The role of gene expression in complex trait heritability

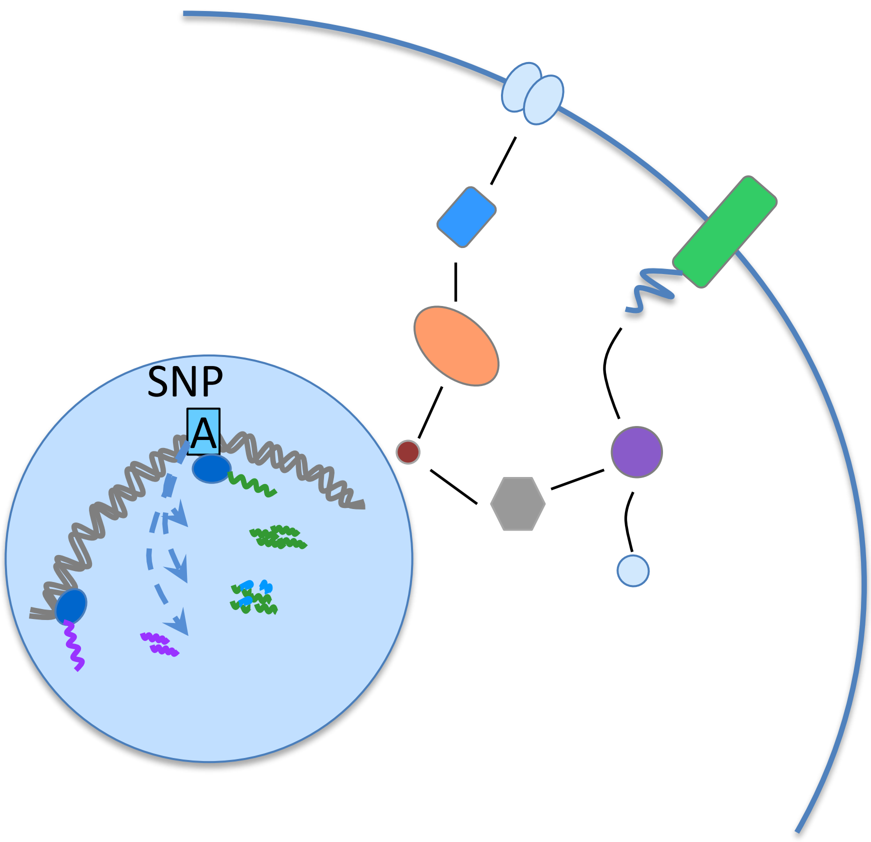
Alexis Battle

Motivation and introduction

How can we use gene expression and epigenetics to help us understand complex trait genetics?

Majority of trait-associated variation is non-coding.

Common hypothesis is that most of these function by altering gene expression.



Motivation and introduction

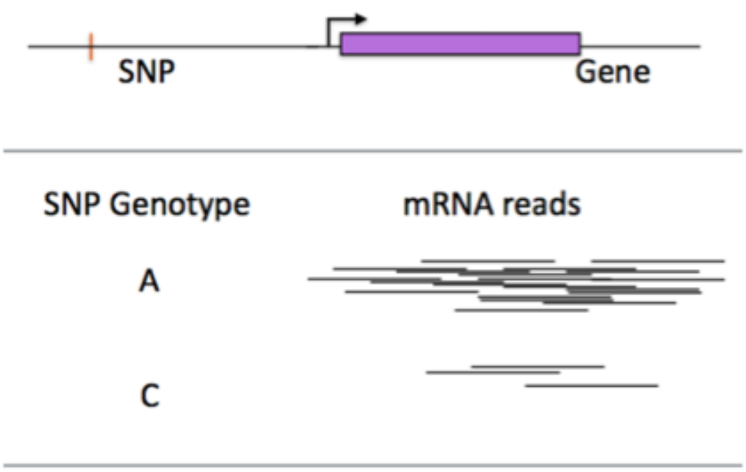
Using expression and epigenetic data to inform missing heritability:

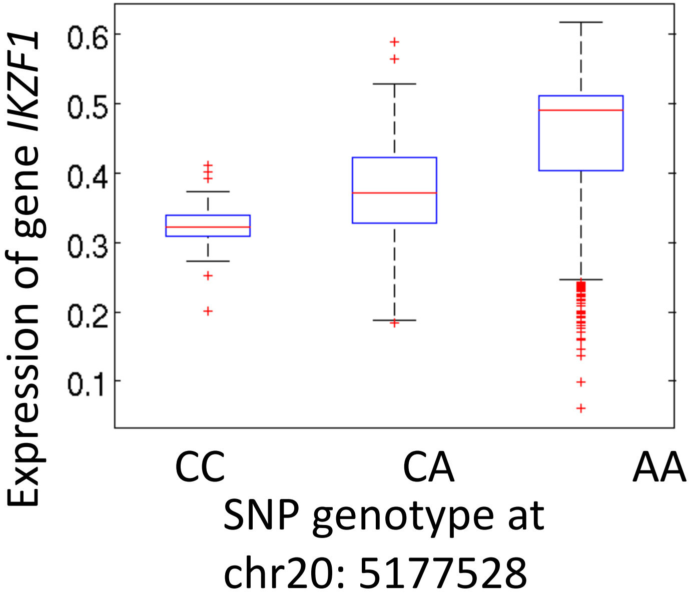
- Quantify contribution of this important component of trait heritability?
- Explain mechanism?
- Increase power to detect trait-associated variants (or build good predictors)?

1. Genetics of gene expression

Genetic variants affect gene expression

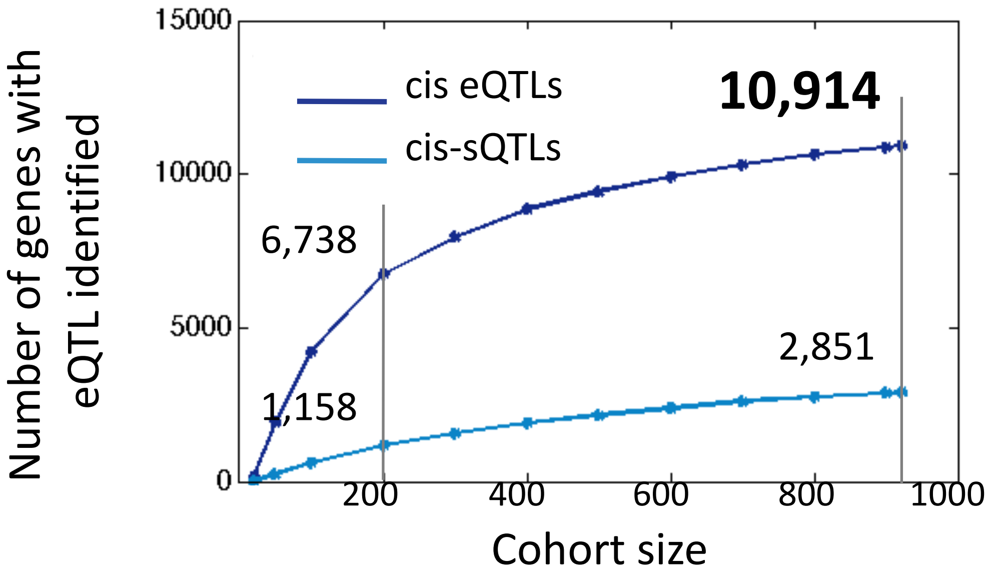
<u>eQTL (expression Quantitative Trait Locus) analysis</u>:
Association between genotype and RNA expression levels





Prevalence of eQTLs

Cis-eQTLs have now been identified for nearly every human gene, with numerous large studies available



Battle, Genome Research, 2014

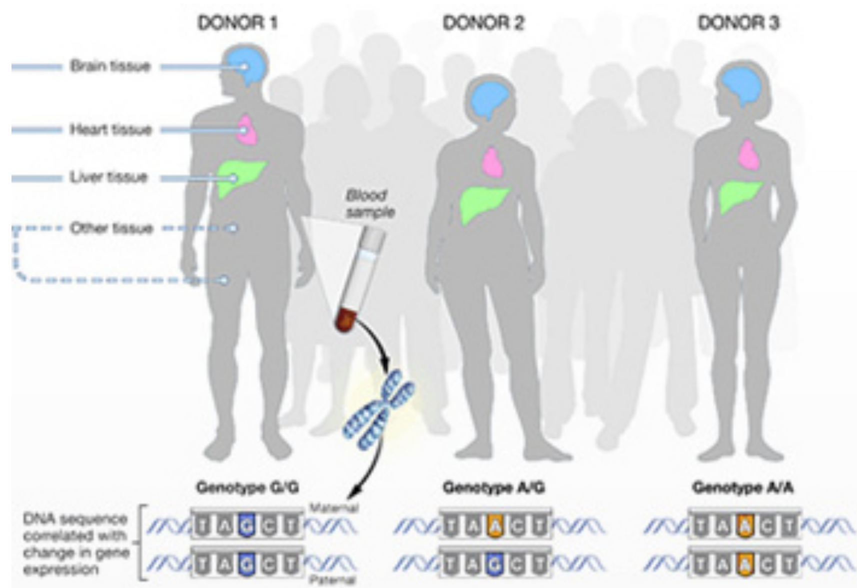
Large-scale eQTL analyses

- DGN: 922 whole blood RNA-seq
- GEUVADIS: 462 LCL RNA-seq
- MUTHER: 850, several tissues, microarray and later RNA-seq
- Wright et al, 2014: 2,752 twins, whole blood microarray
- Westra et al, 2013: meta-analysis of 5,311 whole blood microarray samples

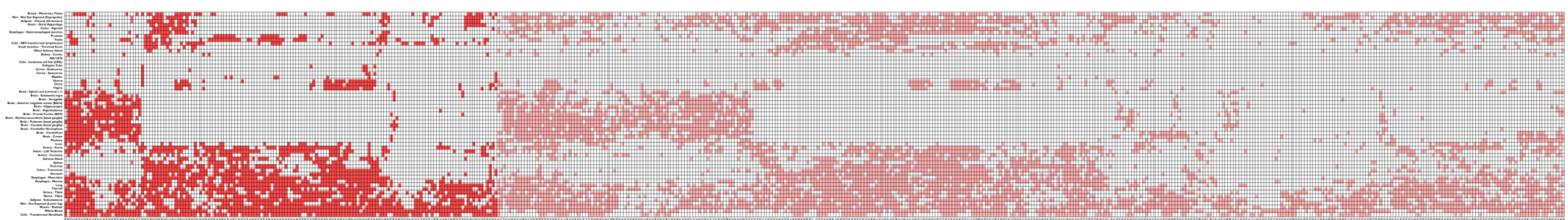
GTEx Project

GTEx Consortium v6p data

- 449 genotyped donors
- 7051 gene expression samples
- 42 post-mortem tissues
  - 31 solid-organ tissues
  - 10 brain subregions



