Summer Institutes of Statistical Genetics, 2023

Module 6: GENE EXPRESSION PROFILING

Greg Gibson and Peng Qiu

Georgia Institute of Technology

Lecture 10: eQTL ANALYSIS

greg.gibson@biology.gatech.edu

http://www.cig.gatech.edu

Expression QTL analysis

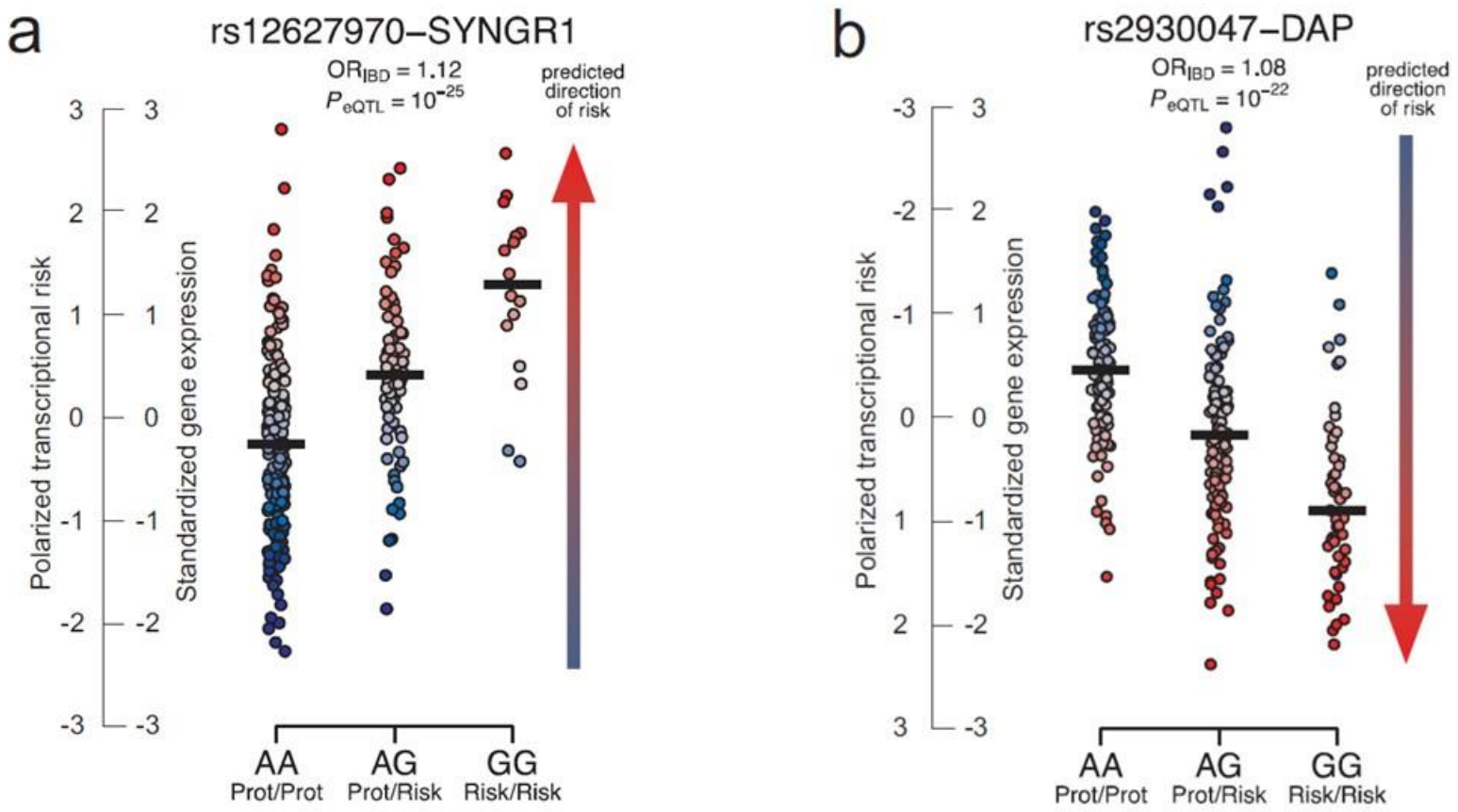
- The architecture of transcription maps genotype onto phenotype

- Expression QTL (eQTL) are QTL that modulate transcript abundance in pedigrees or crosses
- Expression SNP (eSNP) are SNPs that associate with transcript abundance in cohort studies
- GWAS variants and eSNP “often” colocalize, but it is not that simple

- At least 10% of transcripts differ in abundance between any two strains of most organisms; more than 50% across a species

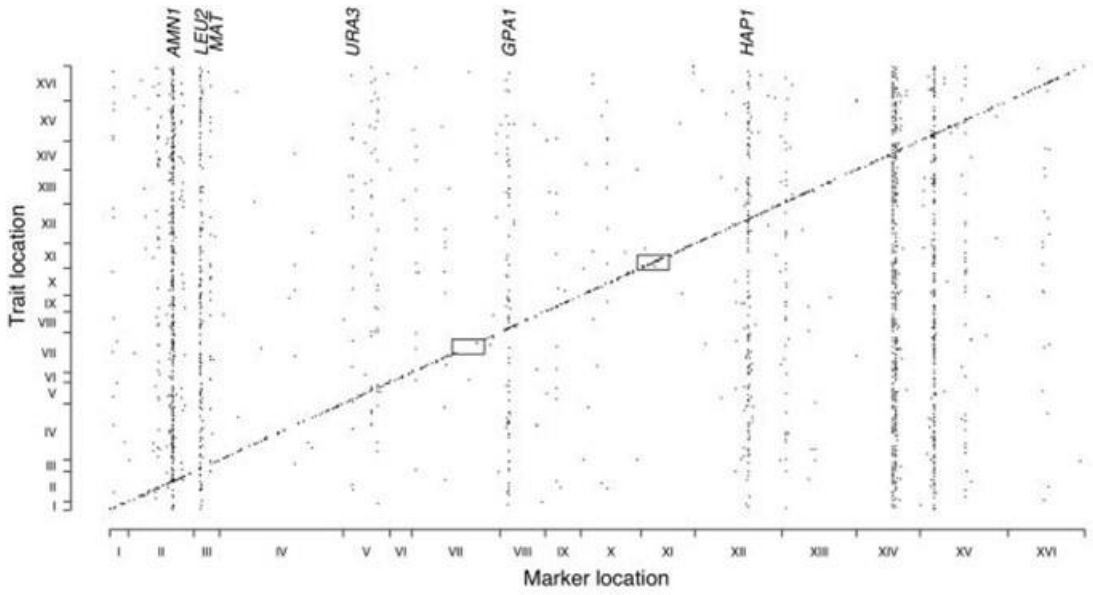
- Estimates of heritability of transcription also suggest that transcription generally shows a similar genetic component as visible traits
- One prominent eSNP may have the largest effect, but typically multiple variants at a locus will independently regulate the transcript, and overall trans-effects explain more of the variance

A couple of eSNPs



cis and *trans* eQTL

- Strong tendency for eQTL to be in *cis* to the actual gene
- Occasionally *trans*-eQTL clustered in hotspots
- Ever-larger eQTL studies refine resolution and increase number of discoveries

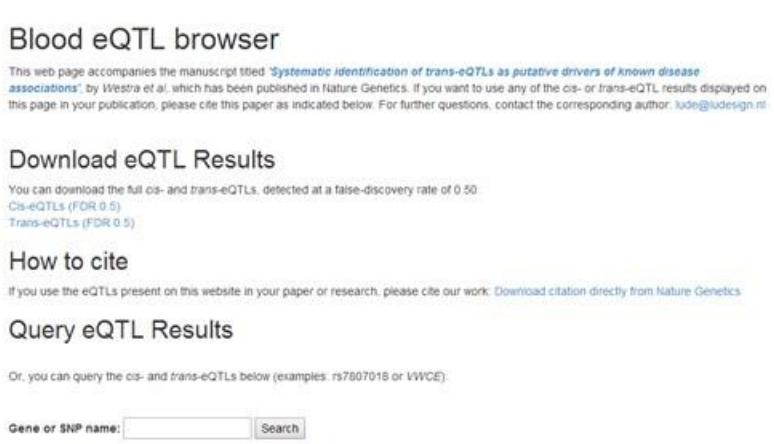


Yeast: Ronald and Akey (2007) *PLoS ONE* **2**: e678

Mice: Schadt, Friend et al (2003) *Nature* **422**: 297-302

Meta-analysis

http://genenetwork.nl/bloodeqtlbrowser/



Blood eQTL browser

This web page accompanies the manuscript titled *'Systematic identification of trans-eQTLs as putative drivers of known disease associations'*, by *Westra et al.* which has been published in Nature Genetics. If you want to use any of the *cis*- or *trans*-eQTL results displayed on this page in your publication, please cite this paper as indicated below. For further questions, contact the corresponding author: lude@ludesign.nl

