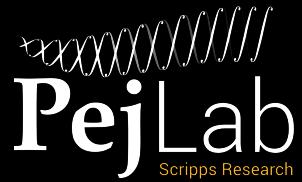


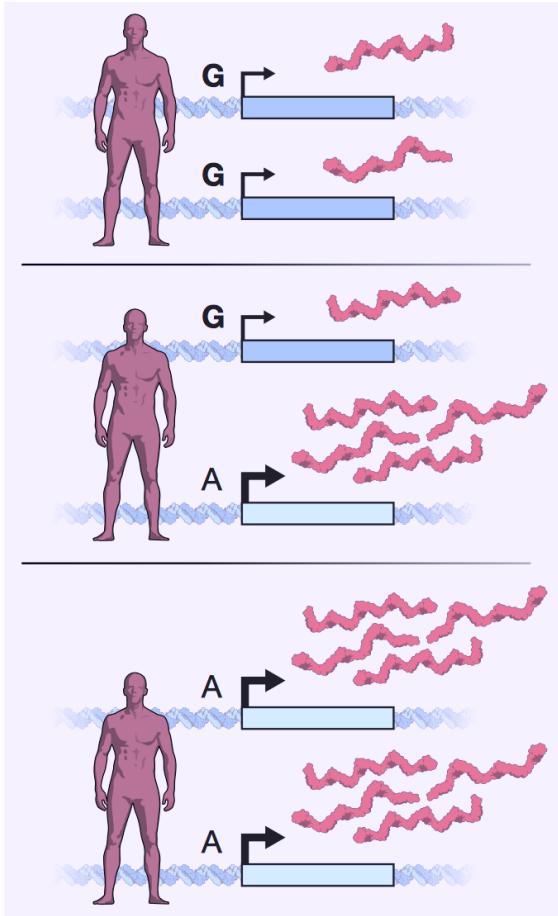
# Estimating the effect size for independent eQTLs



