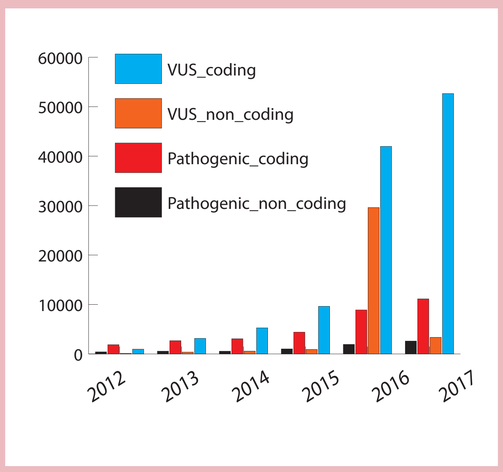
Understanding genotype phenotype relationships in human diseases using an edgetic approach

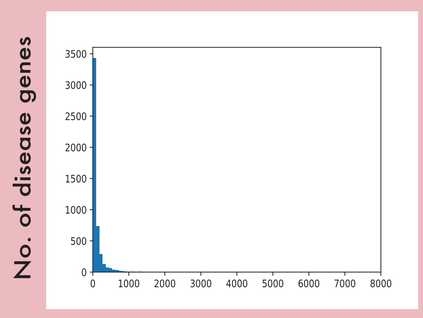
Georges Coppin[1,2,3], Anupama Yadav[1,2], Dylan Markey[1], David De Ridder[1], Kerstin Spirohn[1], Tong Hao[1], Jochen Weile[1,3,4], Atina Cote[1,3,4], Michael A. Calderwood[1,2], David E. Hill[1,2], Frederick P. Roth[1,4,5,6], Jean-Claude Twizere[3], Marc Vidal[1,2]

[1]Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, DFCI, Boston, MA, [2]Department of Genetics, Harvard Medical School, Boston, MA, [3]Molecular Biology of Diseases, Groupe Interdisciplinaire de Génomique Appliquée (GIGA) and Laboratory of Viral Interactomes, University of Liège, Liège, Belgium, [4]Donnelly Centre, University of Toronto, Toronto, Canada, [5]LTRI, Mount Sinai Hospital, Toronto, Canada, [6]Department of Molecular Genetics and Computer Science, University of Toronto, Toronto, Canada

Drastic increase in variants of uncertain significance (VUS)[1,2]



Highly skewed literature coverage of disease genes



Urgent need to functionally and molecularly characterize variants in disease genes

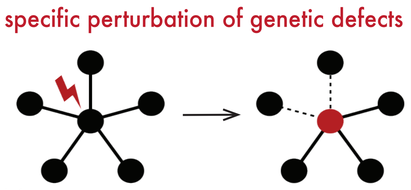
In silico tools[3]
Comprehensive
Predictive

Functional assays[3]
Powerful
Low throughput

Network based approaches

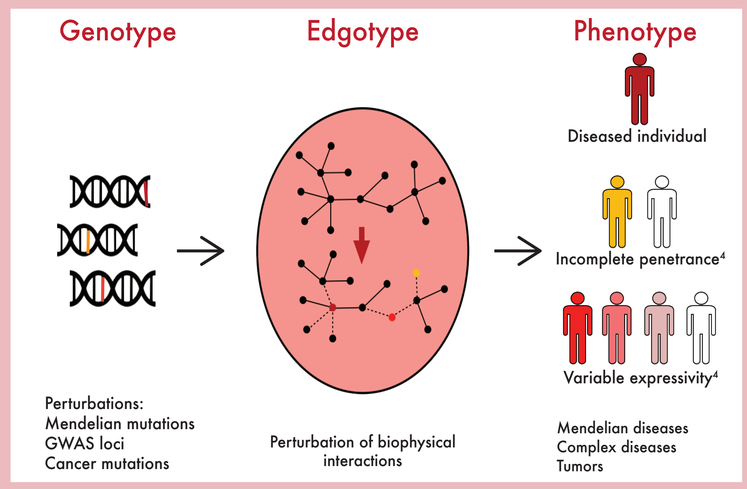
EDGETICS: Edge specific perturbation of genetic defects

Experimental
High throughput
Systematic



Biomolecule (Node) — Physical interaction (Edge)

Network based approach to understand genotype phenotype relationships



High throughput cloning of disease alleles

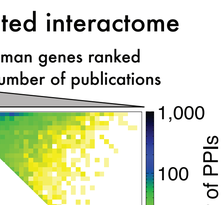
Variant selection

Disease variant archives:
Clinvar[1]
HGMD[5]

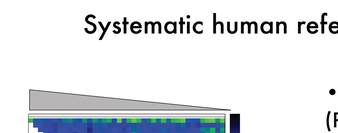
Diverse populations:
1000 genomes[6]
ExAC[7]