44 tissues

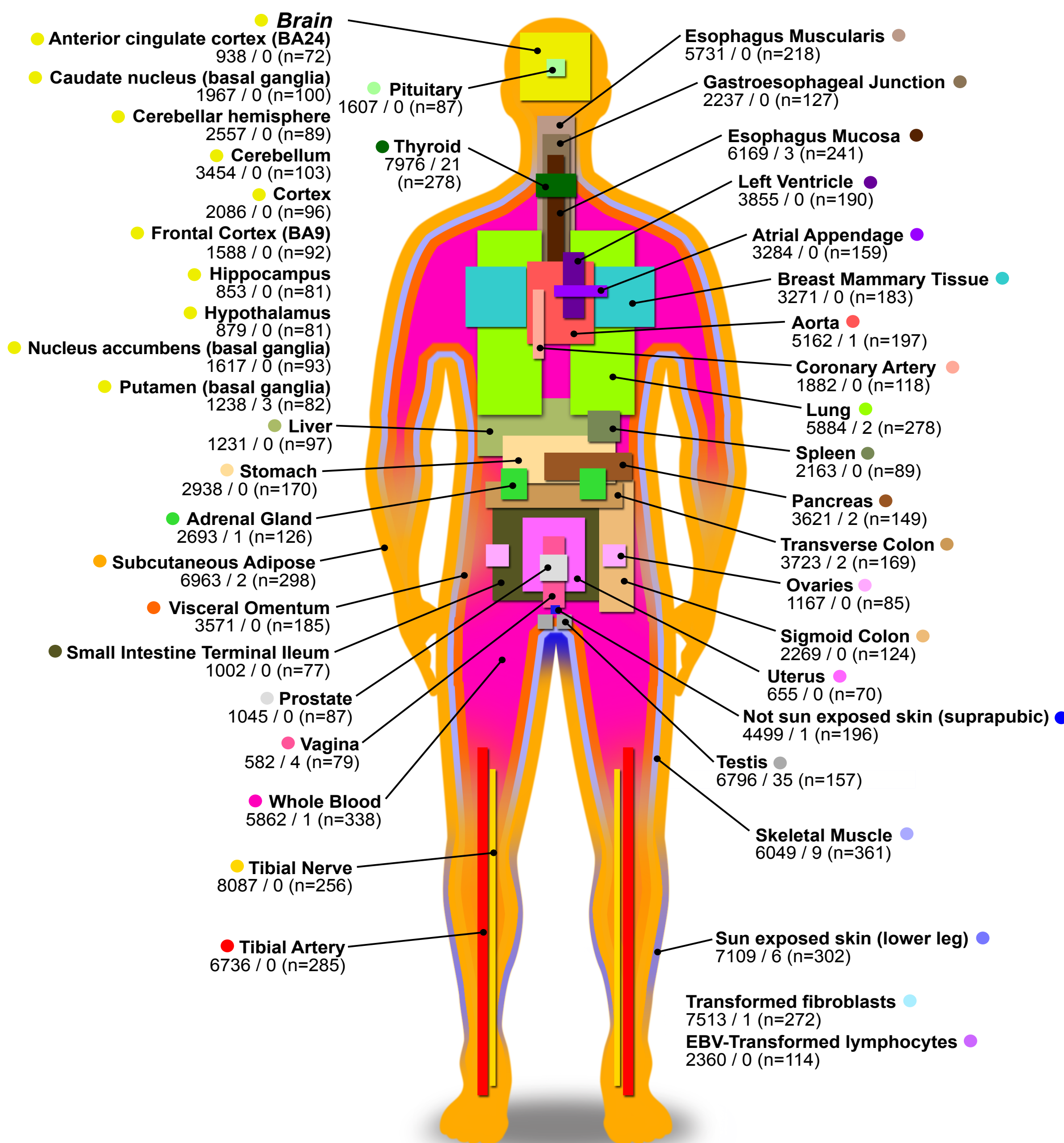


116 individuals (WGS)

449 individuals (RNA-seq + genotype)

The GTEx Consortium, Nature 2017

Genetic effects across human tissues



Total unique eQTL genes:

Cis: 19,725 (FDR 5%)

Trans: 93 (FDR 10%)

Most cis per tissue:

8,087 Tibial nerve (N=256)

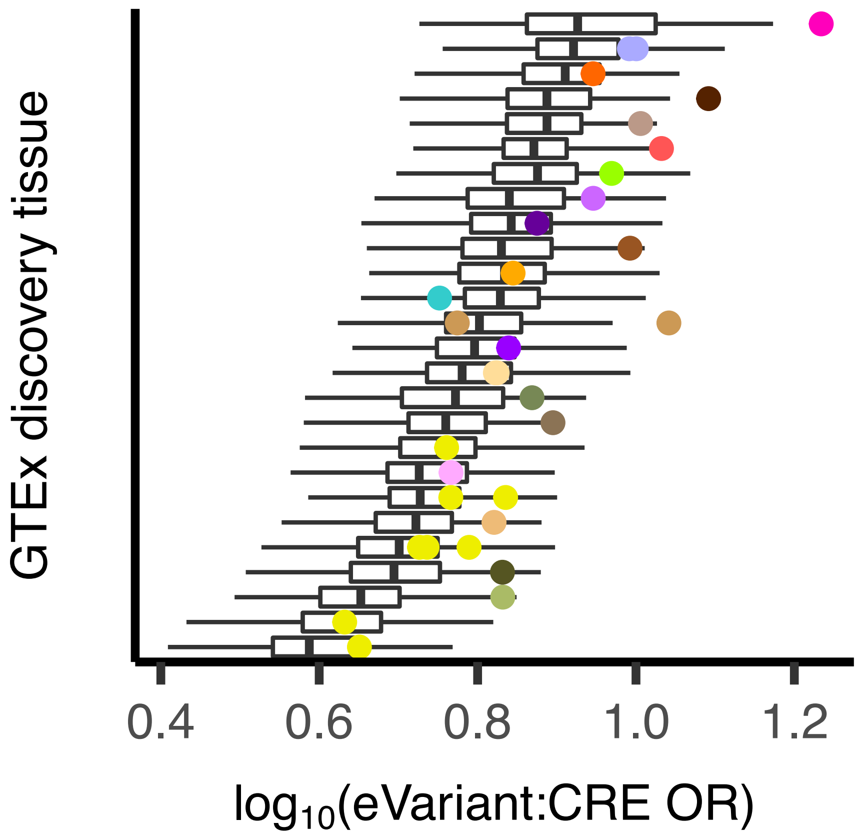
Most trans per tissue:

35 Testis (N=157)

The GTEx Consortium, Nature 2017

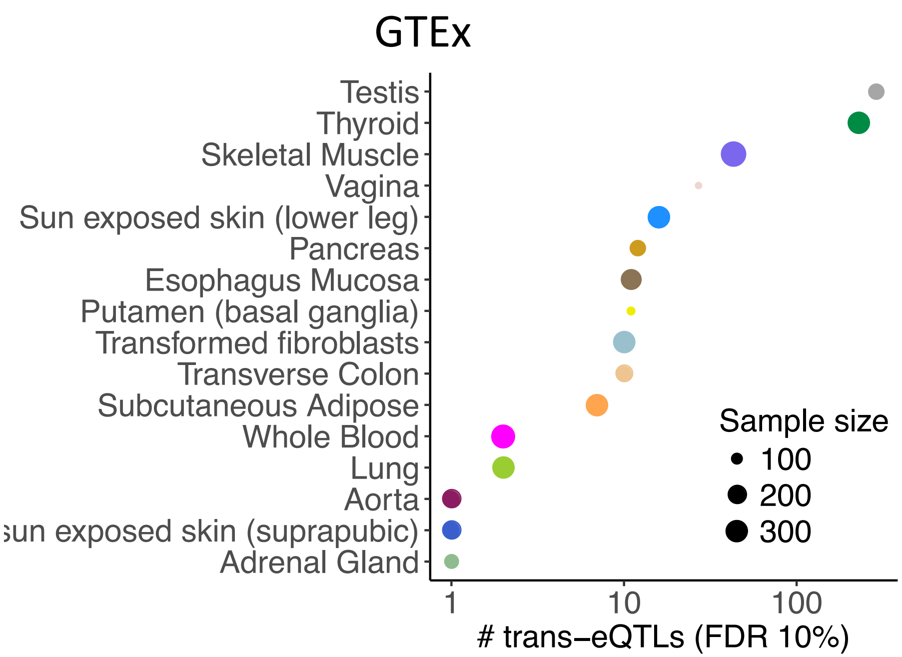
Characterizing eQTLs across tissues

- Cis-eQTL variants fall in tissue-specific regulatory elements (from Roadmap Epigenomics)



The GTEx Consortium, Nature 2017

Trans-eQTLs



Large studies:

Westra et al (N=5,311, using GWAS variants only)
ALSPAC (N=869)
MUTHER (N=850)
DGN (N=922)
Framingham (N=5257)

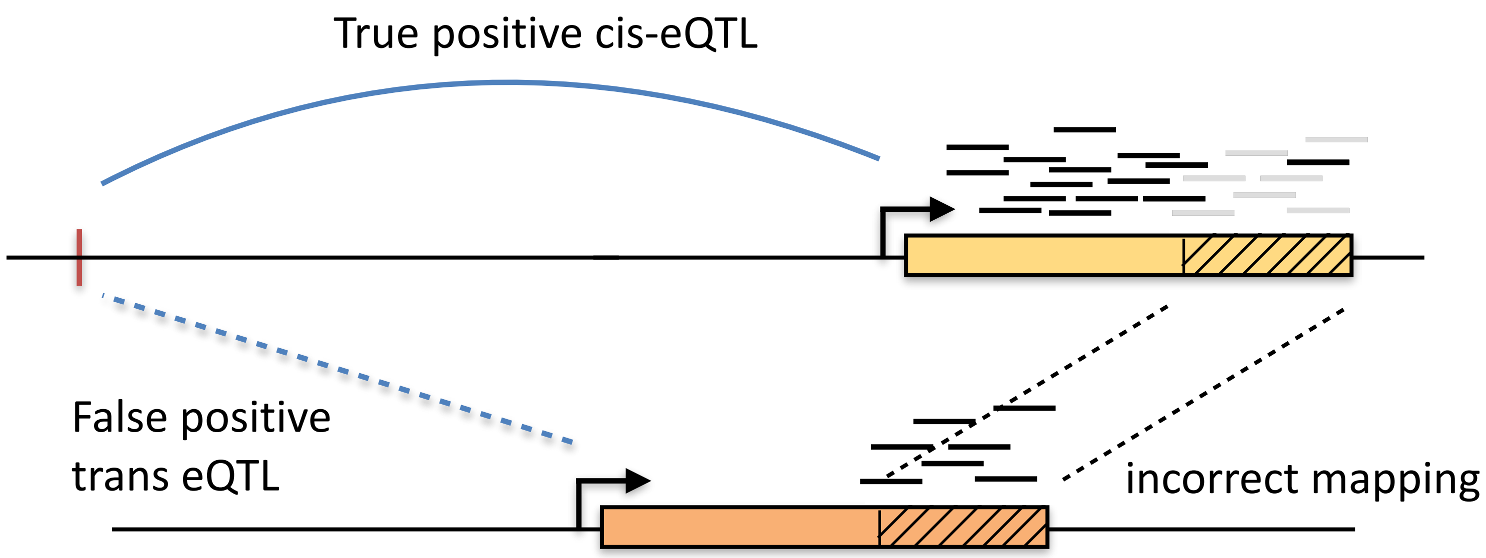
Studies report wildly different # hits (10s–10000s)

Replication and validation remains poor

We remain underpowered at current sample sizes

Challenges for trans-eQTL detection

- Power
- False positives from many sources e.g. over and under correcting confounders (Dahl et al, 2017)
- Mapping error (similar to probe cross-hybrid.)



Slide adapted from Yuan He

Heritability of gene expression

Despite eQTLs being pervasive, estimates for heritability of gene expression are modest

- Average over genes ranging from 0.09 to 0.3 (Price et al, 2008/2011, Wright et al 2014, Wheeler et al 2016, MUTHER)
- Informs need for greater power to detect trans-eQTLs

