Evidence of epistasis in regions of long-range linkage disequilibrium across five complex diseases in the UK Biobank and eMERGE datasets

Pankhuri Singhal,[1] Yogasudha Veturi,[1] Scott M. Dudek,[1] Anastasia Lucas,[1] Alex Frase,[1] Kristel van Steen,[2] Steven J. Schrodi,[3] David Fasel,[4] Chunhua Weng,[4] Rion Pendergrass,[5] Daniel J. Schaid,[6] Iftikhar J. Kullo,[6] Ozan Dikilitas,[6] Patrick M.A. Sleiman,[7] Hakon Hakonarson,[7] Jason H. Moore,[8] Scott M. Williams,[9] Marylyn D. Ritchie,[1,*] and Shefali S. Verma[10,*]

Summary

Leveraging linkage disequilibrium (LD) patterns as representative of population substructure enables the discovery of additive association signals in genome-wide association studies (GWASs). Standard GWASs are well-powered to interrogate additive models; however, new approaches are required for invesigating other modes of inheritance such as dominance and epistasis. Epistasis, or non-additive interaction between genes, exists across the genome but often goes undetected because of a lack of statistical power. Furthermore, the adoption of LD pruning as customary in standard GWASs excludes detection of sites that are in LD but might underlie the genetic architecture of complex traits. We hypothesize that uncovering long-range interactions between loci with strong LD due to epistatic selection can elucidate genetic mechanisms underlying common diseases. To investigate this hypothesis, we tested for associations between 23 common diseases and 5,625,845 epistatic SNP-SNP pairs (determined by Ohta's D statistics) in long-range LD (>0.25 cM). Across five disease phenotypes, we identified one significant and four near-significant associations that replicated in two large genotype-phenotype datasets (UK Biobank and eMERGE). The genes that were most likely involved in the replicated associations were (1) members of highly conserved gene families with complex roles in multiple pathways, (2) essential genes, and/or (3) genes that were associated in the literature with complex traits that display variable expressivity. These results support the highly pleiotropic and conserved nature of variants in long-range LD under epistatic selection. Our work supports the hypothesis that epistatic interactions regulate diverse clinical mechanisms and might especially be driving factors in conditions with a wide range of phenotypic outcomes.

Introduction

Genome-wide scans are a foundation for understanding complex traits because they elucidate how genomic variation affects phenotypic variation. However, the nature of biological systems suggests that relationships between genotypes and phenotypes are often more complex than can be detected with the methods usually employed.[1] Extant phenotypic variation is a consequence of evolutionary processes and environmental effects, resulting in allele-frequency changes within a population.[2] The explained phenotypic variation is due to a combination of additive and non-additive effects that together define broad-sense heritability.[3] Non-additive effects, including higher-order interactions or epistasis, can be interpreted as dependencies or complex relationships between genes or other sources of genetic variation that influence the presentation of a phenotype. Studies in model organisms demonstrate that epistatic interactions are a key factor driving phenotypic complexity, but the role of epistasis in phenotypic determination in humans remains elusive.[4,5] When studies that statistically test for interaction of genetic variants do identify higher-order interactions, they are often hard to replicate for reasons such as but not limited to model instability, insufficient model complexity as a result of missing variables, limited statistical power in replication datasets, changes in allele frequency, variation in contextual factors, and lack of interpretability of identifiable models.[6] Hence, the role of non-additive effects in the context of disease mechanisms remains a challenge to elucidate.

Using evolutionary processes to model genetics of complex traits can enrich our ability to detect epistasis.[4] Evolutionary processes can produce genomic patterns of variation such as linkage disequilibrium (LD) or non-random associations of alleles.[7] Because the patterns observed are due to past events, leveraging LD patterns to recapitulate evolutionary processes can enhance understanding of biological mechanisms. For example, strong LD ($R^2 > 0.8$) observed in conserved genomic regions across ancestry groups might indicate a functional relationship among variants of fundamental cellular phenomena.[8] Loci can

[1]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [2]Department of Human Genetics, Katholieke Universiteit Leuven, ON4 Herestraat 49, 3000 Leuven, Belgium; [3]Laboratory of Genetics, School of Medicine and Public Health, University of Wisconsin, Madison, WI 53706, USA; [4]Columbia University, New York, NY 10027, USA; [5]Genentech, San Francisco, CA 94080, USA; [6]Mayo Clinic, Rochester, MN 55902, USA; [7]Children's Hospital of Pennsylvania, Philadelphia, PA 19104, USA; [8]Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; [9]Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA; [10]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
*Correspondence: marylyn@pennmedicine.upenn.edu (M.D.R.), shefali.setiaverma@pennmedicine.upenn.edu (S.S.V.)
https://doi.org/10.1016/j.ajhg.2023.03.007.

remain in strong LD for many reasons, including physical proximity and functionality as a "supergene." The mode of inheritance known as a supergene occurs as a result of genomic rearrangement that strives to preserve or lock beneficial alleles across more than one gene.[9] This phenomenon of multiple tightly linked loci regulating a system of discrete phenotypes has been observed across the animal kingdom in functionally related genes that clearly contribute to a shared phenotype.[10,11] Conversely, the phenomenon whereby a single gene influences multiple phenotypic traits is referred to as pleiotropy. Recent work has found pleiotropy to be highly prevalent,[12] if not ubiquitous, in human genotype-phenotype mapping. Pleiotropy and epistasis are inherent properties of biomolecular networks and are critical to understanding the genetics underlying common human disease.[13]

Ohta's D statistics[14] were developed to parse LD in order to determine the contribution of epistasis from population subdivision. By partitioning the LD between a pair of loci into components within and between populations, one can estimate the components attributable to differences in allele frequencies among subpopulations and to epistatic selection. Other theoretical evaluations of LD for structured populations do not take into account population subdivision, and they have been designed to address specific evolutionary circumstances rather than a general model as in Ohta's statistics.[15] SNPs in strong LD are typically pruned out of genomic analyses to reduce the burden of testing "redundant" variants.[16] Genome-wide association studies (GWASs) leverage LD by testing associations between phenotypes and tag SNPs, which function as identifiable proxies for causal SNPs. Experiments in model systems have shown that variants under high selection in evolutionarily conserved regions undergo epistatic selection, possibly as a means of genomic regulation.[17] Although the levels of evolutionary conservation in many non-coding regions are comparable to those of protein-coding regions, non-coding regions have a higher abundance of small-effect-size variants that can have a significant cumulative impact on phenotypes.[18] Other studies hypothesize a role for epistatic selection in these highly conserved regions.The consequences of epistatic selection in these regions might affect structure, function, and the evolution of proteins through physical interactions, as well as changes in long-range regulatory activity by way of three-dimensional chromatin conformation.[19–21]

In humans, the biological mechanisms related to regions under epistatic selection remain unknown, and thus there has been a long-standing debate about the role of epistasis in disease systems. If these disease systems are assumed to be in flux, then evolution can be thought of as continuously "tinkering" or adding variation to solve problems.[22] Phillips et al. describe this as something of a house of cards, in which removing one central component can bring the entire house down, given that one locus may be interacting with many other genes.[23] This intrinsic structural dependency predicated on an iterative accumulation of small changes can be considered the byproduct of 3.5 billion years of descent with modification, as opposed to an intricate piecemeal system.[24] Some alleles have stronger effects on the overall genetic system, and these can be captured as "main effects" in statistical models. In contrast, other alleles contribute more subtle effects through modes of regulation such as transcription, splicing, and epistasis.[23] Univariate and multivariate approaches with interaction effects can be used in tandem to yield greater insights into genetic disease mechanisms than can be obtained with either approach alone.

There has been little study of the effects of epistatic interactions between regions of long-range LD, especially in cases where the LD spans chromosomes. Previous work has found an association between tightly linked SNPs in interchromosomal interactions and aging-related phenotypes such as premature death.[25] Genomic interactions in three-dimensional space are influenced by chromosome topology and transcriptional programs.[26,27] Recent work in mouse and human cell models has shown that certain three-dimensional interchromosomal interactions are a prerequisite for proper physiological gene-expression programs and therefore exhibit conservation.[28]

Thus, the main objective of this study was to uncover associations of disease phenotypes with long-range epistatic interactions between evolutionarily conserved regions in strong LD. On the basis of the pervasiveness of epistasis and what is known about epistatic selection and interchromosomal interactions, we hypothesize that uncovering long-range, high-LD interactions due to epistatic selection can help uncover genetic mechanisms underlying common diseases.[29]

Methods

Selection of epistatic SNP pairs in the UKBB dataset

The UKBB contains genotype data for a total of 488,377 participants and electronic health records (EHRs) for nearly 400,000 participants. At the time of recruitment, participants provide information about their sociodemographic, lifestyle, and health-related factors. Physical measures (such as blood pressure and anthropometry) are also collected from all participants upon recruitment. UKBB genomic data are based on genome build GRCh37 (released in 2009). UKBB genotype data are imputed to the Haplo-type Reference Consortium (HRC) panel.[30]

To identify SNP pairs that were in LD as a result of epistatic selection, we calculated Ohta's D statistics for genome-wide SNP pairs in the European-ancestry population of the 1000 Genomes phase III dataset (n = 503), across subpopulations EAS (East Asian), SAS (South Asian), AFR (African), EUR (European), and American (AMR). As a first step, we calculated pairwise LD across all SNP pairs and selected the pairs with $R^2 > 0.3$ by using PLINK v. 2.0. Next, we determined which SNP pairs were in long-range LD by selecting pairs that are in LD and are located at least 250,000 base pairs (approximately 0.25 cM) apart for intrachromosomal models. We considered all independent SNPs for each chromosome for interchromosomal models. We then filtered the results to include only the SNP pairs with strong LD by setting a threshold

of $R^2 \geq 0.7$ and minor-allele frequency >0.1 for both variants. This yielded 186,119 unique SNPs under epistatic selection; these made up 7,586,336 pairwise SNP-SNP models when we considered long-range interactions and LD. These models included 817,892 interchromosomal models. We then tested the pairwise models for epistatic selection by calculating Ohta's D statistics by using the *ohtadstats* R package,[31] which provided D and D′ statistics (D^2_{IT}, D^2_{IS}, D^2_{ST}, D'^2_{IS}, and D'^2_{ST}) for all models along with the ratios of d2is_mat to d2st_mat (ratio1) and dp2st_mat to dp2is_mat (ratio2). Finally, we selected the models for which both ratio1 and ratio2 were greater than 1, suggesting epistatic selection. This yielded 5,625,845 SNP-SNP models comprising 136,019 unique SNPs. We then tested the association of these models with phenotypes in the UKBB dataset.

Testing of long-range epistatic SNP-SNP models for phenotypic associations

We tested the 5,625,845 SNP-SNP models for associations with 23 complex-disease outcomes with minimum cases of n = 100 in the UKBB European ancestry population of unrelated individuals (n = 384,331). Phenotype definitions for each disease were based on the presence or absence of ICD9/ICD10 codes in EHRs and inclusion criteria outlined by PheCode. We conducted epistasis association tests by using the FastEpistasis module in PLINK v1.9 to identify epistatic SNP-SNP models that were significantly associated with each phenotype. FastEpistasis is a software tool that computes tests of epistasis for a large number of SNP pairs as an efficient parallel extension to the PLINK epistasis module. Epistatic effects are tested by normal linear regression of a binary response on the marginal effects of each SNP and an interaction effect of the SNP pair, where SNPs are coded as additive effects, taking values 0, 1, or 2. The test for epistasis reduces to testing whether the interaction term is significantly different from zero.

