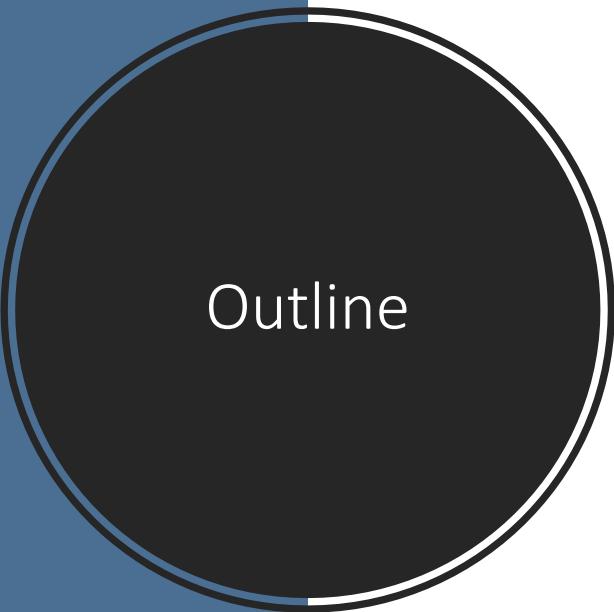


Maryam Labaf

Computational Science PhD Program
University of Massachusetts Boston

Effect of SNPs in cis-regulatory motifs on penetrance of pathogenic mutations

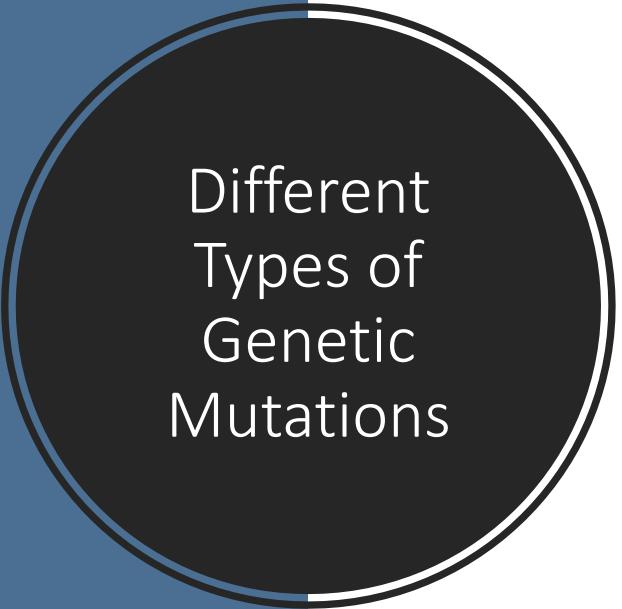


Outline

- Genetic Mutation
- Biological Definitions
- Genomic Variation
- Effect of SNPs in cis-regulatory motifs on penetrance of pathogenic mutations

Different Types of Genetic Mutations

	DNA Level	RNA Level	Protein Level
No mutation	CTC CTC CTC	GAG GAG GAG	Glu – Glu - GLu
Point Mutation	CTC CTC CAC	GAG GAG UG	Glu – Glu - Val
Frame –Shift Mutation	CTC C CTC CTC	GAG GGA GGA G	Glu – Gly - Gly
No mutation	ACA	UGU	Cys
Non-Sense Mutation	ACT	UGA	STOP!
Missense Mutation	ACC	UGG	Trp
• Silent mutations • Conservative mutation • Non-conservative mutation	CCA , CCG, CCT, CCC → Gly		
	Glu → Asp		
	Ser → Phe		



Different Types of Genetic Mutations

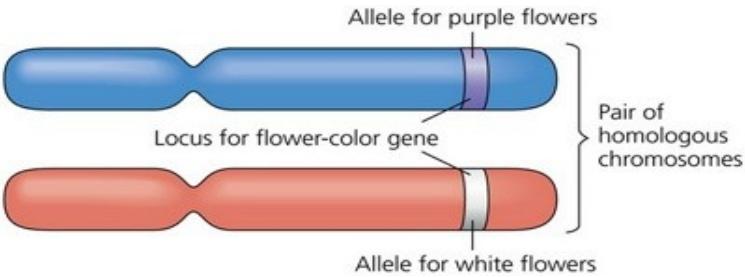
- **Synonymous mutations:**
 - Point mutations i.e. only one miscopied of DNA nucleotide
 - The mutated codon has the same meaning as the original codon
 - Amino acid does not change and not affects protein
 - No real role in the evolution of species
- **Nonsynonymous Mutations:**
 - Insertion or deletion of a single nucleotide in the sequence during transcription
 - Causes a frameshift mutation which throws off the entire reading frame of the amino acid sequence
 - Changes the resulting protein that is expressed
 - Be a lethal mutation if it happens near the beginning and entire protein is changed
 - Non-sense mutation - causes stop codon