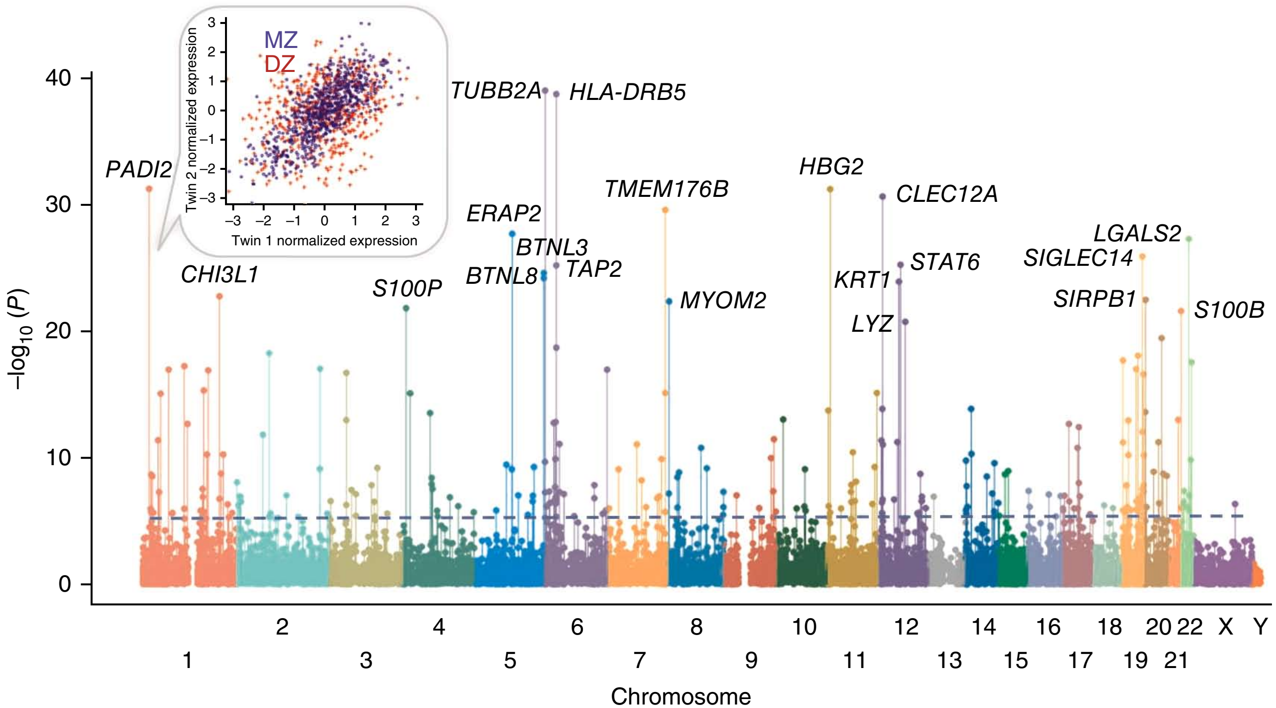


Figure from Wright et al NG 2014

Heritability of gene expression

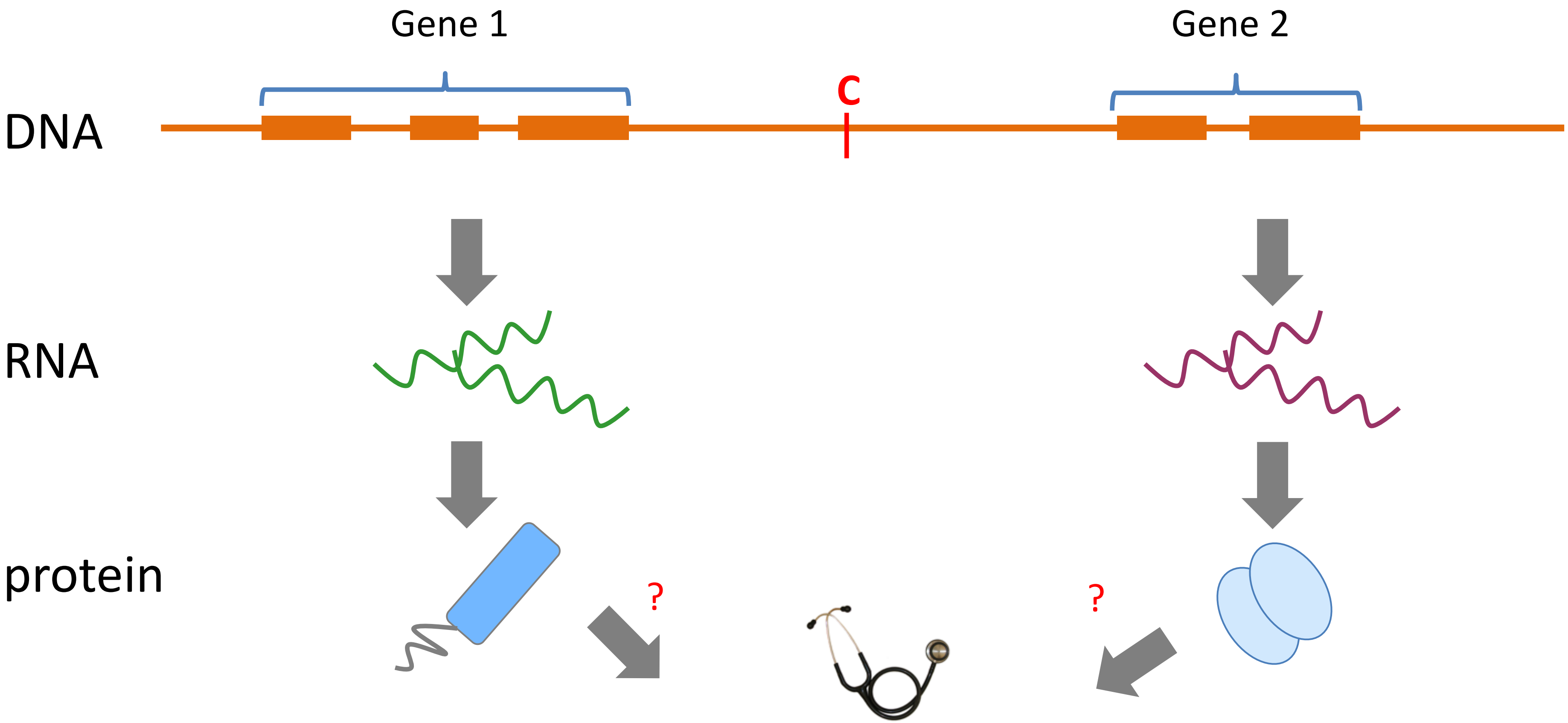
- Trans effects contribute much more to gene expression heritability than cis
- h^2_{cis}/ h^2 estimates range from 10-40%
  - Price et al 2011
  - Wright et al 2014
  - Grundberg et al 2012
- Varies by tissue, population, power, method
- h^2_{cis} sparse (Wheeler et al 2016), trans often mediated by cis effects

2. Connecting expression and epigenetics to complex traits

eQTLs and complex disease genetics

Help interpret GWAS variants (especially non-coding):

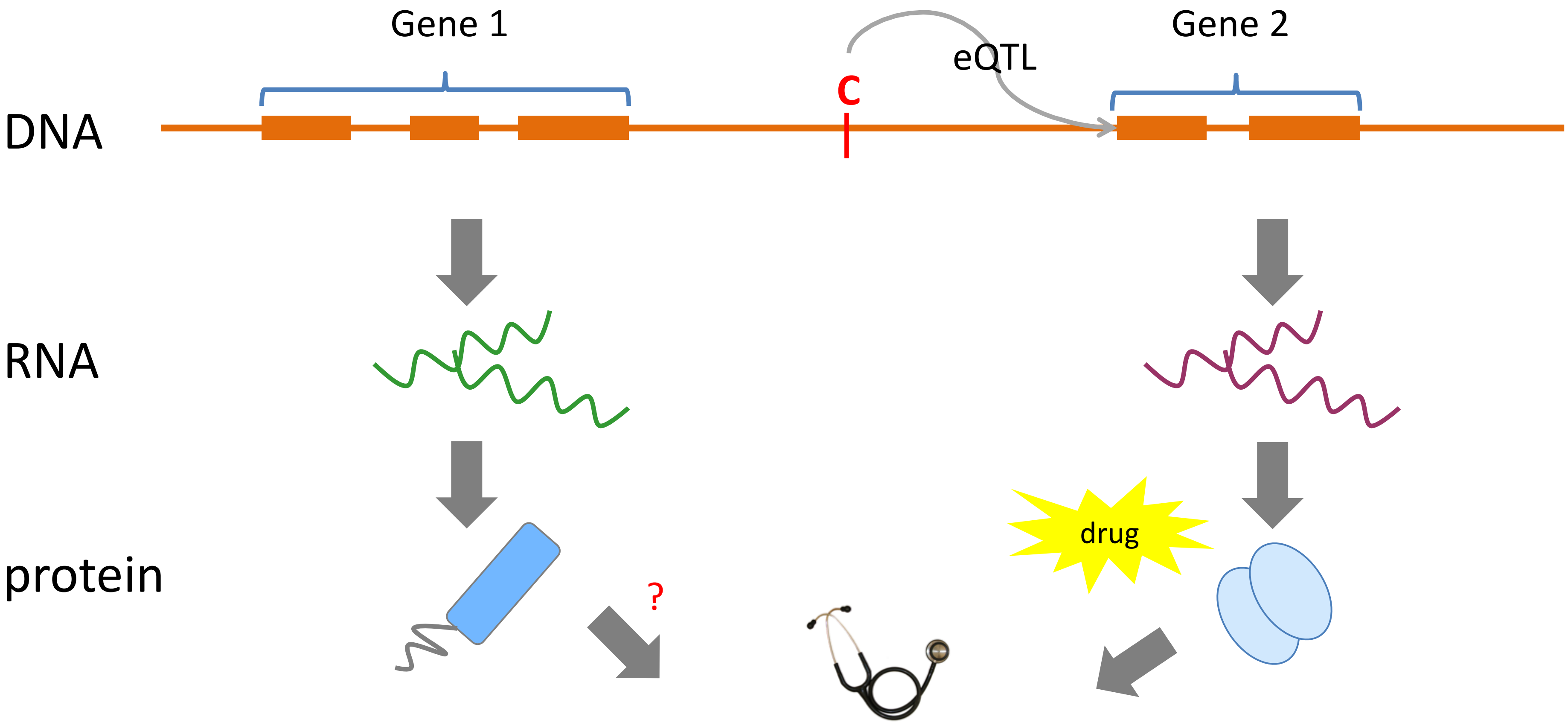
- understand mechanism
- guide interventions



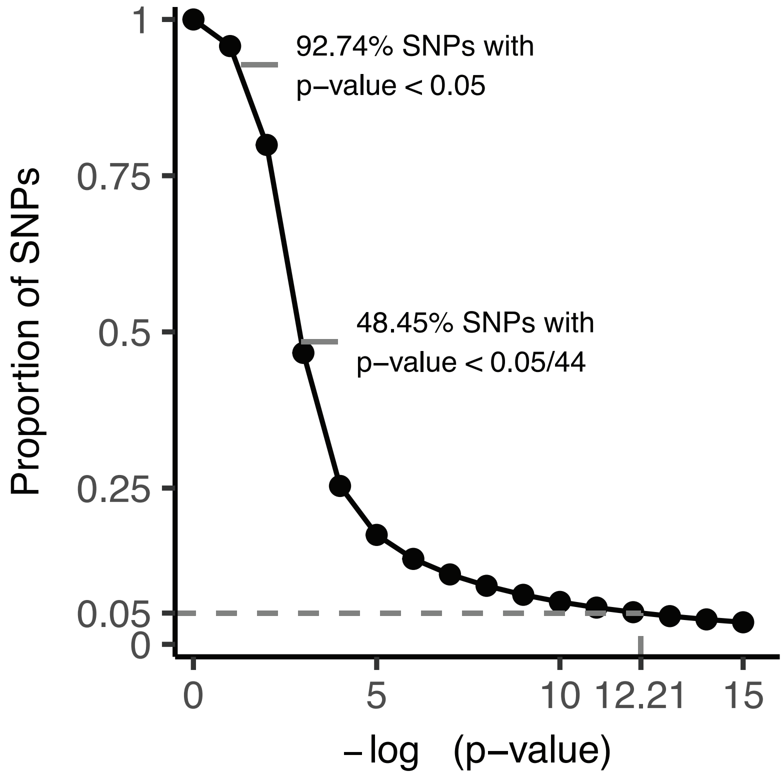
eQTLs and complex disease genetics

Help interpret GWAS variants (especially non-coding):

