vanderbilt University, Nashville, TN.

Many large-scale whole-exome and whole-genome sequencing projects have recently been compiled by the Exome Aggregation Consortium (ExAC) to produce a large reference database of coding variation in human populations. Analysis of loss-function (LoF) variants in this database revealed 3,230 genes with a significant depletion of LoF variants. Despite this evidence of biological importance and many missense variants in these genes, 72% of t hem have no associated human disease phenotype in ClinVar or OMIM. We hypothesize that low-frequency missense variants that cause less severe perturbations of gene function may be of clinical relevance. Phenome-wide association studies (PheWAS) leverage the wealth of phenotypic information contained in electronic medical records (EMR) to evaluate the clinical impact of genetic variants. While most PheWAS have been conducted using common variants, aggregate association statistics, like the burden test and the sequence-kernel association test (SKAT), can extend this approach to evaluate the collective impact of multiple variants within a genomic region. Using gene-based SKAT-O association tests and exome chip genotyping of 21,957 individuals of European ancestry from Vanderbilt's EMR-linked biobank (BioVU), we apply PheWAS as a tool for discovering the clinical impact of low-frequency missense variation in LoF-intolerant genes.

2990T

PheCLC: A novel statistical method for phenome-wide association studies. *H. Zhu, S. Zhang, Q. Sha.* Michigan Technological University , Houghton, MI.

Over the last decade, genome-wide association studies (GWAS) have been widely performed to identify genetic associations for many complex diseases. Typically, GWAS use a phenotype-to-genotype strategy, starting with a particular phenotype that is associated with genetic variants across the genome, and over 1,000 GWAS have been published linking thousands of statistically significant genetic variants to hundreds of human diseases and traits. A common limitation of GWAS is that they focus on only a single phenotype or a small set of phenotypes at a time. As a complement to GWAS, phenome-wide association studies (PheWAS) use a genotype-to-phenotype approach, beginning with a genotype to test for associations over a broad range of phenotypes. PheWAS were first demonstrated with electronic health record (EHR) data in 2010 and have already demonstrated their capacity to discover genetic association related to a wide range of diseases. In this article, we derived a novel and powerful multivariate method, which we referred as PheCLC, to test the association between a genetic variant with large numbers of phenotypes. Suppose that there is a certain number of phenotypic categories containing different phenotypes. PheCLC first calculates the p-values for testing the variant of interest and the phenotypes within each phenotypic category using a clustering linear combination method recently proposed by our group. Then, it combines the p-values obtained from the first step using the method similar to adaptive Fisher's combination method [Liang *et al*., 2016, Scientific Reports 6]. We perform extensive simulation studies to compare the PheCLC method with other existing methods. The results show that our proposed PheCLC method controls the type I error rates very well and has outstanding performance over other methods.

2991F

Simulation study on different sample sizes for rare-variant association analysis. *X. Zhang[1], A.O. Basile[2], S.A. Pendergrass[3], M.D. Ritchie[3].* 1) The Huck Institute of the Life Sciences, The Pennsylvania State University, University Park, PA; 2) Department of Biochemistry, Microbiology and Molecular Biology, The Pennsylvania State University, University Park, PA; 3) Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, PA.

The development of new sequencing techniques and statistical methods for rare variant analyses provide opportunities for identifying the impact of rare variation on traits and outcomes. However, it is important to characterize the impact of sample size, case numbers, and the balance of the numbers of case vs controls for burden and dispersion based low frequency variant association methods. For example, Phenome-Wide Association Studies (PheWAS) may have a wide range of case and control numbers and sample sizes across hundreds of diagnoses and traits, and with the application of these methods to rare variants it is important to understand the strengths and limitations of the analysis. We investigated the effect of these study design scenarios (sample size, case number, and ratio of cases to controls) on type I error and power for regression and SKAT based approaches. We conducted a large scale simulation of randomly selected low-frequency protein-coding regions through eleven different balanced sample size and twenty-one unbalanced sample size scenarios. We also explored the statistical performance of different minor allele frequency thresholds and effect sizes. All simulated datasets were evaluated using BioBin, a software that performs rare variant analysis burden and dispersion tests. Our simulation results demonstrate that using an unbalanced sample size has an overall higher type I error rate for both burden and dispersion tests compared with balanced samples. Interestingly, the burden (regression) test has overall higher type I error for balanced samples while SKAT has higher type I error for unbalanced samples. For scenarios that include various unbalanced sample sizes with a constant control sample size, when increasing the number of case samples, the type I error decreases and power increases. We also observed that case numbers larger than 200 yield higher power with relatively controlled type I error for both tests. Our results provide the landscape of type I error and statistical power for a wide range of sample sizes, which provide important insights for rare variant association study designs. These results can serve as a benchmark for making decisions about future PheWAS study design for rare variant studies.

2992W

DESCEND: Expression distribution deconvolution in scRNA-seq and characterization of transcriptional bursting and expression dispersion. *J. Wang[1], N. Zhang[1], M. Li[2], A. Raj[3], J. Murray[4].* 1) the Wharton School, University of Pennsylvania, Philadelphia, PA, USA; 2) Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA; 3) Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA; 4) Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA.

Individual cells vary greatly in their gene expression levels, reflecting the dynamism of transcription across cells in a cell population. Single-cell RNA-seq (scRNA-seq) enables the quantification and comparison of much richer properties of gene expression beyond the mean, such as dispersion, burstiness, and other features of the gene expression distribution across cells. However, scRNA-seq data is very noisy, and the observed distribution is usually a poor representation of the true distribution. Here we describe DESCEND (DEconvolution of Single Cell ExpressioN Distribution), a method that deconvolves the true cross-cell gene expression distribution from observed scRNA-seq read counts. DESCEND adopts the "G-modelling" nonparametric empirical Bayes distribution deconvolution framework (Efron 2016). Given cell-specific technical noise model, DESCEND recovers the distribution of the unobserved true expression levels and allows adjustment for covariates such as cell size or cell cycle. The accuracy of DESCEND is evaluated through RNA FISH data generated from the same cell population. It shows that DESCEND can reliably deconvolve the true gene expression distribution, leading to the improved characterization of features such as transcriptional bursting and expression dispersion. Reliable deconvolution of the gene expression distribution allows improved downstream analyses. As a first example, we show through the analysis of the Klein et al. (2015) embryonic stem cell differentiation data that dispersion measurements, such as the GINI coefficient, calculated from the recovered gene expression distribution, can be used to identify early differentiation marker genes that are otherwise missed in differential expression analysis based on the mean expression. We then apply DESCEND to characterize transcriptional bursting among cells of a fixed cell-type in data from Zeisel et al. (2015). Bursting is quantified by two parameters: active fraction (fraction of cells where the gene is expressed) and active intensity (mean expression level when the gene is expressed). We show that transcriptome-wide, in all cell types analyzed, cell size is positively correlated with both active fraction and active intensity, which is confirmed by reanalysis of the RNA FISH data. Finally, DESCEND characterizes in detail changes in active fraction and active intensity between cell types in mouse brain, after controlling for differences in cell size.