Biological Definitions

wild-type sequence
ATCTTCAGCCATAAAAGATGAAGTT

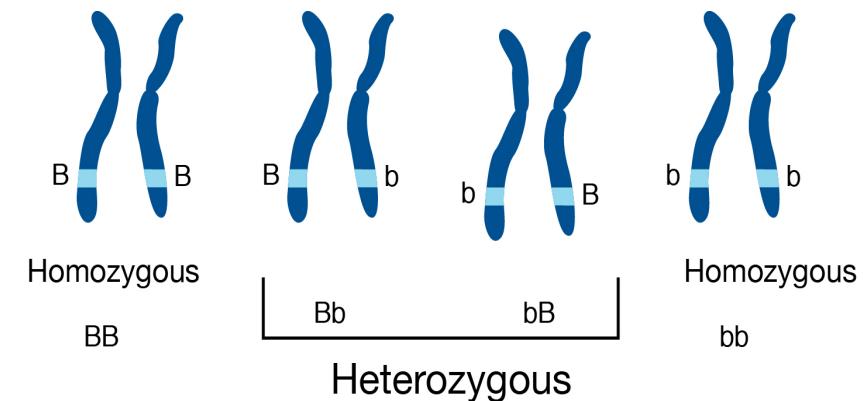
3 bp deletion
ATCTTCAGCCAAAGATGAAGTT

4 bp insertion (orange)
ATCTTCAGCCATATGTGAAAGATGAAGTT

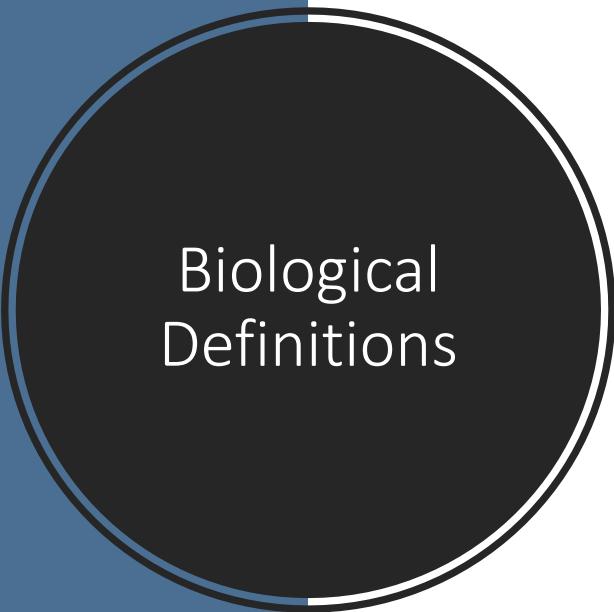


- **Zygosity**
 - A **homozygote** possesses two identical copies of the same allele at a locus
 - A **heterozygote** possesses two different alleles at a locus

- **Alleles**
 - Variant form of a given gene
 - Different alleles can lead to different phenotype



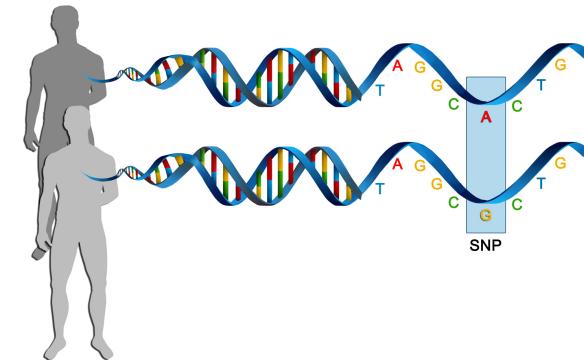
- **Indel** is a genetic difference created by insertion or deletion of a base pair or a longer DNA segment.



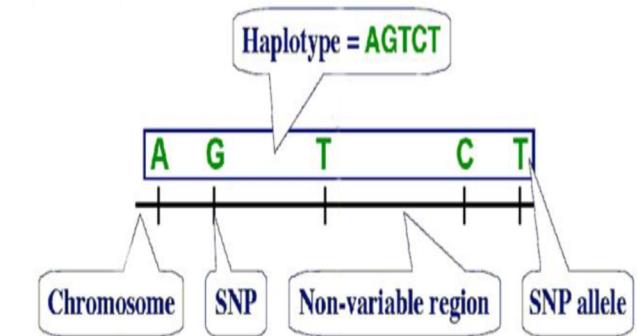
Biological Definitions

- **Rare Variant** is a genetic difference present in <1 % of the alleles in the population.
- **Polymorphism** is a genetic difference present in >1% of the alleles in the population.
- **Penetrance** is the proportion of individuals carrying a particular variant (allele) of a gene (genotype) that also express an associated trait (the phenotype).
- **Expression quantitative trait loci (eQTLs)** are genomic loci that explain all or a fraction of variation in expression levels of mRNAs.

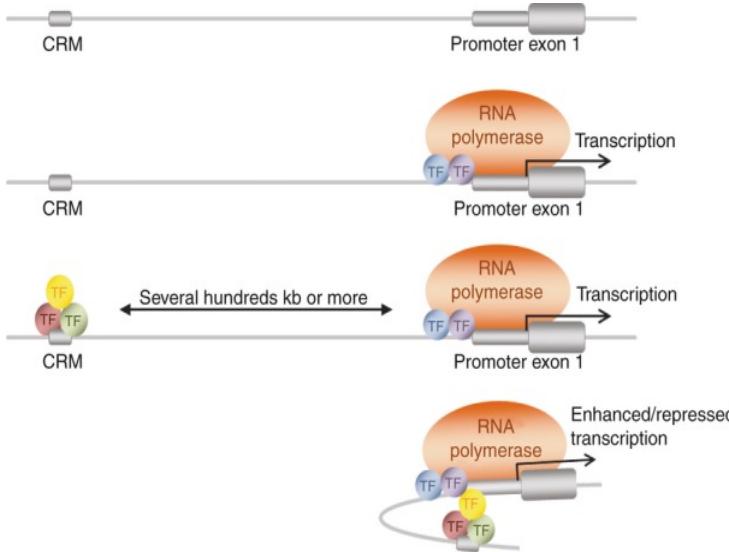
Biological Definitions



- **Single-Nucleotide Polymorphism (SNP)** is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population.