Nava Ehsan  
Pejman Mohammadi's lab

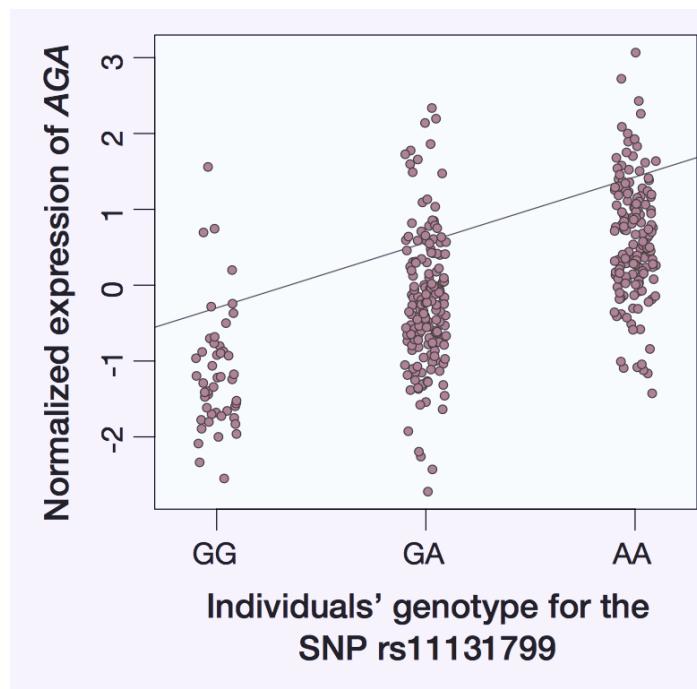
July 22, 2020

# eQTL

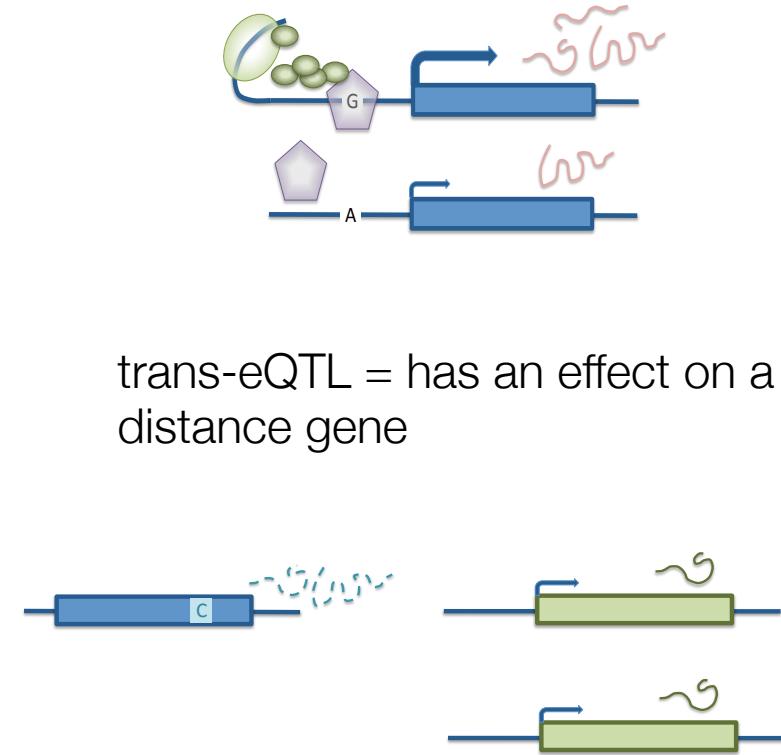


eQTL = expression Quantitative Trait Loci

An association study, where the phenotype is gene expression level

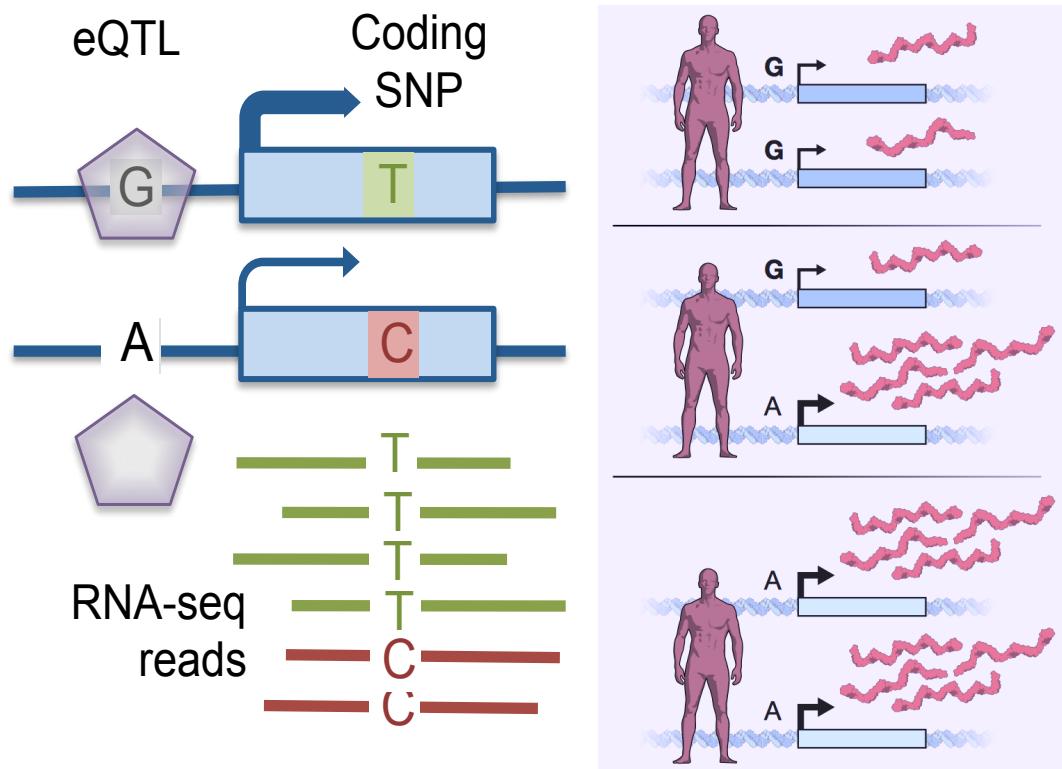


Cis-eQTL = Proximal allelic effects in 1Mb window around the gene



trans-eQTL = has an effect on a distance gene

# ASE

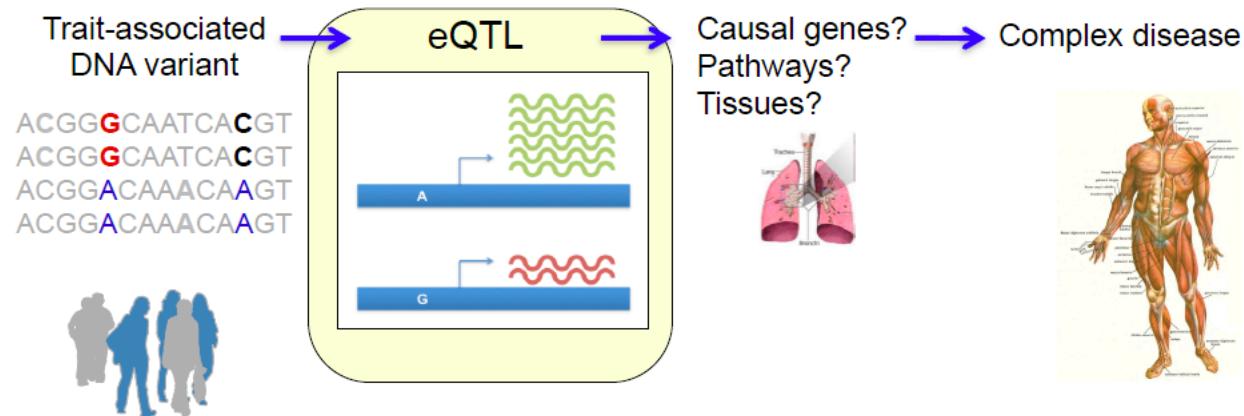


ASE = Allele-Specific Expression

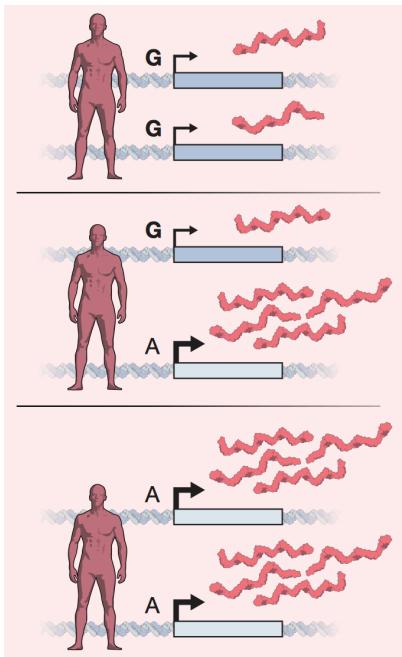
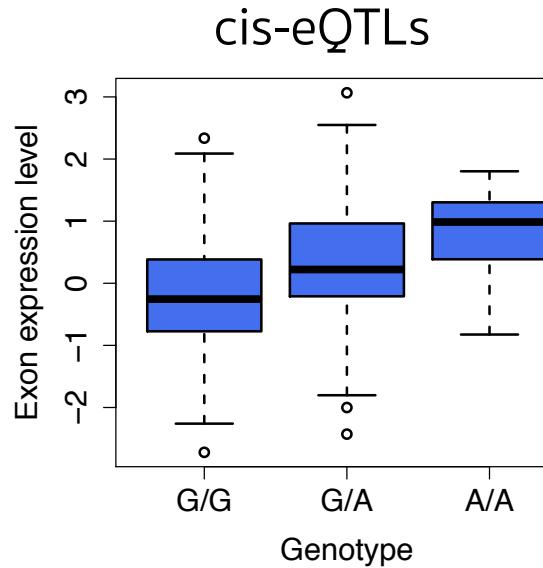
if an individual is heterozygous for a regulatory variant, then the two copies of the gene will show different level of expression

# Why do we care

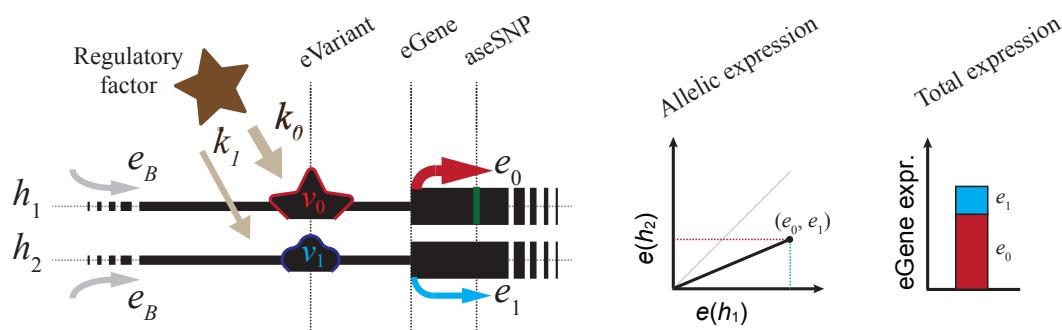
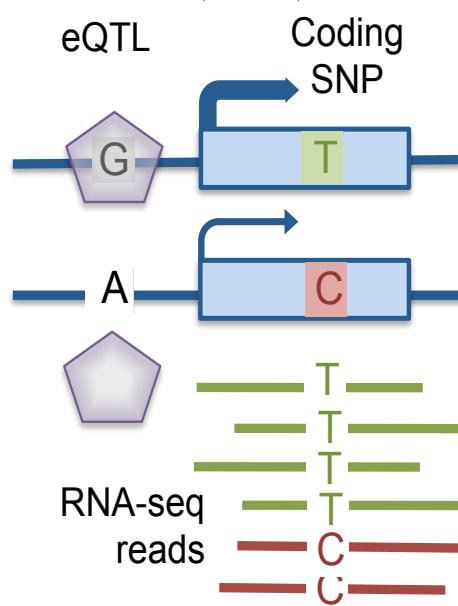
**Variants in noncoding regions  
may be affecting disease via their effect on gene expression**



# Two complementary ways to study regulatory vars



Allele-specific expression (ASE)



# The GTEx Consortium atlas of genetic regulatory effects across human tissues

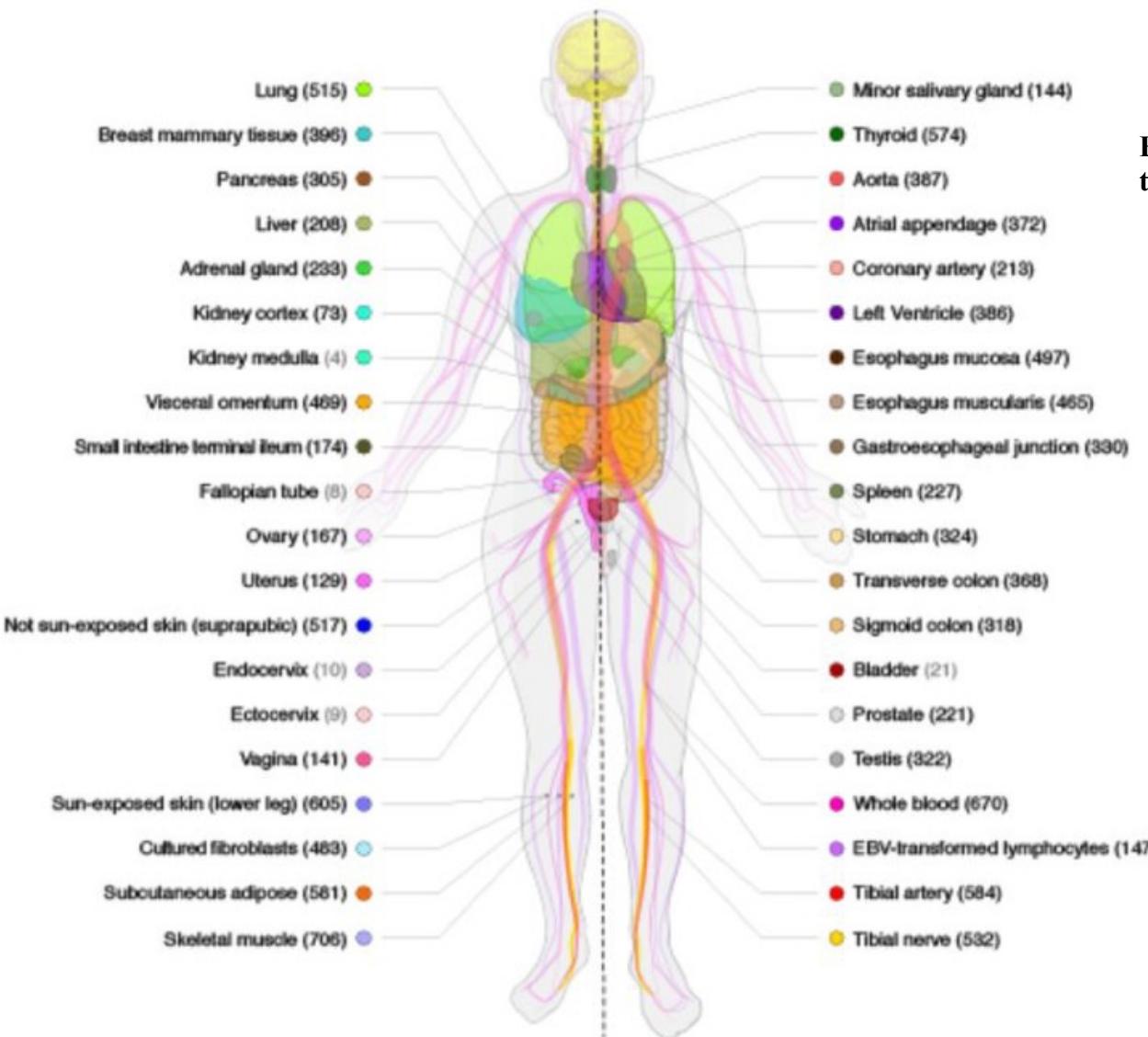
QTL analysis:  
15,201 samples,  
838 donors  
49 tissues,  
samples from at least  
70 individuals