Mapping of SNP-SNP models to cytoband regions

For biological interpretability, we mapped all 5,625,845 SNP-SNP models to cytoband regions to produce cytoband-cytoband (cyto-cyto) models. Cytoband annotation was done with the UCSC Genome Browser build 37 SNP-to-cytoband map files. We also annotated the SNPs to genes by using the software tool Biofilter.[32] In addition to mapping the SNPs to the closest upstream or downstream gene, we used the UCSC browser to manually determine which nearby genes were likely to engage in long-range interactions. We then tested the significance of the cyto-cyto models with Bonferroni correction to a threshold of 1.1×10^{-6} on the basis of the total number of unique cyto-cyto mappings (n = 44,860 unique cyto-cyto pairs). The 5,625,845 SNP-SNP models were binned into the unique cyto-cyto pairs, and the model in each bin with the smallest p value meeting the Bonferroni threshold was selected as a significant cyto-cyto pair. A total of 15 cyto-cyto pairs based on the UKBB data were determined to be significant across 8 of 23 phenotypes.

Testing for replication in the eMERGE dataset

The eMERGE is a consortium of 12 United States academic medical centers that have contributed EHR and genotype data of approximately 100,000 individuals into a central repository. The eMERGE phase III genomic data uses genome build GRCh37 (released in 2009). All eMERGE samples have been imputed to the HRC panel via the Michigan imputation server. We extracted phenotype data in the form of diagnosis codes from the EHRs; individuals were defined as cases or controls on the basis of the occurrence or absence of ICD9/10 code(s) grouped as per PheCode criteria. We extracted disease status from the eMERGE data for the eight phenotypes that had significant associations with epistatic SNP pairs in the UKBB data. We tested all SNP-SNP models that mapped to the 15 statistically significant cyto-cyto models for associations with 23 disease phenotypes in the eMERGE dataset of unrelated individuals of European ancestry. We tested all SNP-SNP models that mapped to each significant cyto-cyto model because the causal SNPs are not known. We used a Bonferroni correction based on the 15 tested cyto-cyto models (0.05/15). One of the 15 cyto-cyto models reached statistical significance after adjustment for multiple hypothesis testing with a p value threshold of 3.5×10^{-3}.

Network analysis

To probe molecular mechanisms that might explain the epistatic interactions between the top gene pairs from our analysis, we created biological-process gene networks with HumanBase.[33] We selected the interacting genes, the biological process terms that best described each gene in the pair one at a time, and all five data types to generate molecular networks of interaction. We implemented a 0.1 threshold for minimum interaction confidence and a maximum-number-of-genes threshold of 15 per network. For each gene in the network, a score was generated to indicate the average weight of connections to the epistatic gene pair; we refer to these genes to as "query genes." Focusing on the first- and second-degree neighbors interacting with the query genes, we further evaluated the networks in which the query genes had the highest connection scores. Genes interacting with both query genes are green nodes, and genes interacting with one of the two query genes are yellow nodes. Networks were generated for each top model gene pair. Networks for molecular interactions in relevant biological processes between *FLRT2* × *PDE4D* and *FLRT2* × *RBFOX1* are in Figures S2A–S2D. Relevant biological processes for *TRPS1* or *LRRC4C* were not available in HumanBase, so no networks were generated for this interaction.

Investigation of pleiotropic results

The replicating cyto-cyto models (Table 2) from UKBB and eMERGE were evaluated for pleiotropy. For each of the models, we looked in the rest of the UKBB and eMERGE results to see whether any other phenotypes were also significantly associated in at least one of these datasets. This gave us a set of phenotypes that were different from the ones shown in Table 2 but that were also significantly linked to the same models in either UKBB or eMERGE. Of these phenotypes, we did a deeper literature search to find a clinical link between the model and phenotype in Table 2 and the newly identified phenotype. On the basis of this, we reported one model and one locus of interest with literature support: 19q13.33 and 19p13.2, linked to MS and AD, and 14q31.3, linked to T2D and psoriasis.

Investigation of epistatic interactions between essential genes

To determine whether genes within our near-significant and significant top models were considered "essential" genes, we used the Full Spectrum of Intolerance to Loss-of-Function (FUSIL) database[34] and essential-gene annotations.[35] We identified the functions and associated diseases for each essential gene by conducting a literature search.

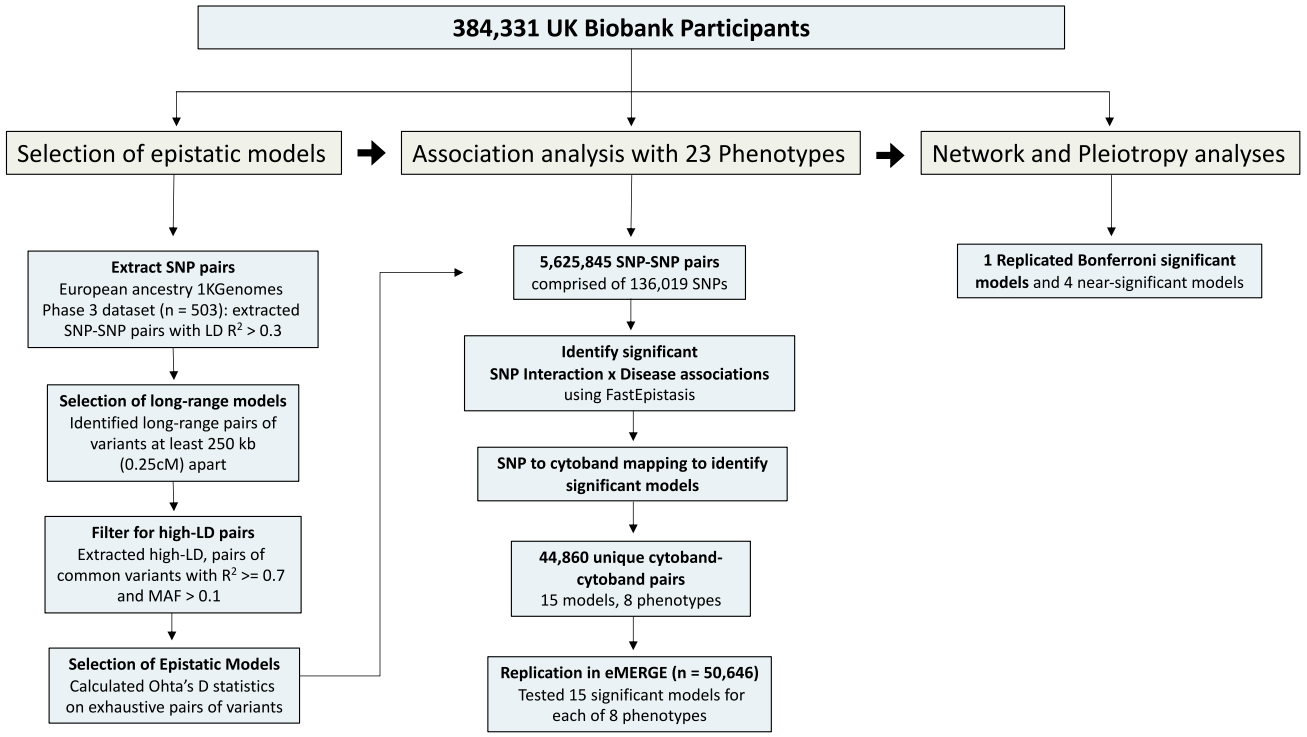

Figure 1. Study workflow
The first part of this study extracted epistatic SNP-SNP interaction models from the 1000 Genomes dataset on the basis of Ohta's D statistics. The second part of the study tested associations of each pair of SNPs across 23 complex phenotypes in the UKBB dataset via FastEpistasis. Significant models were tested for replication in the eMERGE dataset. Replicating models were characterized further in the third part of the study.

Main effects analysis of top models

We tested each SNP in pairs showing significance across the eMERGE and UKBB datasets in univariate tests for association with their corresponding phenotypes. For this, we first identified proxy SNPs that were in LD (R2 > 0.5) with, and located within 1 MB in either direction of, each significant SNP, by using the 1000 Genomes LD panel as a reference. Including sex, age, and the first five principal components as covariates, we then used PLINK v. 1.90Beta4.5 to perform a logistic regression for each SNP. Principle components for both datasets were generated from the individuals in each dataset. A Bonferroni threshold on the number of SNPs (the sum of the significant SNPs and the proxy SNPs) tested per phenotype was calculated based. We identified potential nearby association signals by using LocusZoom plots.[36]

Results

Study overview

A workflow schematic of the study is provided in Figure 1. From a cohort of 384,331 individuals in the UK BioBank (UKBB), 23 different case-control cohorts were created for a range of complex diseases on the basis of the PheCode system.[37,38] We selected phenotypes from diverse disease domains, including cardiovascular, neurological, immune, rheumatic, pulmonary, ocular, gastrointestinal, dermatologic, and neoplastic diseases to investigate the role of epistasis in independent pathologies affecting different tissues. We considered minimum case count when selecting phenotypes. The mean age of the individuals in the UKBB dataset was 57.07 years (standard deviation = 8.07; 55.3% female). Except for breast cancer, all the phenotypes selected for our analysis included both males and females. We tested for associations between the 23 disease phenotypes and 5,625,845 epistatic SNP-SNP models. Models were determined by pairwise tests using Ohta's D statistics (described in detail in methods). Although other methods do leverage biological information to identify epistatic interactions, they lack the evolutionary insights that Ohta's statistics provide on the basis of linkage-disequilibrium architecture. Associations that were significant after Bonferroni correction were tested for replication in the eMERGE consortium dataset (n = 50,646; 50.5% female; mean age 68.15 years, standard deviation = 18.96 years).[39] All individuals in the eMERGE and UKBB cohorts were of European ancestry. Information about the case-control count for each phenotype can be found in Table S1. Further details of the analysis are provided in the methods section.

Association testing of epistatic interactions with complex phenotypes

A total of 5,625,845 epistatic, long-range, high-LD SNP-SNP models was tested for association with each of 23 different phenotypes in the UKBB dataset via the FastEpistasis test, which models marginal and interaction effects of SNPs.[40]

Table 1. SNP-SNP interactions and disease associations identified by FastEpistasis test in the UKBB dataset

Phenotype	Chr A	SNP A	Chr B	SNP B	p value
Acute pulmonary heart disease	12	12:6287428	10	10:93068004	9.60×10^{-6}
AD	6	6:32648500	6	6:32364667	6.47×10^{-7}*
COPD	15	15:81121350	3	3:16237372	2.06×10^{-6}
T2D	16	16:7788521	14	14:86160697	2.29×10^{-7}*
Essential hypertension	12	12:112230036	12	12:111962581	4.89×10^{-6}
Fibromyalgia	20	20:45562611	13	13:39174411	1.79×10^{-6}
Glaucoma	9	9:133944198	7	7:119526478	7.68×10^{-7}*
Hepatic infection	13	13:39174411	8	8:30793179	1.57×10^{-6}
Herpes	1	1:78446761	1	1:78092479	2.42×10^{-7}*
Hyperplasia of prostate	15	15:38271345	4	4:24167431	1.10×10^{-7}*
Hyperplasia of hrostate	17	17:66168247	8	8:63904818	2.46×10^{-7}*
Idiopathic hroctocolitis	3	3:51689306	3	3:51215148	2.14×10^{-6}
Iron deficiency	1	1:201268216	1	1:201002649	4.94×10^{-7}*
Iron deficiency	22	22:45594814	4	4:156510640	5.27×10^{-8}*
Ischemic heart disease	12	12:6287428	10	10:93068004	1.06×10^{-5}
MDD	15	15:58475374	2	2:26628863	5.68×10^{-6}
Cancer of digestive organs	4	4:48783587	4	4:48493237	3.48×10^{-6}
MS	19	19:49194880	19	19:13514610	2.96×10^{-7}*
MS	6	6:32648500	6	6:32364667	4.96×10^{-10}*
Pancreatitis	20	20:45562611	13	13:39174411	1.79×10^{-6}
Psoriasis	12	12:114954298	9	9:130002630	4.53×10^{-7}*
Psoriasis	14	14:86260029	5	5:58499153	2.87×10^{-7}*
Psoriasis	4	4:79179009	3	3:134787497	1.64×10^{-7}*
Schizophrenia	11	11:39275165	8	8:116432183	5.53×10^{-7}*
SCZ	9	9:95674613	5	5:146534330	3.30×10^{-7}*
Tonsilitis	3	3:164060391	3	3:163803419	1.42×10^{-6}
Ulcerative colitis	3	3:51689306	3	3:51215148	2.17×10^{-6}
Viral infection	12	12:105965951	6	6:67707841	2.76×10^{-6}
Breast cancer	4	4:155839500	4	4:150941758	1.02×10^{-5}

Associations between 5,625,845 SNP pairs and 23 phenotypes were tested via FastEpistasis in the UKBB dataset. For Bonferroni correction, SNPs were mapped to cytoband regions, and adjustment was made for the total number of unique cytoband-cytoband pairs ($0.05/44{,}860 = 1.1 \times 10^{-6}$). An asterisk denotes significant p values after Bonferroni correction. Shown here are the top SNP-SNP models of each unique chromosomal pairing across all phenotypes.