2993T

A novel approach for parsing distribution of polygenic risk. *L. Almasy[1,2].* 1) DBHi, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Genetic risk scoring (GRS) relies on utilizing GWAS results to estimate aggregate genetic burden rather than pointing to any particular variant or gene. GRS has been widely applied and studying these aggregate genetic effects has provided insight into the architecture of multiple human phenotypes. However, GRS accounts for only a portion of the known genetic component to various complex disorders. It has been speculated that non-additive genetic components and interactions account for some of this 'missing heritability'. We propose a novel approach for assessing the impact of GRS distribution, independent of burden. We hypothesize that similar burdens of risk alleles may differentially impact liability depending on how they are distributed, with individuals whose burden of 'hits' are concentrated in fewer pathways having higher risk than those of similar GRS whose risk alleles are spread evenly across all pathways. To quantify this clustering within GRS, we construct a score, VARD, that summarizes in a single number the extent to which risk alleles distribute non-randomly across biological pathways. For each pathway, x, we count the number of possible GRS risk alleles in that pathway, POS_x. Risk alleles are defined as alleles with GWAS OR > 1 or by sign of effect for quantitative phenotypes. For each individual, we tally the observed number of risk alleles they carry in pathway x, OBS_x. VARD for each individual is the variance across pathways in the ratio of OBS_x to POS_x. Alternatively, OBS_x and POS_x may be weighted by each SNP's GWAS effect size to reflect differing contributions of each variant to risk. To the extent that the risk alleles carried by an individual are spread evenly across pathways in proportion to the number of GRS SNPs in that pathway, we expect VARD to be small, with little variation in OBS_x/POS_x across pathways. When an individual carries more hits in some pathways than would be expected by chance, VARD is larger. Similar to the way the GRS summarizes aggregate burden of risk, VARD summarizes aggregate distribution of risk, without singling out specific pathways.

2994F

Statistical framework for integrating biological knowledge to accelerate discovery from GWAS data. *S. Bhattacharjee, S. Biswas, S. Pal.* National Institute of Biomedical Genomics, Kalyani, West Bengal, India.

Genome-wide Association Studies (GWAS) have successfully identified thousands of genetic polymorphisms (SNPs) to be associated with various complex diseases. For most traits, the discovered variants tend to map near genes that are clustered in few biological pathways. It therefore appears naïve to perform an unbiased GWA scan without taking pathway/network information into account. Post-hoc enrichment analysis helps in the biological interpretation of already discovered variants; it does not improve the power for novel discovery. It is well-recognized that typical GWAS are under-powered to detect all associated variants due to a huge multiple-testing burden. This is particularly true for variants that are relatively less common and/or those with weaker effects that could have been missed in GWA studies conducted thus far. Pathway/network information, if properly harnessed, has the potential to improve the power of GWAS drastically. We have developed a flexible statistical framework for incorporation of prior knowledge in various forms such as pathways and SNP annotations into GWAS. It enables 'pathway-guided SNP prioritization' thus allowing a genome-wide search to gain substantial power when the associated SNPs cluster in a few biological pathways. The method is data adaptive in the sense that in the worst case, i.e., when the pathway or other prior knowledge is non-informative, the method would behave similar to unbiased search without incurring power loss. Our framework is computationally efficient and allows incorporation of large amount of prior knowledge from various sources. Using a novel whole-genome simulation technique we simulated a large number of causal SNPs with varying degrees of connectivity in pathways. We demonstrate using simulations and real GWAS datasets that the method correctly controls type-1 error at the genome-wide scale, at the same time improving power considerably. We have developed an user-friendly R package named 'GKnowMTest' (Genomic Knowledge-guided Multiple Testing) to be freely available, where p-values of SNPs from a GWAS can be re-weighted using pathways and other annotation databases.

2995W

Improved methods to estimate functional enrichment from genome-wide summary association data. *K. Burch[1], G. Kichaev[1], B. Pasaniuc[1,2,3].* 1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA.

The majority of risk variants identified through GWAS of complex diseases lie in noncoding regions of the genome, rendering biological interpretation of these loci difficult. This has motivated several large-scale efforts to extensively annotate regulatory features in a range of tissue and cell types. Methods that jointly model functional genomic data and GWAS can reveal novel biological insights about the genetic architecture of complex traits; for example, recent work has shown that the heritability of complex diseases is enriched in regulatory features of relevant tissue and cell types. However, existing methods that jointly analyze GWAS and functional annotations either use only genome-wide significant regions, assume one causal SNP per region, do not account for LD, or make prior assumptions about the distribution of SNP effect sizes. In addition, because there are thousands of highly correlated annotations publicly available, it is still unclear how to effectively leverage these data for a particular analysis. Here, we propose a method for estimating functional enrichments from genome-wide GWAS summary data that explicitly models LD and accounts for collinearity in annotations. The novelty of our method lies in the use of the truncated eigendecomposition of the LD matrix to transform the GWAS z-scores such that their underlying LD structure is removed, after which a penalized regression method is used to estimate the effects of the annotations on the de-correlated effect sizes. The resulting functional enrichments can be used to identify the most informative tissues and regulatory features for a given trait, predict the true effect sizes of causal SNPs, increase association power, and improve fine-mapping accuracy.

2996T

HiREPRO: Evaluating Hi-C data REProducibility via Regression. *C. Crowley[1], P. Giusti-Rodríguez[2], P. Sullivan[2,4], M. Hu[3], Y. Li[1,2].* 1) Department of Biostatistics, University of North Carolina Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina Chapel Hill, Chapel Hill, NC; 3) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH; 4) Department of Medical Epidemiology and Biostatistics, Karolinksa Institutet, Stockholm, Sweden.

Hi-C technology is a powerful tool to measure genome-wide chromatin interactions, providing an unbiased view of chromatin spatial originations. Similar to other types of next generation sequencing technologies, such as RNA-Seq and ChIP-Seq, it is critical to rigorously evaluate the reproducibility of Hi-C data among multiple biological replicates. One unique feature of Hi-C data is that the chromatin interaction frequency between two genomic loci shows a dramatic decrease pattern with the increase of the 1D genomic distance between them. Therefore, it is necessary to control 1D genomic distance when evaluating data reproducibility. To achieve this goal, we propose HiREPRO, a zero-inflated negative binomial semi-parametric regression model to quantify Hi-C data reproducibility, while simultaneously adjusting for 1D genomic distance and multiple other confounding factors. HiREPRO provides both point and variance estimates of Hi-C data reproducibility, enabling straightforward practical interpretation. We compared HiREPRO with two other existing methods, HiCRep and HiC-Spector, using multiple deeply sequenced real Hi-C datasets in both human cell lines and primary tissues. We found that HiREPRO not only makes valid assumptions in real Hi-C data, but also provides rigorous statistical inference on Hi-C data reproducibility. More interestingly, the reproducibility estimated by HiREPRO can be used as the discrepancy metric between any two Hi-C datasets, and facilitates the downstream clustering and differential interaction analysis among multiple Hi-C datasets.

2997F

Powerful and robust cross-phenotype association test for case-parent trios. *T. Fischer[1], Y. Jiang[2], K.A. Broadaway[1,3], K.N. Conneely[1,3], M.P. Epstein[1,3].* 1) Dept of Human Genetics, Emory University, Atlanta, GA; 2) Dept of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 3) Center for Computational and Quantitative Genetics, Emory University, Atlanta, GA.