GTEx Consortium  
*BioRxiv*, 2019

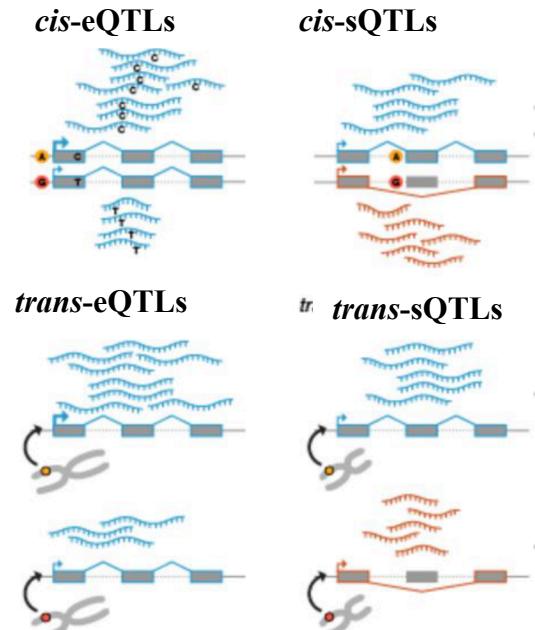
Ferraro et al.  
*BioRxiv*, 2019

Castel et al.  
*BioRxiv*, 2019

Olivia et al.  
*BioRxiv*, 2019

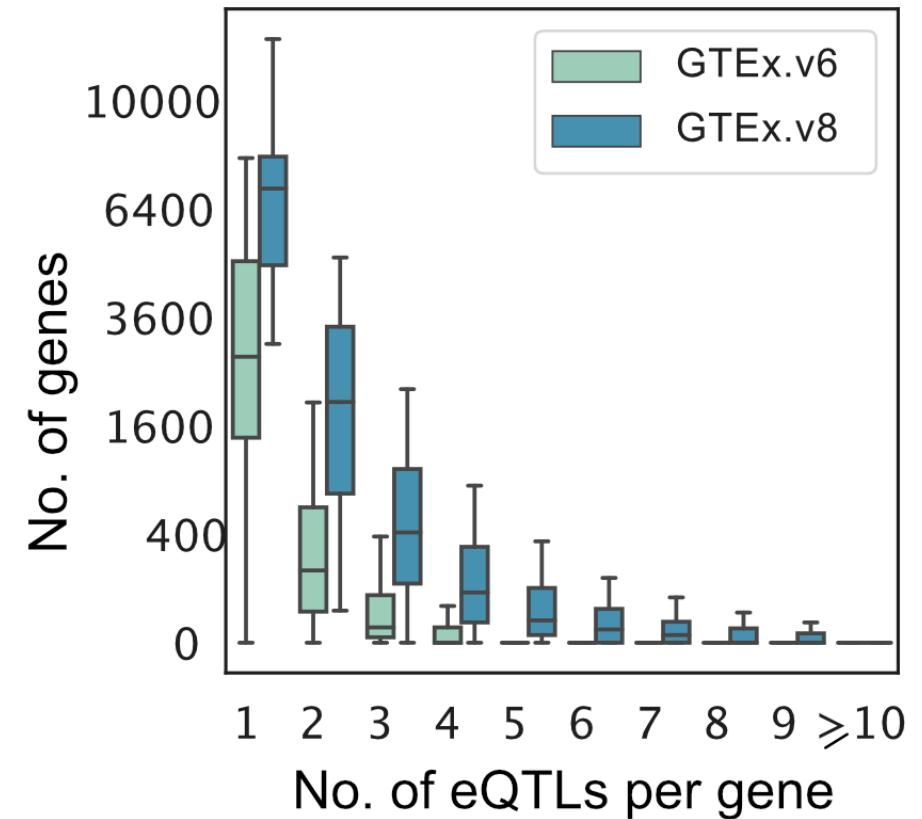


Expression quantitative trait loci (eQTLs)      Splicing quantitative trait loci (sQTLs)



# Many genes have more than 1 eQTL

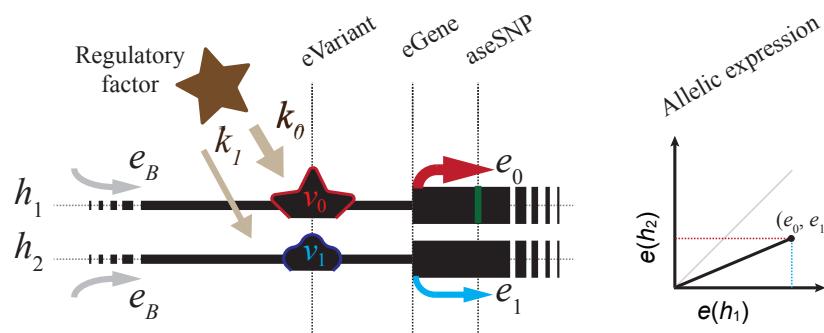
- About 70% of the genes have more than one eQTL in at least one tissue
- Average of the genes associated with more than one eQTLs in each tissue :
  - GTEx V6 : 10%
  - GTEx V8 : 26%



# Extension to multi-eQTL model

**Allelic Fold Change (aFC): Relative strength of the Alt. allele and Ref. allele**

## Single-variant

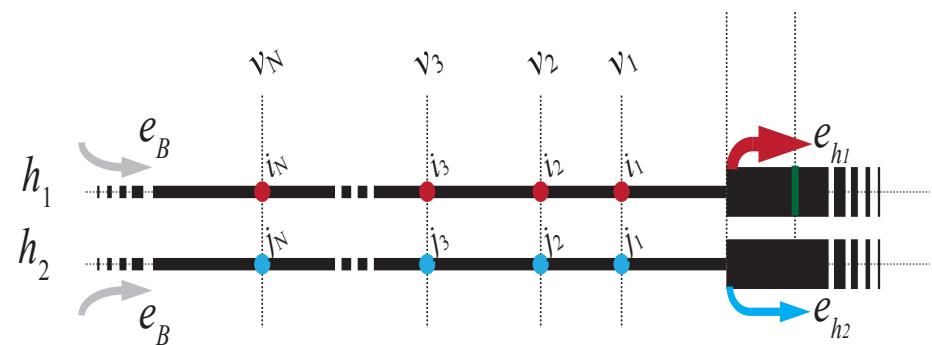


$$e_i = k_i \cdot e_B$$

$$\delta_{1,0} = k_1/k_0$$

Fold change due to  
a single eQTL

## Multiple-variants

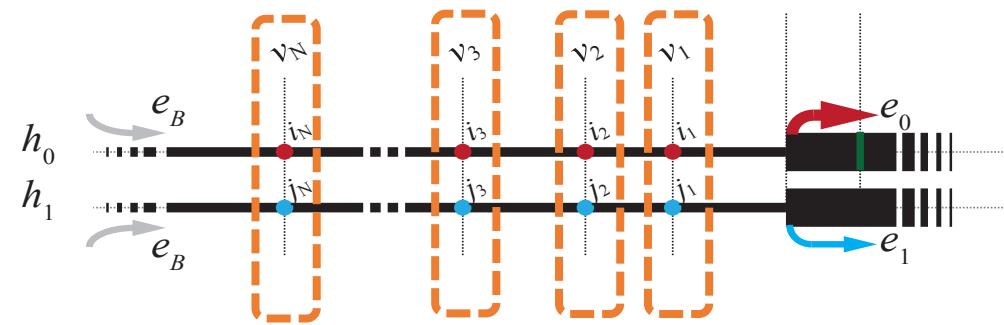


$$\delta_{< i_1, \dots, i_N > < j_1, \dots, j_N >} = \prod_{n=1}^N \delta_{i_n, j_n}^{v_n}$$