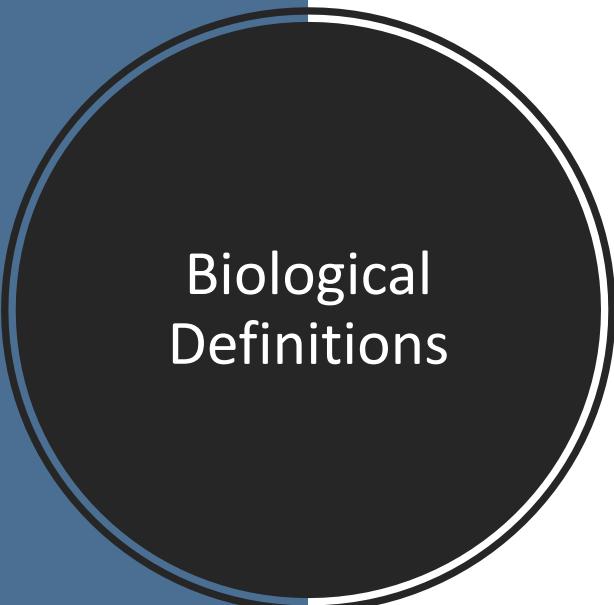


- **Haplotype** is a set of DNA variations, or polymorphisms, that tend to be inherited together.



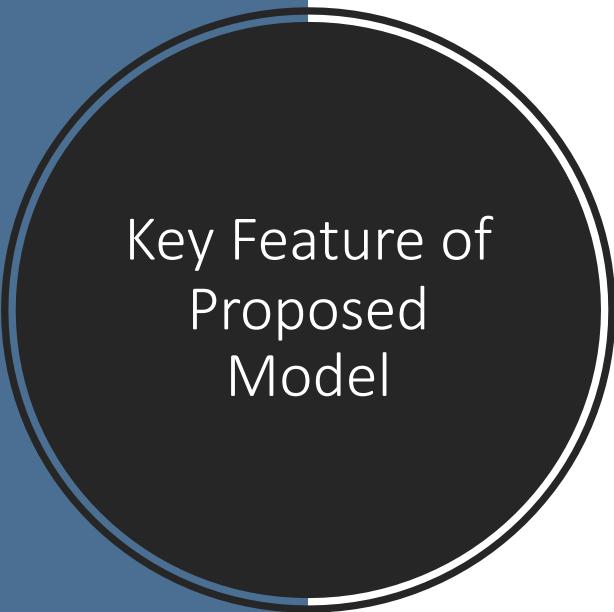
Schematic representation of how a cis-regulatory module can enhance transcription.

- **Cis-regulatory Module** is a stretch of non-coding DNA, 100-1000 DNA bs in length, where a number of transcription factors can bind and regulate expression of nearby genes and regulate their transcription.



Biological Definitions

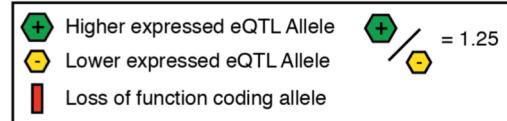
- Variants in non-coding regions that can affect regulation are called **regulatory variants**.
- They can alter regulatory elements, such as enhancer, transcription factor binding sites and DNA methylation regions.
- They can affect disease by involving the distribution of genomic elements that regulate gene expression



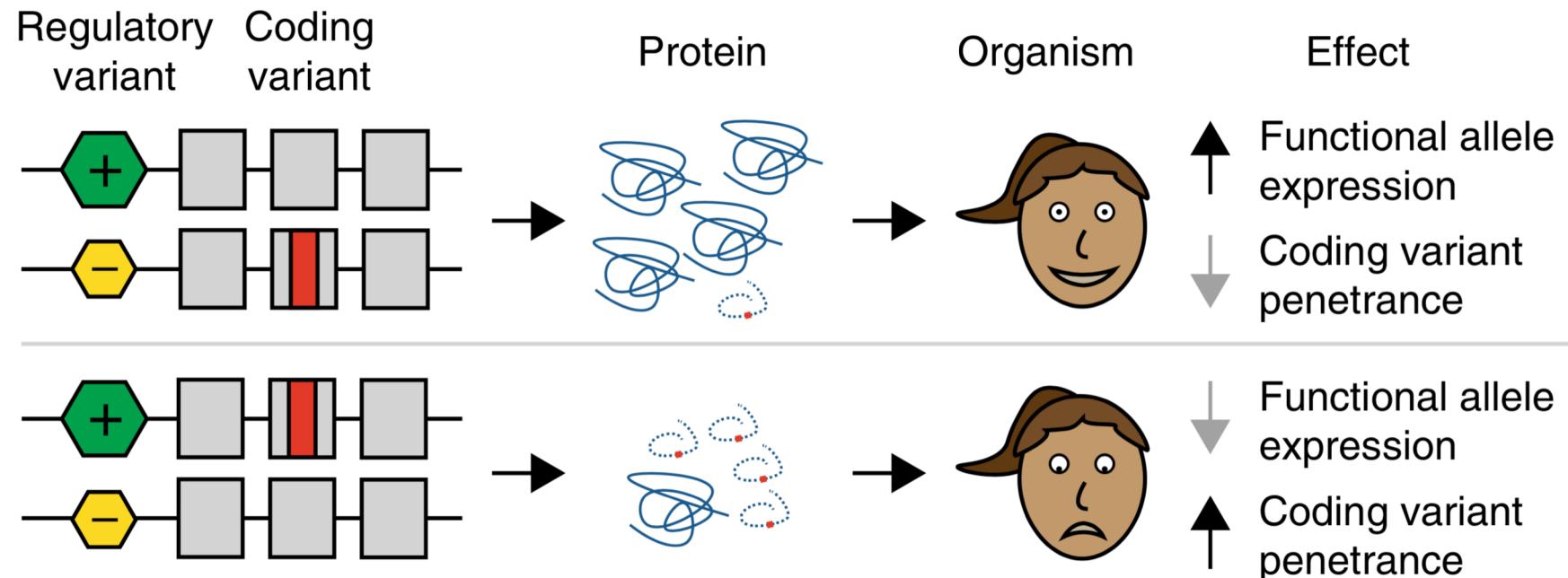
Key Feature of Proposed Model

- **How cis-regulatory variations modify the penetrance of coding variants in their target genes via the joint effects of these variants on the final dosage of functional gene product, depending of their haplotype combinations ?**