Download eQTL Results

You can download the full *cis*- and *trans*-eQTLs, detected at a false-discovery rate of 0.50:
Cis-eQTLs (FDR 0.5)
Trans-eQTLs (FDR 0.5)

How to cite

If you use the eQTLs present on this website in your paper or research, please cite our work: Download citation directly from Nature Genetics

Query eQTL Results

Or, you can query the *cis*- and *trans*-eQTLs below (examples: rs7607018 or *VWCE*):

Gene or SNP name: Search

NATURE GENETICS | LETTER

日本語要約

Systematic identification of *trans* eQTLs as putative drivers of known disease associations

Harm-Jan Westra, Marjolein J Peters, Tõnu Esko, Hanieh Yaghootkar, Claudia Schurmann, Johannes Kettunen, Mark W Christiansen, Benjamin P Fairfax, Katharina Schramm, Joseph E Powell, Alexandra Zhernakova, Daria V Zhernakova, Jan H Veldink, Leonard H Van den Berg, Juha Karjalainen, Sebo Withoff, André G Uitterlinden, Albert Hofman, Fernando Rivadeneira, Peter A C 't Hoen, Eva Reinmaa, Krista Fischer, Mari Nelis, Lili Milani, David Melzer + *et al.*

eQTL meta-analysis on 5,311 individuals replicated in 2,775 more

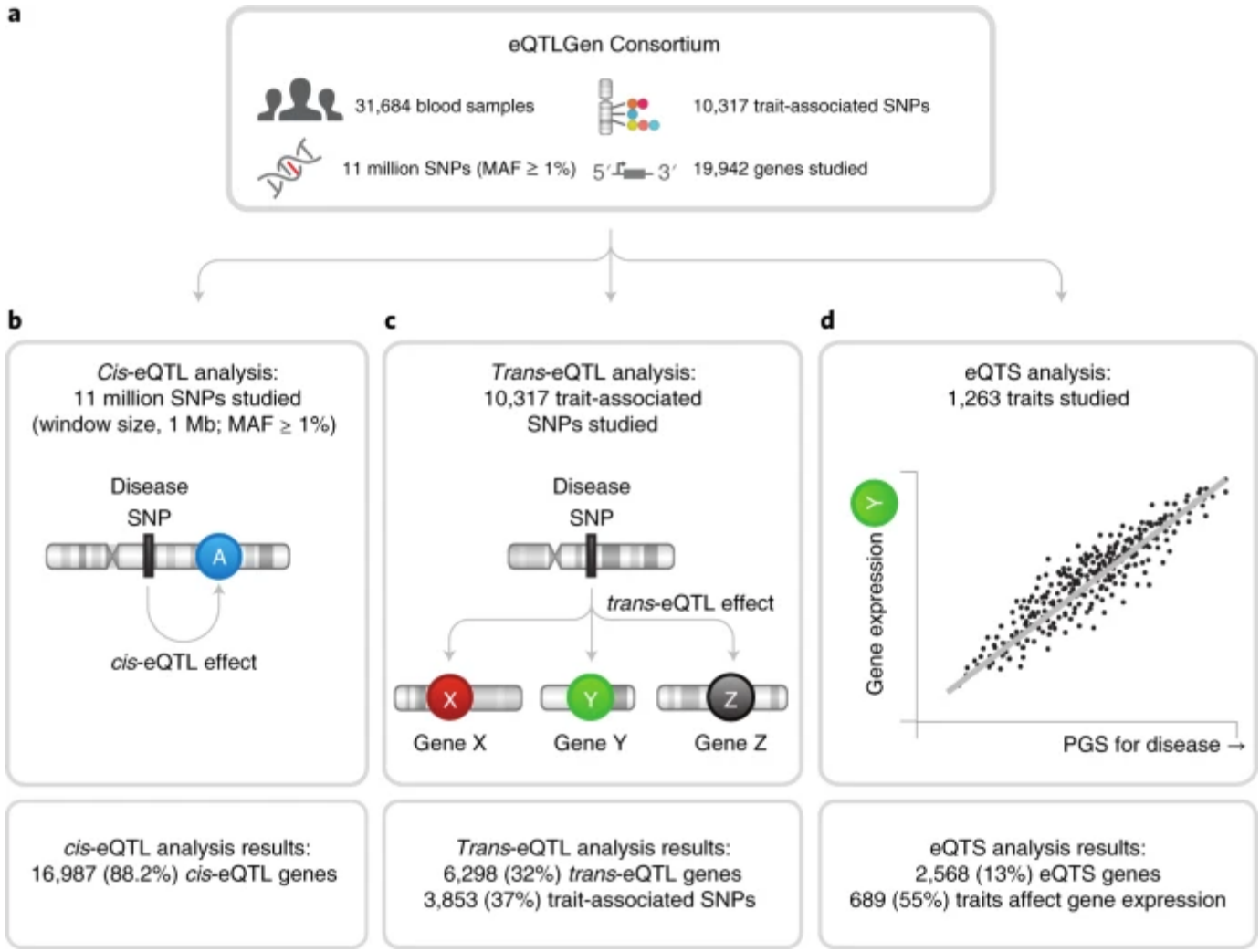
Found trans-eQTL for 233 SNPs at 103 loci many of which are also disease QTL

Also generates local cis-eSNPs for almost half the genome

Westra et al. (2013) *Nature Genetics* **45**: 1238–1243

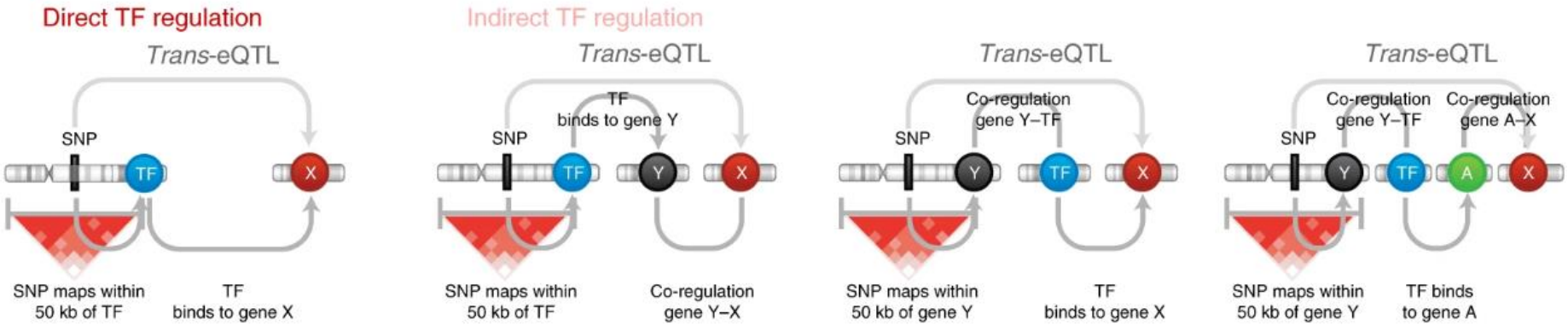
eQTL-Gen Consortium

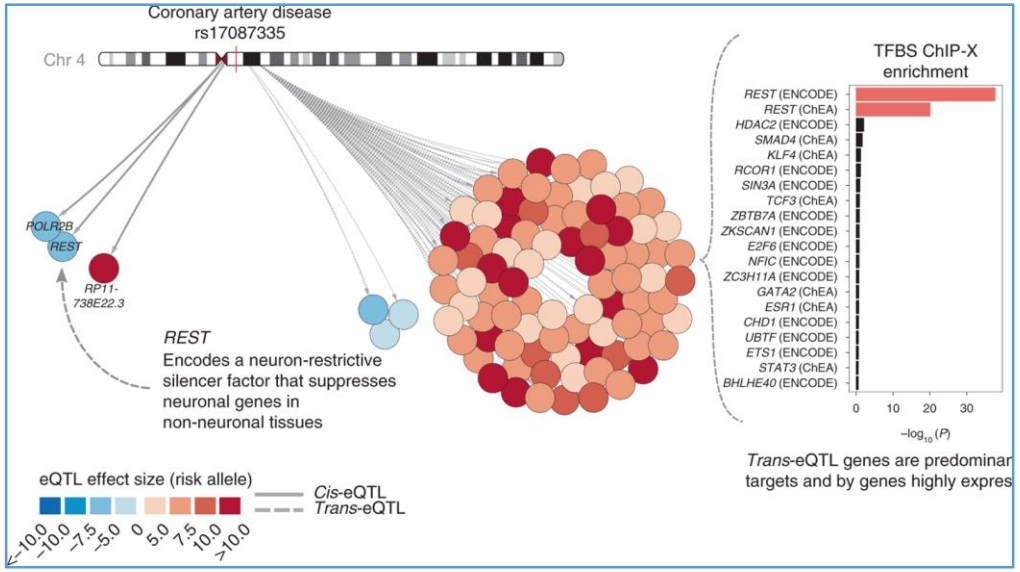
eQTL meta-analysis on 31,684 individuals from 37 cohorts