Fold change due to  
multiple eQTLs

# Cal. The net effect of multiple eQTLs using a single eQTL model

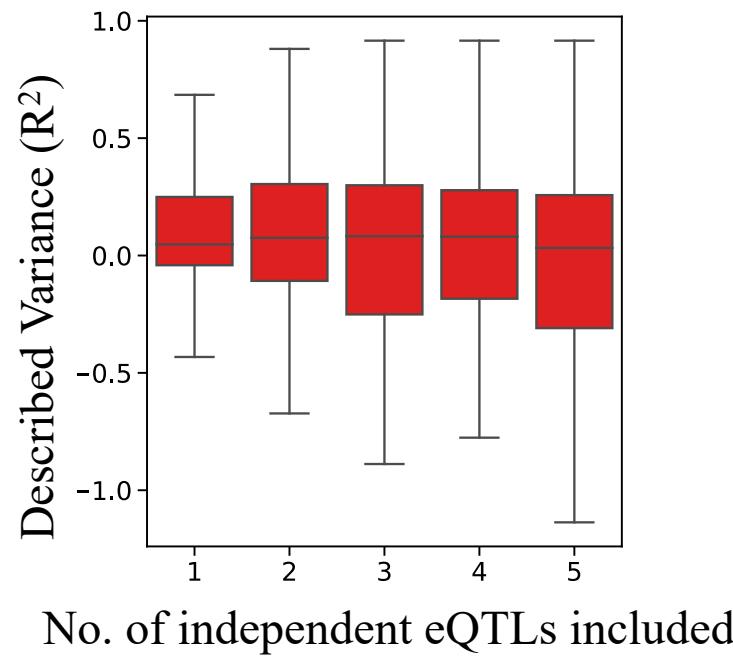
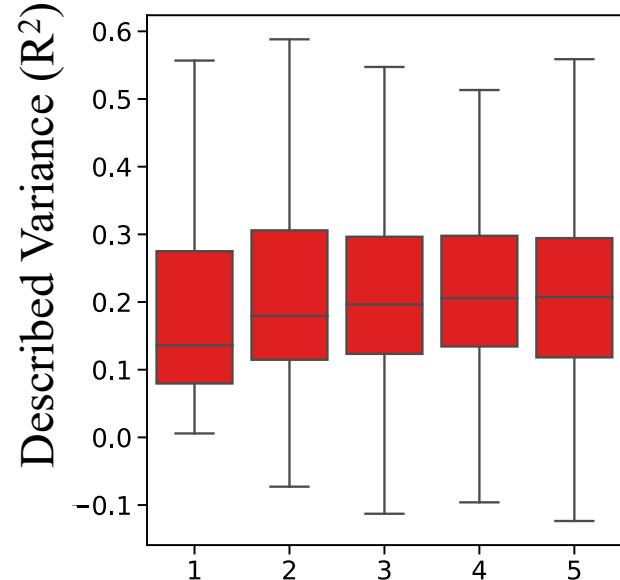
- Current eQTL effect sizes in GTEx are derived from a univariate model



$$\log_2 \exp = \log_2(2^{s.h0} + 2^{s.h1}) + \log_2(c_0) + \epsilon$$

# Need for a multi-variant effect size model

- We evaluate the performance of GTEx v8 published effect sizes in predicting expr. and ASE
  - These are derived by a single variant model
- We find including more eQTLs does not increase the prediction accuracy

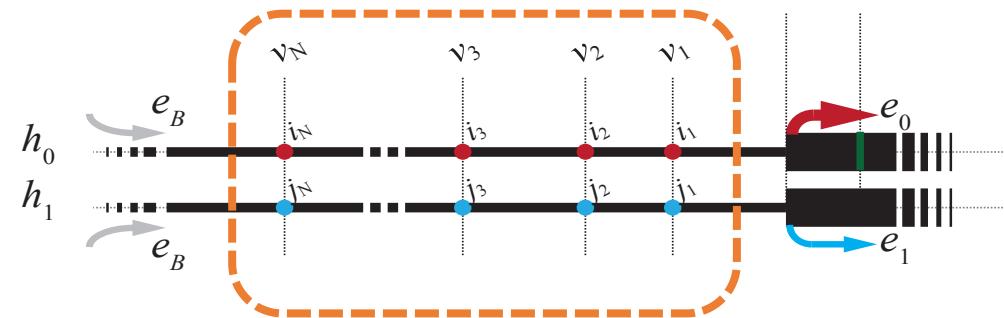




Bence Kotis

# Cal. effect size for multiple eQTLs

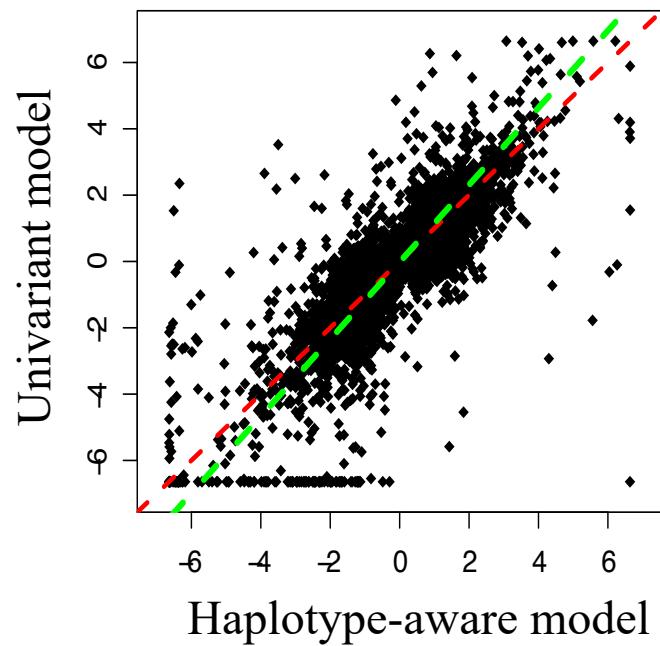
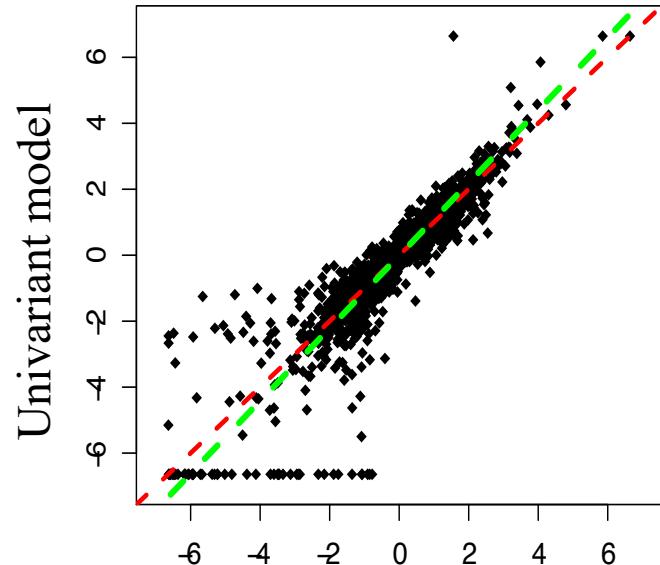
- We used phased genomes
- We included all known variants within each gene to calculate aFCs simultaneously in a haplotype-aware model