- understand mechanism
- guide interventions



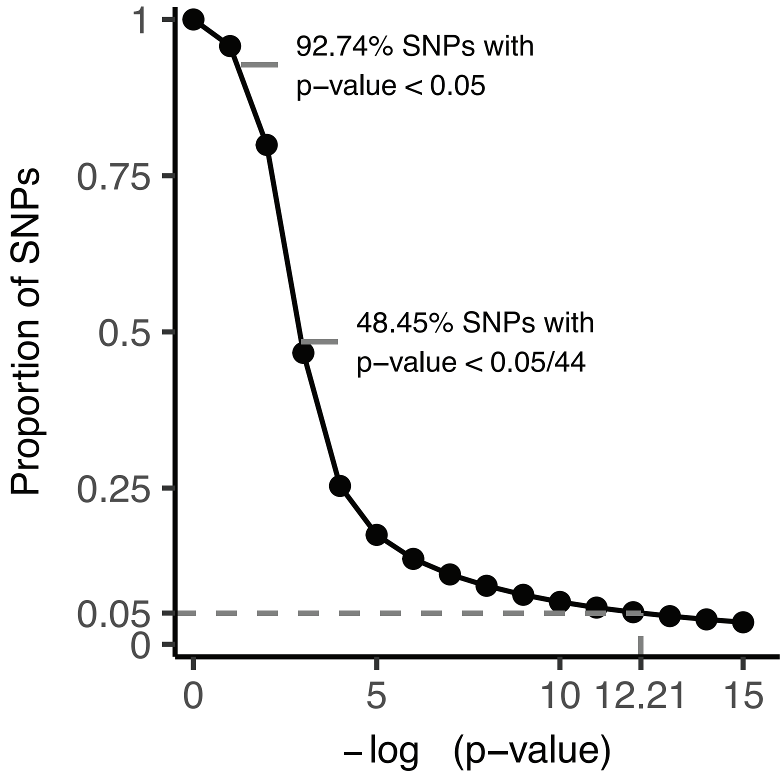
Most SNPs are eQTLs



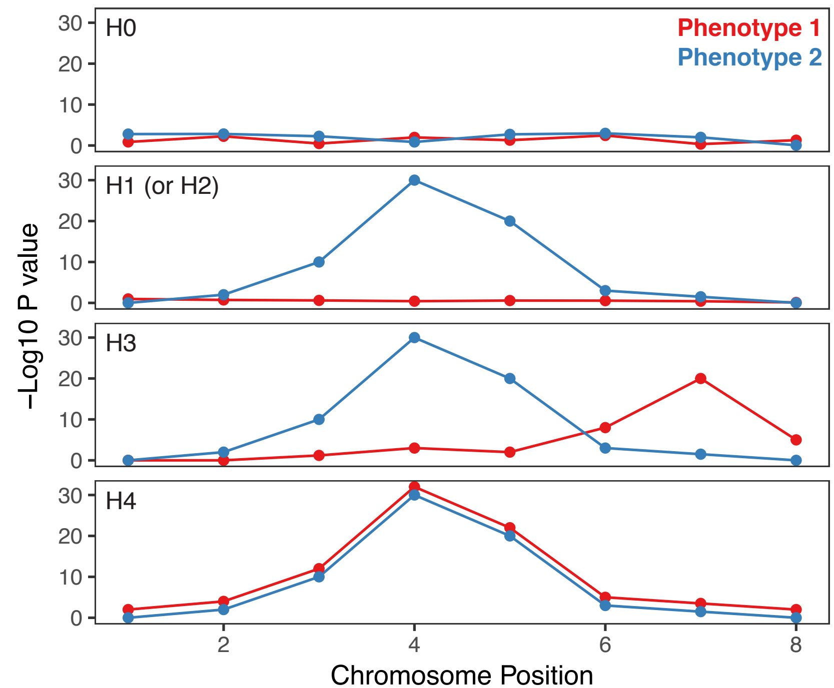
*But…*most of these just tag functional variants

Slide adapted from Casey Brown, UPenn

Most SNPs are eQTLs



Co-localization analysis:



*But…*most of these just tag functional variants

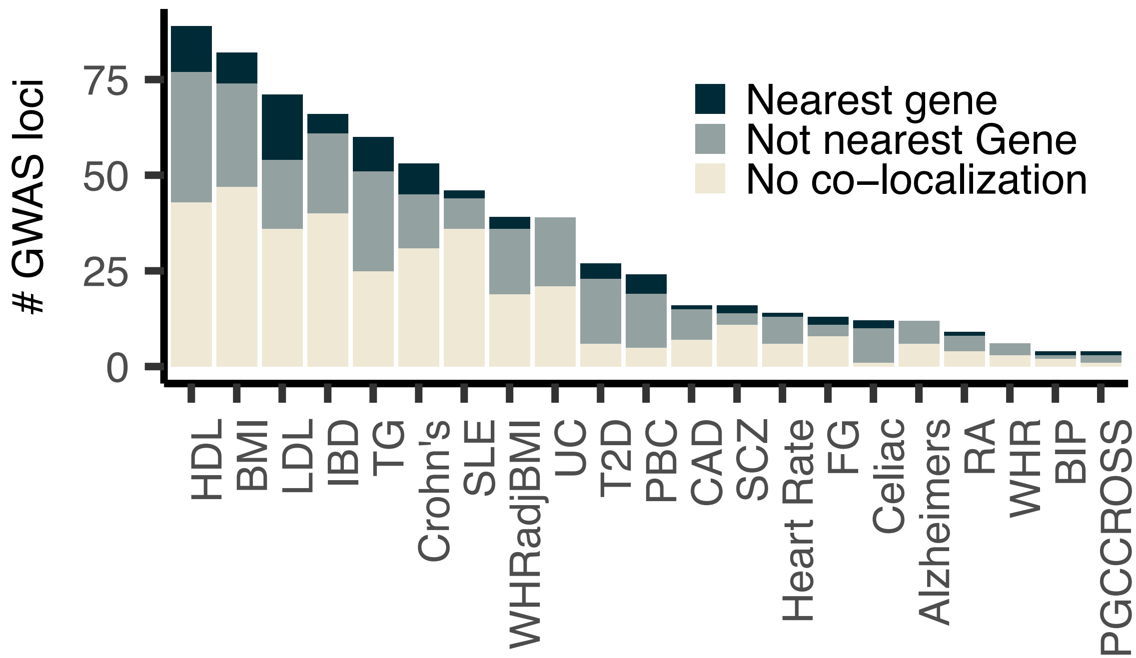
Need to evaluate whether underlying causal variants are actually shared (co-localization)

Slide adapted from Casey Brown, UPenn

eQTLs and complex disease genetics

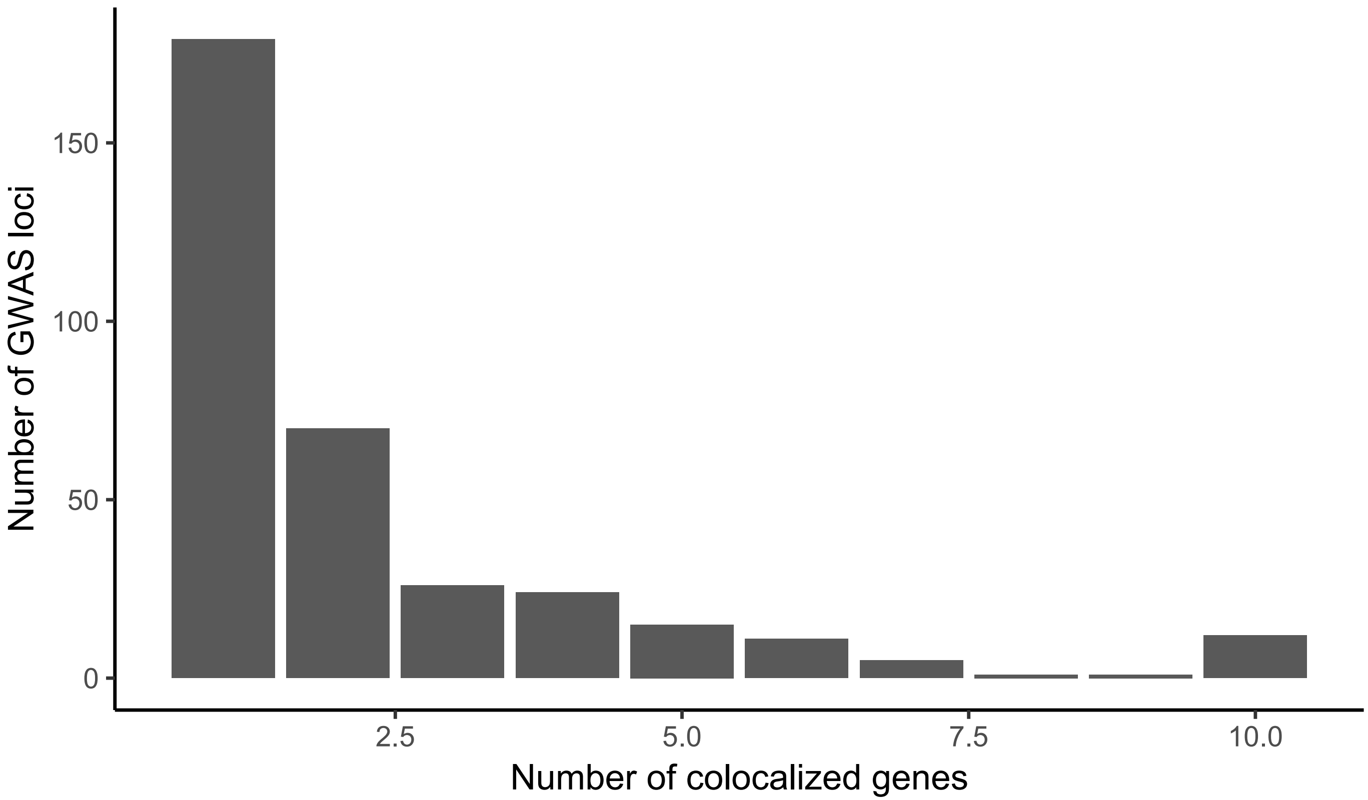
~50% of genetic variants associated with human disease co-localize with an eQTL

- compared to 92% simply associated $p < 0.05/44$ (still enriched over background)