Specific disease cohorts:
Parkinsons
Asthma-COPD

Cloning

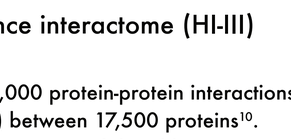


18,000 variants in > 2,100 genes

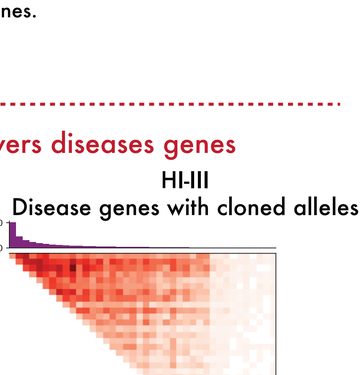


Clones are sequence confirmed using long read sequencing to ensure a single nucleotide change

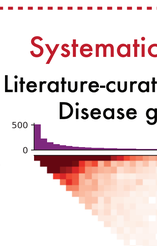
Highly versatile gateway cloning system[8]



Allows testing of mulitple molecular interactions and phenotypes[9]

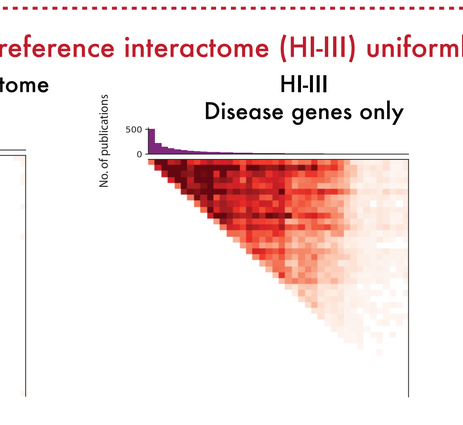


At least one allele cloned for 40 percent of diseases



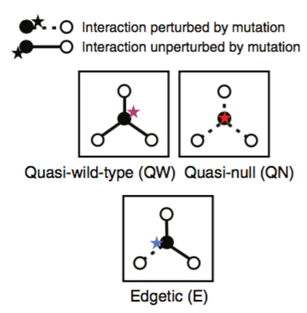
Diseases with cloned alleles
Diseases without cloned alleles

Cloned alleles uniformly spread across disease categories



Disease categories from MeSH

Different perturbation patterns of disease, uncertain and benign variants

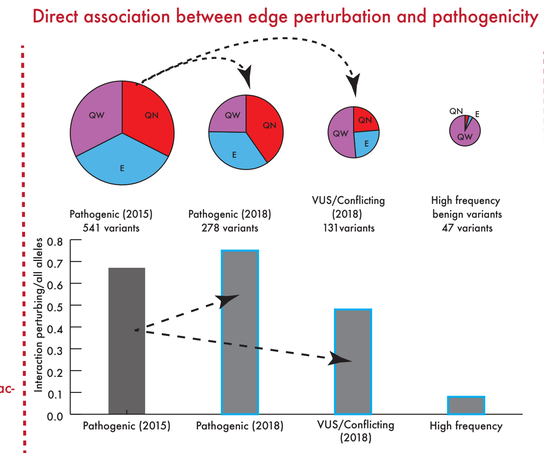


The cloned variants are being tested for multiple molecular interactions including protein-protein interactions, protein-DNA interactions, localization, protein-chaperone interactions, as well as specific functional assays.

Based on the protein interaction profiles, the variants are divided into three categories: QW, QN and E[12].
Majority of the disease causing variants perturb most or specific interactions whereas most of the benign variants retain all WT interactions. VUS, as expected, are a mixture of disease causing and benign variants.

An example compares edgetic profiles between disease variants and VUSs of the DNA damage response protein MLH1, mutations in which cause both hereditary and somatic forms of colorectal cancer.

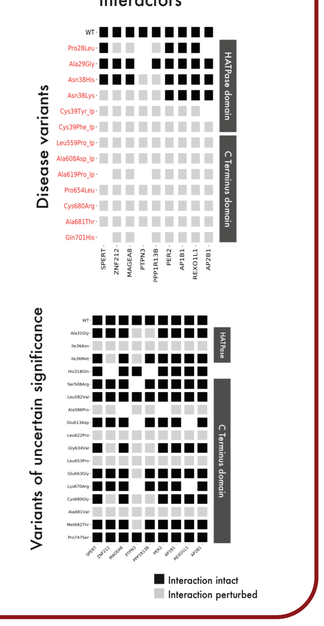
Direct association between edge perturbation and pathogenicity



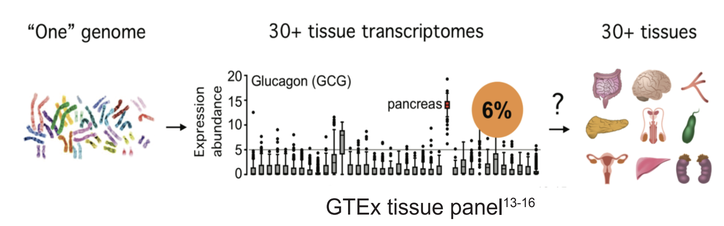
Clinvar is systematically reannotating genetic variants based on emerging data and expert pannels[1].
Some varaints, considered pathogenic in 2015 by Clinvar and HGMD, have been reannotated to VUS or conflicting.