$$\log_2 \exp = \log_2(2^{s.h0} + 2^{s.h1}) + \log_2(c_0) + \epsilon$$

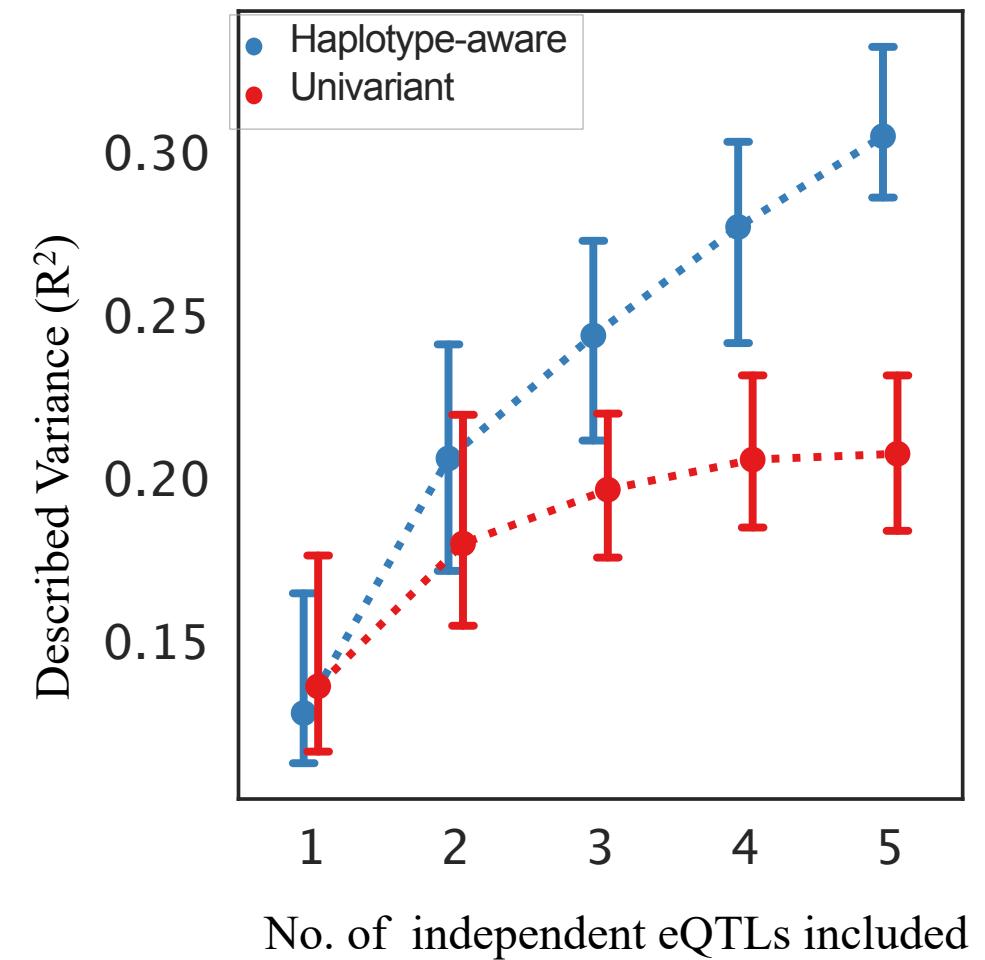
# The effect size estimates are generally correlated

- The effect sizes estimated from haplotype-aware model were correlated to the current effect sizes from GTEx v8 using univariate model



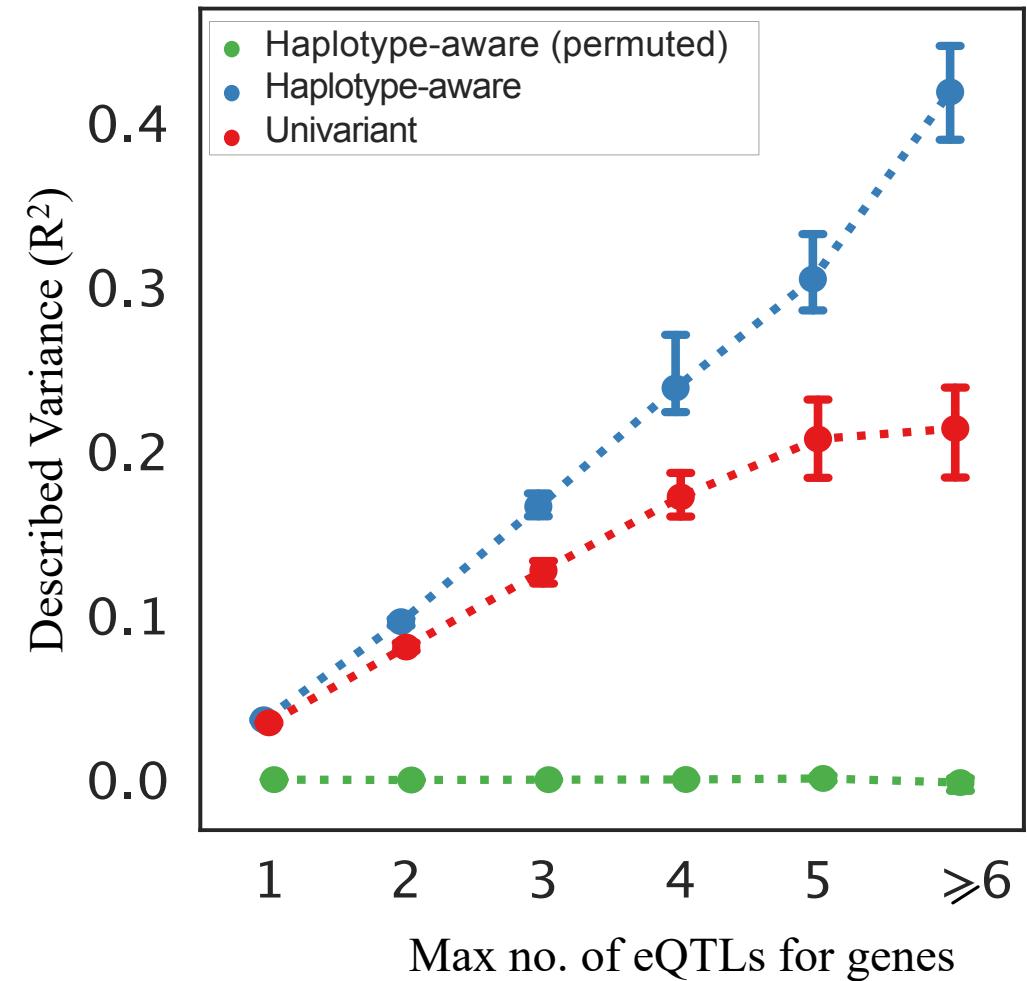
# Predicting gene expression

- The effect size estimates derived by haplotype-aware model yield progressively more accurate predictions as more regulatory variants are considered.