The GTEx Consortium, Nature 2017

Deciphering mechanism

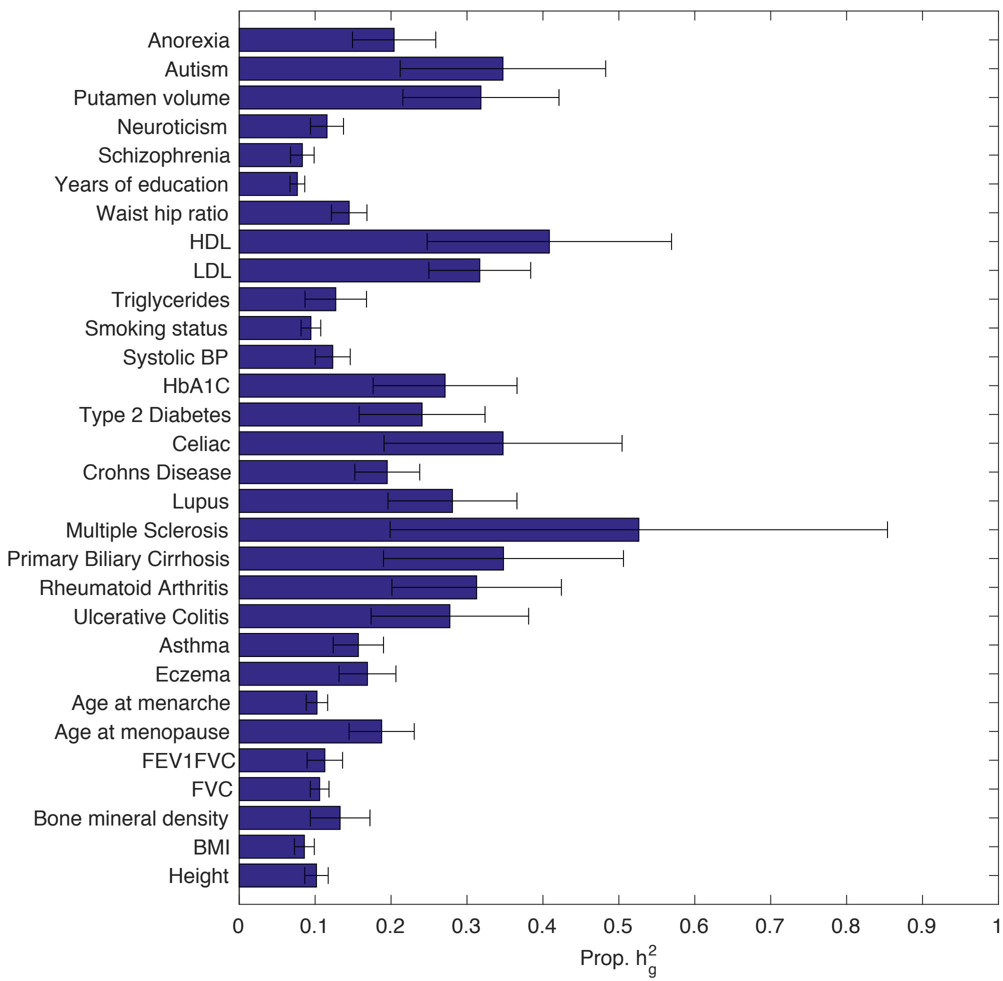


53% of co-localized GWAS loci have > 1 target gene, ambiguity remains

Slide adapted from Casey Brown

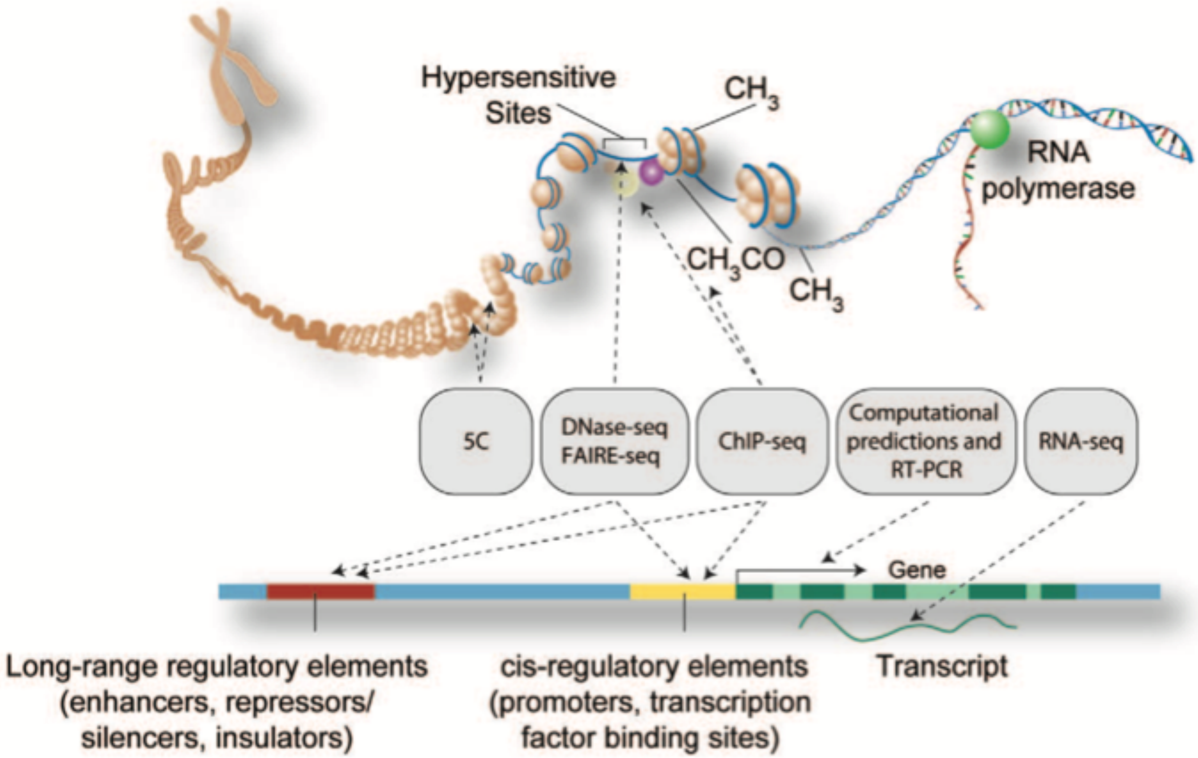
eQTL data informs heritability

GE co-score regression indicates cis-eQTLs explain mean 21% of h^2 across a set of complex traits



O'Connor et al. bioRxiv, 2017

Epigenetic data

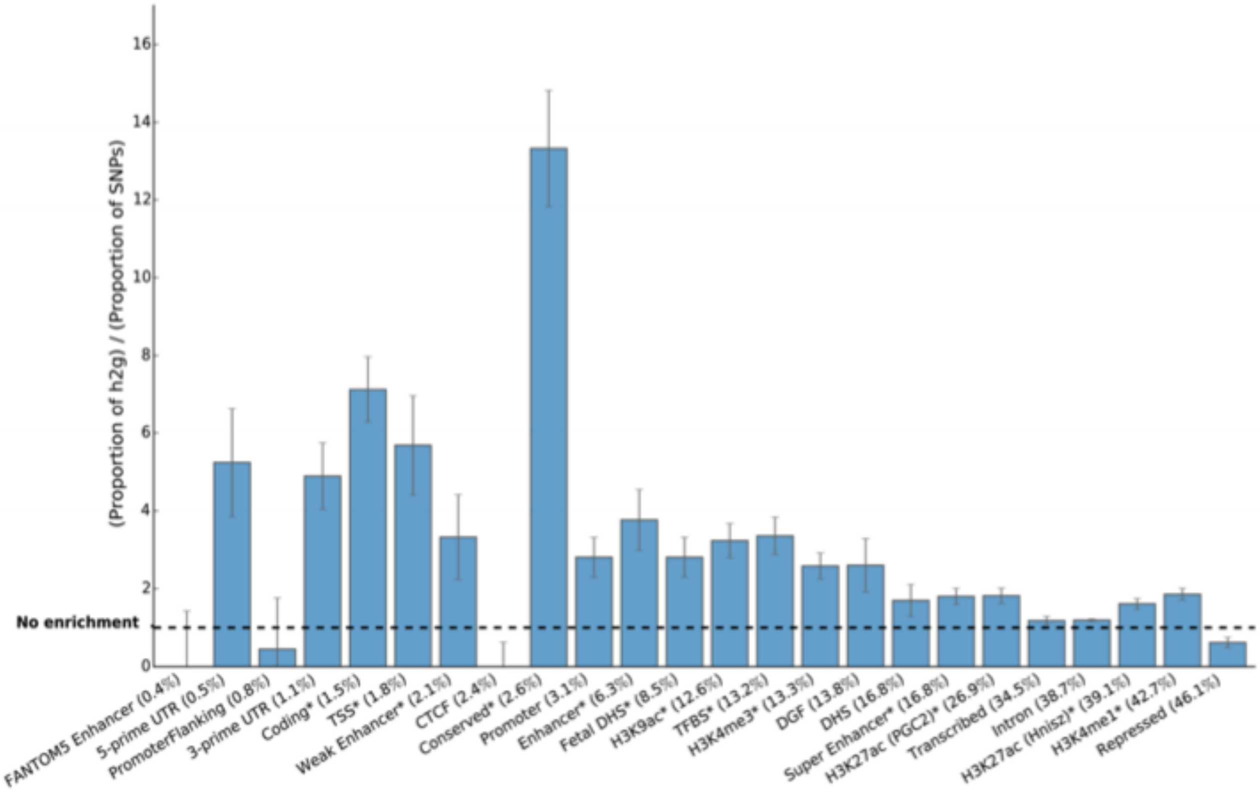


- ENCODE, Roadmap Epigenomics
- Regulatory elements: promoters, enhancers
- Transcription factor binding sites
- CpG sites
- ChromHMM

ENCODE Project Consortium. Plos Biology 2011.

Epigenetic data informs heritability

LD score regression, related approaches partition h^2

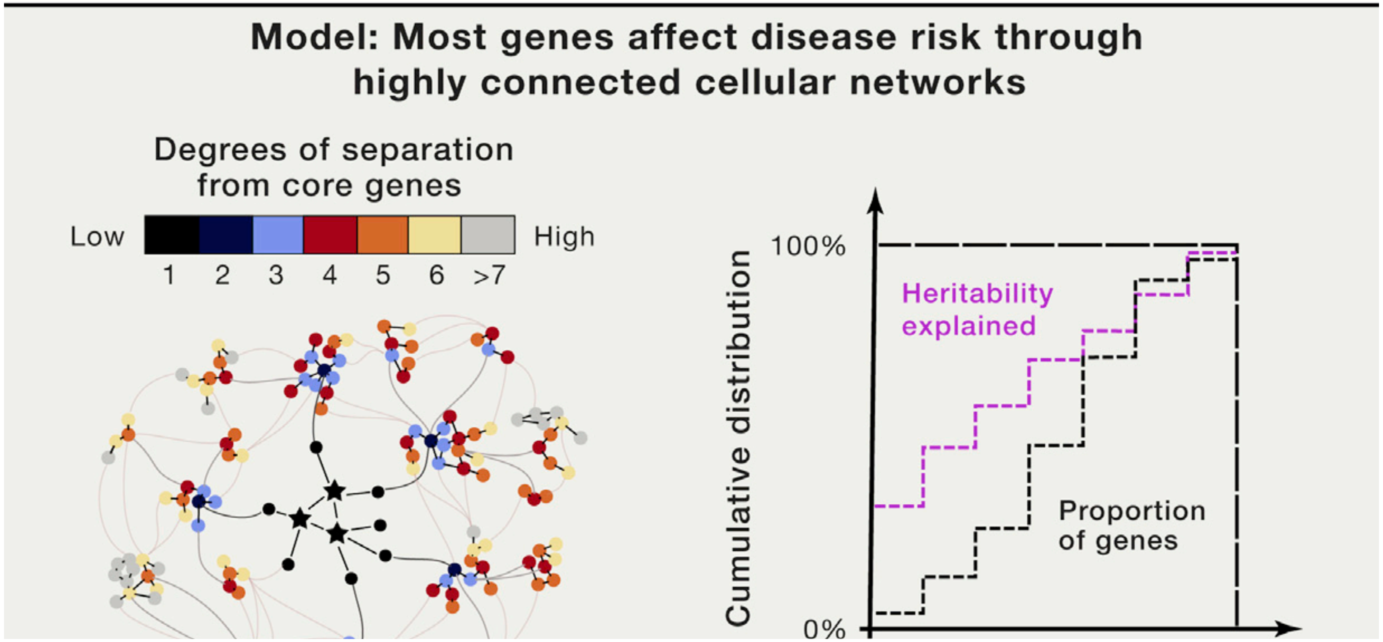


Large scale epigenetic data (Roadmap, ENCODE) enable analysis, indicate contribution of gene regulation

Figure from Finucane, NG, 2015

Ommigenic model

- Most/all expressed genes in disease-relevant cell types affect trait



- Highlights potential role of eQTLs, trans effects

Boyle et al., Cell, 2017

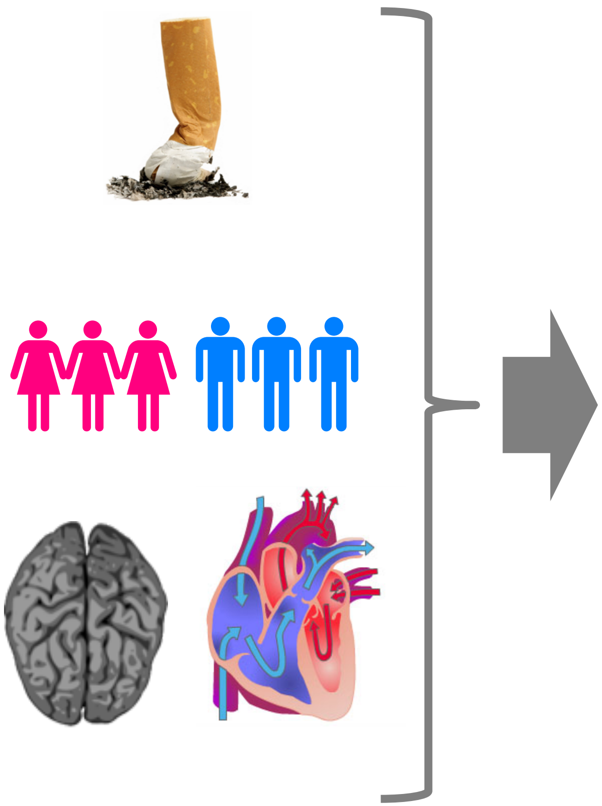
3. Complex effects of genetic variation on gene expression

What are we missing?