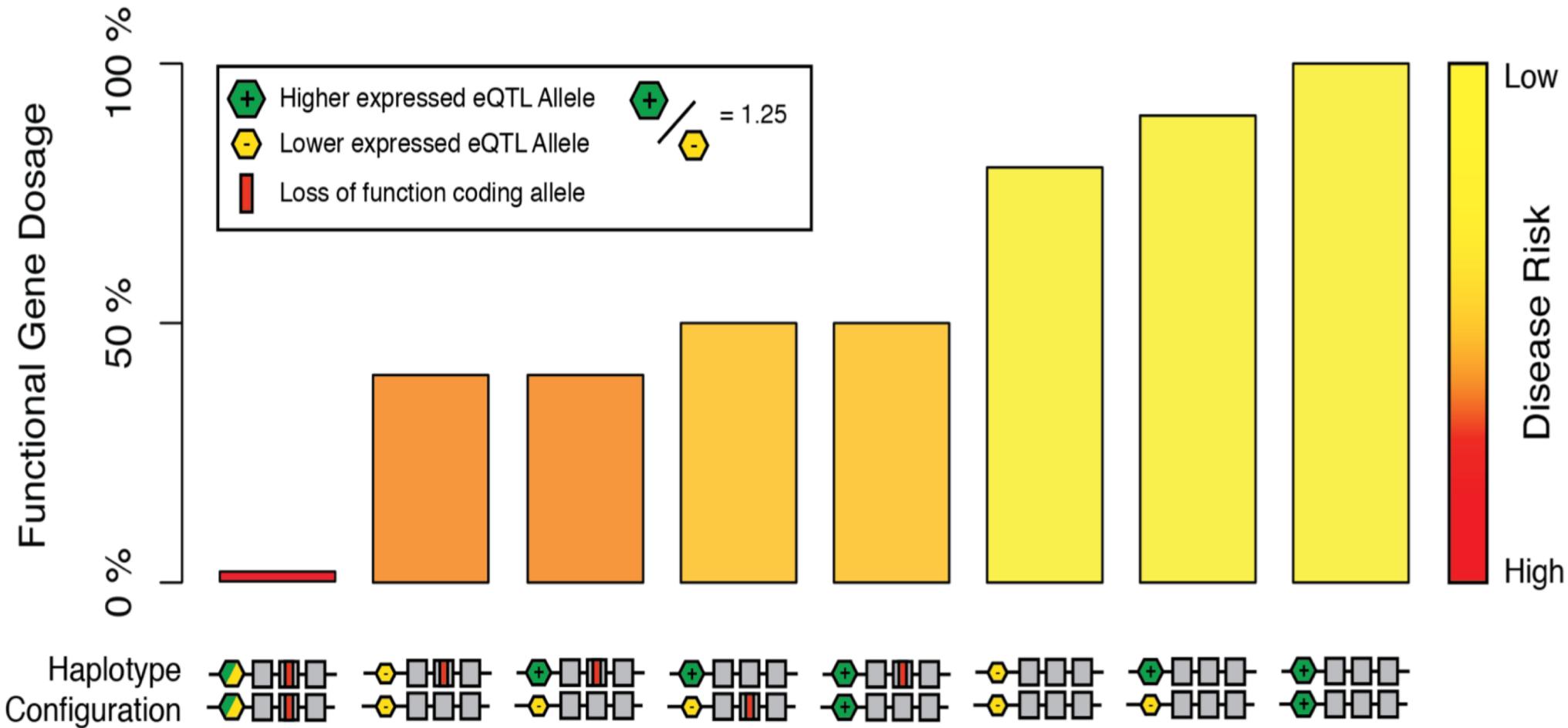
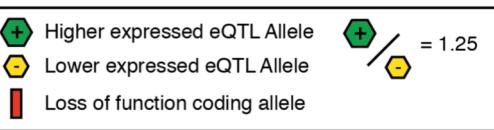
Key Features of Proposed Model

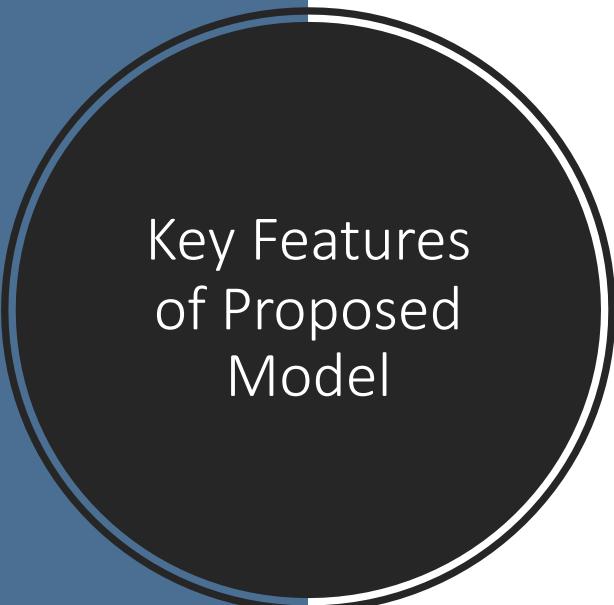


Regulatory Variants as Modifiers of Coding Variant Penetrance.