FastEpistasis is capable of parallelizing high-dimensional pairwise tests for genome-wide SNPs in the PLINK module, thus reducing computational time substantially. We mapped each SNP to a chromosome cytoband region forming what we refer to as a *cytoband-cytoband* (cyto-cyto) model. Each chromosome arm is divided into regions, or cytogenetic bands (cytobands), that can be seen under a microscope when specific stains are used.[41] The SNPs mapped to cytobands are labeled according to their distance from the centromere on the p or q arm of the chromosome. We evaluated the epistatic associations under the framework of cyto-cyto models because the underlying hypothesis of epistasis was driven by identifying regions of chromosomal interaction rather than specific participating SNPs. Interpretation of specific SNP-SNP interactions is difficult because of local LD substructures; however, these are implicitly considered in SNP-to-cytoband mapping. After applying Bonferroni correction to adjust for the number of unique cyto-cyto models ($0.05/44{,}860 = 1.1 \times 10^{-6}$), we selected the SNP-SNP model most significant for each unique cytoband-cytoband mapping in cases where multiple models reached statistical significance. In the UKBB data, 15 out of 44,860 unique cyto-cyto models spanning 9 of the 23 phenotypes reached statistical significance

Table 2. Cytoband-cytoband models that were significant or near significant in both the UKBB and eMERGE datasets

			UKBB				eMERGE				
Phenotype	**Cyto A**	**Cyto B**	**SNP A**	**Gene A**	**SNP B**	**Gene B**	**p value**	**SNP A**	**Gene A**	**SNP B**	**Gene B**	**p value**
T2D	p13.3	q31.3	16:7788521	*RBFOX1*	14:86160697	*FLRT2*	2.29×10^{-7}	16:6426717	*RBFOX1*	14:86186113	*FLRT2*	0.001344**
MS	q13.33	p13.2	19:49194880	*FUT2*	19:13514610	*CACNA1A*	2.96×10^{-7}	19:49196722	*FUT2*	19:13514610	*CACNA1A*	0.003466*
MS	p21.32	p21.32	6:32648500	*HLA-DQB1*	6:32364667	*BTNL2, HCG23*	4.96×10^{-10}	6:32665629	*MTCO3P1*	6:32397863	*BTNL2*	0.003616*
Psoriasis	q31.3	q11.2	14:86260029	*FLRT2*	5:58499153	*PDE4D*	2.87×10^{-7}	14:85403658	*LOC100421611*	5:58499153	*PDE4D*	0.004205*
SCZ	p12	q23.3	11:39275165	*RPL18P8*	8:116432183	*TRPS1*	5.53×10^{-7}	11:41251442	*LRRC4C*	8:116351905	*TRPS1*	0.007327*

The SNP-SNP model with the lowest p value from each unique cytoband-cytoband bin was tested in the eMERGE dataset if it met Bonferroni significance in the UKBB dataset. A total of 15 models (* in Table 1) were tested for replication. One reached significance in the eMERGE dataset (**) at a significance threshold of p value = 3.3×10^{-3}, and four reached near-significance (*) at a threshold of p value = 0.01.

(Table 1). All UKBB FastEpistasis summary statistics can be found in the supplemental files.

We next tested all SNP-SNP pairs mapping to each of the 15 significant cyto-cyto models for associations with their corresponding phenotypes in the eMERGE dataset. The Bonferroni-corrected threshold for eMERGE was adjusted for the 15 cytoband-cytoband models tested (p value = $0.05/15 = 3.3 \times 10^{-3}$). Only one of the 15 cyto-cyto models reached statistical significance in both UKBB and eMERGE datasets; this model was associated with type 2 diabetes. Several other models met our near-significance p value threshold of 0.01 in eMERGE. Two associated with multiple sclerosis, one associated with psoriasis, and one associated with schizophrenia. The statistically significant and near-significant models are referred to hereafter as "top models" and are shown in Table 2. A graphical representation of the top models is shown in Figure 2, highlighting the intrachromosomal and interchromosomal interactions between genes in the cytoband regions. All eMERGE top-model FastEpistasis summary statistics can be found in supplemental files. Below we discuss these models in detail.

Type 2 diabetes

Chromosome 16:p13.3 × chromosome 14:q31.3 (eMERGE p value = 0.001344) was a statistically significant interaction model that was associated with type 2 diabetes. The interacting SNPs in the UKBB dataset map to the intergenic region 66,427 bp downstream of *FLRT2* (MIM: 604807) and the intergenic region 25,181 bp downstream of *RBFOX1* (MIM: 605104), respectively. *RBFOX1*, one of three mammalian paralogs of the *RBFOX* gene family, regulates tissue-specific alternative splicing and post-transcriptional regulation in the brain and heart.[42] *FLRT2* is part of the FLRT gene family encoding membrane proteins involved in the regulation of cell adhesion and repulsion, cell migration, cell signaling, and axon guidance.[43] *FLRT2* interacts dynamically with various proteins, especially during developmental events.[43]

Psoriasis

FLRT2 (chromosomal region 14:q31.3) was also found to be associated with near significance with psoriasis through an interaction with the intronic region of *PDE4D* (MIM: 600129) on chromosomal region 5q11.2 (chromosome 5:q11.2; eMERGE p value = 0.004205). PDE4D is part of the cyclic nucleotide phosphodiesterase family of enzymes that hydrolyze the intracellular second messenger cAMP, a key signal transduction molecule in numerous biological processes.[44] *RBFOX1* has been associated with neurodegenerative and cardiometabolic traits.[45]

Schizophrenia

An interaction between *TRPS1* (MIM: 604386) on chromosomal region 8q23.3 and *LRRC4C* (MIM: 608817) on region 11p12 was associated with schizophrenia with near significance (eMERGE p value = 0.007327). The SNP on chromosome 8 is 531,181 bp upstream of *TRPS1,* whereas the SNP on chromosome 11 falls in intron 1 of *LRRC4C*. TRPS1 is a transcriptional repressor that binds

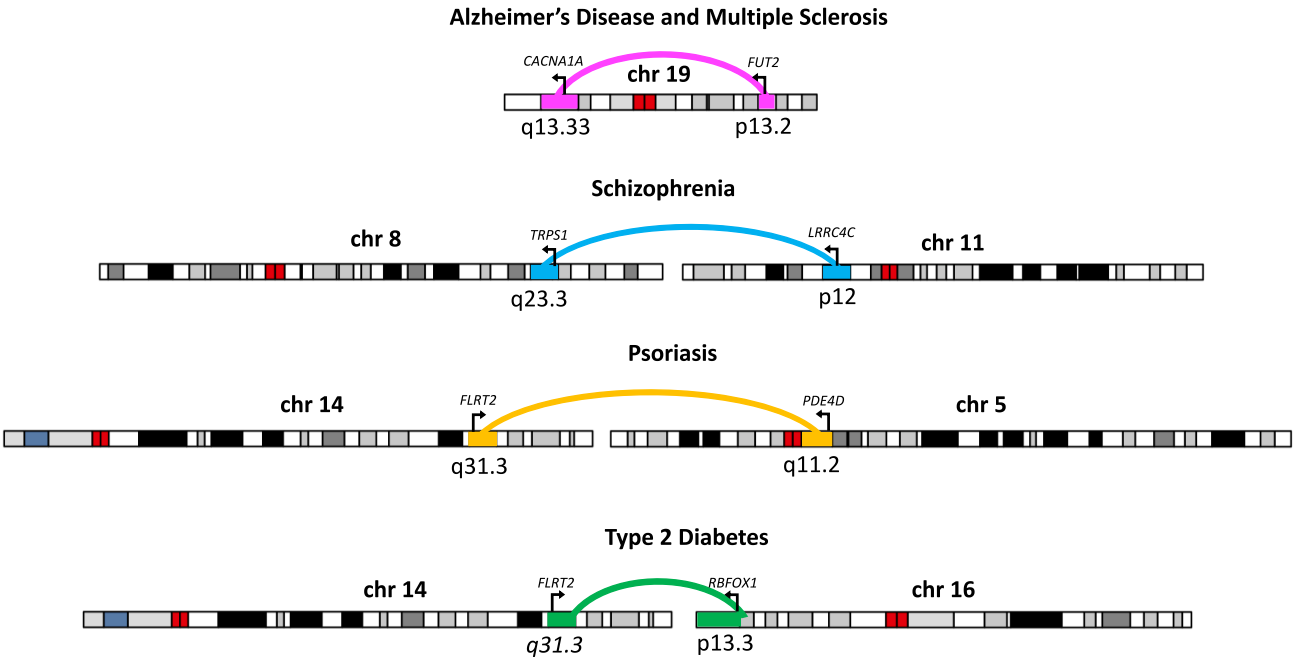

Figure 2. Models that were significant on the basis of the UKBB data and replicated in the eMERGE data
Interchromosomal and intrachromosomal interactions and disease associations of the top models that replicated in both datasets are shown.

to GATA-regulated genes during different stages of embryonic development to influence chondrocyte proliferation and differentiation.[46] *LRRC4C* encodes a post-synaptic adhesion molecule that binds with the conserved family of netrin G ligand (NGL) proteins to regulate synaptic organization.[47]

Multiple sclerosis

A near-significant association of MS with intrachromosomal interaction model for chromosomal region 6:p21.32 (eMERGE p value = 0.003616) mapped to the human leukocyte antigen (*HLA)* region (chromosome 6:p21.3), which is well characterized in the literature for its dynamic role encoding cell-surface proteins responsible for regulation of the immune system.[48] MS, as well as numerous other conditions, including Alzheimer's disease (AD), type 1 diabetes, and rheumatic heart disease, have been linked to the *HLA* region.[49–52] This highlights the strong genetic effect that combinations of alleles in the *HLA* region have on disease susceptibility and protection. These results are promising and could serve to substantiate previous findings; however, a focus on results on the *HLA* region is out of the scope of this study, given the current challenges in sequencing the *HLA* region for interpretation. A second near-significant intrachromosomal interaction model between *CACNA1A* (MIM: 601011) in chromosomal region 19q13.33 and *FUT2* (MIM: 182100) in chromosomal region 19p13.2 was associated with MS (eMERGE p value = 0.003466). *FUT2* determines blood-group secretor status. Being homozygous for the inactive "non-secretor" allele confers susceptibility and resistance to certain infections[53] *FUT2* non-secretor status has been shown to result in significantly increased lymphocyte infiltration levels during infection.[54]

Network analysis to predict molecular mechanisms linking epistatic gene pairs in MS

Three biological-process gene networks for the intrachromosomal interaction between *FUT2* and *CACNA1A* (chromosome 19) are shown in Figure 3. *CACNA1A* is part of a family of genes that provide instructions for making calcium channels, so we generated networks for both "calcium-ion transport" and "calcium-mediated signaling" pathways. *FUT2* and *CACNA1A* both had connection scores of 0.46 in the "calcium-ion transport" network and 0.24 in the "calcium-mediated signaling" network, suggesting that calcium-ion transport better describes the functional context of their interaction. This is further reflected by the high number of high-confidence interactions between primary neighbor nodes connected to both genes in the "calcium-ion transport" network, in contrast to the lower number and lower confidence of interactions between primary neighbor nodes and query genes in the "calcium-mediated signaling" network. Because *FUT2* is responsible for the composition and functional properties of glycans in bodily secretions, we also generated a "proteoglycan biosynthetic process" network; however, the low confidence scores and the low number of primary neighbor nodes connecting both genes in this network do not seem to support a molecular-interaction hypothesis for *FUT2* and *CACNA1A* in the proteoglycan biosynthetic process as strongly. Networks were generated for top models in Figure S2.