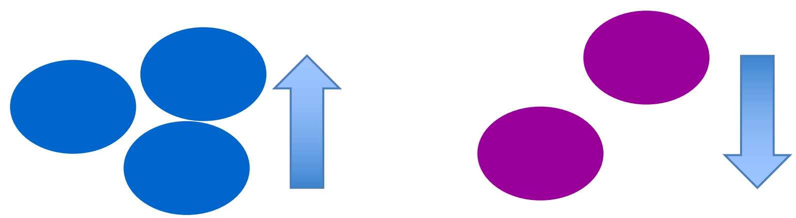
- Most studies are done on steady-state total expression measurements at a single adult or post-mortem time point
- Disease-relevant states include different developmental stages, environmental exposures, cell types
- Other variant classes and regulatory effects

Context-specificity

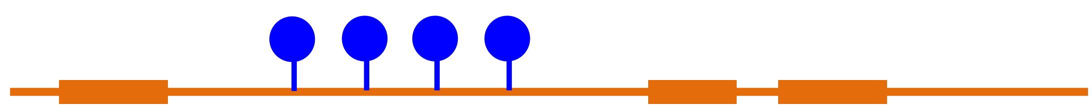
Many factors can *modulate* regulatory effects



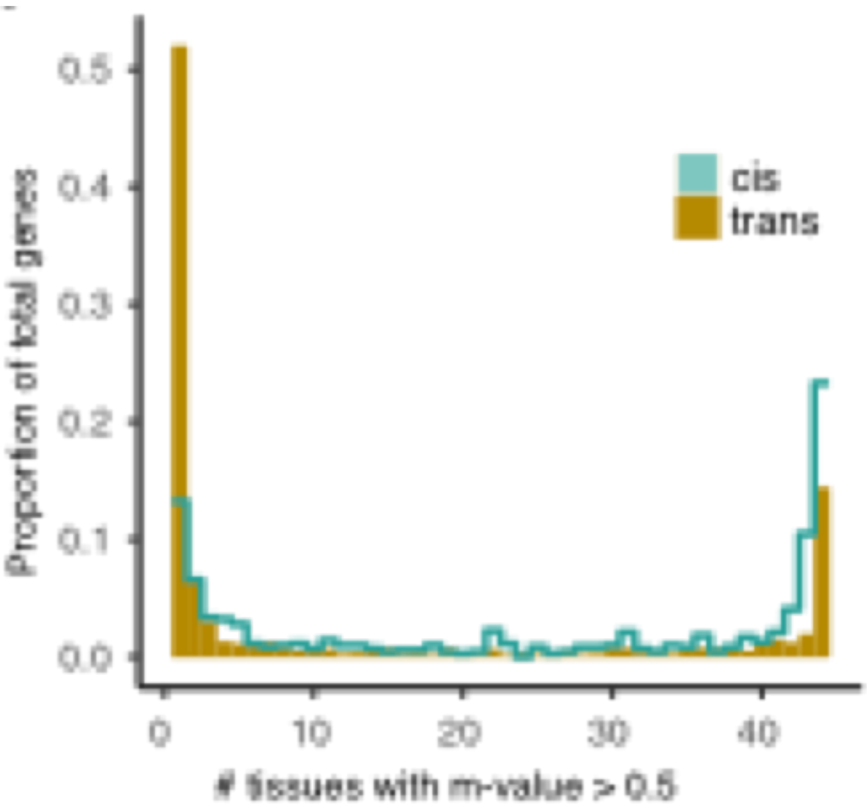
Altered transcription factor abundance

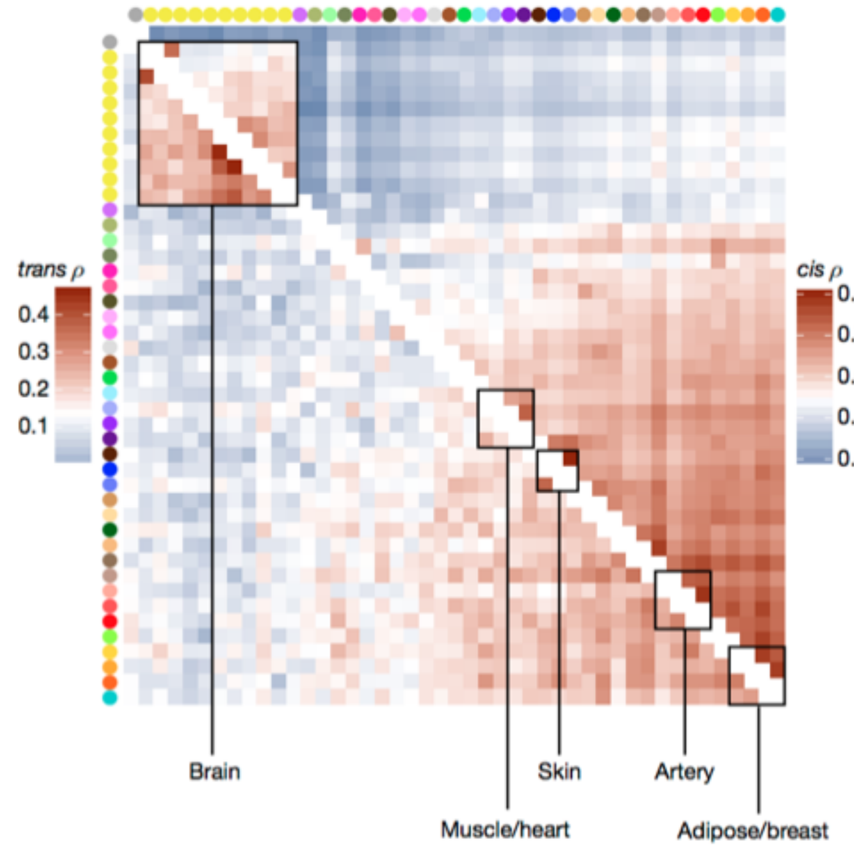


Epigenetic changes



GTEx tissue-specificity of cis and trans

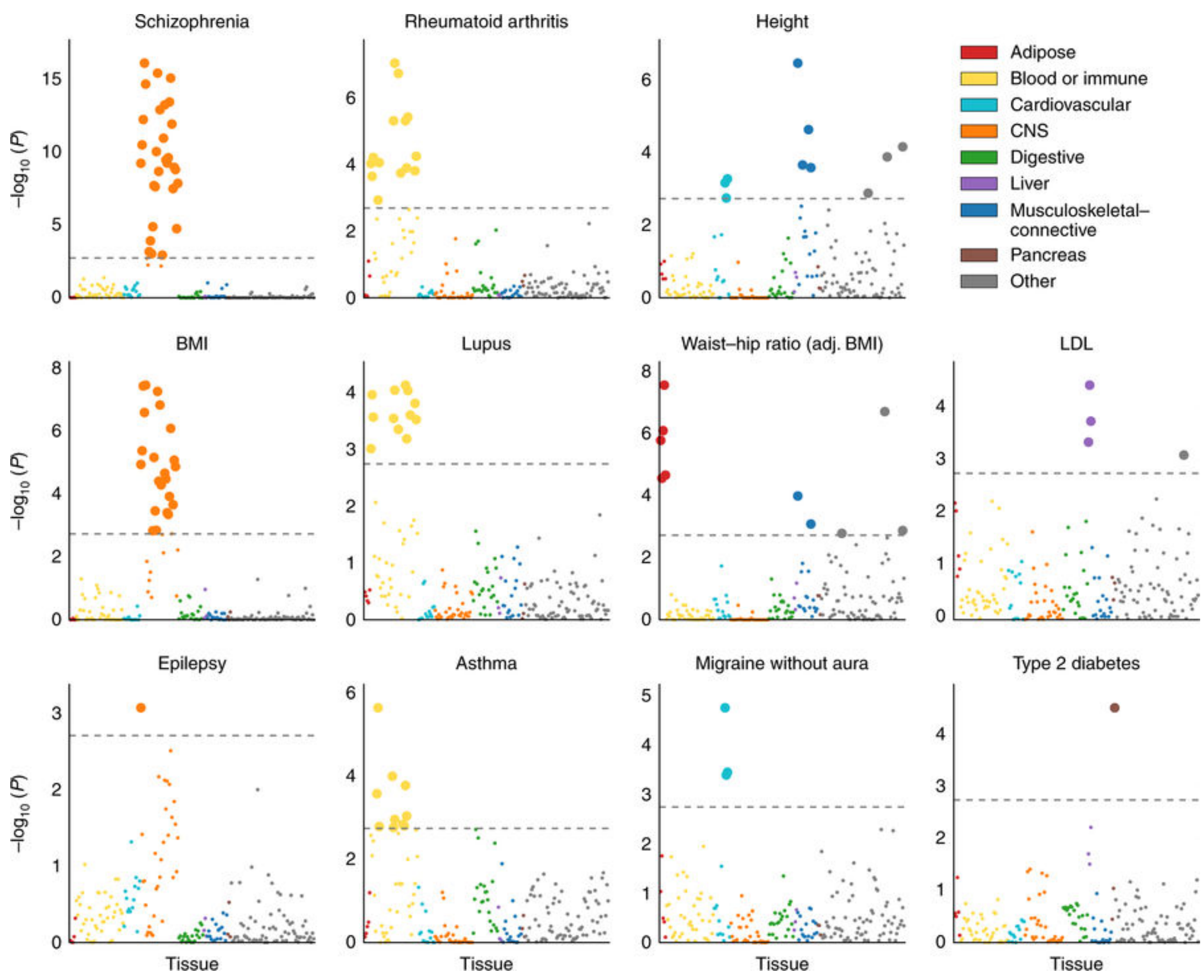




Trans eQTLs appear more highly tissue-specific than cis-eQTLs

The GTEx Consortium, Nature 2017

Tissue specificity and heritability



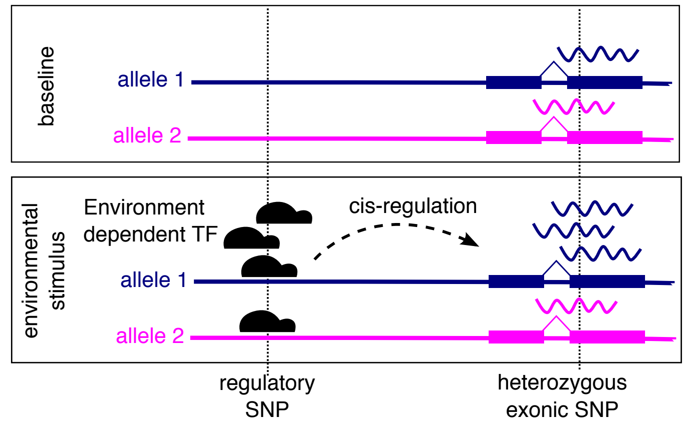
From Finucane et al, NG, 2018

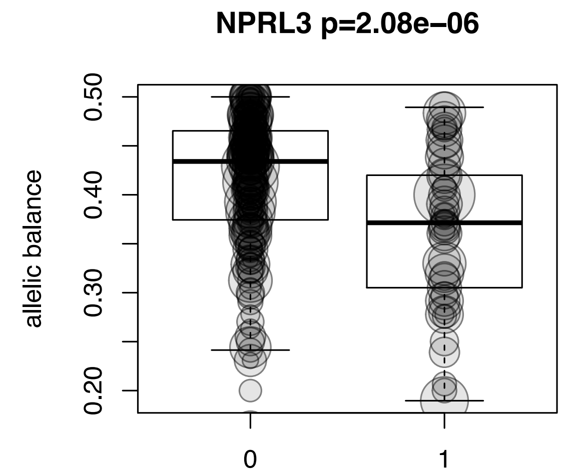
Detecting context-specific QTLs

Many other contexts beyond tissue:

- Recent work explores QTLs in diverse environments, such as infection response
  - Fairfax et al, Science 2014
  - Lee, Science 2014

- Methods for identifying allelic response from RNA-seq data





BP meds and NPRL3: related to genes involved in homeostasis of fluid volume

Knowles et al, NM, 2017

Diverse variants and readouts

- Diverse genetic variant classes, enabled by improved variant calling and methods
  - Structural variants
  - Repeats

- Diverse molecular phenotypes important to h^2:
  - Alternative splicing (Li et al, Science 2016)
  - Translation, protein abundance (Wu et al, 2013 and Battle et al,2015)
  - Epigenetic changes including chromatin accessibility, histone modifications, methylation, etc (McVicker 2013, Grubert 2015, Banovich 2014…)

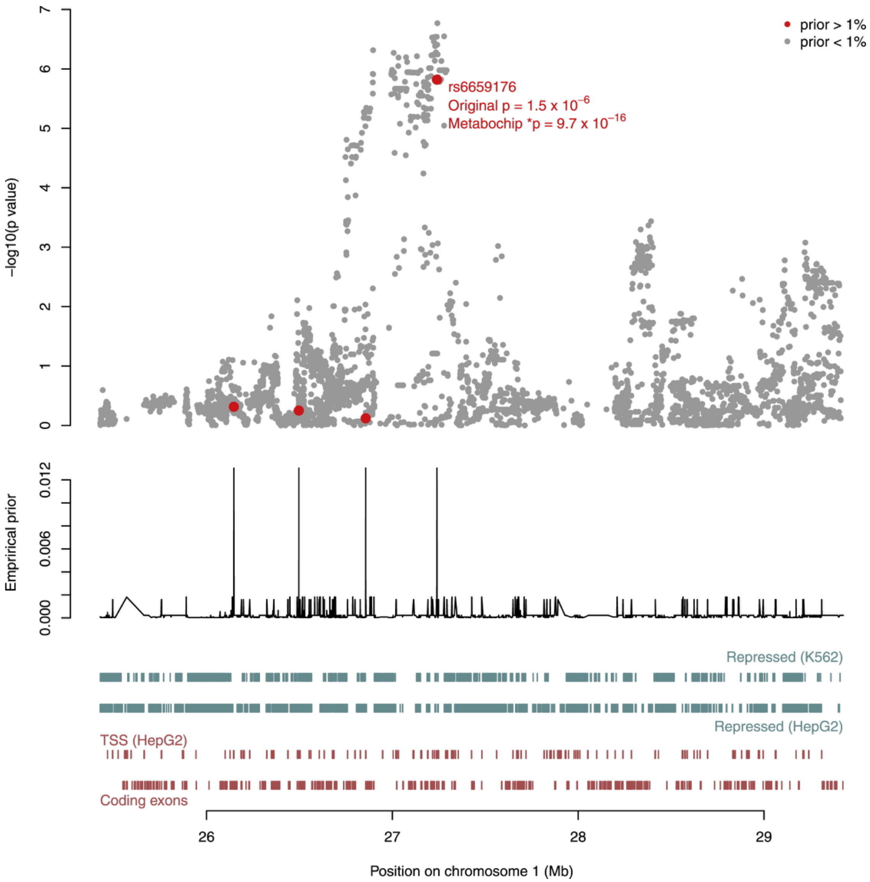

4. Further possibilities

Detecting more?