- An heterozygous individuals for both a regulatory variant and a pathogenic coding variant with two possible haplotype configuration.
- Lower expressed haplotype → Decreases penetrance of the coding variant.
- Higher expressed haplotype → Increases penetrance of the coding variant.





Key Features of Proposed Model

- **Common regulatory variants** typically have such **low effects** on gene dosage that in the absence of coding variants, they do not cause severe disease or substantial reduction of fitness.
- Higher expressed eQTL allele increases gene expression by 1.25x, and disease risk increases non-linearly with decreasing gene dosage.
- This model is plausible for many genes with dosage-sensitivity.



Hypothesis

- ❖ Purifying selection has depleted penetrance increasing haplotype combinations of regulatory and coding variants in the general population.



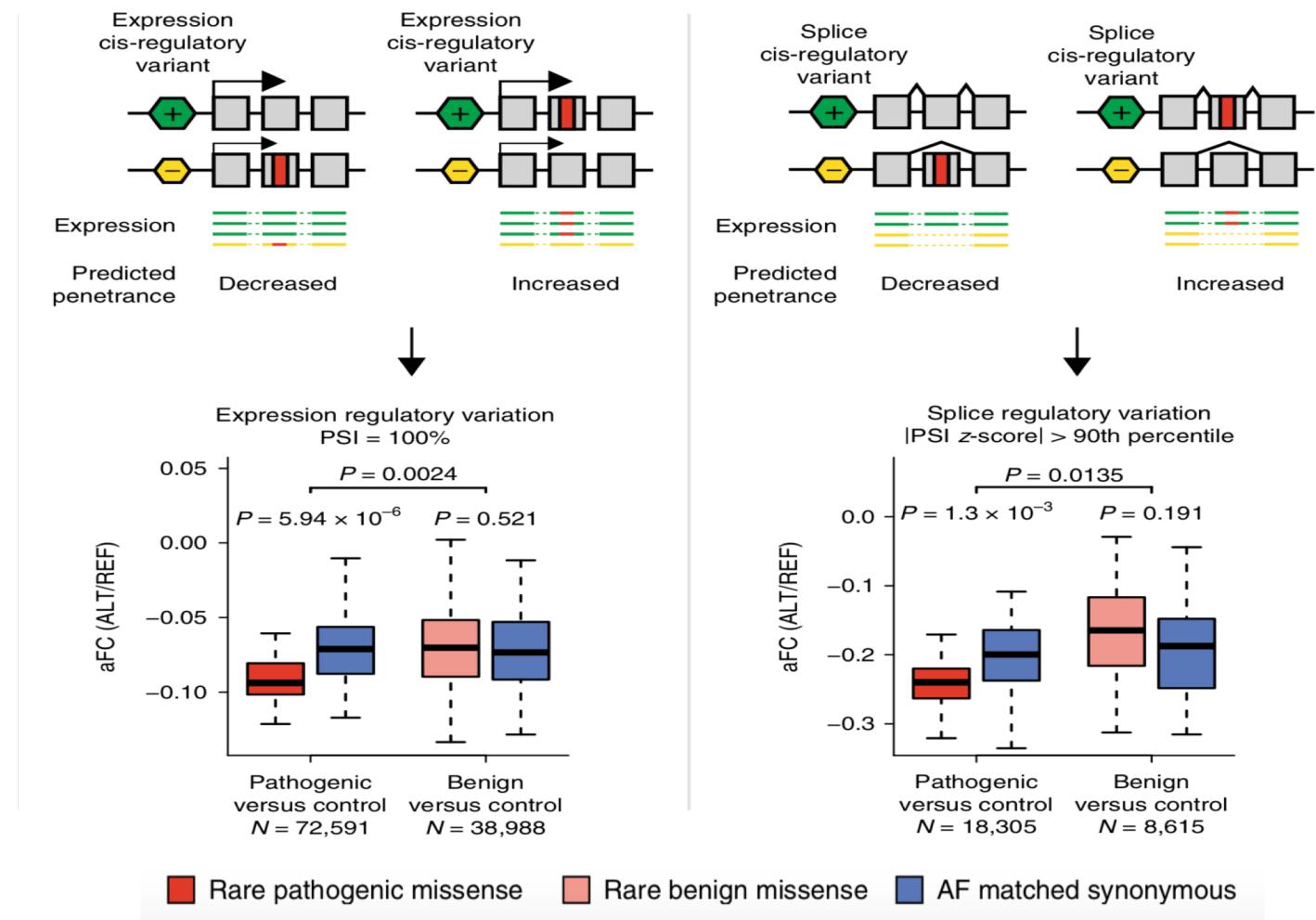
Validation and Method

- **Data :**
 - Exome seq. and SNP array data from GTEx – represents general population with lack of individuals with severe genetic disease.
 - Use CADD score to predict the pathogenicity of variants.
 - Missense variants with CADD > 15 → potentially pathogenic
 - Synonymous variants with CADD < 15 → controls
 - Rare variants → MAF² < 1%
- **Analysis:**
 - Measure the regulatory haplotype of coding variants using allelic expression (AE) data generated from whole exome seq.
 - Calculate the expression of coding variant minor alleles
 - Compare the expression of missense variants to alleles frequency matched synonymous controls using a paired Wilcoxon signed rank test.