The ratio of interaction perturbation (QN + E) is directly correlated with pathogenicity.

Edgetic profiles of variants in MLH1 Interactors

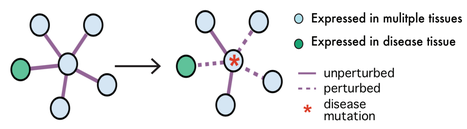


Tissue specificity of diseases caused by uniformly expressed genes

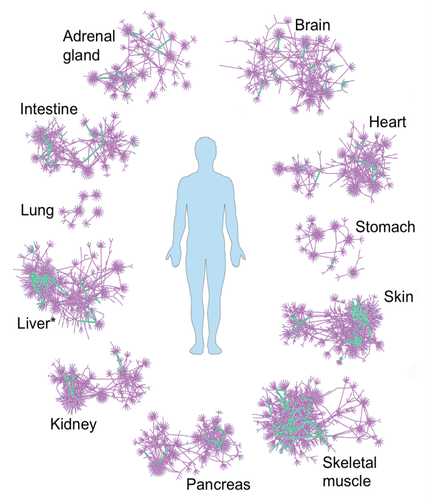


- Most of the Mendelian diseases are tissue specific.
- For >80% of tissue specific diseases, not a single causal gene is expressed preferentially in the diseased tissue[10].

Model:
Variants in uniformyl expressed-genes might cause tissue specfic diseases by perturbing interactions with tissue specific genes.



Constructing tissue specific interactome



~30,000 edges involving disease genes in the systematic human reference protein interactome

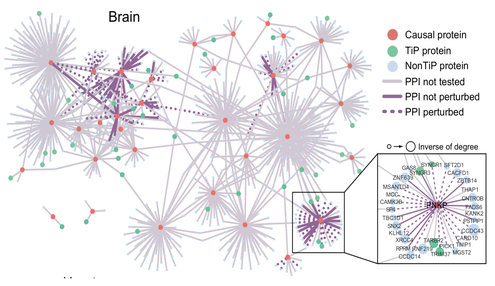
Literature-curated interactome

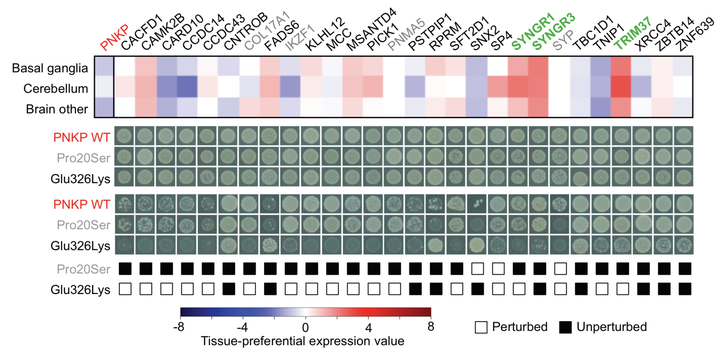
Systematic human reference interactome (HI-III)

- 70,000 protein-protein interactions (PPIs) between 17,500 proteins[10].
- Interactions validated by sequencing and other orthogonal assays.
- Uniform, unbiased coverage of the human genome[10,11].
- 30,000 PPIs involving 50% of disease genes.

Systematic human reference interactome (HI-III) uniformly covers diseases genes

Edgetic effects of variants in uniformly expressed genes causing tissue specific diseases[10]





Funding:
This work was funded by Fonds de la Recherche Scientifique – FNRS (FRIA) and grants P50 HG004233 and U41 HG001715

References: [1]Landrum *et al NAR* 2018, [2]Manolio *et al Cell* 2017, [3]Starita *et al TAJHG* 2017, [4]Cooper *et al Hum genet* 2013, [5]Stenson *et al Human mutation* 2003, [6]The thousand genome project consortium *Nature* 2015, [7]Lek *et al Nature* 2016, [8]Walhout *et al Methods in enzymology* 2000, [9]Sahni *et al Current opinion in genes and development* 2013, [10]**Luck et al bioRxiv 2019**, [11]Rolland *et al Nature* 2014, [12]Sahni *et al Cell* 2015, [13]The GTEx Consortium *Science* 2015, [14]Sonawane *et al Cell Reports* 2017, [15]Paulson *et al BMC Bioinformatics* 2017, [16]Hamosh *et al NAR* 2006