Can expression and epigenetic data help detect more variants or explain more heritability?

New methods integrate diverse data to learn and apply priors to GWAS analysis and prediction scores

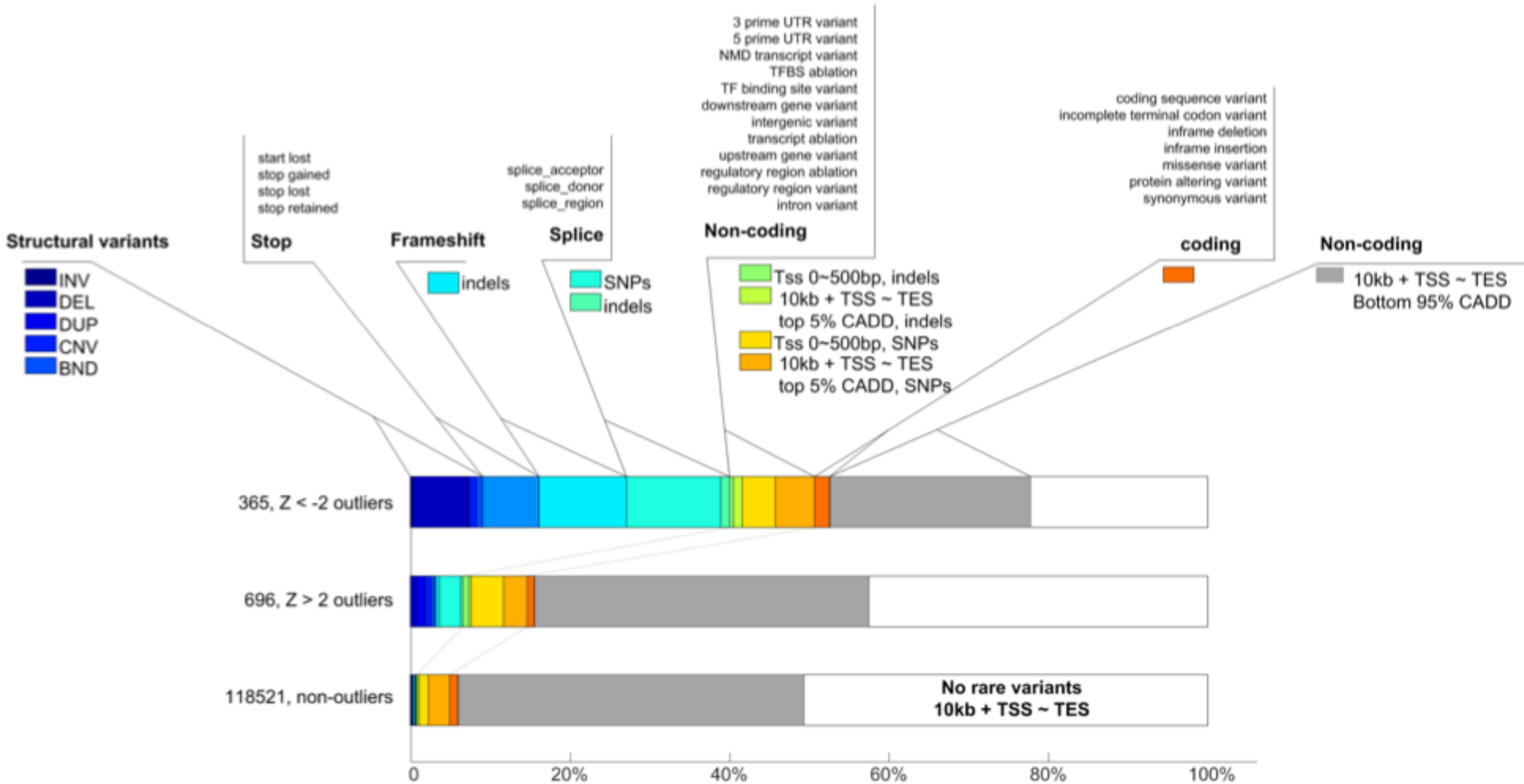
- Pickrell AJHG 2014 estimates 5% increase in loci detectable
- Marigorta NG 2017



Pickrell, 2014

Rare variants

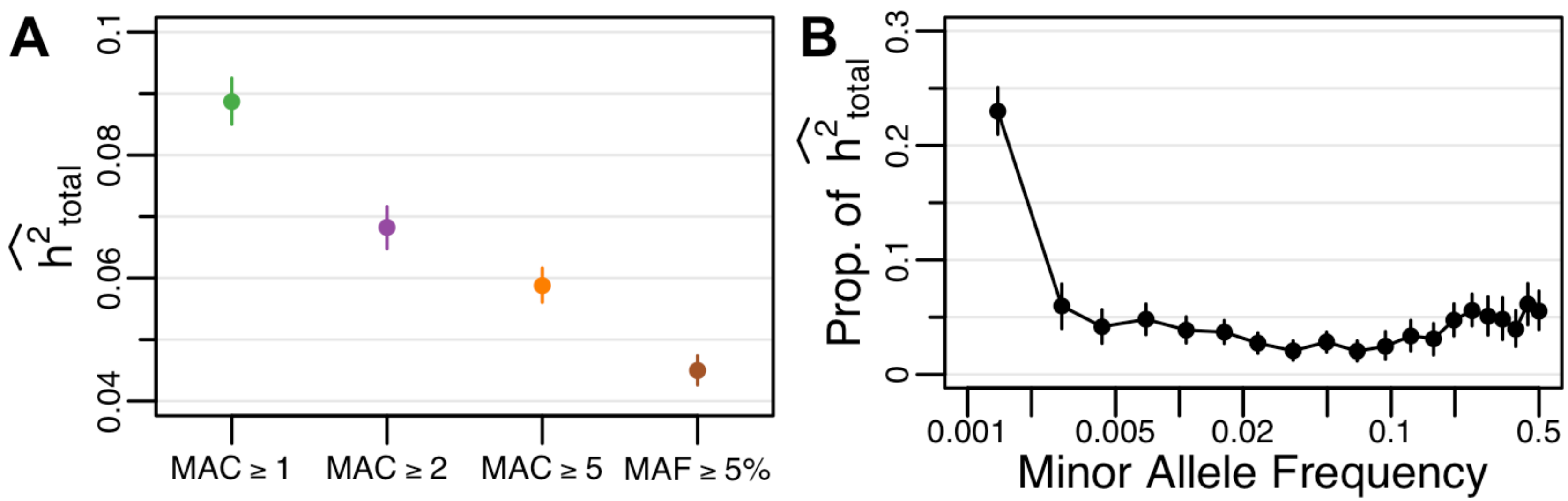
Recent work emphasizes importance of rare variation in driving extreme expression levels



Li et al, Nature, 2017

Rare variants

Preprint (Hernandez et al 2017) suggests rare variants explain a large fraction of heritability of gene expression



5. Conclusions

Progress – what we've learned

- Genetics of gene expression:
  - Prevalence of genetic variants affecting gene expression
  - Large catalogs of cis-QTLs, diverse contexts, variants, mol phenotypes
- Connections to complex traits:
  - Better data and methods provide better estimate of contribution of expression to h^2, and interpretation of individual variants (MR, etc)
  - **Current estimates indicate gene expression contribute sizeable but not majority fraction to trait h^2**
- Contribution of expression, epigenetic data to explaining missing h^2?
  - Modestly improved power for identifying individual GWAS hits through informed priors, potential for better prediction
- Improved interpretation and mechanism

Why delve deeper into expression?

- Help determine when and how much to invest in WGS, expression, epigenetic data

- To continue understanding implicated
  - Genes
  - Tissue and cell types
  - Epigenetic and other regulatory mechanisms

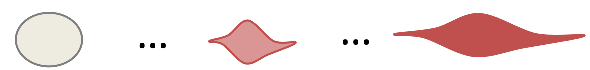
- Challenges and caveats
  - Ambiguity: many variants affect multiple genes
  - Interpretability: missing relevant cell types
  - Power: trans-eQTLs also require large sample sizes

Ongoing effots

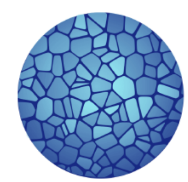
Scaling up eQTL studies, finding trans:

- eQTLGen: meta-analysis of all available whole blood expression data including over 30,000 samples
- GTEx v8: 1,000 individuals, WGS, over 50 tissues

Environment and dynamic QTLs



Single cell analysis - Human Cell Atlas, etc

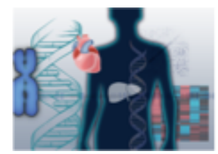


Integrated analysis connecting epigenetic and expression data for improved resolution, disambiguation, power

Methods

Acknowledgements

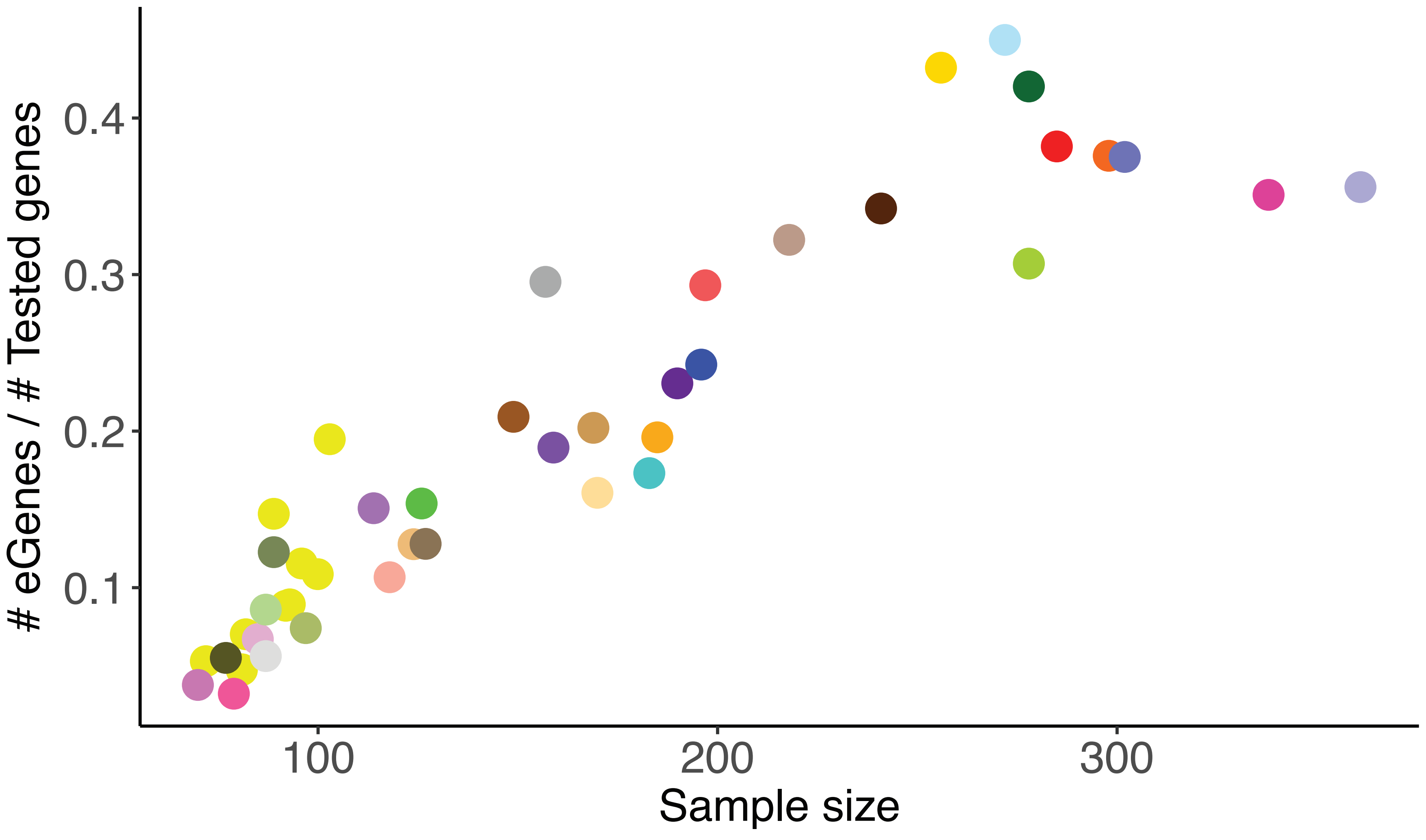
GTEx Consortium
Casey Brown
Barbara Engelhardt
Stephen Montgomery
Ira Hall