There has been increasing interest in identifying pleiotropic genes within the human genome that influence multiple diverse phenotypes. In the presence of pleiotropy, joint testing of these phenotypes is not only biologically meaningful but also statistically more powerful than univariate analysis of each separate phenotype accounting for multiple testing. While many cross-phenotype association tests exist, the majority of such methods assume samples comprised of unrelated subjects and therefore are not applicable to family-based designs, including the valuable case-parent trio design. In this study, we describe a robust gene-based association test of multiple phenotypes collected in a case-parent trio study. Our method is based on the kernel distance covariance (KDC) method, where we first construct a similarity matrix for multiple phenotypes and a similarity matrix for genetic variants in a gene; we then test the dependency between the two similarity matrices. The method is applicable to either common variants or rare variants in a gene and resulting tests from the method are by design robust to confounding due to population stratification. We evaluated our method through simulation studies and observed that the method is substantially more powerful than standard univariate testing of each separate phenotype. We also applied our method to phenotypic and genotypic data collected in case-parent trios as part of The Genetics of Kidneys in Diabetes (GoKinD) study and identified a genome-wide significant gene demonstrating cross-phenotype effects that was not identified using standard univariate approaches.

2998W

Learning causal networks of molecular phenotypes with Mendelian randomization. *A.Q. Fu[1,2,3], M.B. Badsha[1,2,3].* 1) Department of Statistical Science, University of Idaho, Moscow, ID; 2) Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID; 3) Center for Modeling Complex Interactions, University of Idaho, Moscow, ID.

It has been challenging both experimentally and computationally to elucidate the regulatory (or causal) relationships among genes, to identify the genuine target of an eQTL from multiple associated genes, or to establish causality between different biological processes (such as transcription versus DNA methylation). Correlation (or association) is often used to represent a potentially causal relationship, but similar levels of correlation can arise from different causal mechanisms. Is it possible to go beyond correlation and learn causality directly from genomic data collected in vivo? Is it possible to establish causality for many molecular phenotypes simultaneously? Here we present MRPC, a novel algorithm based on Mendelian Randomization (MR), which uses both genotypes (SNPs, indels, and copy number variation) and molecular phenotypes (e.g. gene expression and DNA methylation), and efficiently learns a causal regulatory network (or a causal graph, which has directed edges); MR views genetic variants in a natural population as perturbations randomly performed by Nature. Our algorithm is a variation of the classical PC algorithm for learning causal graphs, and conducts a series of statistical tests for marginal and conditional independence. However, unlike other PC algorithms, our algorithm incorporates MR and fixes several issues with existing implementations, such as R packages bnlearn and pcalg, both of which have been widely used. Our method further controls the false discovery rate and reduces the impact of outliers. We performed extensive simulations and demonstrated superior performance of MRPC to pcalg and bnlearn. For example, on 1000 data sets simulated for a complex graph of 22 nodes and 23 edges with moderate signal strength and sample size, the mean distance (with standard deviation) between the inferred graph and the truth is only 2.21 (0.65) for MRPC, but 6.80 (2.08) for bnlearn and 8.71 (2.07) for pcalg. We applied MRPC to the SNP and gene expression data in yeast to replicate some of the published gene regulatory networks, to the copy number variation, gene expression and DNA methylation data of the breast cancer patients in TCGA to study the regulatory relationships among cancer genes and between transcription and methylation, as well as to the SNP and gene expression data in GEUVADIS to identify the true target of eQTLs from multiple associated genes. MRPC is implemented in an R package and freely available on GitHub.

2999T

A test for Hardy-Weinberg equilibrium in structured populations. *W. Hao, J.D. Storey.* Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ.

Hardy-Weinberg Equilibrium (HWE) is a general and far-reaching principle in population genetics which often serves as an assumption in statistical applications. In studies with genome-wide genotyping data such as genome-wide association studies (GWAS), it is a common procedure to test markers for HWE as a quality control step with the goal of identifying genotyping errors. A significant test provides evidence that HWE may not hold but does not determine the reason for the violation, which could be the result of effects other than genotyping error such as population structure or selection, so practitioners often employ *ad hoc* procedures to try to ensure that only genotyping errors are removed. We propose a procedure for testing for Hardy-Weinberg Equilibrium that is conditional on population structure called the "structural HWE" (sHWE) test. The sHWE test utilizes a formulation of population structure that provides a flexible approach to parameterizing allele frequencies, which can be fit using many models of structure. sHWE can be applied on a marker-to-marker basis to determine which markers violate HWE while accounting for the confounding effects of population structure. Further, the genome-wide joint distribution of sHWE p-values can be used to assess a global goodness-of-fit of the model of population structure. This allows one to choose optimal values of tuning parameters such as the latent dimensionality of a model. We demonstrate our method on several publicly available global human datasets. We first analyze these datasets independently, showing how the sHWE procedure allows one to choose the dimensionality of the population structure model. Then, we compared SNPs that are misspecified with respect to the population structure model between datasets and technologies, showing that the sHWE procedure identifies SNPs affected by genotyping errors and that results are replicable between datasets.

3000F

PolyGEE: A generalized estimating equation approach to the efficient and robust estimation of polygenic effects in large-scale association studies. *J. Hecker[1], D. Prokopenko[2], C. Lange[1,2], H. Loehlein Fier[1,3].* 1) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, USA; 3) Working Group of Genomic Mathematics, Bonn, Germany.

In order to quantify the polygenic effects, i.e. undetected genetic effects, in large-scale association studies, we propose a novel generalized estimating equation (GEE) based estimation framework. We develop a marginal model for single-variant association test statistics of complex diseases that generalizes existing approaches as LD Score regression and is applicable to population-based designs, to family-based designs or to combinations of both. We utilize and extend the standard GEE approach so that the parameters of the proposed marginal model can be estimated based on working-correlation/linkage-disequilibrium (LD) matrices from external reference panels. Our method achieves substantial efficiency gains over standard approaches while being robust to misspecification of the LD structure, i.e. the LD structure of the reference panel can differ substantially from the LD structure in the study population. In simulation studies and in applications to population-based and family-based studies, we evaluate and illustrate the features of the proposed GEE-framework. Our results suggest that the approach can be up to 100% more efficient than existing methodology. .

3001W

Estimating genetic correlations in functionally annotated regions using genome-wide summary association statistics. *D.L. Kassler[1], A.L. Price[1,2], H. Finucane[1,2].* 1) Epidemiology Dept, Harvard T. H. Chan School of Public Health, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.

Identifying functional annotations of the genome for which genetic correlation between traits is higher may provide more specific etiological insights. We present a new method, stratified cross-trait LD score regression, that combines features of both stratified LD score regression (Finucane et. al. 2015 Nat Genet) and cross-trait LD score regression (Bulik-Sullivan et. al. 2015 Nat Genet) to estimate genetic correlation for annotated regions using only GWAS summary association statistics. We regress the product of z-scores for two traits on annotation-level LD-scores, interpreting the regression coefficients as an annotation-level contribution to genetic covariance. Normalizing the weighted sum of regression coefficients by the root product of the weighted sums of per-annotation heritabilities (which can be computed using stratified LD-score regression), where weights are given by the amount of overlap between annotations, yields estimates of genetic correlation for individual annotations. Standard errors are calculated using a weighted jackknife. Early results from simulated data are promising, demonstrating the efficacy of LD score regression in recovering genetic covariance and of using the resulting regression coefficients to estimate genetic correlation for functional annotations with high heritability. We simulated traits with heritability enrichment in three overlapping annotations of different sizes, with and without enrichment of genetic correlation in two of the annotations. Our method was able to estimate genetic correlation with roughly 90% confidence interval coverage of the true value (for 95% confidence intervals), and an average bias of less than 0.01 in estimates of enriched correlation. Type 1 error was controlled. We present the results of further simulation studies, along with results from applying stratified cross-trait LD score regression to studies of various traits and annotations.