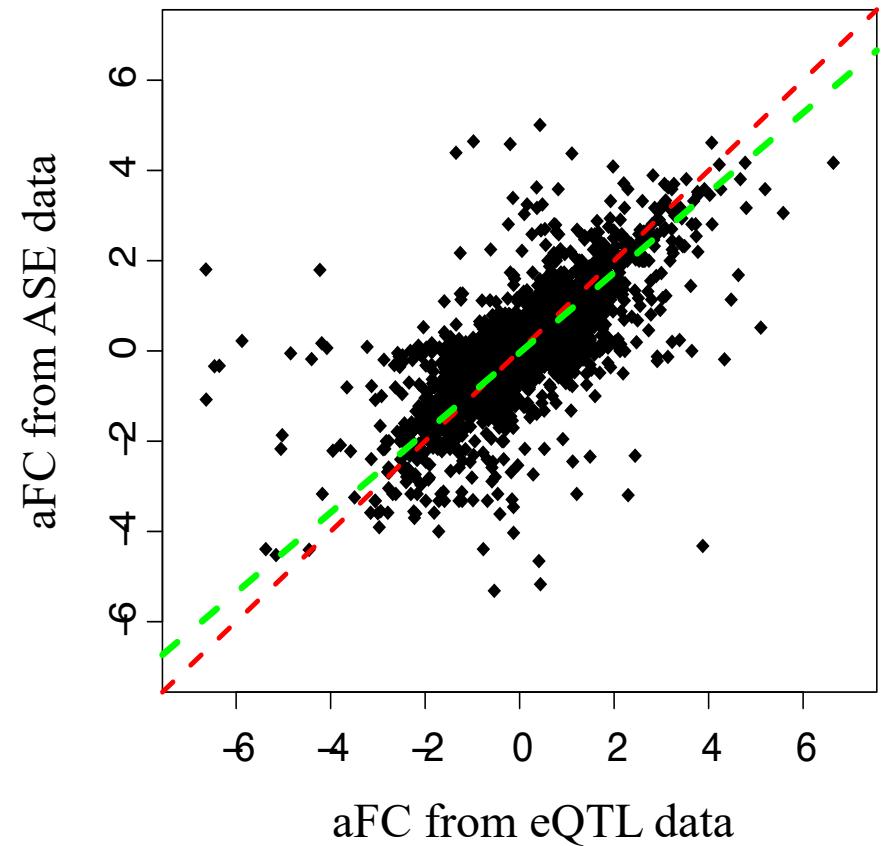
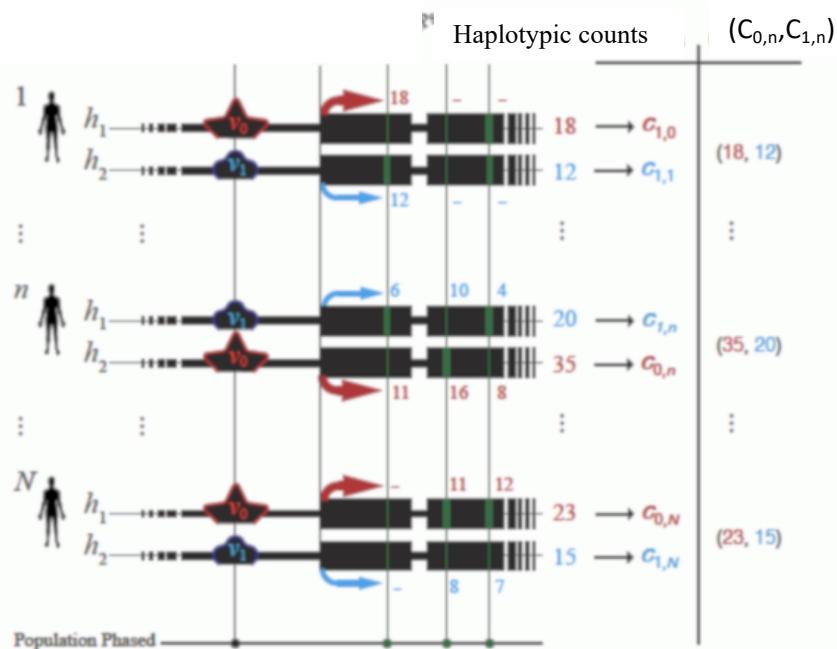
# Predicting gene expression

- Haplotype-aware model improves gene expression prediction accuracy significantly.
- The amount of variation that could be explained by all the known eQTLs, increases with increase in number of variants.



# eQTL data vs. ASE data at population level

- The estimated effect size of an eQTL is highly consistent with observed allelic imbalance among individuals heterozygous for it at the population-level



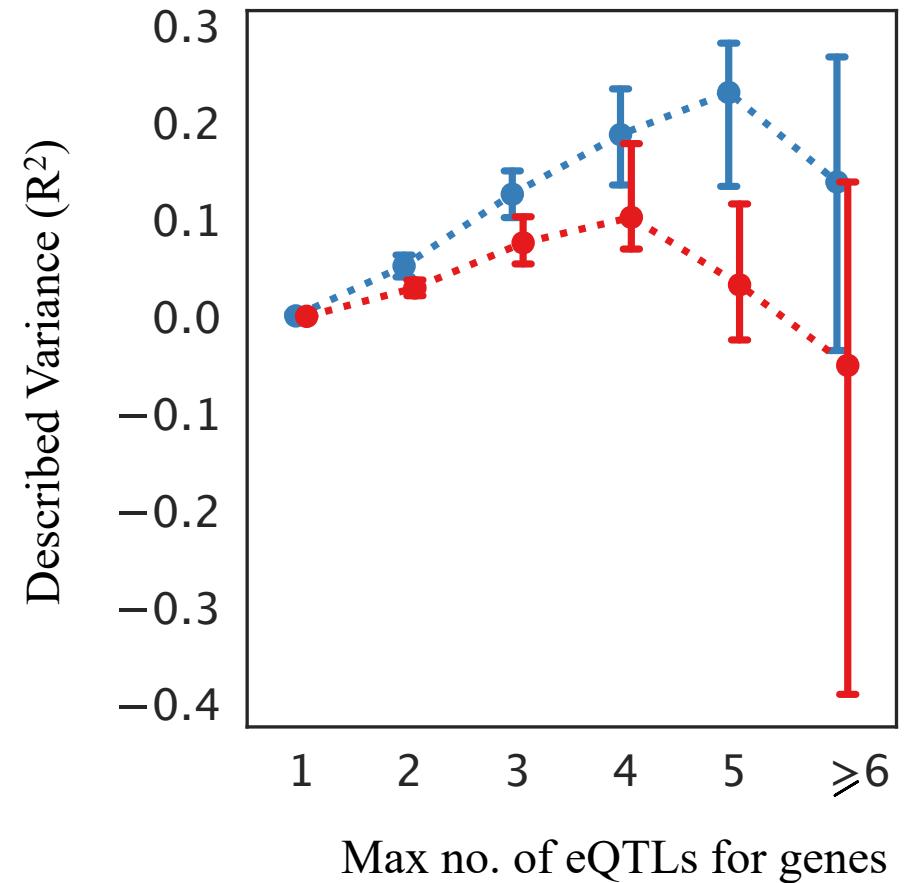
# eQTL data vs. ASE data at Individual level

## The good news:

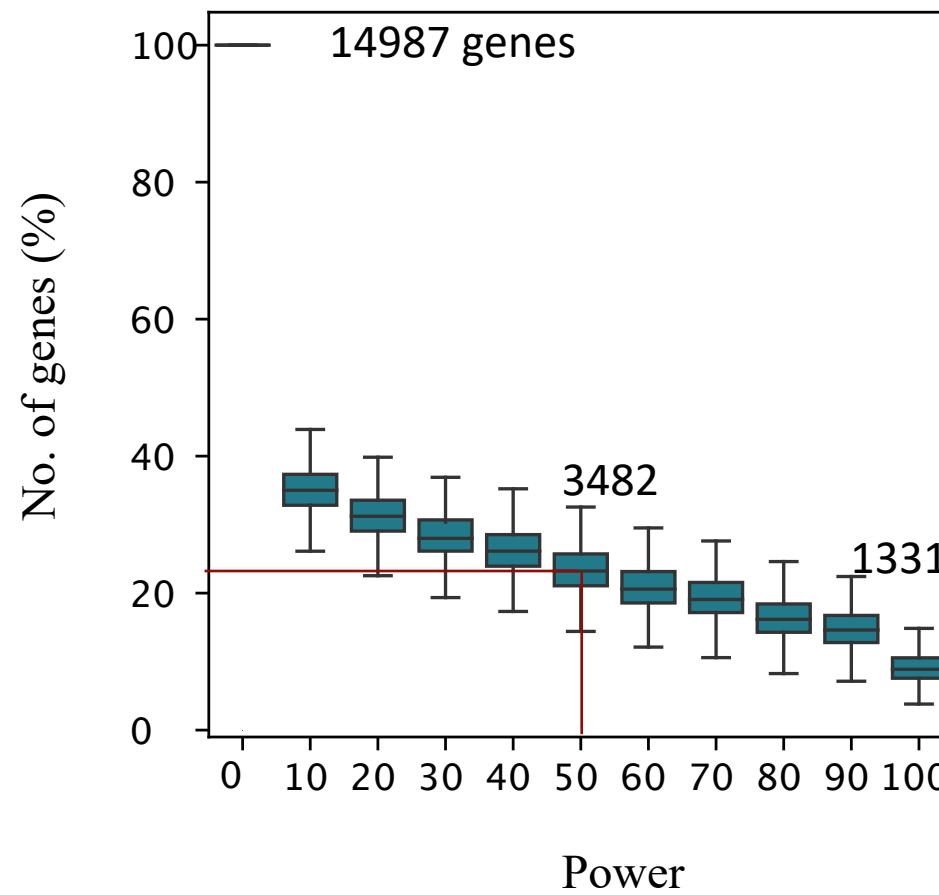
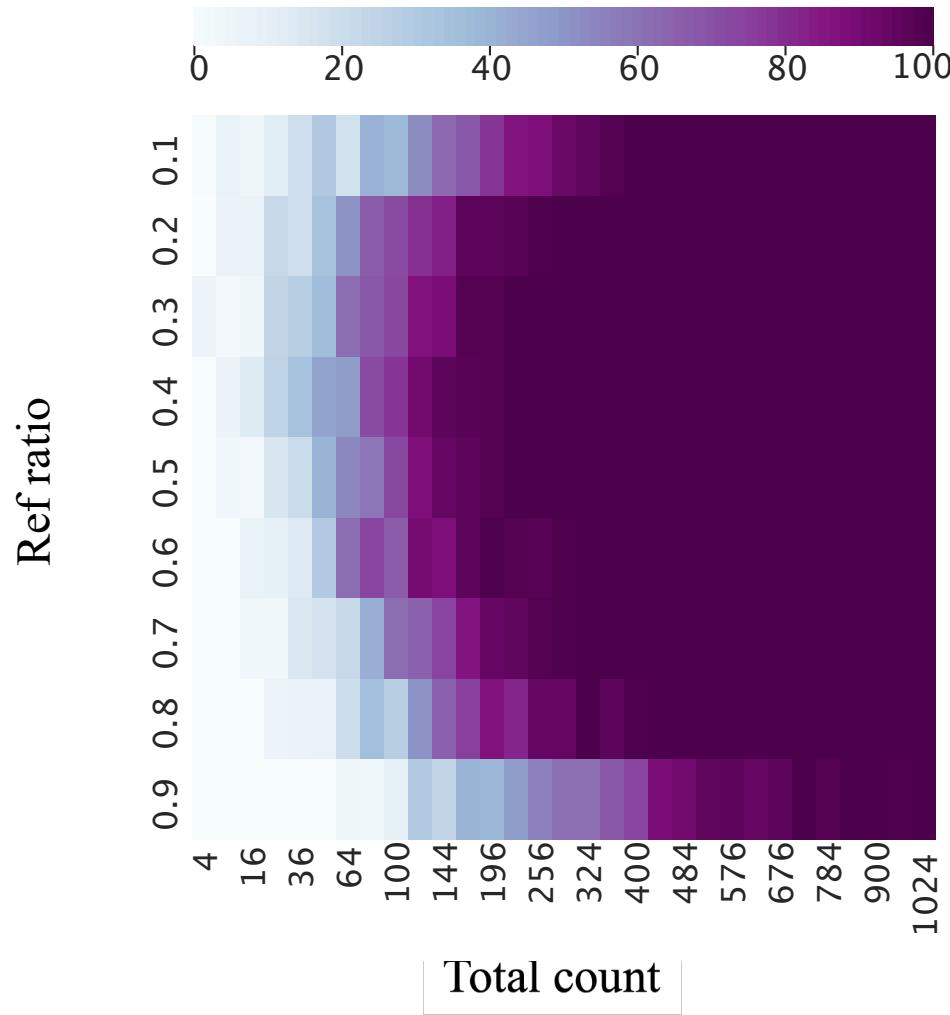
- The haplotype-aware model outperforms the univariate model