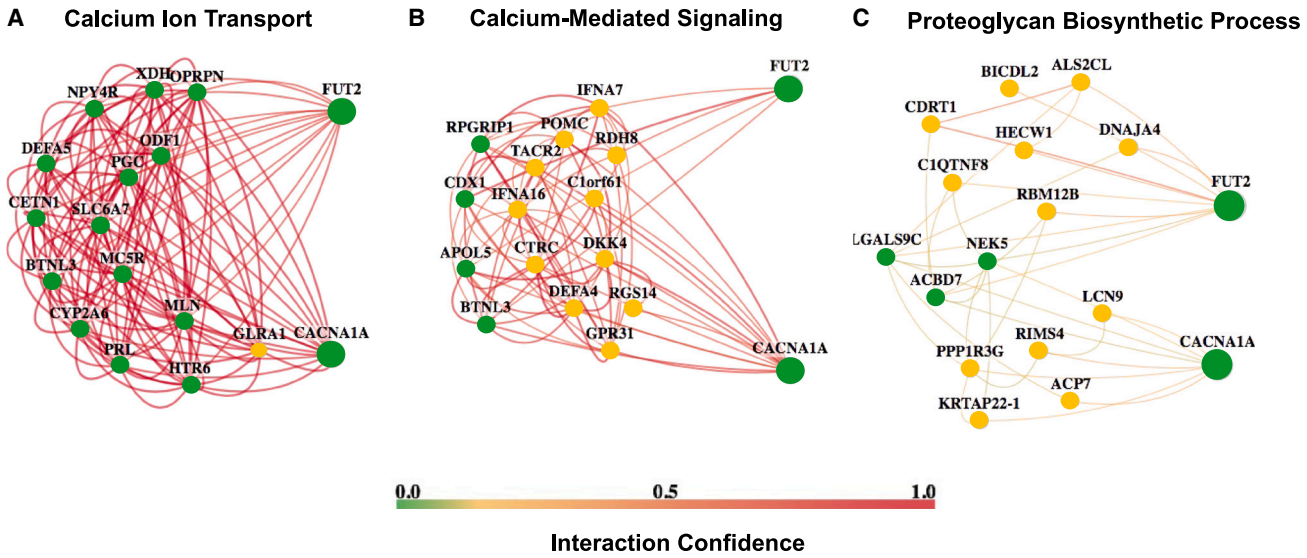

Figure 3. A biological-function network analysis of *FUT2* and *CACNA1A* interaction

Biological-function networks are shown for different selected processes potentially involving the *FUT2* × *CACNA1A* interaction (near-significant association with MS and AD). The interaction confidence scale reflects the strength of the edge weights. *FUT2* and *CACNA1A* are both green nodes, as are all nodes directly connected to both query genes (primary neighbors). Yellow nodes are secondary neighbors, connected to one query gene or the other.

(A) The calcium-ion transport process has many first-degree nodes common to *FUT2* and *CACNA1A*, as well as high-confidence edges, suggesting a strong fit as a process common to both genes.

(B) The calcium-mediated signaling network has more secondary neighbors, suggesting a less strong fit.

(C) The proteoglycan biosynthetic process network has few primary neighbors to both genes, and the edges have low interaction confidence, suggesting this is less likely to be a biological process linking *FUT2* and *CACNA1A*.

Evaluating models of pleiotropy

We identified two epistatic models that were associated with more than one disease and thus exhibited a potentially pleiotropic effect, as discuss below.

Chromosome 19 in Alzheimer's disease and multiple sclerosis

The chromosome 19 interaction between *CACNA1A* and *FUT2* is also a pleiotropic model, associated with Alzheimer's disease in addition to MS. In the UKBB dataset, the association of this model with AD was near significant, with p value = 1.79×10^{-6} (see supplemental information). In eMERGE, the model was statistically significantly associated with AD (p value = 1.27×10^{-4} (see supplemental information). *CACNA1A* is one of the strongest known genetic risk factors for a variety of neurodevelopmental and neurodegenerative conditions, including familial AD, trinucleotide repeat conditions, and SCZ; however, its role in the pathogenesis of these conditions is largely unknown.[55,56] The SNP in chromosomal region 19q13.33 falls in an intergenic region between *FUT2* and pseudogene *SEC1P*.

Chromosome 14q31.3 in type 2 diabetes and psoriasis

Our findings implicate *FLRT2* in two interchromosomal interactions, a psoriasis-associated interaction with *PDE4D* (chromosome 5) and a T2D-associated interaction with *RBFOX1* (chromosome 16). Although the shared etiology of diabetes and psoriasis is unknown, a breadth of work has shown overlap between T2D and psoriasis by way of shared lipid abnormalities, heightened insulin resistance, and cardiovascular risk biomarkers.[57] To evaluate whether this region is predisposed to pleiotropic activity, we identified five Mendelian traits and syndromes (Figure 4) linked to genes within 14q31.3 on the basis of the Online Mendelian Inheritance in Man (OMIM) database.[58] Krabbe disease was (KRB [MIM: 245200]) linked to *GALC* (MIM: 606890); Leber congential amaurosis 3 and retinitis pigmentosa 94 (LCA3 [MIM: 604232]) were linked to *SPATA7* (MIM: 609868); intellectual development disorder (MRT56 [MIM: 617125]) was linked to *ZC3H14*, and Bardet-Biedl Syndrome 8 (BBS8 [MIM: 615985]) and retinitis pigmentosa 51 (MIM: 613464) were linked to *TTC8* (MIM: 6081320.

Univariate analysis to test whether interacting SNPs function as main effects

Individual SNPs in each significant cyto-cyto model were tested in a univariate regression framework to determine whether they function as main effects for their respective phenotype(s). In addition to the SNPs used for determining the p values for the cyto-cyto models, we tested any proxy SNPs that were also in LD ($R^2 \geq 0.5$) and within 1 MB upstream or downstream of the representative SNPs. The plots in Figures S1A–S1J depict the results of the univariate analyses of the SNPs in each model; the chromosome containing the lowest p values for each model is shown. Bonferroni correction was based on the number of proxy SNPs tested. No statistically significant main effects were found, although the results for chromosome 6 overlapped with GWAS hits for various traits, as expected given the dynamic nature of the HLA region. Summary

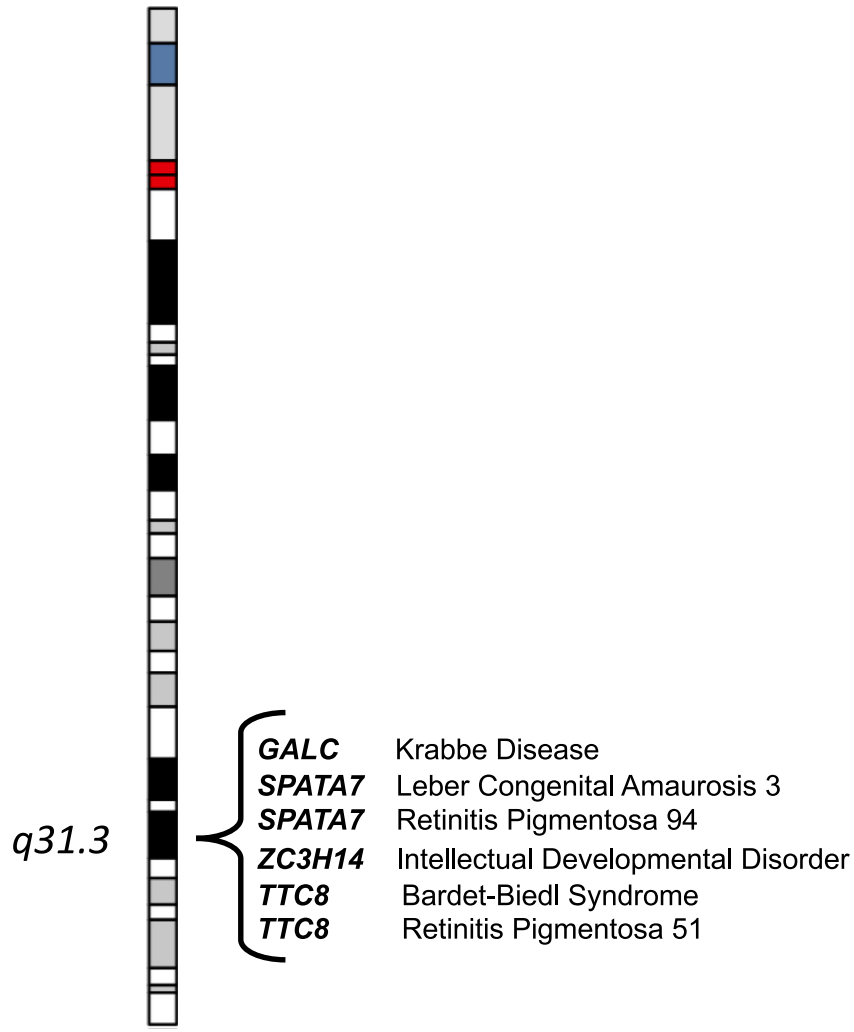

Figure 4. Evidence of pleiotropy on chromosomal region 14q31.3
Chromosomal region 14q31.3 interacts with region 16p13.3 to affect T2D and with region 5q11.2 to affect psoriasis. To determine a clinical basis for the pleiotropy observed on chromosomal region 14q31.3, we used Online Mendelian Inheritance in Man (OMIM) to identify Mendelian conditions known to be associated with this region.

statistics for main-effects analysis can be found in the supplemental information.

Epistasis in essential gene families

Essential genes are critical for the survival of organisms under most conditions and commonly drive cell growth and proliferation. The vast majority of human genes are non-essential but still confer some degree of selective advantage. Out of 19,850 known human genes, 3,915 are considered essential genes.[35] Essential genes are likely to encode hub proteins that are widely expressed in most tissues, making them well-suited as dynamic genes active in the interactome.[59] On the basis of a previous study of essential genes by Ji et al. and the Full Spectrum of Intolerance to Loss-of-Function (FUSIL) database, we determined whether the genes in our top models function as essential genes. Six of the seven genes in our top models, *LRRC4C* being the exception, have been classified as essential genes.[34,35] Figure 5 depicts the essential genes, categorized by their functional designations from FUSIL, as well as their essential functions in growth and development.

Discussion

We hypothesized that long-range epistatic interactions in chromosomal regions with high LD are implicated in mechanisms underlying complex diseases. To test this, we looked for associations between 23 complex disease phenotypes and 5,625,845 epistatic pairs of SNPs with strong, long-range LD. Although we tested associations for specific pairs of SNPs, we interpreted the results in the context of the cytoband regions containing the SNPs. There is substantial debate in the genomics community about how to best link non-coding and intergenic SNPs to corresponding functional genes. Our approach using cytoband regions enabled us to make functional hypotheses about non-coding or intergenic SNPs on the basis of the nearby genes that were most likely to have significant effects on the phenotypes in question. Interestingly, none of the SNPs in our top models had significant effects on the phenotypes of interest in univariate analyses (Figure S1), indicating that the phenotypic associations were driven by interactions between the SNPs rather than by the main-effect contribution of each SNP alone. The genes linked to top models seem to be particularly enriched for dynamic roles as part of large, conserved gene families active during development.