Collaborators
David Knowles
Jonathan Pritchard
Yoav Gilad

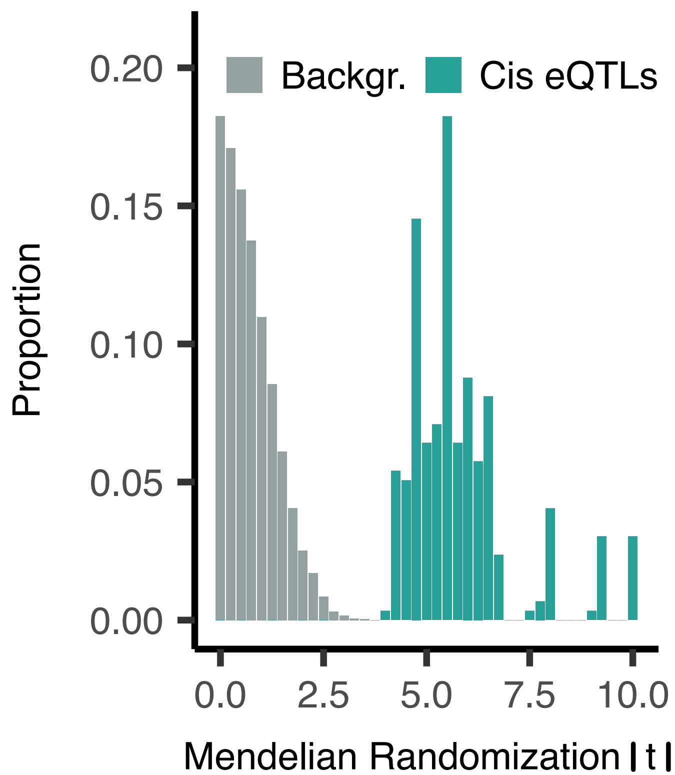
Funding sources
NIH, NHGRI, NIMH
R01 HG008150 R01 MH101814
Searle Scholar Fund

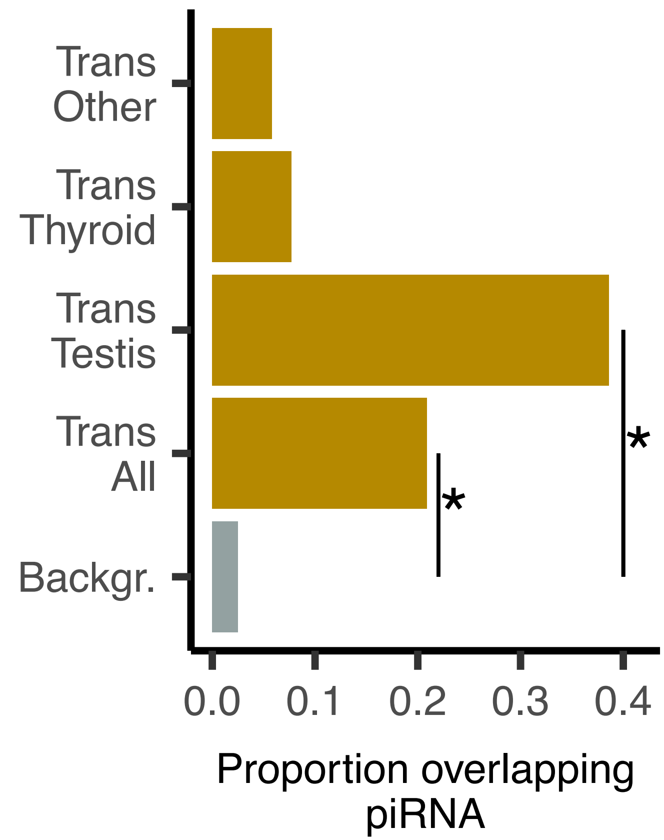
Cis-eQTLs remain to be discovered



GTEx trans-eQTLs

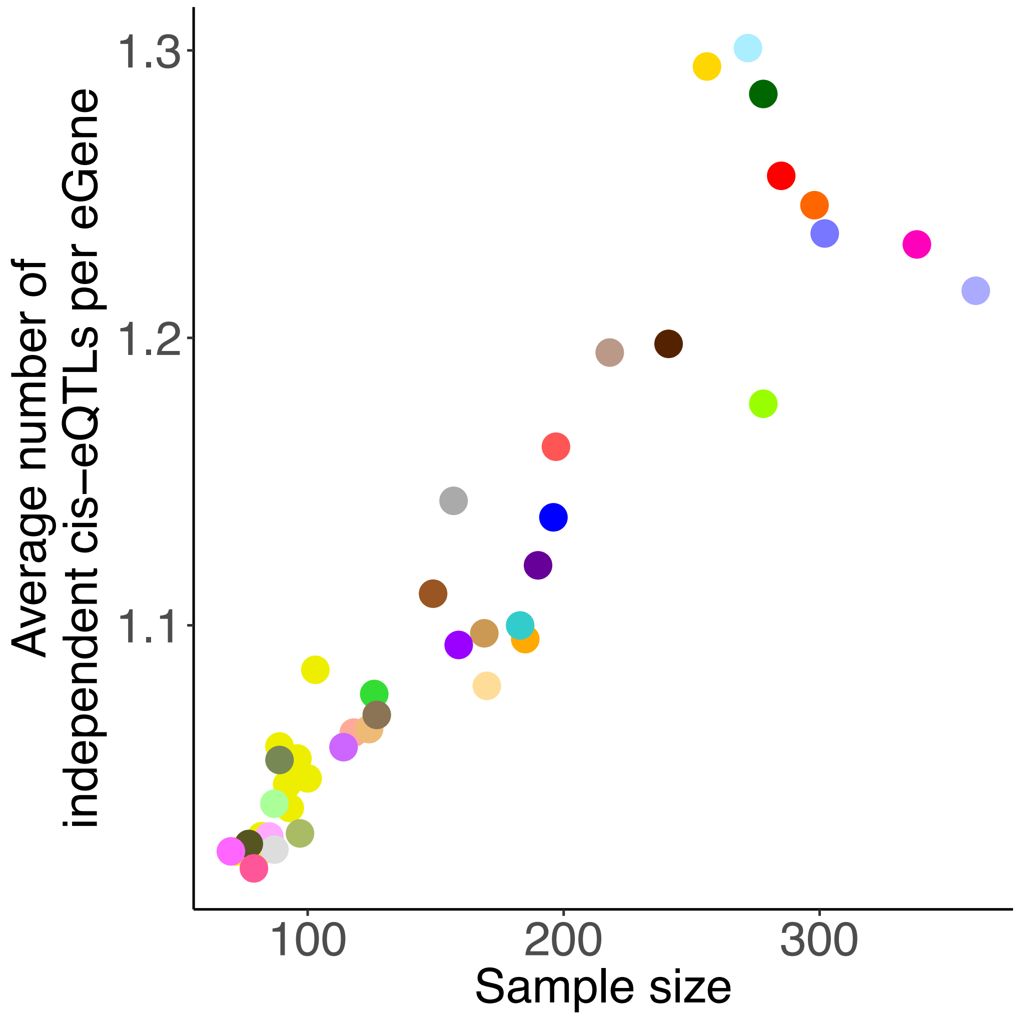
- Trans-eQTL often coincide with cis-eQTLs
- Tissue-specific mechanisms identified





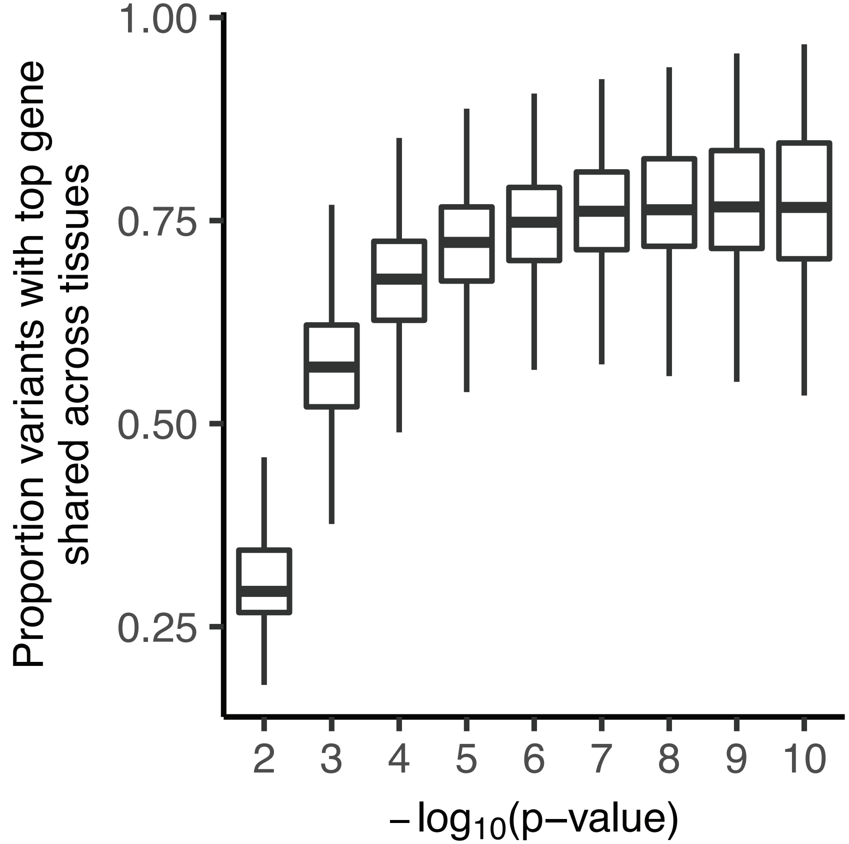
The GTEx Consortium, Nature 2017

Multiple independent SNPs per gene



Variants associated with many genes

Cis-eQTL variants have multiple gene targets, particularly once considering multiple tissues



Progress – what we've learned

- Genetics of gene expression:
  - Understand prevalence of cis-eQTLs
  - Improved eQTL catalogs based on larger studies
  - Complexity: context-specificity, allelic heterogeneity, multiple gene targets
  - Coverage of diverse variant classes and molecular phenotypes including alternative splicing
  - Rare variant effects on gene expression

Progress – what we've learned

- Connections to complex traits:
  - Better epigenetic data and eQTL catalogs provide better estimate of contribution of expression to h^2
  - Improved methods:
    - Co-localization, fine-mapping
    - Mendelian randomization approaches
    - LD-score regression and related approaches tailored for utilizing expression and epigenetic data
  - **Current estimates indicate gene expression contribute sizeable but not majority fraction to trait h^2**

Progress – what we've learned

- Contribution of expression and epigenetic data to explaining missing h^2?
  - Modestly improved power for identifying individual GWAS hits through informed priors
  - Potential improvements for prediction

- Improved interpretation and mechanism
  - Identified target genes of individual GWAS hits
  - Identified relevant tissues and cell types in aggregate

Challenges and caveats

- Ambiguity – many variants affects multiple genes in cis, in multiple tissues

- When missing the relevant cell types, genes, or environments current methods are not always interpretable

- Trans-eQTLs should be major component, but they are largely uncharacterized due to power

Key questions?

- How much heritability is explained by expression
- How much heritability is explained by epigenetics?
  - And is that all reflected in expression if measured in right tissue, right time point, right context?
- Limitations of current data?
- Limitations of current methods?
- Can expression/epigenetic data HELP explain missing heritability