## The bad news:

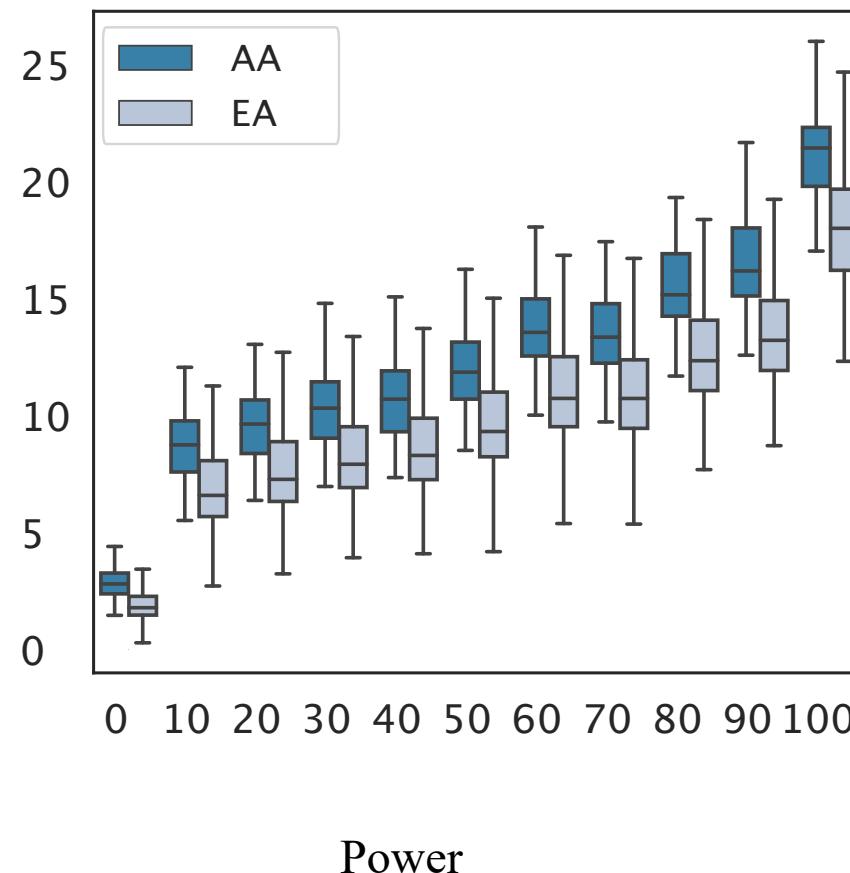
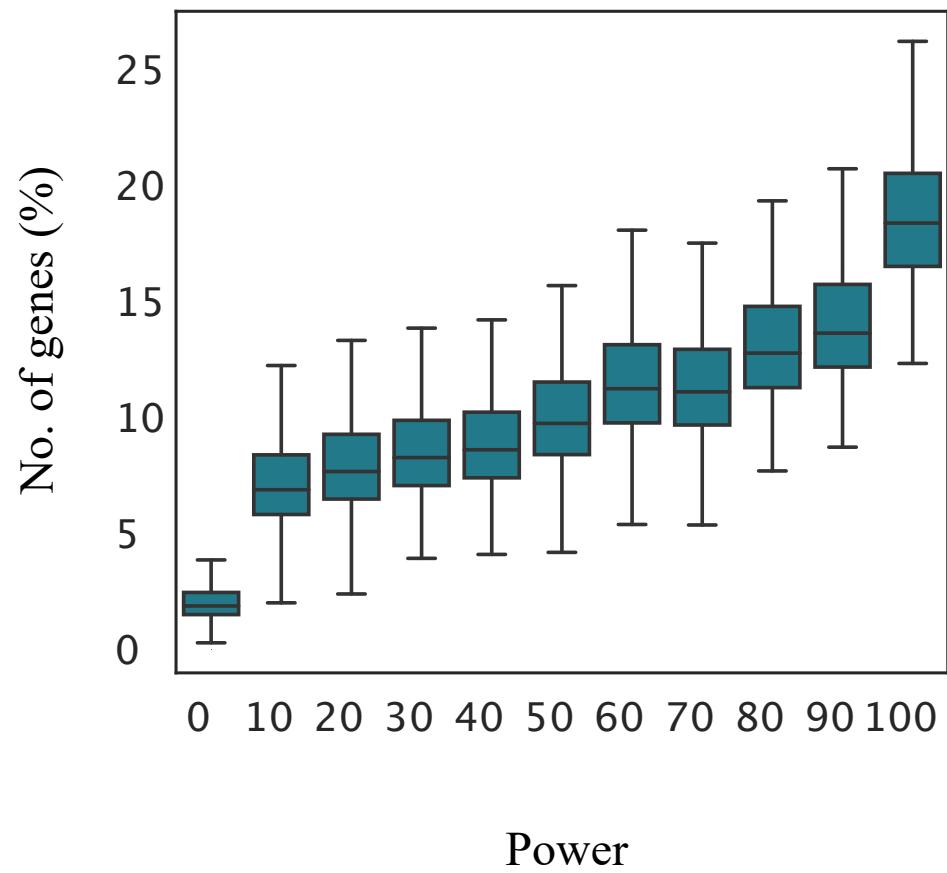
- Known eQTLs describe a small fraction of ASE variation at individual level



For what fraction of the cases the current eQTL data  
“fully” describes ASE signal?