Mechanistic insights on genes with linked complex disorders

FLRT2 and RBFOX1 in type 2 diabetes

Previous studies support a role for *RBFOX1* in regulating gene expression in beta cells through neuron-like alternative-splicing regulation.[60,61] One explanation for this is that it could be because a highly innervated pancreas causes increased neuron-specific transcriptional programs. Another explanation could be that neurons share many phenotypic traits with pancreatic beta cells given that both use similar exocytotic machinery to secrete insulin and neurotransmitters.[60,62] Because FLRT2 is known to

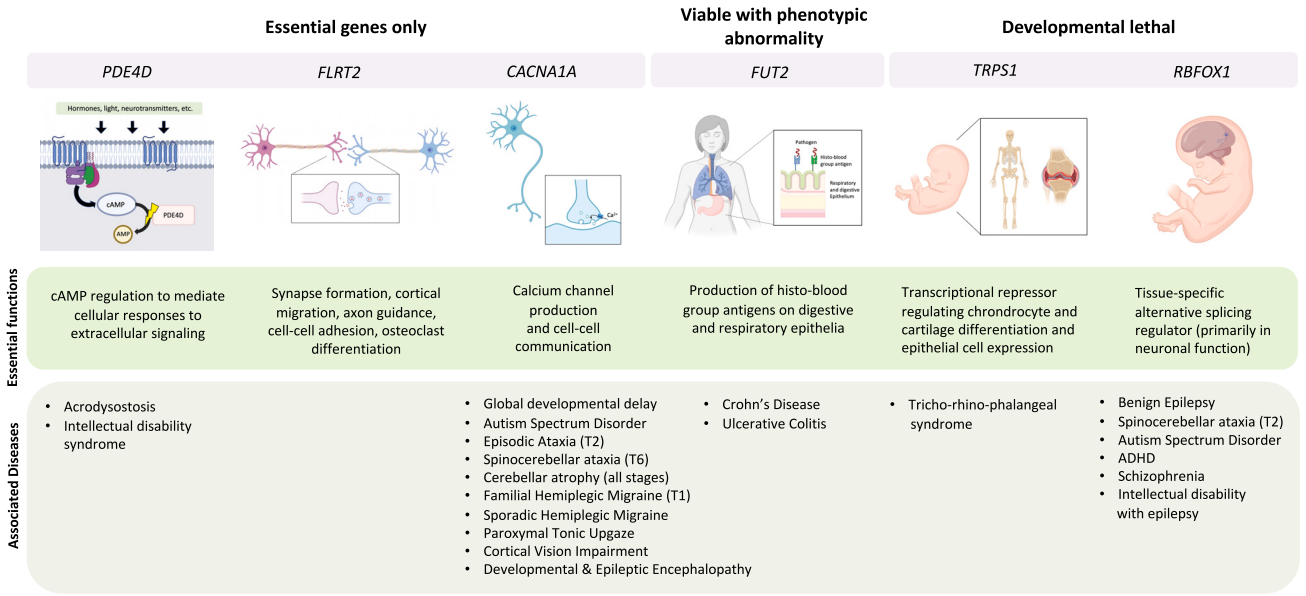

Figure 5. Epistatic interactions map to essential-gene families
Six out of seven genes from our significant and near-significant results function as essential genes; *LRRC4C* is not shown because it is not considered an essential gene. *FUT2* is also viable but has phenotypic abnormality. *TRPS1* and *RBFOX1* are also developmental lethal. Molecular-mechanism panels were created with BioRender.com.

interact with proteins such as ADGRL3, FGFR2, and UNC5D to mediate various cell-signaling pathways, we hypothesize that the interaction between chromosomes 14 and 16 regulates transcriptomic activity in T2D through an alternative-splicing program.[63]

FLRT2 and PDE4D 3 in psoriasis

Individuals with psoriasis are known to have elevated expression of *PDE4D* mRNA in peripheral-blood mononuclear cells.[64] Apremilast, an oral small-molecule inhibitor of PDE4D, is used for treating psoriasis and other chronic inflammatory disorders such as asthma and Behçet's disease through inhibition of the Th17 pathway.[65–67] Our results corroborate the established proinflammatory link between *PDE4D* and psoriasis; however, the mechanistic basis for an association between a *PDE4D-FLRT2* interaction and psoriasis is unknown.

Given the role of *FLRT2* in modulating cortical migration during the development of the nervous system, Akita et al. proposed an analogous guidance function for *FLRT2* in vascular development.[68] *FLRT2* is one of the primary ligands that binds and activates the adhesion G-protein-coupled receptor *LPHN2*, which acts as a repulsive guidance receptor that controls blood-vessel structure and function in model systems.[69] *FLRT2* has also been identified as an autoantigen, or cell-surface target, of anti-endothelial cell antibodies in the vascular systems of individuals with systemic lupus erythematosus.[70] Once *FLRT2* activates *LPHN2,* it elicits the synthesis of cAMP, which is hydrolyzed by PDE4D.[44,71] Our results together with the previous findings suggest that interaction between *PDE4D* and *FLRT2* has a proinflammatory effect in individuals with psoriasis and that it might act through an autoimmune pathology in the vascular system. Further functional experiments are needed to clarify the mechanism.

TRPS1 and LRRC4C in schizophrenia

There is no previous evidence that *TRPS1* and *LRRC4C* interact to influence schizophrenia; however, *LRRC4C* has been implicated in brain disorders, including schizophrenia, bipolar disorder, autism spectrum disorder, and developmental delay.[72,73] Mouse models have shown that mice lacking NGL-1 exhibit hyperactivity and anxiolytic-like behavior as a result of the widespread excitation of neurons in the brain. This suggests that *LRRC4C* plays a role in the suppression or dampening of neuronal activity.[47] We hypothesize that *TRPS1* acts as a repressor of *LRRC4C*; this hypothesis is supported by a previous analysis of GTEx bulk tissue-expression data[74] showing that *LRRC4C* (ENSG00000148948.7) is highly expressed in brain tissues, whereas *TRPS1* (ENSG00000104447.12) is expressed at lower levels in brain tissue than in any other tissues.

FUT2 and CACNA1A in multiple sclerosis

Previous work highlights the role of lymphocyte-mediated calcium-influx patterns observed in autoimmune conditions, including MS.[75] Thus, a genetic interaction between *FUT2* and *CACNA1A* could elucidate a calcium-dependent autoimmune mechanism. We conducted post-hoc network analysis by using HumanBase to identify pathways or functions that would explain the context of the molecular interaction between *CACNA1A* and *FUT2* (Figure 2). The gene connections in networks generated for known pathways involving each gene (calcium-ion transport, calcium-mediated signaling, and proteoglycan biosynthetic process) suggested that the "calcium-ion transport" network best explains the *FUT2* and *CACNA1A*

interaction. This supports previous findings of ion-channel dysfunction in MS.[76] The primary neighbors connecting both genes in the "calcium-ion transport" network provide supporting evidence for various mechanistic hypotheses of MS. *BTNL3*, associated with high-density lipoprotein (HDL) cholesterol levels, is posited to play a role in reduced brain atrophy and demyelination in MS.[77,78] Unexpectedly, the "calcium-ion transport" network also included *OPRPN* and *ODF1*. *OPRPN* encodes the PROL1 protein, which functions in penile erection. *ODF1* encodes the protein that forms the outer dense fibers surrounding the sperm tail.[79] Variation in both genes has been linked to infertility in men, and previous work has shown a higher incidence of infertility in men with MS.[80] *MC5R* in the "calcium-ion transport" network encodes the melanocortin 5 receptor, which exerts immunomodulatory effects by converting primed T cells to regulatory T (T_{reg}) cells.[81] Reduced T_{reg} cell signaling in chronic inflammation can lead to an increase in the number of autoimmune antigen-presenting cells that ultimately cause a self-destructive central nervous system environment in MS. Our network analysis does not define direct links between the aforementioned genes and MS, but it does shed light on potential molecular components of the respective phenotypes seen in MS.

Pleiotropy and epistasis as drivers of subtle phenotypic heterogeneity in related conditions

Despite substantial evidence in support of pleiotropy, there is no empirical basis for pleiotropy in humans.[82,83] Our analysis yields top interaction models that are associated with multiple phenotypes, suggesting a potential pleiotropic etiology for certain clinical pathologies. We hypothesize molecular mechanisms that could explain our results supporting epistatic and pleiotropic relationships.

CACNA1A × FUT2 *in Alzheimer's disease and multiple sclerosis*

Chromosome 19 has long been linked to AD in the GWAS literature as a result of variation in genes such as *APOE* (19q13.32), *ABCA7* (19q13,3), and *CD33* (19q13.41).[84] We found chromosome 19's *CACNA1A* and *FUT2* interaction to potentially play a pleiotropic role in mediating AD and MS.

Located within 2 MB upstream of *CACNA1A* are *TRMT1* and *LDLR*, which are both linked to AD and MS.[85–87] The SNP in the intron of *CACNA1A* might regulate that gene in tandem with the SNP in chromosomal region 19q13.33, or the interaction might be regulating other genes via longer-range effects. *FUT2* has been associated with autoimmune-dysregulation markers such as vitamin B deficiency, elevated cholesterol levels, type 1 diabetes, and dysregulated gut microbiota, all of which have been linked to AD and MS.[77,88–92] Previous work has aimed to detect overlap between autoimmune pathways related to MS pathology and blood lipid pathways related to AD pathology, but no common mechanism for MS and AD pathologies has been identified.[93] We hypothesize that a shared cardiometabolic and autoimmune molecular etiology underpins both AD and MS through the *CACNA1A* and *FUT2* interaction. Functional experiments are needed for validation and elucidation of the mechanism.

FLRT2 in type 2 diabetes and psoriasis

Chromosome 14q31.3, encoding *FLRT2,* was identified as a top locus in interaction models for psoriasis and diabetes in our FastEpistasis analysis. Given the previously identified comorbid relationship and shared etiological factors between psoriasis and diabetes, we evaluated chromosomal region 14q31.3 to identify association with other cardiometabolic and autoimmune conditions. Using the OMIM database, we identified five Mendelian traits.

Bardet-Biedl syndrome (BBS) is a pleiotropic ciliopathy with a wide range of clinical variability. Although BBS individuals have a high propensity toward T2D due to obesity, it is unclear whether diabetes is a comorbidity or a risk independent of obesity. However, recent mouse models of BBS do support the dysregulation of the immune and hematopoietic systems as obesity-independent drivers of T2D.[94] Krabbe disease, a rare autosomal-recessive lysosomal storage disease, has phenotypic overlap with MS in that its hallmark is demyelination caused by the buildup of unmetabolized lipids throughout the central nervous system.[95] Retinitis pigmentosa (RP), a genetic retinopathy that causes vision loss over time, presents varying types of vision loss (e.g., night vision, central vision, or color vision) and severity in different individuals. How lipid dysregulation plays a role in RP disease pathology is not well understood. A previous study found RP to be a result of abetalipoproteinemia.[96,97] An understanding of energy metabolism in retinal cells is critical to uncovering the vascular changes that drive the different stages of RP.[98] Leber congenital amaurosis (LCA) is a group of rare monogenic diseases that frequently result in rapid or progressive vision loss accompanied by other symptoms such as intellectual disability, hearing loss, and cataracts. Lipid changes are not known to be associated with LCA.

The genetic association between chromosomal region 14q31.3 and these conditions suggest that lipid dysregulation, a known link in BBS and Krabbe disease and an unknown link in RP and LCA, may be the mechanism connecting psoriasis and T2D and driving some of these conditions. We hypothesize that epistatic interactions can "modify" and manifest variable expressivity such that the nature of the same condition can vary among individuals. For example, some individuals with BBS who have alleles supporting an interaction involving *FLRT2* may have a higher risk of developing diabetes than individuals with BBS who do not have those alleles. We can only speculate as to why alleles in chromosomal region 14q31.3 might predispose individuals to cardiometabolic dysregulation in the context of complex and Mendelian conditions. However, our data support the hypothesis that genetic heterogeneity manifests as phenotypic heterogeneity and epistatic interactions may explain some portion of the inter-individual variability.