3002T

Estimating the contribution of gene-environment interactions to phenotypic variance. *V. Laville[1], Y.J. Sung[2], M. Province[3], J. Gauderman[4], D. Chasman[5], D.C. Rao[2], H. Aschard[1] on behalf of the CHARGE Gene-Lifestyle Interactions Working Group.* 1) Institut Pasteur, Paris, France; 2) Division of Biostatistics, Washington University School of Medecine, St. Louis, MO, USA; 3) Division of Statistical Genomics, Washington University School of Medecine, St. Louis, MO, USA; 4) Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA; 5) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA.

Several genome-wide Gene-by-Environment (GxE) interaction analyses have been published with the aim of identifying single genetic variants interacting with environmental exposure in human traits and diseases. Conversely, little has been done to estimate the overall contribution of GxE interactions to heritability. Indeed, GxE are often overlooked because of their limited impact on prediction models although previous work showed it may lead to more effective strategies for prevention and treatment. Following the Occam's razor principle, contribution of interaction effects to phenotypic variance is commonly estimated using an orthogonalized model where GxE contribution corresponds to the phenotypic variance explained on top of marginal genetic effects. In this model the genetic additive variance (whose standardized value is the narrow sense heritability) is measured at the mean of E, while GxE contribution to phenotypic variance (h^2_x) captures changes in heritability that would occur with changes in the exposure variance. We propose an alternative strategy which consists of estimating the contribution of GxE to heritability (h^2_{INT}) as the difference between the in-sample genetic additive variance and the expected genetic variance at a baseline value of the exposure. For simplicity, we first focused on quantitative outcome assuming independence between genes and the environment and independence between genetic and interaction effects. We first demonstrate analytically that $h\_h^2_x \leq h^2_{INT}$. Performing series of simulations, we observed that $h\_h^2_x$ is very small in most scenarios even when GxE explain the majority of the phenotypic variance. On the other hand, h^2_{INT} correctly infers the contribution of GxE terms to the outcome variance. To allow the application of our method in large-scale data, we first developed an approach to approximate both parameters (h^2_x and h^2_{INT}) using only genome-wide GxE analysis summary statistics. We also considered the special case of binary exposure where genetic additive variance can be compared between strata. We applied these strategies on real summary data from gene-smoking interactions in blood pressure in ~70,000 individuals from the CHARGE Gene-Lifestyle Interactions Working Group. Preliminary results highlighted differences up to 5.5% in genetic heritability between exposed and unexposed individuals although not significant ($P = 0.19$).

3003F

A generalized permutation testing method for binary trait association in structured samples. *J. Mbatchou[1], M.S. McPeek[1,2].* 1) Department of Statistics, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL.

   In testing for association between a phenotype and a predictor based on genetic or genomic data, permutation testing can be useful when a null distribution for the test statistic is not readily available. For example, such situations arise in the analysis of rare variants using adaptive tests, or in the context of correcting for multiple comparisons in genome-wide association studies. The naïve permutation approach on the phenotype vector relies on the exchangeability of the entries under the null hypothesis, an assumption that is generally not valid in the presence of population structure or related individuals. In the context of quantitative traits, a permutation-based method has been developed that incorporates the presence of important covariates and correlation in the data (Abney et al, 2002). We extend this approach to binary traits, where we account for the binary nature of the trait and the correlation structure present in the data through a quasi-likelihood framework, and generate phenotypic replicates using a transformation of the residuals that results in uncorrelated entries (hence second-order exchangeability). This approach can be used to generate phenotypic replicates in a wide range of applications involving humans or other model organisms. Simulation analyses are performed to evaluate the type 1 error rate and power performance of the proposed approach. We illustrate the application of this method in the assessment of genome-wide significance for association mapping in dogs.

3004W

Integrative analysis of eQTL and GWAS summary statistics to identify functional relationships. *J. Morrison, J. Marcus, N. Knoblauch, X. He, M. Stephens.* Human Genetics, University of Chicago, Chicago, IL.

   Genome-wide association studies (GWAS) have successfully identified thousands of associations between genetic variants and complex traits. However, the functional mechanisms underlying most of these associations remain unclear. One way to help elucidate functional mechanisms is to combine GWAS results with gene expression data. For example, if a disease-associated variant is also associated with expression of a particular gene, then this gene might play a role in disease etiology. However, this pattern could also be caused by a pleiotropic SNP that influences disease risk through another mechanism. Methods have been proposed by Barbeira et al. (2016) and Gusev et al. (2016) that use cis (close to gene) expression quantitative trait loci (eQTLs) to impute the heritable component of gene expression in GWAS studies and estimate the correlation between imputed expression and phenotype. Imputing expression eliminates some environmental sources of noise in measured expression levels, but there are still many non-causal mechanisms that produce the shared genetic effects detected by these methods. For example, disease risk and gene expression might both depend on a third factor but not be causally related. We propose a new statistical approach based on explicit modeling of the bivariate distribution of GWAS and eQTL effects (both cis and trans) with the goal of disentangling causal and non-causal mechanisms. The intuition behind this method is that, if expression of a gene causally affects a phenotype, then every eQTL for that gene should also be associated with phenotype. Further, the eQTL and phenotype effect sizes should be correlated. Non-causal mechanisms leading to correlation result in only a small portion of eQTLs being associated with the phenotype. In simulations, our method correctly controls type 1 error and has dramatically better power than the Sherlock method of He et al. (2013) which uses the same intuition. Importantly, Sherlock and the imputation-based correlation methods can be mislead by the presence of one or two pleiotropic SNPs with strong effects in both the GWAS and the eQTL study, making them prone to false positives. By analyzing patterns across all SNPs, we avoid this undesirable feature. In real GWAS and eQTL data, our method detects known relationships between traits without using the SNPs with the strongest effects, indicating that our method can be valuable even the absence of genome-wide significant associations.