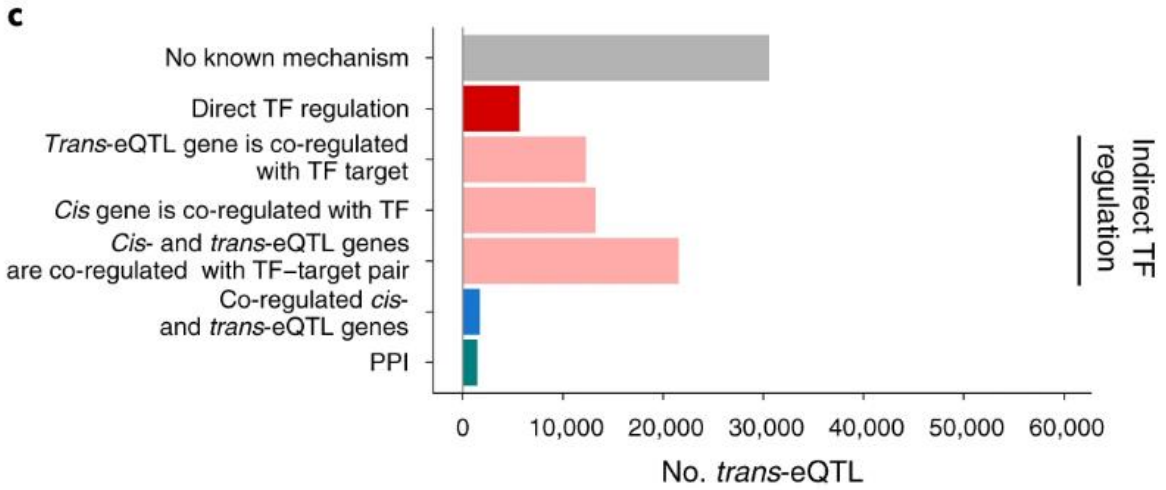


Võsa et al. (2021) *Nature Genetics* **53**: 1300–1310

Mechanisms of trans-eQTL effects







Võsa et al. (2021) *Nature Genetics* **53**: 1300–1310

Single cell eQTL: The OneK1K Study

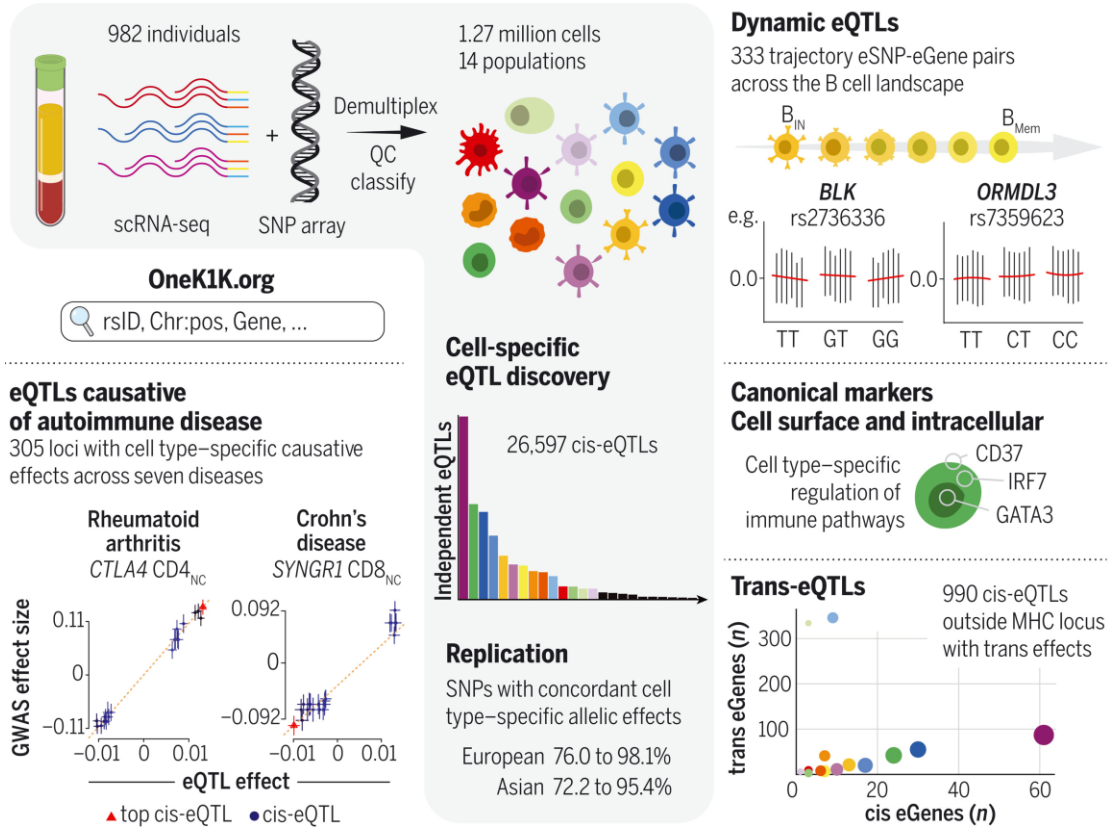
1.27M peripheral blood immune cells of 14 types, from 982 adult donors -> 26,597 *cis*-eQTL and 990 *trans*-eQTL

19% of the eQTL co-localize with an immune disorder GWAS association; MR confirms causal effects for 305 loci

333 of 1988 eQTL in B cells change during maturation from naïve to memory, 66% of which found by interaction

Majority of effects cell-type specific, but
common effects tend to be concordant

High level of replication in Asian dataset



Yazar et al (2022) *Science 3***76**: abf3041

GTEx (Genotype-Tissue-Expression Project)