Epistasis regulates function in conserved essential gene families

Given the fundamental cellular roles that essential genes play during and after development, epistasis might be a highly conserved mode of genetic regulation, especially in the context of disease etiology. Epistatic selection might be occurring at higher rates among essential genes, giving rise to diverse clinical outcomes beyond the development phase. Figure 5 demonstrates the essential-gene status of genes in our top models. Six out of seven genes are essential genes; variation in *FUT2* is considered developmentally viable but phenotypically abnormal, and variation in *RBFOX1* and *TRPS1* is considered developmentally lethal. When we evaluated the diseases associated with each essential gene, we found that all the essential genes except *FLRT2* are associated with at least two or more related conditions that differ only by a subset of symptoms. A prime example is *CACNA1A*, which is implicated in a variety of neurological conditions, ranging from autism spectrum disorder and cerebellar atrophy to epileptic encephalopathy.[99] Similarly, *FUT2* has been associated with inflammatory bowel conditions, including Crohn disease and ulcerative colitis.[100] Syndromes are sets of complex symptoms that often co-occur, indicating a specific condition. On the basis of the results of this study, we hypothesize that epistatic variation is a driving factor that differentiates certain sets of overlapping symptoms into distinct pathologies, much like syndromes. These results complement our findings in Figure 4, that epistatic selection in a region could be modulating related etiologies (much like 14q31.3) and leading to phenotypic heterogeneity of the same underlying mechanism(s). Functional studies are needed to test the molecular interactions of the essential genes in our top models to further characterize their role beyond development and in adulthood disease.

Study limitations and future directions

In this study, we propose biological mechanisms that could explain the statistical association of pairwise epistatic SNP models with complex diseases. Functional validation of the molecular mechanisms proposed remains as a future work. *In vitro* knock-out experiments in relevant clinical cell lines or *in vivo* experiments in disease-specific animal models with orthologous gene pairs might yield additional information on the interacting genes and their influence on phenotypic manifestation of the associated disease markers. Understanding the effects of over- or under-expression of genes as a result of manipulation of interacting SNP/gene pairs can shed light on the interactions within and across biological pathways, and this understanding can be translated into health insights into disease etiology. We also acknowledge that higher-order interactions beyond pairwise SNP-SNP models might play an important role in complex disease etiology. Our study focused on genome-wide pairwise interactions; higher-order tests were out of scope as a result of computational limitations.

Conclusion

We leveraged Ohta's D statistics to (1) identify epistatic interactions between distant genomic regions that exhibit strong LD—such regions are typically excluded from genomic analyses—and (2) test association of these interactions with a range of complex diseases. After testing associations of 5,625,845 SNP-SNP pairs with complex diseases, we identified five interactions with disease associations that replicated in the UKBB and eMERGE datasets (Table 2). Most of the interactions were interchromosomal, which is surprising given that most long-range genomic and high-throughput chromosome conformation analyses are limited to single chromosomes.[101] Associations between specific epistatic interactions and conditions, including T2D, psoriasis, MS, schizophrenia, and AD, were identified. (Figure 2). Furthermore, we identified epistatic interactions with a pleiotropic basis. In particular, chromosomal region 14:q31.3 had long-range interactions with chromosome 5 (associated with psoriasis) and chromosome 16 (associated with T2D). We conclude that psoriasis and T2D most likely share an etiology based on dysregulation of lipid metabolism with an autoimmune component. Our post-hoc analysis of chromosomal region 14q31.3 showed that five other Mendelian conditions with lipid dysregulation as a potential shared mechanism are also associated with the same cytoband (Figure 4). We conclude from these findings that epistatic interactions might function at a subtle level to modify phenotypic presentation. For example, some individuals with the interaction between chromosome 14 and chromosome 5 might present a lipid-autoimmune phenotype leading to psoriasis. Similarly, individuals with the interaction between chromosome 14 and chromosome 16 might present a lipid-cardio metabolic phenotype leading to T2D, whereas other individuals might present lipid-ocular conditions such as RP. Different variants within the interacting genes might function differently, adding another layer of fine modulation. This is just one example of how epistatic genetic variation might drive variation of disease phenotypes.

Our epistatic gene models appeared to be enriched with members of highly conserved gene families with dynamic and essential roles. For example, *TRPS1* has been associated with a rare autosomal condition called Tricho-rhino-phalangeal syndrome type 1 (Figure 5), which causes highly variable symptoms including hypermobility, craniofacial abnormalities, and hyperhidrosis.[102] We hypothesize that epistatic gene interactions modulate the combination of symptoms that are expressed in among individuals. Our results suggest that epistasis can be thought of as "fine-tuning," in which interactions manifest variable expressivity such that the same pathology can present as distinct conditions in different individuals. This is known to occur in Mendelian traits, but there is not a concrete basis for this phenomenon in complex traits.

It is likely that a combination of main effects, small-effect variants, and interaction effects determines disease

pathology. In this study we evaluated interaction effects and marginal effects of the same loci and determined that associations with disease were driven by interaction effects only. Further work to model these effects in tandem is needed to provide a more complete understanding of disease risks and mechanisms. Our findings highlight the challenges that remain in leveraging genomics in human health. A stronger emphasis on evaluating genetic interactions is needed if we are to better understand clinical risk associations and inform personalized medicine solutions.

Data and code availability

This study did not generate data. All data produced are available online. eMERGE network phase III data can be accessed through dbGaP (study ID phs001584.v2.p2). UKBB data were accessed through protocol number 32133. Supplemental tables of summary statistics are available to download at https://tinyurl.com/6ej3z554.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.03.007.

Acknowledgments

The authors would like to acknowledge their funding sources. M.D.R. is supported by R01 HG010067 and U01 AG066833. J.H.M. is supported by R01 LM010098 and U01 AG066833. P.S. is supported by F31 AG069441-01. eMERGE Network (phase III): This phase of the eMERGE Network was initiated and funded by the NHGRI through the following grants: U01HG8657 (Group Health Cooperative/University of Washington); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701 (Vanderbilt University Medical Center, serving as the coordinating center); U01HG8676 (Mass General Brigham and Broad Institute); and U01HG8664 (Baylor College of Medicine). The authors would also like to thank Stephen Schaeffer and the participants of the Epistasis Discovery in Genetics and Epidemiology (EDGE) meeting over the years for their insightful discussion and feedback that has guided this project.

Declaration of interests

The authors declare no competing interests.

Received: December 12, 2022
Accepted: March 7, 2023
Published: April 6, 2023

Web resources

OhtaDstats R package, https://github.com/pfpetrowski/OhtaDStats
PLINK v. 1.9, https://www.cog-genomics.org/plink/
PLINK v. 2.0, https://www.cog-genomics.org/plink/2.0/
UK Biobank, http://www.ukbiobank.ac.uk/

References

1. Moore, J.H. (2003). The ubiquitous nature of epistasis in determining susceptibility to common human diseases. In Human Heredity, pp. 73–82. https://doi.org/10.1159/000073735.
2. Draghi, J. (2019). Phenotypic variability can promote the evolution of adaptive plasticity by reducing the stringency of natural selection. J. Evol. Biol. *32*, 1274–1289. https://doi.org/10.1111/jeb.13527.
3. Sella, G., and Barton, N.H. (2019). Thinking about the evolution of complex traits in the era of genome-wide association studies. Annu. Rev. Genomics Hum. Genet. *20*, 461–493. https://doi.org/10.1146/annurev-genom-083115.
4. Mackay, T.F., and Moore, J.H. (2014). Why epistasis is important for tackling complex human disease genetics. Genome Med. *6*, 124. https://doi.org/10.1186/gm561.
5. Lehner, B. (2007). Modelling genotype-phenotype relationships and human disease with genetic interaction networks. J. Exp. Biol. *210*, 1559–1566. https://doi.org/10.1242/jeb.002311.
6. Greene, C.S., Penrod, N.M., Kiralis, J., and Moore, J.H. (2009). Spatially uniform relieff (SURF) for computationally-efficient filtering of gene-gene interactions. BioData Min. *2*, 5. https://doi.org/10.1186/1756-0381-2-5.
7. Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. *9*, 477–485. https://doi.org/10.1038/nrg2361.
8. Guryev, V., Smits, B.M.G., van de Belt, J., Verheul, M., Hubner, N., and Cuppen, E. (2006). Haplotype block structure is conserved across mammals. PLoS Genet. *2*, e121–e1118. https://doi.org/10.1371/journal.pgen.0020121.
9. Thompson, M.J., and Jiggins, C.D. (2014). Supergenes and their role in evolution. Heredity *113*, 1–8. https://doi.org/10.1038/hdy.2014.20.
10. Jeong, H., Baran, N.M., Sun, D., Chatterjee, P., Layman, T.S., Balakrishnan, C.N., Maney, D.L., and Yi, S. v (2022). Dynamic molecular evolution of a supergene with suppressed recombination in white-throated sparrows. Elife *11*, e79387. https://doi.org/10.7554/elife.79387.
11. Joron, M., Frezal, L., Jones, R.T., Chamberlain, N.L., Lee, S.F., Haag, C.R., Whibley, A., Becuwe, M., Baxter, S.W., Ferguson, L., et al. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature *477*, 203–206. https://doi.org/10.1038/nature10341.
12. Chesmore, K., Bartlett, J., and Williams, S.M. (2018). The ubiquity of pleiotropy in human disease. Hum. Genet. *137*, 39–44. https://doi.org/10.1007/s00439-017-1854-z.
13. Tyler, A.L., Asselbergs, F.W., Williams, S.M., and Moore, J.H. (2009). Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. Bioessays *31*, 220–227. https://doi.org/10.1002/bies.200800022.
14. Ohta, T. (1982). Linkage disequilibrium due to random genetic drift in finite subdivided populations. Proc. Natl. Acad. Sci. USA *79*, 1940–1944. https://doi.org/10.1073/pnas.79.6.1940.
15. Ohta, T. (1982). Linkage disequilibrium with the island model. Genetics *101*, 139–155. https://doi.org/10.1093/genetics/101.1.139.
16. Calus, M.P.L., and Vandenplas, J. (2018). SNPrune: An efficient algorithm to prune large SNP array and sequence