1. Combined Annotation Dependent Depletion (CADD) is a tool for scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome.
2. MAF: minor allele frequency

- Reduction of the allelic expression for rare pathogenic missense variants by 0.7% compare to synonymous control.
- Pathogenic variants are less likely to accumulate on haplotypes where the corresponding exon is more likely to be included in transcripts.

- Pathogenic coding variants less frequently exits in high-penetrance regulatory haplotype combinations.

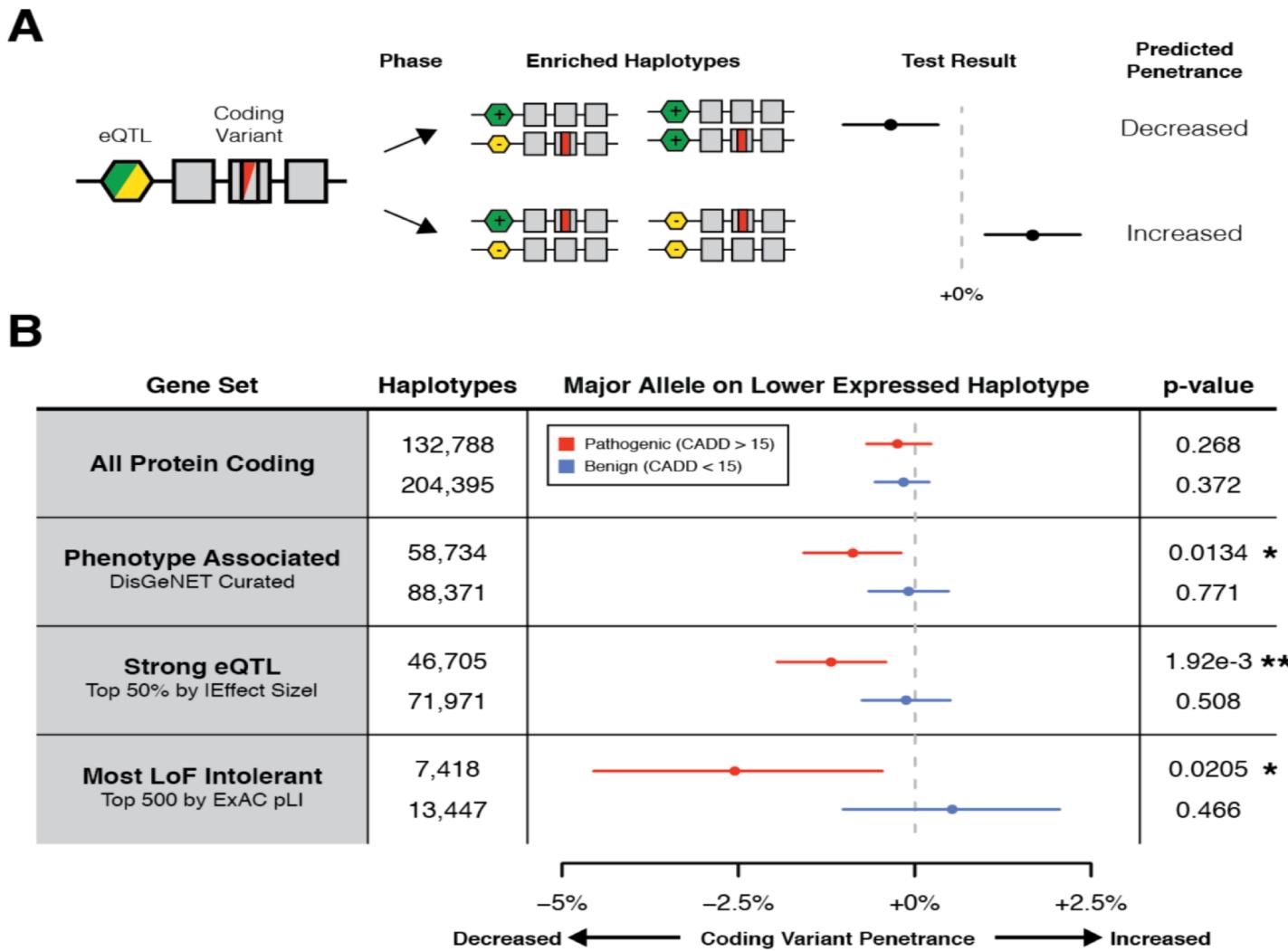


Distribution of Coding Variants on cis-eQTL

- ❖ Check whether the distribution of coding variant on cis-eQTL haplotypes in GTEx shows signs of selection against increased penetrance?!
- **Data:**
 - Haplotype phased genetic data from GTEx – Genome seq.
- **Results:**
 - Analysis of all protein coding → No evidence
 - Set of the genes with known phenotypic association → 0.88 % reduction of rare potentially pathogenic variants
 - Genes with strong eQTL → Significant reduction
 - The strongest effect at the most loss-of-function intolerant genes as measured by ExAC pLI .

Distribution of Coding Variants on cis-eQTL

eQTL haplotype configurations that are predicted to increase pathogenic coding variant penetrance are depleted in the genomes of GTEx individuals



- ❖ In the general population, pathogenic coding variants less frequently exits in high-penetrance regulatory haplotype combination.