3005T

Modeling ancestry-dependent phenotypic variance increases power in multi-ethnic association studies and enables detection of variance effects. *S. Musharoff[1], D. Park[1], J. Galanter[2], S. Huntsman[1], C. Eng[1], E.G. Burchard[1,3], N. Zaitlen[1].* 1) Lung Biology, University of California, San Francisco, San Francisco, CA; 2) Genentech, South San Francisco, CA; 3) Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

Multi-ethnic genetic analyses are challenging due to differential genetic and phenotypic structures, but also provide unique opportunities for unravelling the architecture of complex disease. Here we consider the possibility that populations with differences in phenotypic mean also have differences in phenotypic variance. We first demonstrate that, in addition to GxG and GxE interactions, variable selective pressures directly induce changes of phenotypic variance, which can produce large differences in disease burden. We next develop an analysis method based on the double generalized linear model (ADGLM), which accounts for relationships between ancestry and phenotypic variance in genetic association studies. In realistic simulations with ancestry-variance relationships we show that linear regression and linear mixed models can give inflated or deflated association p-values. For example, simulated population-differentiated SNPs produce test statistics with a λ_{GC} value of 1.56 when analyzed with linear regression, while ADGLM analyses produce a λ_{GC} value of 1.00. Furthermore, ADGLM has better power than both tests, with increases as large as 66%, and gives unbiased parameter estimates. We applied ADGLM the Study of African Americans, Asthma, Genes and Environments (SAGE). We observed a significant association of European ancestry proportion and baseline lung function (FEV_1) variance (p-value < 1.5e-2), implying differential selective pressure for this phenotype between ancestral European and African populations. Furthermore, ADGLM finds significant associations of FEV_1 with global ancestry, while linear regression fails to detect this known association. We next examined a methylation QTL analysis of Puerto Ricans from the Genetics of Asthma in Latino Americans (GALA) study. We detected associations of ancestry with methylation variance at 44 CpG sites (p-values < 1.6e-7), which may be due to ancestry-associated gene-environment interactions. Consistent with simulations, ADGLM identifies significant associations with methylation at 8 of ~320K sites while linear regression finds only one. In summary, we show in simulated and real data that ancestry proportion can affect phenotypic variance, linear regression and linear mixed models are miscalibrated when ancestry is correlated with variance, and ADGLM is more powerful and correctly calibrated in these situations. We conclude that ADGLM is an essential tool for genetic analyses of admixed populations.

3006F

Y chromosome variants associate with height but not disease risk factors: The Ygen consortium. *N. Pirastu[1], I. Gandin[2], Y. Okada[3,4,5], P.K. Joshi[1], P. Gasparini[9], M. Kubo[6], T. Esko[7,8], J.F. Wilson[1], Y-Gen.* 1) Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Midlothian, United Kingdom; 2) Medical Genetics, Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy; 3) Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan; 5) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan; 6) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 7) Estonian Genome Center, University of Tartu, Tartu, Estonia; 8) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 9) Medical Genetics, Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy; Medical Sciences, Chirurgical and Health Department, University of Trieste, Trieste, Italy.

The Y chromosome is perhaps the least studied part of the genome and displays challenging levels of population genetic structure. Despite its involvement in several rare disorders, the impact of Y chromosome variation on common disease is still mostly unknown. We have conducted the largest association study of Y chromosome genetic variation, analysing 41 quantitative traits in 159 cohorts including 5 ethnicities and comprising up to 412,856 men. 68 markers available across all populations have been used to infer haplogroup identity for each studied individual and thus population frequency. In order to detect association between Y chromosome variation and the traits of biomedical or evolutionary interest, we used three different methods: 1) single point association, 2) summary statistic-derived ANOVA, 3) Y-Wide association. Analyses were performed in each population separately and combined by meta-analysis. Results for the three different approaches agreed across traits. We detected strong association across multiple samples only with height and a weaker one with weight as a consequence. The associations were with a haplogroup which is only common in our data in Japan and which is defined by >200 phylogenetically equivalent markers (in perfect LD), making determination of the causal variant challenging. No other significant effects were detected across traits assessing lung, liver, kidney, heart and cognitive function, glycaemic, fertility, inflammatory, anthropometric and ocular traits. Several putative associations of a specific haplogroup to a specific trait were observed in single populations, but these were not replicated. To further empower analysis of the rarer haplogroups, we conducted similar analyses in UK Biobank, but this confirmed our finding that no associations were present. Our results suggest that common variation on the Y chromosome plays only a very minor role, if any, in complex disease risk. Further more detailed study will be required to understand if the population-specific associations are due to private downstream variants or are false positives. .

3007W

Heritability informed power optimization (HIPO) leads to improved methods of discovering genetic association across multiple traits. *G. Qi[1], N. Chatterjee[1,2].* 1) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

Background Genomewide association studies have shown that pleiotropy is a common phenomenon and can be potentially exploited for enhanced detection of susceptibility loci that are implicated across multiple traits. Existing statistical methods typically carry out pleiotropic analysis at the level of individual loci failing to take into account the degree of genetic correlation that could be informed using genomewide analysis.   Methods We propose heritability informed power optimization (HIPO) for conducting powerful pleiotropic analysis using summary-level association statistics across traits. We utilize LD score regression to estimate genotypic and phenotypic variance-covariance matrices associated with the traits and then utilize them to find optimal linear combinations of association coefficients that are expected to yield the highest non-centrality parameter for the underlying test statistics. These optimal weights can take into account difference in sample sizes across traits and can be computed by eigen decomposition of suitable matrices.   Results Simulation studies show HIPO can maintain right type-I error and can have genomewide average χ^2 statistics much higher than those for individual traits. We applied HIPO to conduct cross-trait association analysis across LDL, HDL, triglycerides and total cholesterol using publicly available data from the Global Lipids Genetics Consortium (N=188,577) and identified 25 novel loci which were not identified by any of the individual trait analysis ($p<5\times10^{-8}$, LD-pruning performed with r^2 threshold 0.1 and new region defined to be at 500 Kb apart from known SNPs). A second application of HIPO for cross-disorder (autism spectrum disorder, ADHD, bipolar disorder, major depressive disorder and schizophrenia) using data from Psychiatric Genomics Consortium (33,332 cases and 27,888 controls in total) identified 5 loci at high-level of significance ($p<5\times10^{-7}$) that were not detectable at the same level of significance by either single trait analysis or cross-disorder meta-analysis. Three of the 5 loci showed evidence of replication in the most recent large study of SCZ.   Conclusions HIPO is a powerful method for association signals across genetically correlated traits. It can be potentially extended to high-dimensional phenotypes as a way of dimension reduction to maximize power for subsequent genetic association testing.

3008T

We didn't see this in GWAS: Understanding and fixing unfamiliar problems in association analyses, when pooling whole genome sequence data from multiple studies. *K. Rice[1], X. Zheng[1], S. Gogarten[1], T. Sofer[1], C.A. Laurie[1], C.C. Laurie[1], B. Weir[1], T. Thornton[1], A. Szpiro[1], J. Bis[2], J. Brody[2].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Cardiovascular Health Research Unit, University of Washington, Seattle, WA.

Large-scale association analyses are now underway, using whole genome sequence (WGS) data on thousands of participants. Unlike earlier GWAS work, where data were combined using meta-analyses of summary statistics, in WGS participant-level data from multiple studies is typically pooled, and results are obtained from a single analysis. While there are good reasons for this approach to WGS, not least computational tractability, we describe how it can lead to false-positive results when the equivalent GWAS-style approach would not. Specifically, we consider the impact of differential phenotype variances by study (due to e.g. different measurement protocols) and its interplay with adjustment for relatedness across studies (e.g. allowing random variability proportional to a kinship matrix, or other genetic relatedness matrix). As well as explaining why these issues lead to difficulties in WGS where they did not for GWAS, we describe methods suitable for WGS work – available in straightforward and freely-available software – that used pooled data and provide appropriate control of false-positive results. For both single-SNP and region-based analyses, the problems and their solutions are illustrated with several examples from the NHLBI Trans-Omics for Precision Medicine Whole Genome Sequencing Program, TOPMed.