datasets based on high linkage disequilibrium. Genet. Sel. Evol. *50*, 34. https://doi.org/10.1186/s12711-018-0404-z.
17. Mackay, T.F.C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat. Rev. Genet. *15*, 22–33. https://doi.org/10.1038/nrg3627.
18. Hua, B., and Springer, M. (2018). Widespread cumulative influence of small effect size mutations on yeast quantitative traits. Cell Syst. *7*, 590–600.e6. https://doi.org/10.1016/j.cels.2018.11.004.
19. Schaeffer, S.W., and Miller, E.L. (1993). Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of Drosophila pseudoobscura. Genetics *135*, 541–552. https://doi.org/10.1093/genetics/135.2.541.
20. Navarro-Dominguez, B., Chang, C.-H., Brand, C.L., Muirhead, C.A., Presgraves, D.C., and Larracuente, A.M. (2022). Epistatic selection on a selfish Segregation Distorter supergene - drive, recombination, and genetic load. Elife *11*, e78981. https://doi.org/10.7554/eLife.78981.
21. Huang, W., Richards, S., Carbone, M.A., Zhu, D., Anholt, R.R.H., Ayroles, J.F., Duncan, L., Jordan, K.W., Lawrence, F., Magwire, M.M., et al. (2012). Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc. Natl. Acad. Sci. USA *109*, 15553–15559. https://doi.org/10.1073/pnas.1213423109.
22. Jacob, F. (1977). Evolution and tinkering. Science *196*, 1161–1166. https://doi.org/10.1126/science.860134.
23. Phillips, P.C. (2008). Epistasis - The essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. *9*, 855–867. https://doi.org/10.1038/nrg2452.
24. Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA *104*, 8597–8604. https://doi.org/10.1073/pnas.8597.
25. Kulminski, A.M., Culminskaya, I., and Yashin, A.I. (2013). Inter-chromosomal level of genome organization and longevity-related phenotypes in humans. Age (Dordr) *35*, 501–518. https://doi.org/10.1007/s11357-011-9374-6.
26. Gandhi, M., Evdokimova, V.N., T Cuenco, K., Nikiforova, M.N., Kelly, L.M., Stringer, J.R., Bakkenist, C.J., and Nikiforov, Y.E. (2012). Homologous chromosomes make contact at the sites of double-strand breaks in genes in somatic G 0/G 1-phase human cells. Proc. Natl. Acad. Sci. USA *109*, 9454–9459. https://doi.org/10.1073/pnas.1205759109.
27. Krueger, C., King, M.R., Krueger, F., Branco, M.R., Osborne, C.S., Niakan, K.K., Higgins, M.J., and Reik, W. (2012). Pairing of homologous regions in the mouse genome is associated with transcription but not imprinting status. PLoS One *7*, e38983. https://doi.org/10.1371/journal.pone.0038983.
28. Maass, P.G., Barutcu, A.R., and Rinn, J.L. (2019). Interchromosomal interactions: A genomic love story of kissing chromosomes. J. Cell Biol. *218*, 27–38. https://doi.org/10.1083/jcb.201806052.
29. Koch, E., Ristroph, M., and Kirkpatrick, M. (2013). Long range linkage disequilibrium across the human genome. PLoS One *8*, e80754. https://doi.org/10.1371/journal.pone.0080754.
30. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283. https://doi.org/10.1038/ng.3643.
31. Petrowski, P.F., King, E.G., and Beissinger, T.M. (2019). An R framework for the partitioning of linkage disequilibrium between and within populations. J. Open Res. Softw. *7*, 15. https://doi.org/10.5334/jors.250.
32. Bush, W.S., Dudek, S.M., and Ritchie, M.D. (2009). Biofilter: a knowledge-integration system for the multi-locus analysis of genome-wide association studies. Pac. Symp. Biocomput., 368–379.
33. Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. Nat. Genet. *47*, 569–576. https://doi.org/10.1038/ng.3259.
34. Cacheiro, P., Muñoz-Fuentes, V., Murray, S.A., Dickinson, M.E., Bucan, M., Nutter, L.M.J., Peterson, K.A., Haselimashhadi, H., Flenniken, A.M., Morgan, H., et al. (2020). Human and mouse essentiality screens as a resource for disease gene discovery. Nat. Commun. *11*, 655. https://doi.org/10.1038/s41467-020-14284-2.
35. Ji, X., Kember, R.L., Brown, C.D., and Bućan, M. (2016). Increased burden of deleterious variants in essential genes in autism spectrum disorder. Proc. Natl. Acad. Sci. USA *113*, 15054–15059. https://doi.org/10.1073/pnas.1613195113.
36. Boughton, A.P., Welch, R.P., Flickinger, M., VandeHaar, P., Taliun, D., Abecasis, G.R., and Boehnke, M. (2021). LocusZoom.js: interactive and embeddable visualization of genetic association study results. Bioinformatics *37*, 3017–3018. https://doi.org/10.1093/bioinformatics/btab186.
37. Wei, W.Q., Bastarache, L.A., Carroll, R.J., Marlo, J.E., Osterman, T.J., Gamazon, E.R., Cox, N.J., Roden, D.M., and Denny, J.C. (2017). Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One *12*, e0175508. https://doi.org/10.1371/journal.pone.0175508.
38. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779. https://doi.org/10.1371/journal.pmed.1001779.
39. Stanaway, I.B., Hall, T.O., Rosenthal, E.A., Palmer, M., Naranbhai, V., Knevel, R., Namjou-Khales, B., Carroll, R.J., Kiryluk, K., Gordon, A.S., et al. (2019). The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. Genet. Epidemiol. *43*, 63–81. https://doi.org/10.1002/gepi.22167.
40. Schüpbach, T., Xenarios, I., Bergmann, S., and Kapur, K. (2010). FastEpistasis: A high performance computing solution for quantitative trait epistasis. Bioinformatics *26*, 1468–1469. https://doi.org/10.1093/bioinformatics/btq147.
41. Dolan, M. (2011). The role of the Giemsa stain in cytogenetics. Biotech. Histochem. *86*, 94–97. https://doi.org/10.3109/10520295.2010.515493.
42. Damianov, A., Ying, Y., Lin, C.H., Lee, J.A., Tran, D., Vashisht, A.A., Bahrami-Samani, E., Xing, Y., Martin, K.C., Wohlschlegel, J.A., and Black, D.L. (2016). Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. Cell *165*, 606–619. https://doi.org/10.1016/j.cell.2016.03.040.

43. Li, J., Shinoda, Y., Ogawa, S., Ikegaya, S., Li, S., Matsuyama, Y., Sato, K., and Yamagishi, S. (2021). Expression of FLRT2 in postnatal central nervous system development and after spinal cord injury. Front. Mol. Neurosci. *14*, 756264. https://doi.org/10.3389/fnmol.2021.756264.
44. Ong, W.K., Gribble, F.M., Reimann, F., Lynch, M.J., Houslay, M.D., Baillie, G.S., Furman, B.L., and Pyne, N.J. (2009). The role of the PDE4D cAMP phosphodiesterase in the regulation of glucagon-like peptide-1 release. Br. J. Pharmacol. *157*, 633–644. https://doi.org/10.1111/j.1476-5381.2009.00194.x.
45. Zhao, W.W. (2013). Intragenic deletion of RBFOX1 associated with neurodevelopmental/neuropsychiatric disorders and possibly other clinical presentations. Mol. Cytogenet. *6*, 26. https://doi.org/10.1186/1755-8166-6-26.
46. Wang, Y., Lin, X., Gong, X., Wu, L., Zhang, J., Liu, W., Li, J., and Chen, L. (2018). Atypical GATA transcription factor TRPS1 represses gene expression by recruiting CHD4/NuRD(MTA2) and suppresses cell migration and invasion by repressing TP63 expression. Oncogenesis *7*, 96. https://doi.org/10.1038/s41389-018-0108-9.
47. Choi, Y., Park, H., Kang, S., Jung, H., Kweon, H., Kim, S., Choi, I., Lee, S.Y., Choi, Y.E., Lee, S.H., and Kim, E. (2019). NGL-1/LRRC4C-mutant mice display hyperactivity and anxiolytic-like behavior associated with widespread suppression of neuronal activity. Front. Mol. Neurosci. *12*, 250. https://doi.org/10.3389/fnmol.2019.00250.
48. Dendrou, C.A., Petersen, J., Rossjohn, J., and Fugger, L. (2018). HLA variation and disease. Nat. Rev. Immunol. *18*, 325–339. https://doi.org/10.1038/nri.2017.143.
49. Patsopoulos, N.A., Barcellos, L.F., Hintzen, R.Q., Schaefer, C., van Duijn, C.M., Noble, J.A., Raj, T., Gourraud, P.A., ANZgene, and Gourraud, P.A., et al. (2013). Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet. *9*, e1003926. https://doi.org/10.1371/journal.pgen.1003926.
50. Lu, R.-C., Yang, W., Tan, L., Sun, F.-R., Tan, M.-S., Zhang, W., Wang, H.-F., and Tan, L. (2017). Association of HLA-DRB1 polymorphism with Alzheimer's disease: a replication and meta-analysis. Oncotarget *8*, 93219–93226. https://doi.org/10.18632/oncotarget.21479.
51. Jiang, Z., Ren, W., Liang, H., Yan, J., Yang, D., Luo, S., Zheng, X., Lin, G.-W., Xian, Y., Xu, W., et al. (2021). HLA class I genes modulate disease risk and age at onset together with DR-DQ in Chinese patients with insulin-requiring type 1 diabetes. Diabetologia *64*, 2026–2036. https://doi.org/10.1007/s00125-021-05476-6.
52. Auckland, K., Mittal, B., Cairns, B.J., Garg, N., Kumar, S., Mentzer, A.J., Kado, J., Perman, M.L., Steer, A.C., Hill, A.V.S., and Parks, T. (2020). The human leukocyte antigen locus and rheumatic heart disease susceptibility in South Asians and Europeans. Sci. Rep. *10*, 9004. https://doi.org/10.1038/s41598-020-65855-8.
53. Azad, M.B., Wade, K.H., and Timpson, N.J. (2018). FUT2 secretor genotype and susceptibility to infections and chronic conditions in the ALSPAC cohort. Wellcome Open Res. *3*, 65. https://doi.org/10.12688/wellcomeopenres.14636.2.
54. Santos-Cortez, R.L.P., Chiong, C.M., Frank, D.N., Ryan, A.F., Giese, A.P.J., Bootpetch Roberts, T., Daly, K.A., Steritz, M.J., Szeremeta, W., Pedro, M., et al. (2018). FUT2 Variants Confer Susceptibility to Familial Otitis Media. Am. J. Hum. Genet. *103*, 679–690. https://doi.org/10.1016/j.ajhg.2018.09.010.
55. Psychiatric GWAS Consortium Bipolar Disorder Working Group (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. *43*, 977–983. https://doi.org/10.1038/ng.943.
56. Grosso, B.J., Kramer, A.A., Tyagi, S., Bennett, D.F., Tifft, C.J., D'Souza, P., Wangler, M.F., Macnamara, E.F., Meza, U., and Bannister, R.A. (2022). Complex effects on CaV2.1 channel gating caused by a CACNA1A variant associated with a severe neurodevelopmental disorder. Sci. Rep. *12*, 9186. https://doi.org/10.1038/s41598-022-12789-y.
57. Brazzelli, V., Maffioli, P., Bolcato, V., Ciolfi, C., D'Angelo, A., Tinelli, C., and Derosa, G. (2021). Psoriasis and diabetes, a dangerous association: evaluation of insulin resistance, lipid abnormalities, and cardiovascular risk biomarkers. Front. Med. *8*, 605691. https://doi.org/10.3389/fmed.2021.605691.
58. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. *33*, D514–D517. https://doi.org/10.1093/nar/gki033.
59. Goh, K.-I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., and Barabási, A.L. (2007). The human disease network. Proc. Natl. Acad. Sci. USA *104*, 8685–8690. https://doi.org/10.1073/pnas.0701361104.
60. Juan-Mateu, J., Rech, T.H., Villate, O., Lizarraga-Mollinedo, E., Wendt, A., Turatsinze, J.V., Brondani, L.A., Nardelli, T.R., Nogueira, T.C., Esguerra, J.L.S., et al. (2017). Neuron-enriched RNA-binding proteins regulate pancreatic beta cell function and survival. J. Biol. Chem. *292*, 3466–3480. https://doi.org/10.1074/jbc.M116.748335.
61. Nutter, C.A., Jaworski, E., Verma, S.K., Perez-Carrasco, Y., and Kuyumcu-Martinez, M.N. (2017). Developmentally regulated alternative splicing is perturbed in type 1 diabetic skeletal muscle. Muscle Nerve *56*, 744–749. https://doi.org/10.1002/mus.25599.
62. Arntfield, M.E., and van der Kooy, D. (2011). β-Cell evolution: How the pancreas borrowed from the brain: The shared toolbox of genes expressed by neural and pancreatic endocrine cells may reflect their evolutionary relationship. Bioessays *33*, 582–587. https://doi.org/10.1002/bies.201100015.
63. Wei, K., Xu, Y., Tse, H., Manolson, M.F., and Gong, S.G. (2011). Mouse FLRT2 interacts with the extracellular and intracellular regions of FGFR2. J. Dent. Res. *90*, 1234–1239. https://doi.org/10.1177/0022034511415272.
64. Schafer, P.H., Truzzi, F., Parton, A., Wu, L., Kosek, J., Zhang, L.H., Horan, G., Saltari, A., Quadri, M., Lotti, R., et al. (2016). Phosphodiesterase 4 in inflammatory diseases: Effects of apremilast in psoriatic blood and in dermal myofibroblasts through the PDE4/CD271 complex. Cell. Signal. *28*, 753–763. https://doi.org/10.1016/j.cellsig.2016.01.007.
65. Chen, Y., Li, Z., Li, H., Su, W., Xie, Y., Pan, Y., Chen, X., and Liang, D. (2020). Apremilast Regulates the Teff/Treg Balance to Ameliorate Uveitis via PI3K/AKT/FoxO1 Signaling Pathway. Front. Immunol. *11*, 581673. https://doi.org/10.3389/fimmu.2020.581673.
66. Schett, G., Sloan, V.S., Stevens, R.M., and Schafer, P. (2010). Apremilast: A novel PDE4 inhibitor in the treatment of autoimmune and inflammatory diseases. Ther. Adv. Musculoskelet. Dis. *2*, 271–278. https://doi.org/10.1177/1759720X10381432.