The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans

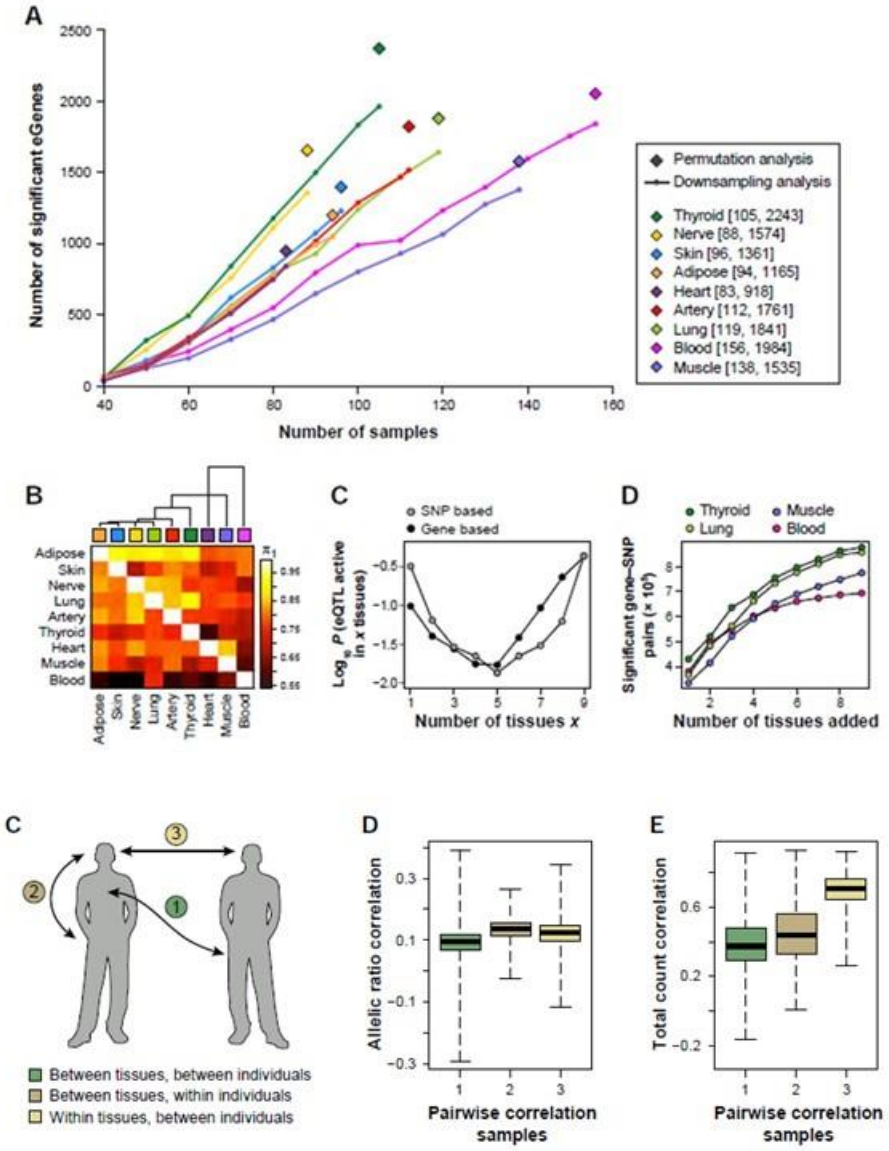
The GTEx Consortium[*,†]

[+] Author Affiliations

[*]Corresponding author: Kristin G. Ardlie (kardlie@broadinstitute.org) or Emmanouil T. Dermitzakis (emmanouil.dermitzakis@unige.ch)

ABSTRACT | EDITOR'S SUMMARY

Understanding the functional consequences of genetic variation, and how it affects complex human disease and quantitative traits, remains a critical challenge for biomedicine. We present an analysis of RNA sequencing data from 1641 samples across 43 tissues from 175 individuals, generated as part of the pilot phase of the Genotype-Tissue Expression (GTEx) project. We describe the landscape of gene expression across tissues, catalog thousands of tissue-specific and shared regulatory expression quantitative trait loci (eQTL) variants, describe complex network relationships, and identify signals from genome-wide association studies explained by eQTLs. These findings provide a systematic understanding of the cellular and biological consequences of human genetic variation and of the heterogeneity of such effects among a diverse set of human tissues.



Science (2015) **9(5)**: e1003486

Some software

PLINK: The basic tool for GWAS
http://pngu.mgh.harvard.edu/~purcell/plink/tutorial.shtml

Matrix eQTL: Ultra-fast eQTL analysis
http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/

GEMMA: Genome-wide Efficient Mixed Model Association (GEMMA)
http://stephenslab.uchicago.edu/software.html#gemma

FMeQTL: Bayesian Joint mapping
https://github.com/xqwen/fmeqtl

DAP: Deterministic Approcimation of Posteriors (Fast Bayesian)
https://github.com/xqwen/dap

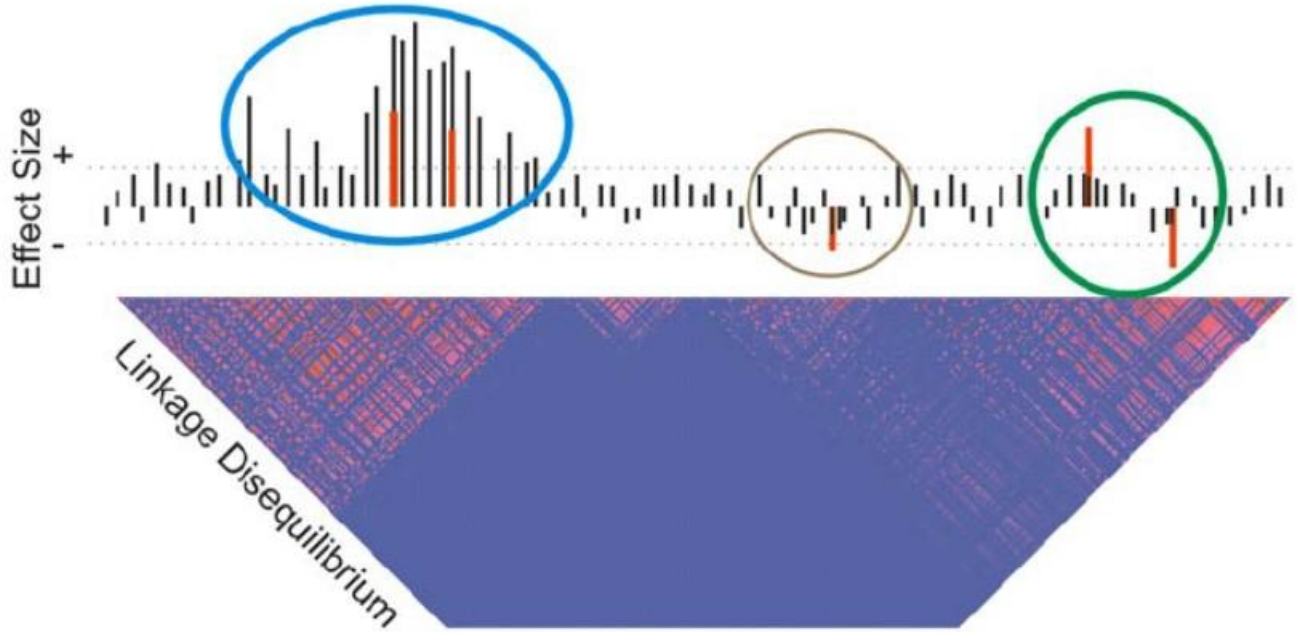
CAVIAR: Bayesian Fine Mapping
http://genetics.cs.ucla.edu/caviar/

Ventham et al (2016) *Nature Communications* **7**: 13507

Why Colocalized Signals do not alone imply Causation

Sampling variance means that we can only map “credible intervals”
Many genes harbor multiple eSNPs, and possibly multiple trait associated SNPs
LD means that multiple sites can interfere with one another in estimation of peak locations

The nearest gene is only sometimes the one affected by a SNP!



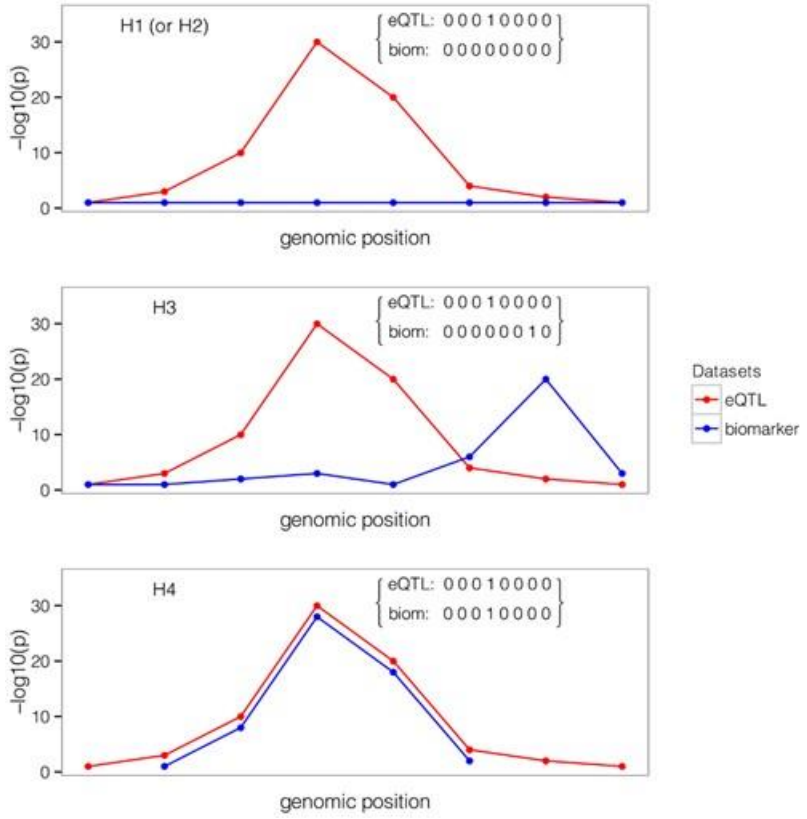
Coloc: A Bayesian test for colocalization of pairs of association signals

H1 is the hypothesis that there is only an eQTL signal at a locus

H2 is the hypothesis that there is only a GWAS signal at a locus.