3009F

Investigating shrinkage methods to improve accuracy of GWAS and PRS effect size estimates. *Y. Ruan, S. Choi, P. O'Reilly.* SGDP, IoPPN, King's College London, London, United Kingdom.

   As the scale and power of GWAS have increased to detect the small genetic effect sizes involved in complex polygenic traits, the results of GWAS, especially SNP effect size estimates, are increasingly utilised for prediction. A popular approach is the application of polygenic risk scores (PRS). However, the effect size estimates are inherently prone to overfit the specific samples on which the GWAS were performed (*e.g.* "the Winner's Curse"). The inflation of effect size estimation reduces the out-of-sample accuracy of prediction based on GWAS results. Many shrinkage methods have been developed to correct for such inflation. As GWAS results and large individual genotype data sets and become widely available, there is greater opportunity to accurately evaluate effect size inflation and compare the performance of different shrinkage methods. Using large-scale real genotype data, we show that SNP effect sizes were overfitted even in relatively large samples. We develop a novel permutation-based shrinkage method and compare its performance with previously developed methods based on local false discovery rates and the LASSO. We evaluate performance across a wide range of phenotypes and investigate different factors (*e.g.* phenotype heritability) that influence their impact on the predictive power of polygenic risk scores.

3010W

A simple, consistent estimator of heritability from GWAS summary statistics. *A. Schork[1], W. Thompson[1,2], A. Schwartzman[2].* 1) Institute of Biological Psychiatry, Roskilde, Denmark; 2) UC San Diego, La Jolla, CA, USA.

   Estimating the total contribution of common genetic variants to various quantitative and disease traits has become a common and insightful investigation into the genetic architecture of complex phenotypes. The current "gold standard" approach for these analyses, GREML analysis using the GCTA software developed by Yang et al., requires access to actual subject genotypes which is not always feasible. Further, recent reports have suggested this "gold standard" model, which assumes a particular form for the distribution of per allele effects, may perform poorly when this assumed distribution is far from the truth. The LD-score regression has been proposed as an alternative that relaxes the dependence on actual genotypes. It estimates heritability from GWAS summary statistics using a similar, but not equivalent set of assumptions, and is also sensitive to deviant effect distributions. We propose a novel estimator for heritability explained by common SNPs that is derived analytically from first principles of fixed effects multiple linear models. As such, our estimator makes no assumptions about the underlying distribution of per allele effects allowing it to remain unbiased across a broad range of scenarios. Further, our estimator requires only GWAS summary statistics and three simple scalar summaries of the SNP LD matrix. In this paper we introduce this estimator and compare its performance to GREML by GCTA and LDSC regression on a wide range of simulated and real phenotypic data.

3011T

GWAS meta-analysis allowing for sample overlap estimated using summary statistics. *S. Sengupta, G. Abecasis.* Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

Meta-analysis leads to an increase in sample size and power in genome-wide association studies, enabling discovery of novel signals and refinement of previously discovered loci. To facilitate successive rounds of meta-analysis, many studies and consortia now make GWAS summary statistics publicly available. However, when published studies or meta-analysis include overlapping subjects, meta-analysis that do not account for this overlap can lead to inflated Type I error. Here, we propose a method to identify overlap among GWAS using summary statistics, estimate the degree of overlap and finally correctly meta-analyze potentially overlapping studies. Our method builds upon and extends previous methods that allow meta-analysis of GWAS studies with known overlap proportions. We first used actual genotypes from 5,000 GWAS samples to create a series of overlapping studies with simulated phenotypes. We then attempted to estimate sample overlap using GWAS summary statistics and to conduct a meta-analysis that accounted for this estimated overlap. Simulation results show that our method provides an accurate estimate of sample overlap (for example, when GWAS studies overlapped by 18% of samples, we estimate overlapping proportions as 17.9% +- 0.3%). We then meta-analyzed GWAS summary statistics accounting for estimated overlap and observed that the mean genomic control was 1.0 (mean genomic control 1.2 when meta-analyzed without considering overlap). We next created another series of overlapping GWAS studies using actual lipids and type II diabetes (t2d) GWAS data. We were able to estimate the overlap proportions (0.14 estimated as 0.15 for lipids, 0.08 estimated as 0.09 for t2d) and meta-analyze the data, obtaining well calibrated statistics (correlations between Z-scores from analysis of datasets with no overlap and Z-scores from analysis of datasets with duplicated samples analyzed using our method were 95.6-97.5%). Finally, we estimated the overlap in GWAS data for schizophrenia (SCZ) from the Psychiatric Genomics Consortium (PGC). Two phases of SCZ European summary statistics (SCZ1 + Swedish cohort, N=33,100; SCZ2 - SCZ1 + Swedish cohort, N=34,158) are available with a Swedish cohort of known sample size (N=11,244) overlapping and the overlap was estimated as 11,566. In conclusion, we have developed a method that allows us to meta-analyze data using summary statistics where different studies may overlap.

3012F

Statistical and population genetics of extreme phenotypes. *O. Soylemez*[1,2,3], *S. Sunyaev*[1,2,3]. 1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Medical and Population Genetics Program, Broad Institute. Cambridge, MA; 3) Harvard Medical School, Boston, MA.

Extreme values of quantitative traits or rare severe phenotypic presentations can be caused by individual mutations of very large effects segregating in Mendelian fashion. Alternatively, such phenotypes may reflect extremes of polygenic or oligogenic genetic inheritance. Beyond importance for basic genetics, this distinction is practically important for the design of genetic studies that can follow either Mendelian or complex trait paradigm. Currently, sequencing studies of rare phenotypes attribute only a fraction of cases to an obvious monogenic cause. Frequently, individuals with extreme phenotype have family members with milder presentation. Thus, it is expected that family history is informative about genetic architecture of extreme phenotypes. To inform our thinking about this problem, we developed a statistical and population genetics modeling framework to analyze the genetic architecture of extreme phenotypes. We performed forward simulations using SLiM2.0 under various demographic history scenarios to generate large-scale DNA sequence variation with a range of distributions of selective effects. Then, we used a liability threshold disease model to simulate pedigrees of individuals with extreme liability. The simulations explored the impact of various parameters - coupling between selective and phenotypic effect, incomplete penetrance, level of polygenicity, prevalence for rare and extreme phenotypes, and familial clustering of the phenotype – on the genetic architecture of extreme phenotypes. We estimated the relative likelihood that an individual with an extreme phenotype carries a mappable Mendelian mutation of large effect as opposed to the likelihood that the individual's presentation represents the high extreme of the phenotypic distribution characterized by polygenic inheritance. Next, we evaluated whether presence of mildly affected family members carries a substantial amount of information about the genetic model underlying the extreme phenotype. We apply the results of this analysis in the studies of extreme, rare and unique phenotypes. Our results inform large-scale sequencing studies of rare genetic conditions, and help characterize the genetic landscape of apparently Mendelian phenotypes.