Disease Risk and Regulatory Modifiers

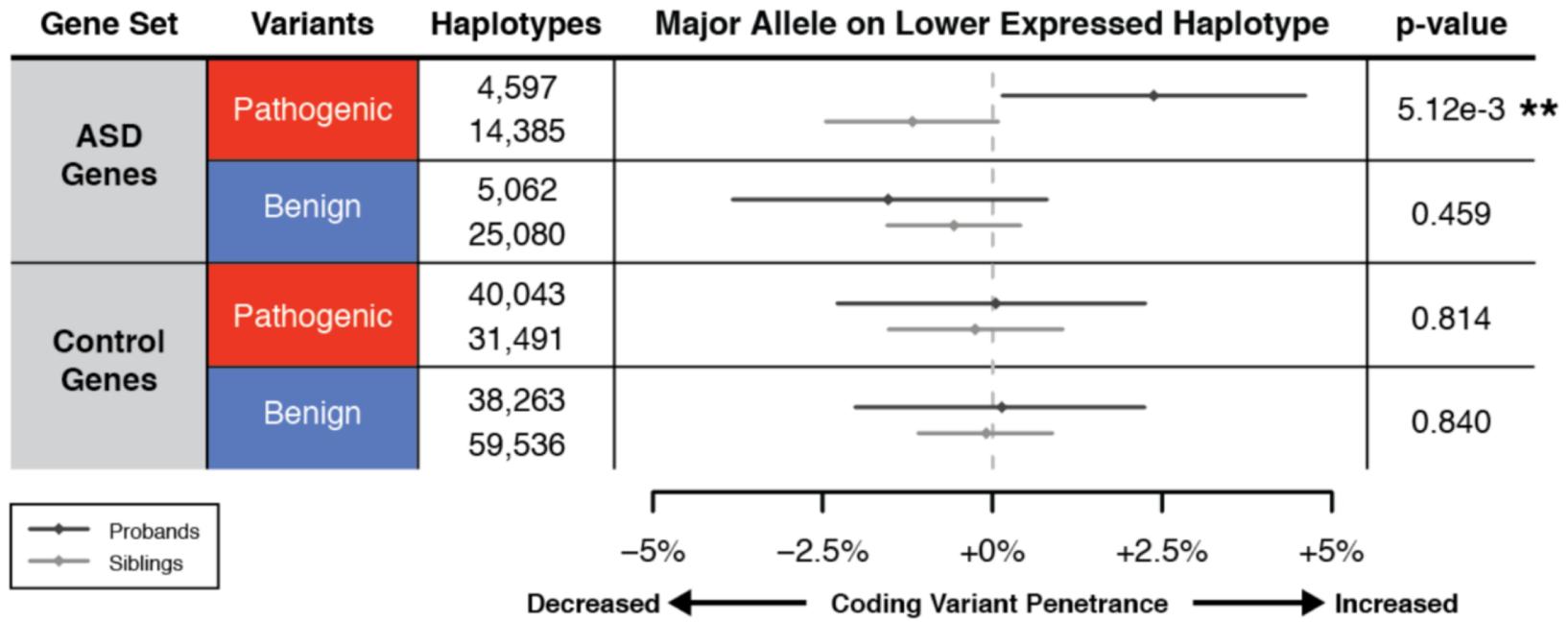
- ❖ **Genetic regulatory modifiers of pathogenic coding variants affect disease risk.**

Analysis of Autism Spectrum Disorders (ASD)

- **Data:** Simons Simplex Collection with genetic data from 2600 simplex families – one child with autism, parents and any unaffected siblings.
- **Filtering:**
 - Combining exome seq. with SNP array data.
 - Common regulatory variants annotated based on the most significant eQTL variant in each gene in GTEx to produce joint regulatory and coding haplotypes for 2304 ASD (cases) and 1712 controls.
- **Analysis:**
 - Comparing the haplotype configuration of rare pathogenic variants between case and controls.
 - To enrich for disease causing variants they analysis only the variants in the cases that are not in unaffected.
- **Observation:**
 - At 493 ASD implicated genes, probands had a significant increase of major coding alleles of rare potentially pathogenic variants found on lower expressed haplotypes as compared to unaffected siblings.
 - No affect for rare potentially pathogenic variants at control gene matched for the rate of coding variant occurrence.