67. Afra, T.P., Razmi, T.M., and Dogra, S. (2019). Apremilast in Psoriasis and Beyond: Big Hopes on a Small Molecule. Indian Dermatol. Online J. *10*, 1–12. https://doi.org/10.4103/idoj.IDOJ_437_18.
68. Akita, T., Kumada, T., Yoshihara, S.i., Egea, J., and Yamagishi, S. (2016). Ion channels, guidance molecules, intracellular signaling and transcription factors regulating nervous and vascular system development. J. Physiol. Sci. *66*, 175–188. https://doi.org/10.1007/s12576-015-0416-1.
69. Camillo, C., Facchinello, N., Villari, G., Mana, G., Gioelli, N., Sandri, C., Astone, M., Tortarolo, D., Clapero, F., Gays, D., et al. (2021). LPHN2 inhibits vascular permeability by differential control of endothelial cell adhesion. J. Cell Biol. *220*, e202006033. https://doi.org/10.1083/jcb.202006033.
70. Shirai, T., Fujii, H., Ono, M., Nakamura, K., Watanabe, R., Tajima, Y., Takasawa, N., Ishii, T., and Harigae, H. (2012). A novel autoantibody against fibronectin leucine-rich transmembrane protein 2 expressed on the endothelial cell surface identified by retroviral vector system in systemic lupus erythematosus. Arthritis Res. Ther. *14*, R157. https://doi.org/10.1186/ar3897.
71. Sando, R., and Südhof, T.C. (2021). Latrophilin GPCR signaling mediates synapse formation. Elife *10*, e65717. https://doi.org/10.7554/eLife.65717.
72. Maussion, G., Cruceanu, C., Rosenfeld, J.A., Bell, S.C., Jollant, F., Szatkiewicz, J., Collins, R.L., Hanscom, C., Kolobova, I., de Champfleur, N.M., et al. (2017). Implication of LRRC4C and DPP6 in neurodevelopmental disorders. Am. J. Med. Genet. *173*, 395–406. https://doi.org/10.1002/ajmg.a.38021.
73. Zhang, Y., Li, D., Zeng, Q., Feng, J., Fu, H., Luo, Z., Xiao, B., Yang, H., and Wu, M. (2021). LRRC4 functions as a neuron-protective role in experimental autoimmune encephalomyelitis. Mol. Med. *27*, 44. https://doi.org/10.1186/s10020-021-00304-4.
74. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.
75. Orbán, C., Biró, E., Grozdics, E., Bajnok, A., and Toldi, G. (2013). Modulation of T lymphocyte calcium influx patterns via the inhibition of Kv1.3 and IKCa1 potassium channels in autoimmune disorders. Front. Immunol. *4*, 234. https://doi.org/10.3389/fimmu.2013.00234.
76. Waxman, S.G. (2002). Ion channels and neuronal dysfunction in multiple sclerosis. Arch. Neurol. *59*, 1377–1380. https://doi.org/10.1001/archneur.59.9.1377.
77. Blumenfeld Kan, S., Staun-Ram, E., Golan, D., and Miller, A. (2019). HDL-cholesterol elevation associated with fingolimod and dimethyl fumarate therapies in multiple sclerosis. Mult. Scler. J. Exp. Transl. Clin. *5*. 2055217319882720. https://doi.org/10.1177/2055217319882720.
78. Liu, N., Guo, Y.-N., Wang, X.-J., Ma, J., He, Y.-T., Zhang, F., He, H., Xie, J.-L., Zhuang, X., Liu, M., et al. (2022). Copy number analyses identified a novel gene: APOBEC3A related to lipid metabolism in the pathogenesis of preeclampsia. Front. Cardiovasc. Med. *9*, 841249. https://doi.org/10.3389/fcvm.2022.841249.
79. Yang, K., Meinhardt, A., Zhang, B., Grzmil, P., Adham, I.M., and Hoyer-Fender, S. (2012). The small heat shock protein ODF1/HSPB10 is essential for tight linkage of sperm head to tail and male fertility in mice. Mol. Cell Biol. *32*, 216–225. https://doi.org/10.1128/MCB.06158-11.
80. Glazer, C.H., Tøttenborg, S.S., Giwercman, A., Bräuner, E.V., Eisenberg, M.L., Vassard, D., Magyari, M., Pinborg, A., Schmidt, L., and Bonde, J.P. (2018). Male factor infertility and risk of multiple sclerosis: A register-based cohort study. Mult. Scler. *24*, 1835–1842. https://doi.org/10.1177/1352458517734069.
81. Taylor, A., and Namba, K. (2001). In vitro induction of CD25+ CD4+ regulatory T cells by the neuropeptide alpha-melanocyte stimulating hormone (alpha-MSH). Immunol. Cell Biol. *79*, 358–367. https://doi.org/10.1046/j.1440-1711.2001.01022.x.
82. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J.F., and Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. *89*, 607–618. https://doi.org/10.1016/j.ajhg.2011.10.004.
83. Paaby, A.B., and Rockman, M. v (2013). The many faces of pleiotropy. Trends Genet. *29*, 66–73. https://doi.org/10.1016/j.tig.2012.10.010.
84. Moreno-Grau, S., Hernández, I., Heilmann-Heimbach, S., Ruiz, S., Rosende-Roca, M., Mauleón, A., Vargas, L., Rodríguez-Gómez, O., Alegret, M., Espinosa, A., et al. (2018). Genome-wide significant risk factors on chromosome 19 and the APOE locus. Oncotarget *9*, 24590–24600. https://doi.org/10.18632/oncotarget.25083.
85. Lu, Y., Yue, D., Xie, J., Cheng, L., and Wang, X. (2022). Ontology specific alternative splicing changes in Alzheimer's disease. Front. Genet. *13*, 926049. https://doi.org/10.3389/fgene.2022.926049.
86. Mailleux, J., Timmermans, S., Nelissen, K., Vanmol, J., Vanmierlo, T., van Horssen, J., Bogie, J.F.J., and Hendriks, J.J.A. (2017). Low-density lipoprotein receptor deficiency attenuates neuroinflammation through the induction of apolipoprotein E. Front. Immunol. *8*, 1701. https://doi.org/10.3389/fimmu.2017.01701.
87. Gopalraj, R.K., Zhu, H., Kelly, J.F., Mendiondo, M., Pulliam, J.F., Bennett, D.A., and Estus, S. (2005). Genetic association of low density lipoprotein receptor and Alzheimer's disease. Neurobiol. Aging *26*, 1–7. https://doi.org/10.1016/j.neurobiolaging.2004.09.001.
88. Schepici, G., Silvestro, S., Bramanti, P., and Mazzon, E. (2019). The gut microbiota in multiple sclerosis: an overview of clinical trials. Cell Transplant. *28*, 1507–1527. https://doi.org/10.1177/0963689719873890.
89. Najafi, M.R., Shaygannajad, V., Mirpourian, M., and Gholamrezaei, A. (2012). Vitamin B(12) deficiency and multiple sclerosis; is there any association? Int. J. Prev. Med. *3*, 286–289.
90. Smyth, D.J., Cooper, J.D., Howson, J.M.M., Clarke, P., Downes, K., Mistry, T., Stevens, H., Walker, N.M., and Todd, J.A. (2011). FUT2 nonsecretor status links type 1 diabetes susceptibility and resistance to infection. Diabetes *60*, 3081–3084. https://doi.org/10.2337/db11-0638.
91. Ellinghaus, D., Ellinghaus, E., Nair, R.P., Stuart, P.E., Esko, T., Metspalu, A., Debrus, S., Raelson, J. v, Tejasvi, T., Belouchi, M., et al. (2012). Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. Am. J. Hum. Genet. *90*, 636–647. https://doi.org/10.1016/j.ajhg.2012.02.020.
92. Hazra, A., Kraft, P., Selhub, J., Giovannucci, E.L., Thomas, G., Hoover, R.N., Chanock, S.J., and Hunter, D.J. (2008). Common variants of FUT2 are associated with plasma vitamin B12 levels. Nat. Genet. *40*, 1160–1162. https://doi.org/10.1038/ng.210.

93. Podbielska, M., O'Keeffe, J., and Pokryszko-Dragan, A. (2021). New insights into multiple sclerosis mechanisms: lipids on the track to control inflammation and neurodegeneration. Int. J. Mol. Sci. *22*, 7319. https://doi.org/10.3390/ijms22147319.
94. Tsyklauri, O., Niederlova, V., Forsythe, E., Prasai, A., Drobek, A., Kasparek, P., Sparks, K., Trachtulec, Z., Prochazka, J., Sedlacek, R., et al. (2021). Bardet-Biedl Syndrome ciliopathy is linked to altered hematopoiesis and dysregulated self-tolerance. EMBO Rep. *22*, e50785. https://doi.org/10.15252/embr.202050785.
95. Mar, S., and Noetzel, M. (2010). Axonal damage in leukodystrophies. Pediatr. Neurol. *42*, 239–242. https://doi.org/10.1016/j.pediatrneurol.2009.08.011.
96. Gouras, P., Carr, R.E., and Gunkel, R.D. (1971). Retinitis pigmentosa in abetalipoproteinemia: Effects of vitamin A. Invest. Ophthalmol. *10*, 784–793.
97. Berson, E.L. (2000). Nutrition and retinal degenerations. Int. Ophthalmol. Clin. *40*, 93–111. https://doi.org/10.1097/00004397-200010000-00008.
98. Fu, Z., Chen, C.T., Cagnone, G., Heckel, E., Sun, Y., Cakir, B., Tomita, Y., Huang, S., Li, Q., Britton, W., et al. (2019). Dyslipidemia in retinal metabolic disorders. EMBO Mol. Med. *11*, e10473. https://doi.org/10.15252/emmm.201910473.
99. Damaj, L., Lupien-Meilleur, A., Lortie, A., Riou, É., Ospina, L.H., Gagnon, L., Vanasse, C., and Rossignol, E. (2015). CACNA1A haploinsufficiency causes cognitive impairment, autism and epileptic encephalopathy with mild cerebellar symptoms. Eur. J. Hum. Genet. *23*, 1505–1512. https://doi.org/10.1038/ejhg.2015.21.
100. Wu, H., Sun, L., Lin, D.-P., Shao, X.-X., Xia, S.-L., and Lv, M. (2017). Association of fucosyltransferase 2 gene polymorphisms with inflammatory bowel disease in patients from Southeast China. Gastroenterol. Res. Pract. *2017*, 4148651. https://doi.org/10.1155/2017/4148651.
101. Park, L. (2019). Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants. Sci. Rep. *9*, 11380. https://doi.org/10.1038/s41598-019-47832-y.
102. Seitz, C.S., Lü, H.-J., Wagner, N., Bröcker, E.-B., and Hamm, H. (2001). Trichorhinophalangeal Syndrome Type I Clinical and Molecular Characterization of 3 Members of a Family and 1 Sporadic Case. Arch. Dermatol. *137*, 1437–1442. https://doi.org/10.1001/archderm.137.11.1437.