3013W

Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. *L. Sun[1,2], B. Chen[1], R.V. Craiu[1].* 1) Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada; 2) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

Due to analytical challenges, the so-called 'whole-genome' association scans routinely omit the X-chromosome, because it differs intrinsically between males and females. Almost a decade ago, Clayton (2008) in his seminal review paper has outlined many related analytical challenges and corresponding remedies, including the assumptions of Hardy-Weinberg equilibrium and equal allele frequencies or sample sizes between female and males. Yet, Wise et al. (2013) noted recently that more than two thirds of published studies do not include X-chromosome, and they made a plea to the research community "toward integrating the X chromosome in genome-wide association analyses". The bottleneck is the lack of sound statistical approaches to address the inherent model ambiguity associated the X-inactivation phenomenon, where one of the two alleles in a female may or may not be silenced. In the absence of biological evidence in favor of one specific model, we consider a Bayesian model averaging (BMA) framework that offers a principled way to account for the inherent model uncertainty. We obtain the posterior distributions generated from Bayesian regression models for analyzing X-chromosome SNPs under the X-inactivation and no X-inactivation assumptions. We combine the estimates from the two models following the BMA principle. We calculate the BMA-based highest posterior density (HPD) region for the parameter of interest. The BMA posterior distribution is directly interpretable as a weighted average for the genetic effect, averaged over the two models, with more weight given to the one with stronger support from the data. To rank multiple SNPs, we calculate Bayes factors comparing the averaged model with the null model of no association for each SNP. We evaluate the performance of the proposed Bayesian approach via extensive simulations and application studies. We show that the proposed Bayesian method provides more feature-rich quantities than frequentist's approaches. We also highlight a few interesting findings, including the importance of using g-priors in this setting and the choice of reference allele when analyzing X-chromosome SNPs.

3014T

Estimating the proportions of additive, dominant and recessive genetic effects. *H. Wu, E. Baker, P.F. O'Reilly.* MRC Centre for SGDP, King's College London, London, United Kingdom.

GWAS have been applied to a huge range of complex traits, but the analyses performed almost always assume additive genetic effects. While these additive models are well-powered for the detection of both additive and dominant effects, they are typically underpowered to detect recessive effects. Therefore, if recessive effects are reasonably common then the power of GWAS may be substantially reduced by testing for additive effects only. This could also lead to reduced power for downstream analyses, such as the application of polygenic risk scores. Here we investigate the statistical power of different inheritance models in GWAS, including the additive, dominant, and recessive models. We analyse the performance of these models via simulation of large-scale genetic data. We also perform the same analysis on real large-scale genetic data across a range of complex traits. Our results suggest that there are a substantial proportion of recessive effects in the genome for the tested complex traits. We conclude that statistical power can be optimised by taking a more nuanced approach to the inheritance model assumed. We estimate the potential increase in susceptibility loci and in the predictive power of polygenic risk score analyses by taking such an approach.

3015F

Scalable Bayesian functional genome-wide association study method with summary statistics. *J. Yang, X. Zhou, G. Abecasis.* Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI.

   Combining standard genome-wide association study (GWAS) association results by single variant tests with the functional annotations of variants can help interpret association results and understand underlying biology. Our previously proposed Bayesian functional GWAS method (bfGWAS) that integrates functional information into GWAS simultaneously evaluate association evidence for genome-wide variants, select likely causal variants to help fine-mapping, and assess the importance of each functional category. Even with parallel computation enabled by bfGWAS, the Markov chain Monte Carlo (MCMC) methods that carry out these computations using individual-level data can require thousands of computer hours for a single GWAS.   To speed up computation, we propose an implementation based on summary statistics for bfGWAS. Since bfGWAS evaluates association evidence based on a joint variable selection linear regression model of genome-wide variants, by using the correlation or linkage disequilibrium (LD) between variants that are jointly evaluated in the model. Thus, we use either an LD matrix estimated from study samples or from a reference sample, such as an appropriate subset of the 1000 Genomes samples. Additionally required summary statistics include the effect sizes by single variant tests, phenotype variance, and sample size. Different from the previously proposed Bayesian Regression with Summary Statistics (RSS) method for GWAS that models the true joint likelihood conditioning on effect sizes and effect-size variances by single variant tests, we derive analytical formulas for conducting the MCMC computations by summary statistics. We applied bfGWAS with summary statistics to the real GWAS data of Age-related Macular Degeneration (~33K samples, ~10M variants). We found that implementing bfGWAS with summary statistics saved up to 90% CPU time while producing the same results as using individual-level data. In addition, we found that using the LD matrix estimated from a reference panel of the same ethnic group as the study samples produced approximately the same results as using the actual LD matrix of the study samples. In conclusion, we successfully implement bfGWAS with summary statistics, which has the same efficiency as using individual-level data, greatly saves computational cost, and makes the meta-analysis of bfGWAS possible by using only summary statistics from individual studies.

3016W

Using imputation for haplotype association. *G.J.M. Zajac, S. Das, E.M. Schmidt, S. Zöllner, G.R. Abecasis.* Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI.

   Haplotype association testing offers many advantages over single-marker tests, especially when there is allelic heterogeneity or epistatic interaction. Powerful and scalable haplotype association tests may reveal new biological insights into complex traits. Here we propose a new GWAS-scale haplotype association test that evaluates whether phenotype similarity is correlated with haplotype similarity as measured by Li and Stephens-based haplotype copying models. We modified our imputation software to calculate a site-specific score for each individual that summarizes affinity between individual haplotypes and all case haplotypes. To search for haplotype association, we search for loci where case haplotypes have systematically higher scores than control haplotypes. To search for marker association, we evaluate the effect of individual genotypes after accounting for this local haplotype similarity score, allowing us to test for the effect of a marker, conditioned on the background haplotype. In both models, we control for population stratification and cryptic relatedness. We tested the method on 97 electronic health record defined cases of primary hypercoagulable state and 583 controls from the Michigan Genomics Initiative, matched on ancestry and age. The samples were genotyped using the HumanCoreExome chip at 600,000 markers. Testing for haplotype association reproduced the GWAS hit in the F5 locus (p-value 1×10^{-3} compared to GWAS p-value 4×10^{-7}). Jointly regressing case-control status on genotype and haplotype score resulted in a stronger signal overall and reduced the signal from many SNPs in high LD with the locus. Our method allows testing for association between a trait and haplotypes in a region rather than the genotype at a specific SNP at large scale. Thus, we expect the method to be useful to find new associations not previously discoverable by GWAS.

3017T

Robustly doubling the sample size: A unifying regression framework for allele-based association test. *L. Zhang[1], L. Sun[1,2].* 1) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada.

Allele-based association test, comparing difference in allele frequency between cases and controls, is locally most powerful. However, the seminal work of Sasieni (1997) highlighted the importance of the Hardy-Weinberg equilibrium (HWE) assumption, thus cemented the current use of genotype-based association test that is robust to HWE. Here, we revisit the problem and aim to develop an alternative allele-based association test that is not only robust to HWE, but also handles quantitative traits as well as various data complications such as genotype uncertainty, correlated individuals from pedigrees, and presence of additional covariates. To this end, we propose a unifying regression framework with *individual allele as the response variable*. We show that the score test statistic derived from this novel regression model contains a correction factor that explicitly adjusts for the departure from HWE. In the simple case of a population-based case-control association study with no genotype uncertainty, we prove that the proposed robust allele-based score test statistic (derived from model regressing individual allele on phenotype) and the standard genotype-based score test statistic (derived from model regressing phenotype on genotype) have *identical form*, despite their apparent differences in the choice of response variable and corresponding sample size. This analytical result has several important implications. It provides theoretical insights on why genotype-based test is robust to HWE, explains why earlier attempts to correcting for the biased allele-based test (e.g. Schaid and Jacobsen 1999) do not lead to improved power, allows decomposition of the individual association and Hardy-Weinberg disequilibrium components in one single regression framework, and leads to the possibility of joint analysis of multiple (continuous and/or binary) phenotypes using allele-based approach, all in the presence of missing data, familial correlation and population heterogeneity. To support our analytical findings, we provide empirical evidence from extensive simulation and application studies.