H3 is the hypothesis that there are two independent eQTL and GWAS signals in linkage.

H4 is the strong hypothesis that the same SNP (not just the locus) is responsible for both the GWAS and eQTL.



Giambartolomei et al (2014) *PLoS Genetics* **10(5)**: e1004383

Examples of H3 and H4

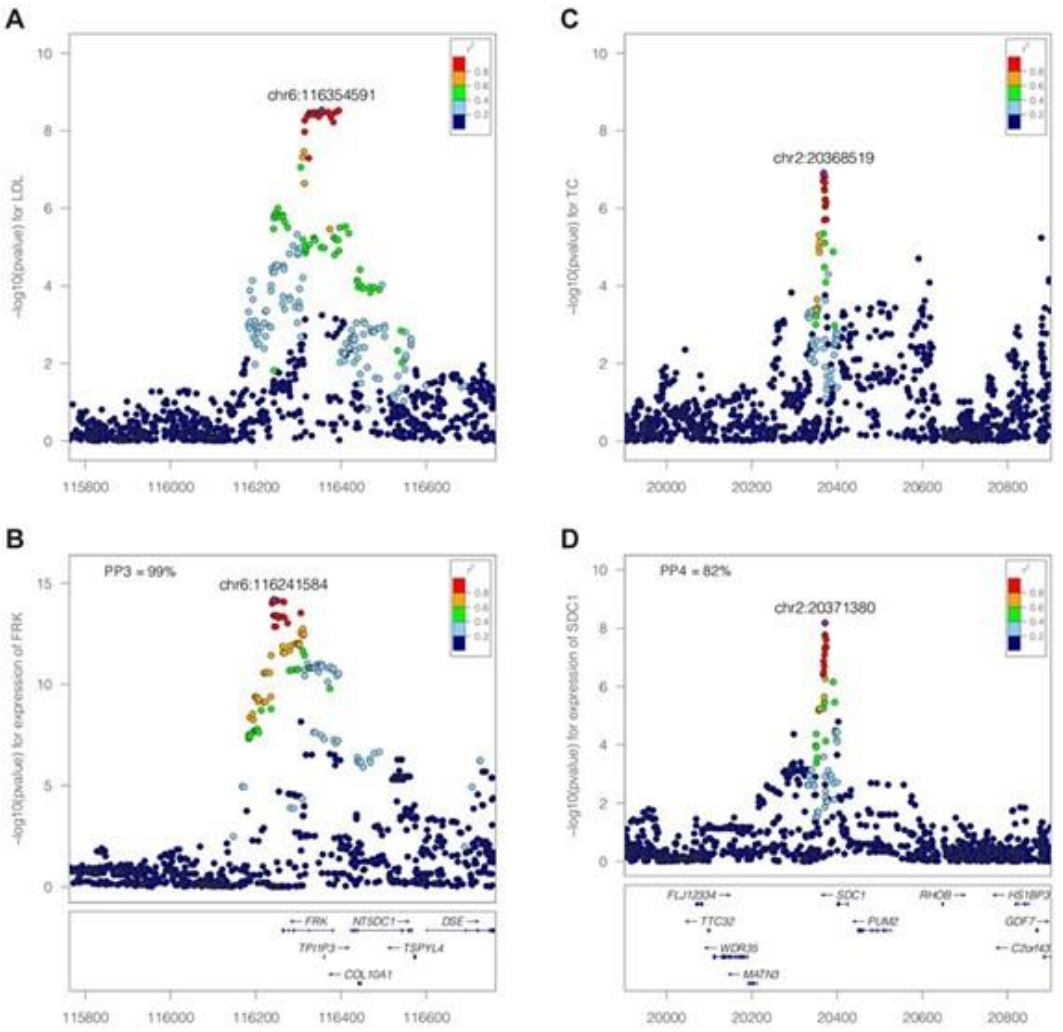
On the left, the profile of association at the FRK locus with LDL (top) is very different from that with *FRK* expression.

H3 is the supported hypothesis.

On the right, even though there are two different peak SNPs, they are in the same strong LD region and the profiles are almost the same for Total Cholesterol and *Soc1* expression.

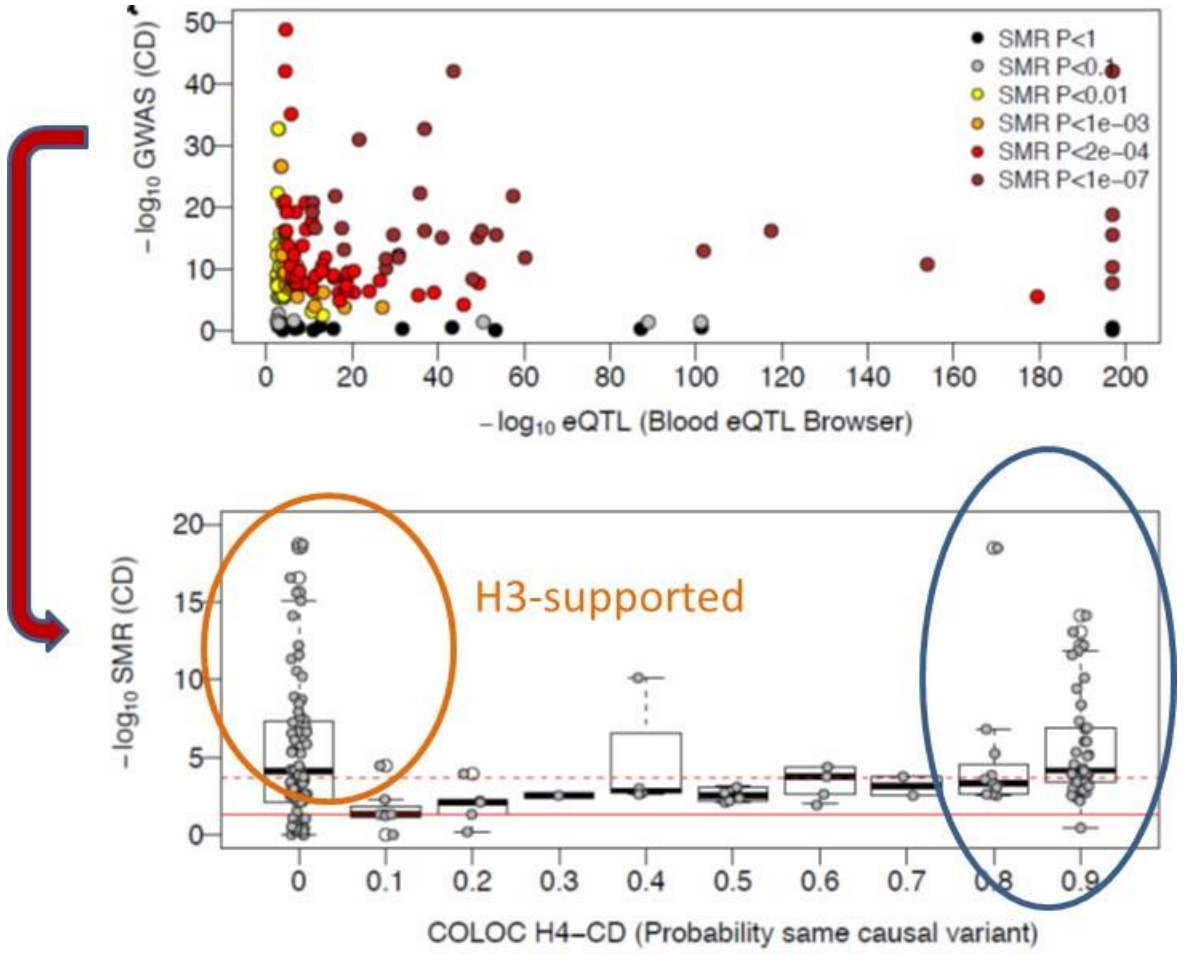
H4 is the supported hypothesis.