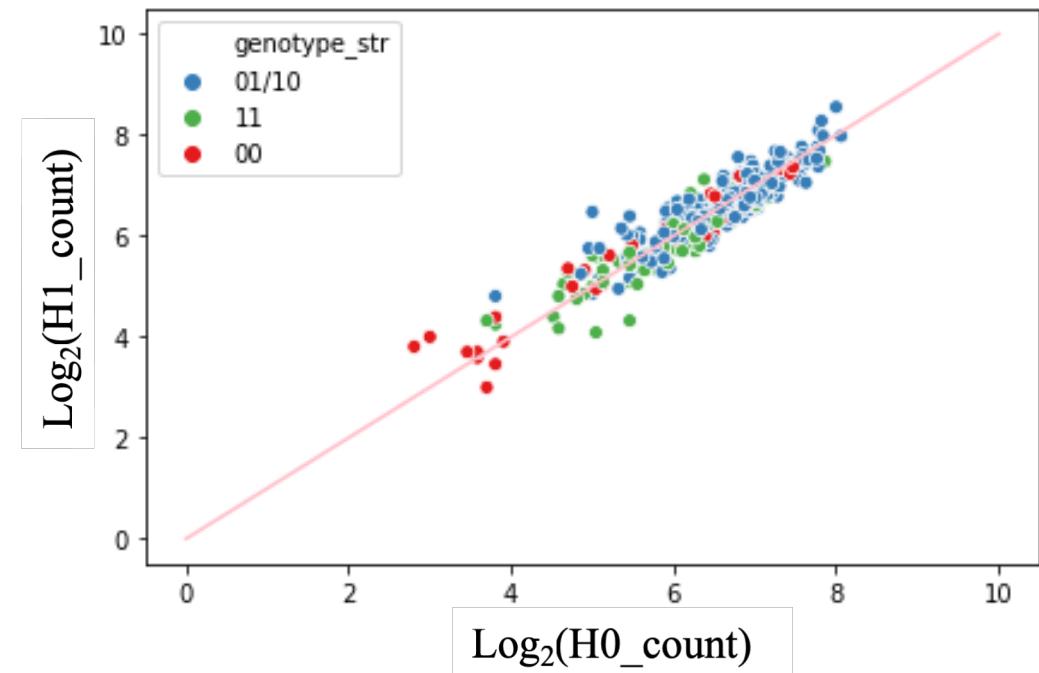
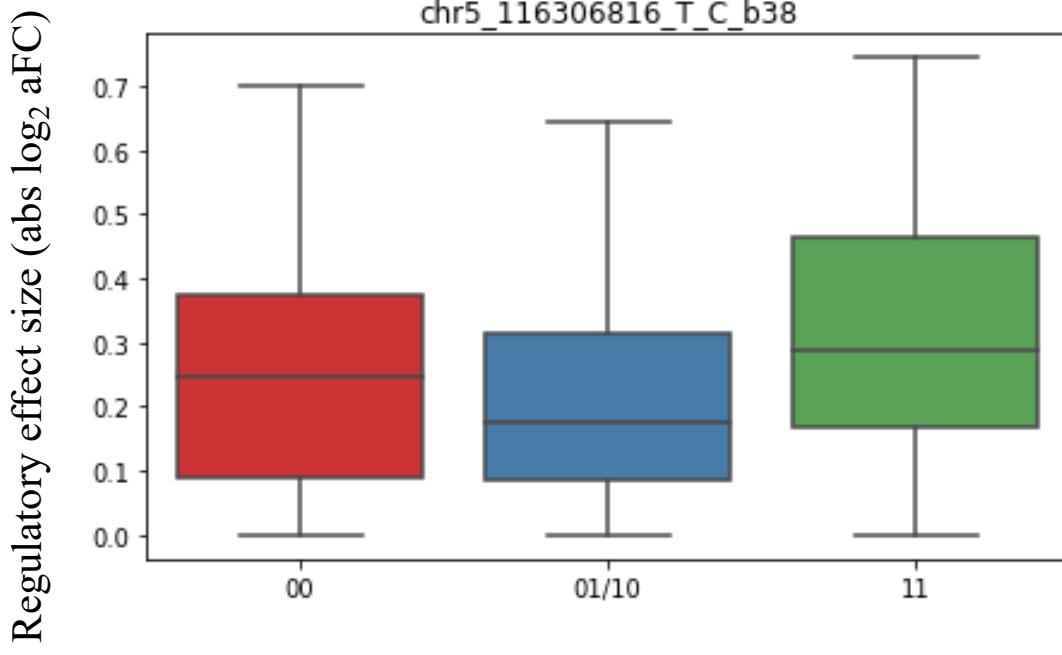


For what fraction of the cases the current eQTL data  
“fully” describes ASE signal?



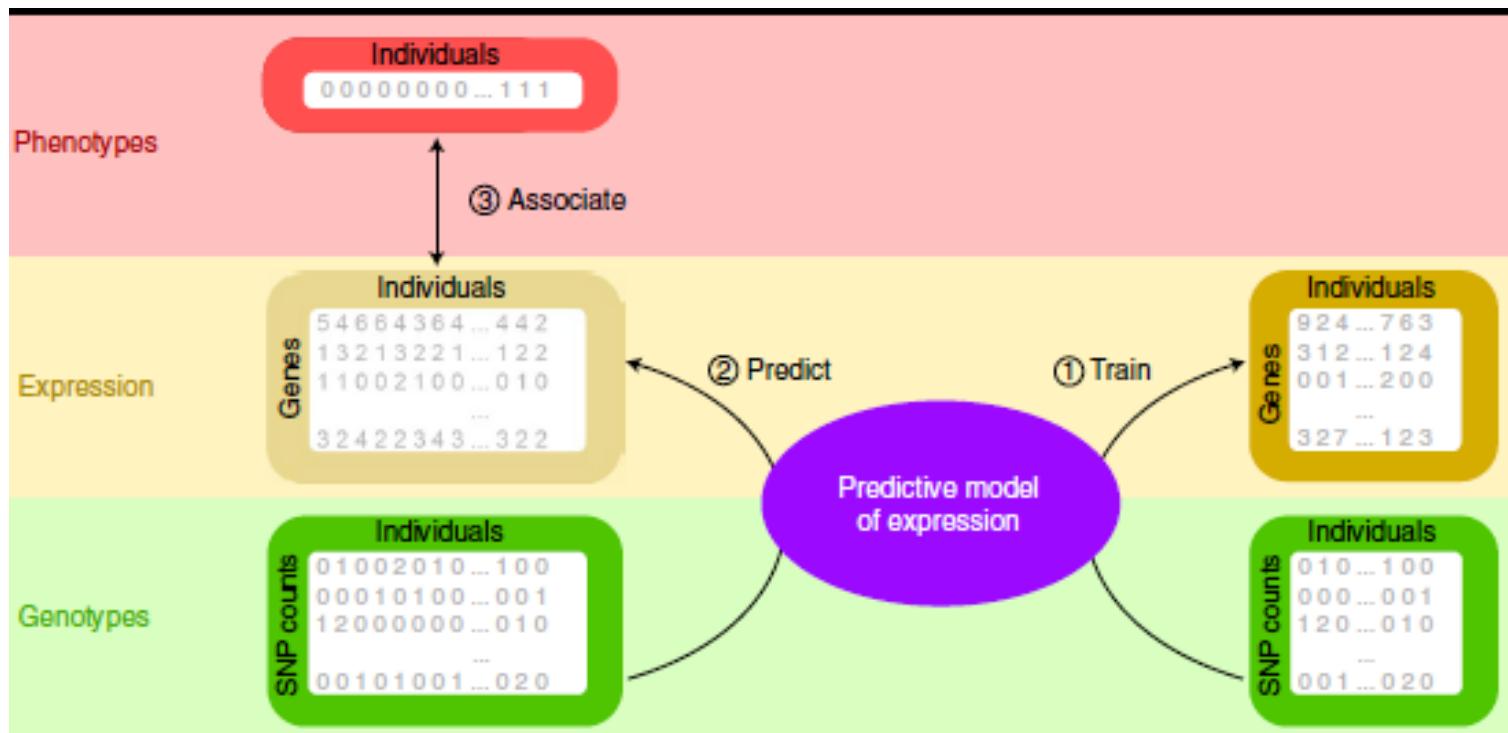
# Future Work

- Identifying eQTLs with or without ASE signals and see if they are functionally different



# Future Work

- Using our haplotype-specific estimates of regulatory effect size to impute gene expression in transcriptome-wide association study and other genetic association tests.



# Acknowledgements

Pejman Mohammadi

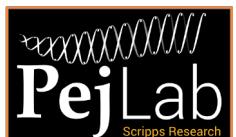
Bence Kotis

Daniel Munro

Marcela Mendoza

Yuren Dong

GTEx Consortium



Thank you