3018F

On simulation design for evaluating type 1 error: What is the 'correct' null model? *T. Zhang[1], L. Sun[1,2].* 1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto; 2) Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto.

When developing any new statistical test for association between an outcome Y and genotype G of a SNP, the first step is to check its accuracy or type 1 error (T1E) control using simulations. This is often achieved by the following standard simulation design, denoted as S0. S0 simulates phenotype values Y0 under the `conceptual' null model of no association, e.g. $Y0 \sim 1 + e$, then independently generates genotype data G and estimate the empirical T1E from $Y0 \sim G + e$; for notation simplicity and w.l.g. additional covariates are omitted from the regression model. Robustness to assumptions such as normality is also typically evaluated using S0 where e would be generated from other distributions. In practice, whole-genome scans study association between an empirical Y that comes from an (unknown) alternative and a large number of Gs, of which the majority are not associated with Y, i.e. under the `empirical' null . This reality can be mimicked by design S1.1 where one simulates Y from an alternative, e.g. $Y1 \sim G + E + G \times E + e$, then it independently generates G^{new} and evaluates T1E from $Y1 \sim G^{new}$; E was used in data simulation but assumed not being collected in practice. Finally, design S1.2 can be used where one permutes simulated Y1 and evaluates T1E from $Y1^{permute} \sim G$, again under the `empirical' null. It is important to note, however, that even if the error term e is normal, marginally Y1 is not in contrast to Y0. This distinction can be consequential particularly for more advanced association tests. For example, the recent scale-tests, developed to inform presence of interaction effects (e.g. Cao et al. 2014 and Soave et al. 2015), are generally more sensitive to the assumption of normality than the traditional location-tests focusing on genetic mean effects. Importantly, even if the departure from normality is generally minor in practice and appears to pass standard diagnostic tools for non-normality, it can noticeably affect T1E evaluation. For example, when using the likelihood ratio test for variance heterogeneity, while simulation design S0 shows the method being accurate, S1.1 and S1.2 suggest otherwise with empirical T1E values as high as 0.07 for the 0.05 nominal level, and 0.002 for 0.001. This is an important observation that could impact many emerging association tests that are designed to improve power by going beyond the first moment.

3019W

Efficiently controlling for unbalanced case-control sampling and sample relatedness for binary traits in PheWAS by large cohorts. *W. Zhou[1], J.B. Nielsen[2], L.G. Fritsche[3], M.B. Elvestad[4], B. Wolford[1], M. Lin[2], K. Hveem[3,5], H.M. Kang[6], G.R. Abecasis[6], C.J. Willer[1,2,7], S. Lee[6].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Internal Medicine, Division of Cardiology, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 3) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway; 4) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 5) Department of Medicine, Levanger Hospital, Nord-Trøndelag Health Trust, 7600 Levanger, Norway; 6) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, 48109, United States of America; 7) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America.

Decreases in genotyping cost allow for large biobanks to genotype all participants, enabling phenome-wide association studies (PheWAS) on electronic health record-based phenotypes. Both widely used linear mixed models and the recently proposed logistic mixed model (GMMAT) approaches perform poorly -- producing large type 1 error rates -- in the analysis of the resulting case-control studies, where there are typically many controls for each case. Linear mixed model approaches are not designed to handle binary traits and extremely unbalanced case-control ratios invalidate asymptotic assumptions of logistic regression. In addition, since GMMAT requires inverting the kinship matrix, it cannot be used for large cohorts. Here, we describe a novel method to allow for analysis of very large samples (>50,000 individuals), for binary traits with unbalanced case-control ratios, and account for sample relatedness. Our method, called SPA-GMMAT, uses the saddlepoint approximation (SPA) to calibrate the distribution of score test statistics from logistic mixed models. Similar to BOLT-LMM, a linear mixed-model method that handles very large samples, our method utilizes state-of-art optimization strategies for large cohorts. Our method can analyze 67,619 samples and 20.5 million variants in ~1 day using 16 threads and computing time scales linearly with increases in both the numbers of samples and markers. Through simulation studies, we show SPA-GMMAT can control for the type 1 error rates even when the case-control ratios are extremely unbalanced (1:100). In contrast, BOLT-LMM and GMMAT tend to produce inflated test statistics especially for less frequent and rare variants. To illustrate the practical advantages of our approach, we apply SPA-GMMAT and BOLT-LMM to binary traits in the Nord-Trøndelag Health (HUNT) Study with substantial sample relatedness. As an example, we conducted genome-wide analyses of 20.5 million variants (MAF > 0.0001) for venous thromboembolism (case:control=2,325:65,294). Four previously reported GWAS loci passed the genome-wide significant threshold by SPA-GMMAT, while 135 loci were significant by BOLT-LMM. The genomic inflation factor (λ_{GC}) at the 0.1 percentile drops from 1.32 (λ_{GC} of BOLT-LMM) to 1.03 (λ_{GC} of SPA-GMMAT). In summary, we propose a novel method that control for unbalanced case-control ratios and sample relatedness for binary traits. It efficiently handles large sample sizes and can be applied to PheWAS by large biobanks.

3020T

A Bayesian framework for transcriptome-wide association studies. *J.D. Rosen[1], Q. Duan[2], L.M. Raffield[3], A.P. Reiner[4,5], M. Hu[6], Y. Li[1,3,7].* 1) Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA; 2) University of Cincinnati Evaluation Services Center, University of Cincinnati, Cincinnati, OH 45221, USA; 3) Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA; 4) Department of Epidemiology, University of Washington School of Public Health, Seattle, WA; 5) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; 6) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195, USA; 7) Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599, USA.

Recently proposed transcriptome-wide association studies (TWAS) methods have been shown to be promising for revealing genes associated with complex traits. TWAS methods utilize a three step process. Specifically, one first trains a predictive model of gene expression based on an eQTL dataset, and then applies this model to large-scale GWAS dataset(s) for which expression measurements are typically unavailable. Finally these predicted expression values are used to conduct association analysis with phenotypic trait(s) of interest. Basic current TWAS methods use point estimates from the predictive model without considering multiple sources of uncertainty, including variable selection uncertainty, gene expression prediction uncertainty, and genotype measurement uncertainty (for example, due to using imputed genotypes in either the training eQTL or testing GWAS dataset, or both). We propose a novel unified Bayesian framework that addresses all of the above challenges. We adopt a Bayesian variable selection and multiple imputation framework that also takes into consideration genotype measurement uncertainty in both training eQTL and testing GWAS datasets. Our simulation studies demonstrate the advantages of our proposed method over existing ones in terms of both power and validity. More interestingly, our preliminary analysis revealed multiple genes associated with blood cell related traits.