Bayesian analysis evaluate each H relative to the other four and generates a confidence level for the most likely one.



Giambartolomei et al (2014) *PLoS Genetics* **10(5)**: e1004383

SMR and coloc are complementary



H4-supported

Limitations of colocalization analyses

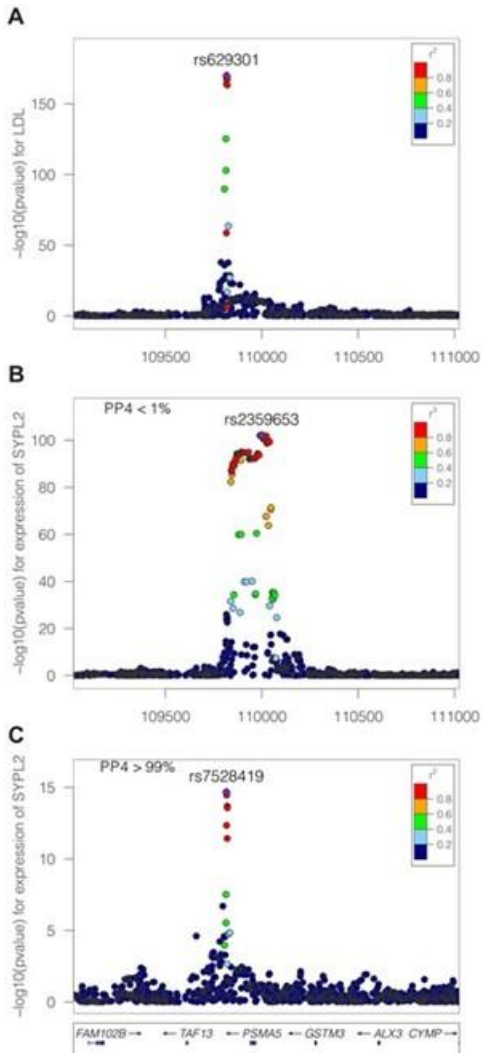
Heavily dependent on statistical power of the contributing analyses, which is generally relatively low

Depends on high quality imputation if the SNPs are not directly typed

Assumes that the GWAS and eQTL are evaluated on the same population (there is no stratification)

A negative result may arise if the incorrect tissue is being studied for the gene expression

Assumes there is a single causal variant at each locus for each effect (which is very unlikely) although this example shows that conditional analysis has the potential to resolve joint effects



Giambartolomei et al (2014) *PLoS Genetics* **10(5)**: e1004383

Joint Mapping

A variety of open source methods are appearing that utilize Bayesian methods to perform joint mapping of eQTL

A statistical framework for joint eQTL analysis in multiple tissues.
Flutre T, Wen X, Pritchard J, Stephens M. *PLoS Genet*. 2013 **9**(5): e1003486.

This paper shows that combining signals across tissues increases power while also allowing assessment of whether the effect sizes are different in different cell types. Implemented in eQTLBMA software.

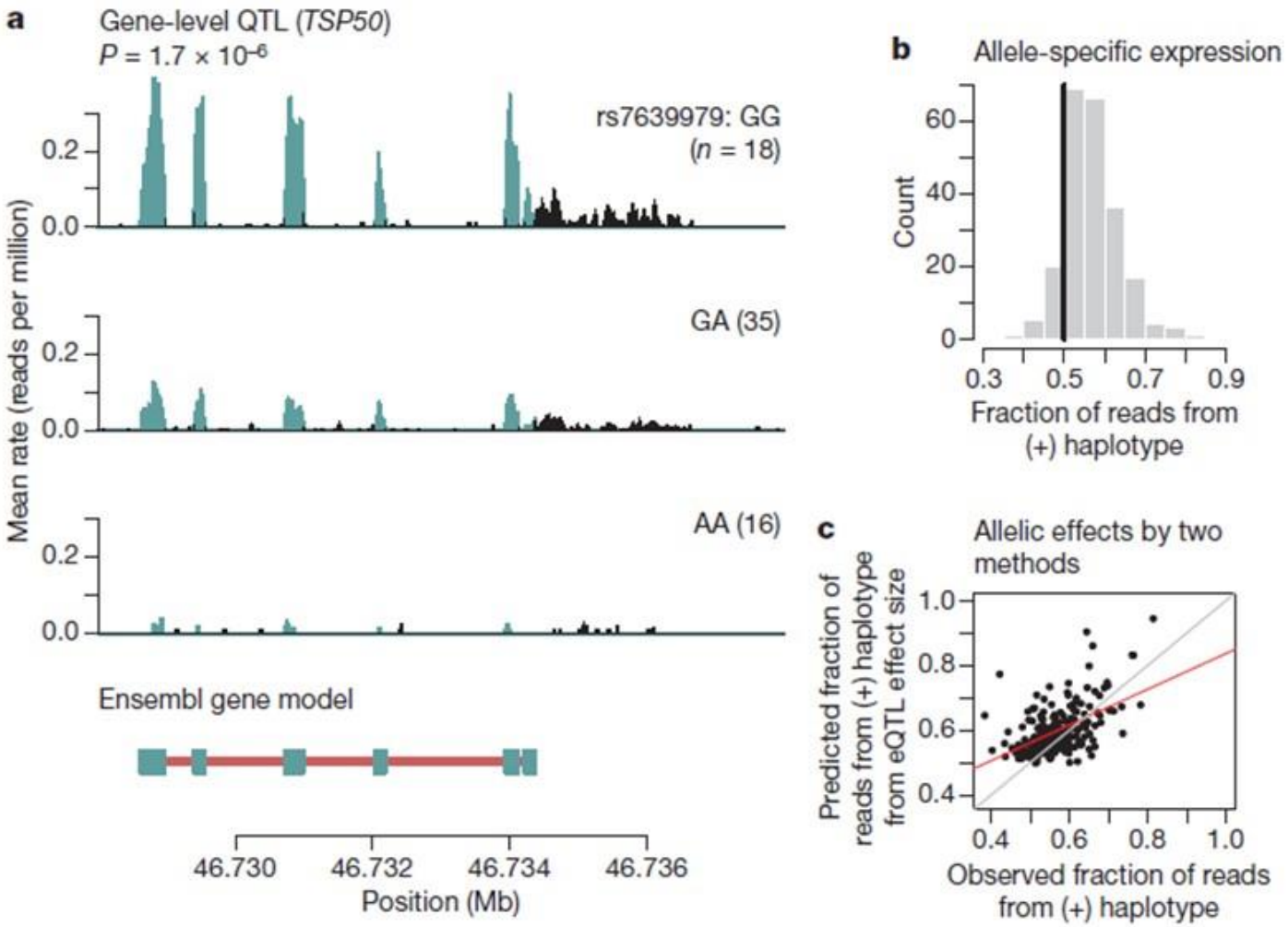
Cross-population joint analysis of eQTLs: Fine mapping and functional annotation.
Wen X, Luca F, Pique-Regi R. *PLoS Genet.* **11**(4): e1005176.

This paper shows that combining signals across populations increases power while also allowing assessment of how incorporating ENCODE data improves resolution. Implemented in FM QTL software.

Efficient integrative multi-SNP association analysis via Deterministic Approximation of Posteriors
Wen X, Lee Y, Luca F, Pique-Regi R. *AM J Hum. Genet.* **98**(6): 1114-1129.