ASD eQTL Analysis

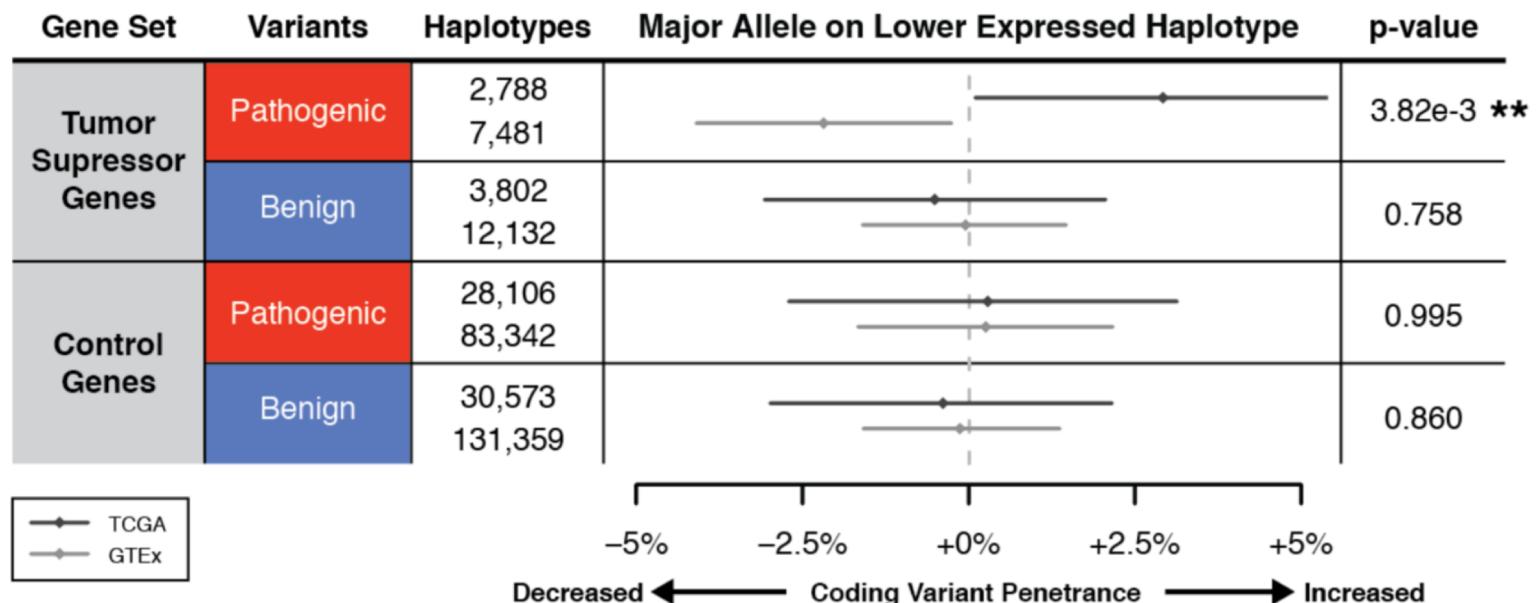
eQTL haplotype configurations that are predicted to increase pathogenic coding variant penetrance are enriched in individuals with ASD vs. control.



- Rare (MAF < 1%) pathogenic (CADD > 15, including missense, splice, and stop gained, and frame-shift)
- Rare benign (CADD < 15, including synonymous and missense) variants

Analysis of Cancer Risk

- **Data:** Use Cancer Genome Altas (TCGA) analysis the role of regulatory modifiers of penetrance in germline cancer risk.
- **Analysis:**
 - The haplotypes of coding variants and common regulatory variants annotated based on the most significant eQTL variant for each **gene in GTEx samples**.
 - Compare with the haplotype configurations of rare potentially **pathogenic variants observed in TCGA individuals**.
- **Observation:** At tumor suppressor genes, TCGA had a significant increase of major coding alleles of rare potentially pathogenic variants found on the lower expressed haplotypes compare to GTEx controls.





Experimental Demonstration

- Experimental study design and use CRISPR/Cas9 to edit a Mendelian missense SNP in *FLCN* (rs199643834).
 - **rs1708629** consider as the most likely causal regulatory variant of the FLCN eQTL.
 - Recovering monoclonal cell lines, genotyping them by targeted DNA-seq and performing targeted RNA-seq of the edited SNP.
 - **Obtaining:**
 - Four colons with a single copy of the Mendelian variant on lower expressed haplotype (snpLOW)
 - Three colons with a single copy on the higher expresses haplotype (snpHIGH)
 - Two monoallelic colons with 3 copies of the alternatives allele
 - Four with only the reference alleles (WT)
 - Quantifying the differential expression of genes in low and high edited SNP expression clones separately.
- ❖ **Result:** Colons with higher expression of the SNP shows a significantly stronger differential expression of both downregulated and upregulated genes compare to lower SNP expression clones.



Main Results

- ❖ eQTL can modify the penetrance of a disease-causing coding variant.
- ❖ Purifying selection is acting on joint regulatory and coding variants haplotypes.
- ❖ Combination of an individual's regulatory and coding variant genotypes has an affect on phenotype.
- ❖ Analyses of autism and cancer cohorts provide direct evidence that regulatory modifiers of coding variants contribute to disease risk.
- ❖ The strength of modified penetrance depends on the functional importance and dosage-sensitivity of the gene, effect size of the regulatory variants that affect expression or splicing, and the type of coding variant



Summary

- **Data:** genomic and genetic data from the Genotype-Tissue Expression Project (GTEx)
- **Observation:**
 - **In general population:** purifying selection has depleted haplotype combinations predicted to increase pathogenic coding variant penetrance.
 - **Cancer and autism:** an enrichment of penetrance increasing haplotype configurations for pathogenic variants in disease-implicated genes, providing evidence that regulatory haplotype configuration of coding variants affects disease risk.
- **Experiments:** A regulatory variant can modify the penetrance of a coding variant by introducing a Mendelian SNP using the CRISPR/Cas9 on distinct expression haplotypes and using the transcriptome as a phenotypic readout.
- **Results:** Joint regulatory and coding variant effects are an important part of the genetic architecture of human traits and contribute to modified penetrance of disease-causing variants.

References

- *S. E. Castel, A. Cervera, P. Mohammadi, F. Aguet, F. Reverter, A. Wplman, R. Guigo, I. Iossifov, A. Vasileva & T. Lappalainen.* Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nature Genetics* volume 50, pages1327–1334 (2018)
- *Cart Tony Fakhry, CAUSAL REASONING AND MACHINE LEARNING MODELS FOR CELLULAR REGULATORY MECHANISMS. PhD Dissertation May 2018*
- *P.M. Visscher, N.R. Wray, Q. Zhang, P. Sklar, M.I. McCarthy, M.A. Brown, J. Yang.* 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*, 101 (2017), pp. 5-22
- *E. Rojano, P. Seona, J. A G Ranea, J. R Rerkins.* Regulatory variants: from detection to predicting impact. *08 June 2018 Briefings in Bioinformatics*, bby039, <https://doi.org/10.1093/bib/bby039>

Ask for...