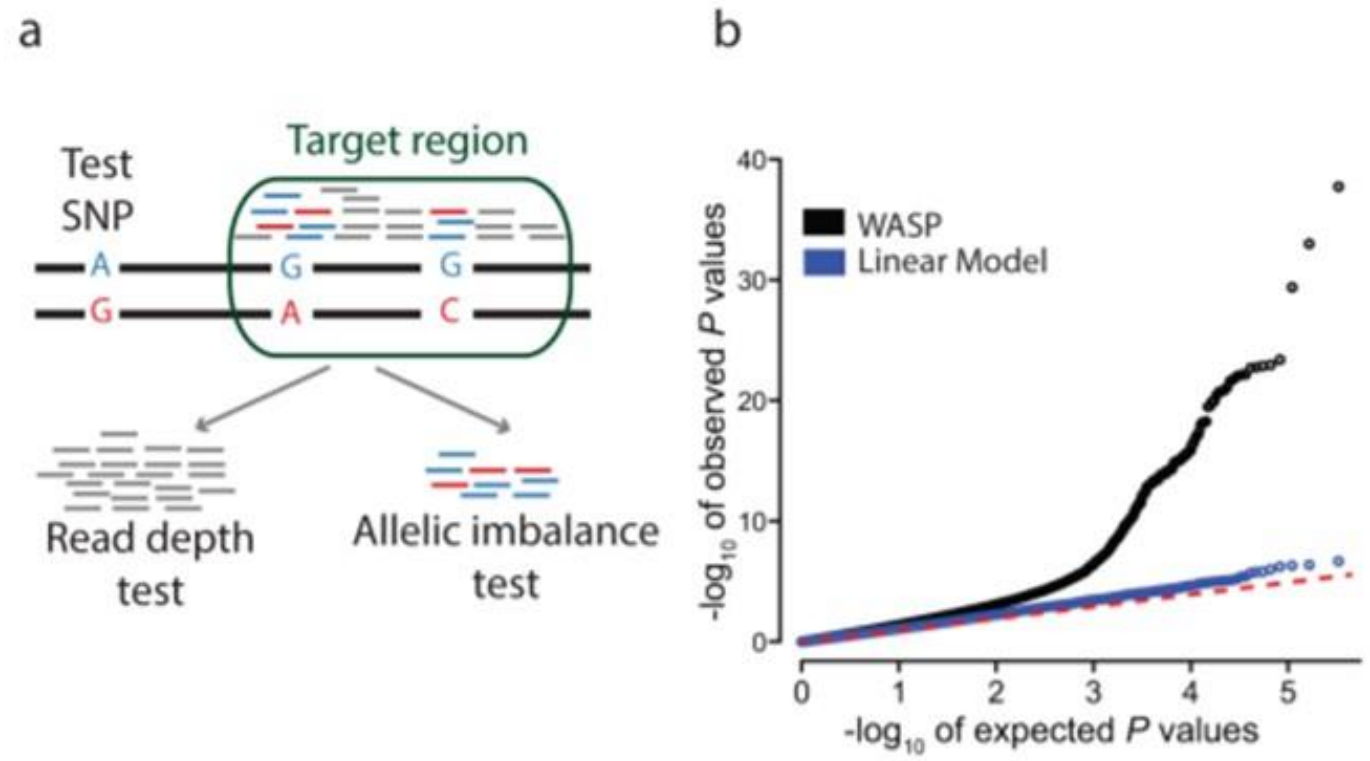
This paper extends the framework for incorporating ENCODE data while allowing for multiple causal variants at each locus. Implemented in DAP software: http://github.com/xqwen/dap/

eQTL and Additivity



Pickrell et al. (2010) *Nature* **464**: 768-772

WASP

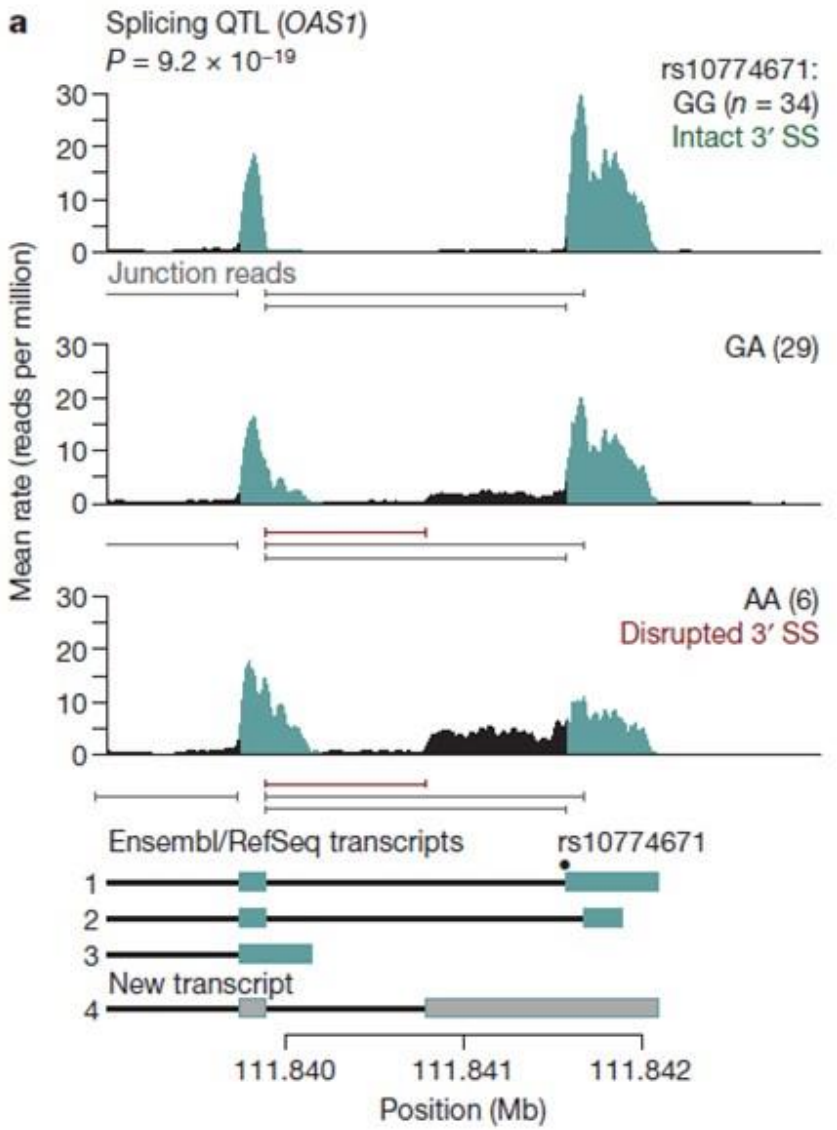


Formulates a CHT: Combined Haplotype Test
“The CHT jointly models two components: the allelic imbalance at phased heterozygous SNPs and the total read depth in the target region”

Van de Geijn, *et al* (2015) *Nature Methods* **12**: 1061-1063

https://github.com/bmvdgeijn/WASP

sQTL (Splicing QTL)

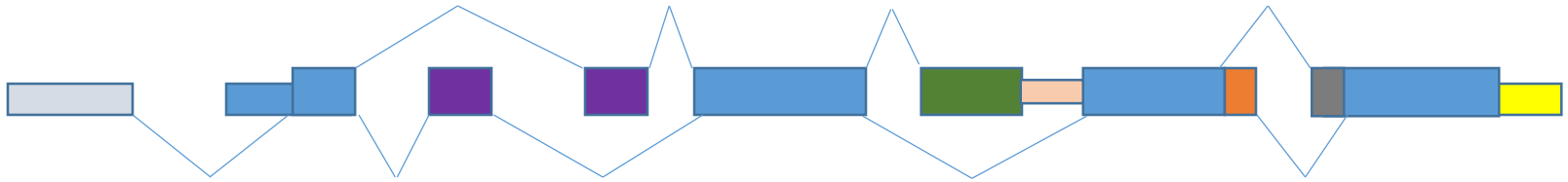


Pickrell et al. (2010) *Nature* **464**: 768-772

MISO: Mixture of Isoforms

http://genes.mit.edu/burgelab/miso/

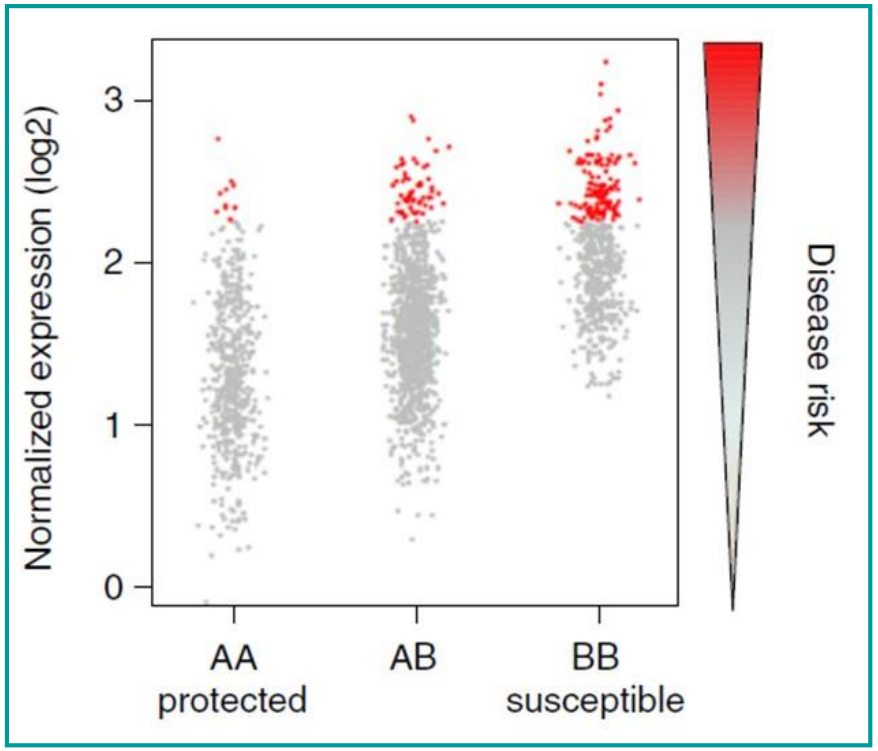
Estimates the Percent Spliced In (PSI, ψ) for various features:

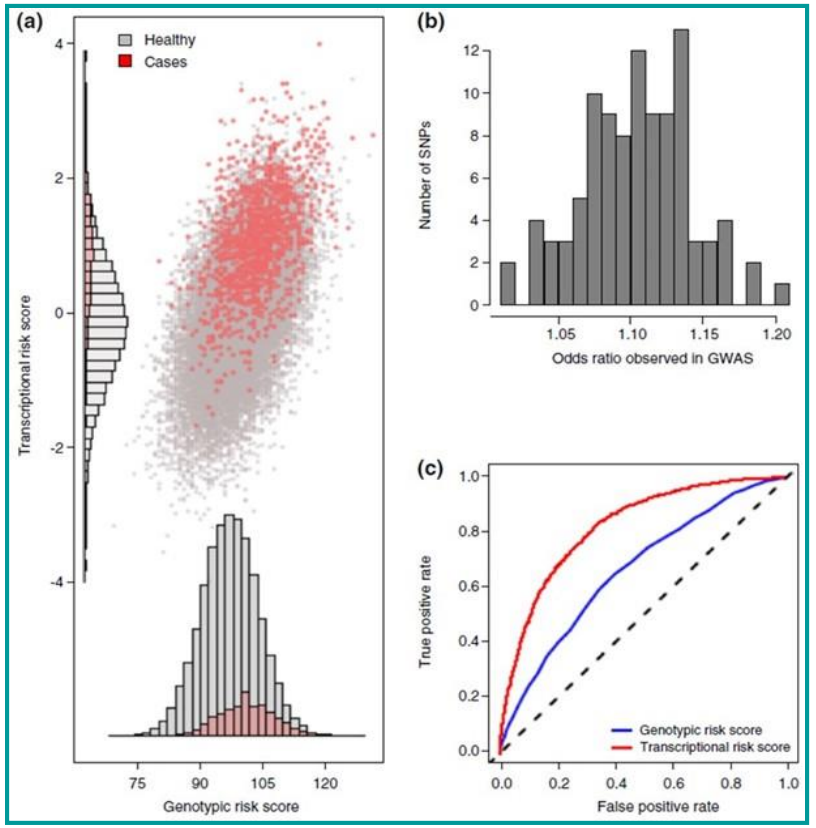


- Alternate First Exon
- Mutually exclusive Exon
- Excluded Exon
- Included Intron
- Alternate Splice Donor
- Alternate Splice Acceptor
- Alternate 3’ end

Katz et al. (2010) *Nature Methods* **7**: 109–115

Transcriptional Risk Scores - theory

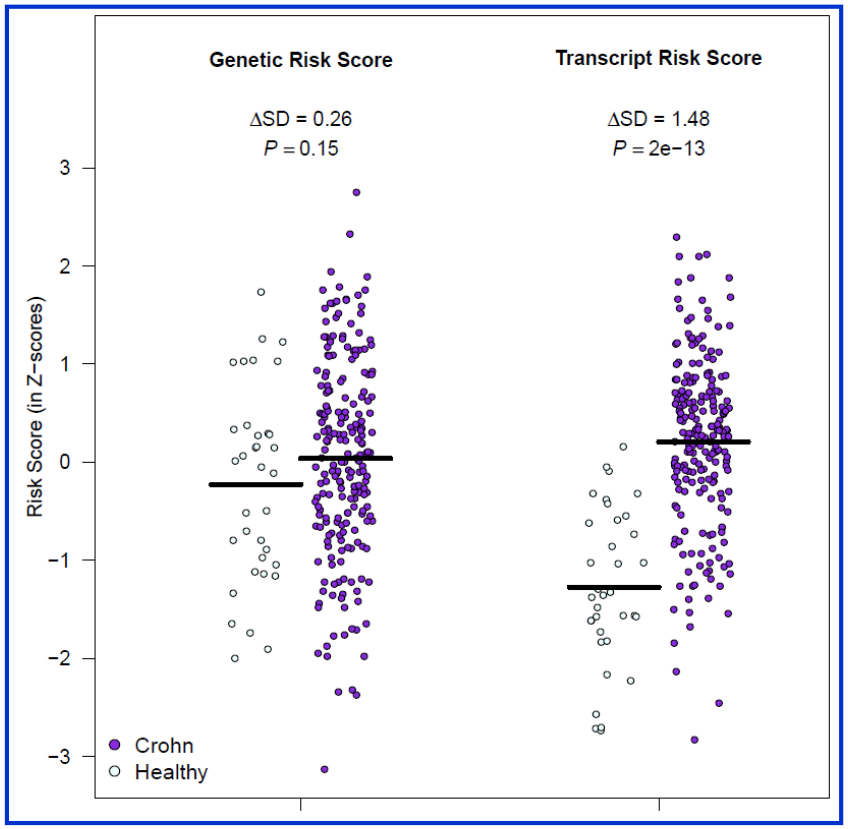




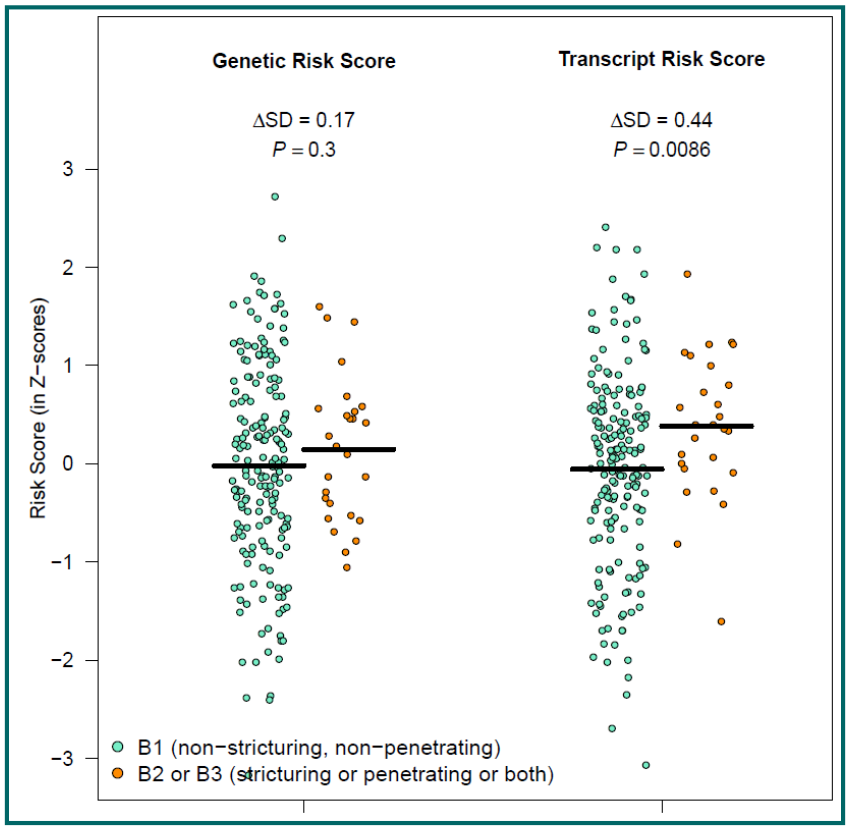
Gibson et al (2015) *Genome Medicine* **7**: 60

Transcriptional Risk Scores for Crohn's Disease

Healthy-Disease



Disease Progression



Marigorta et al (2017) *Nature Genetics* **49**: 1517-1521