

# Clinical Utility of Pharmacogenomic Findings: Beyond Single Variants

Hae Kyung Im, PhD



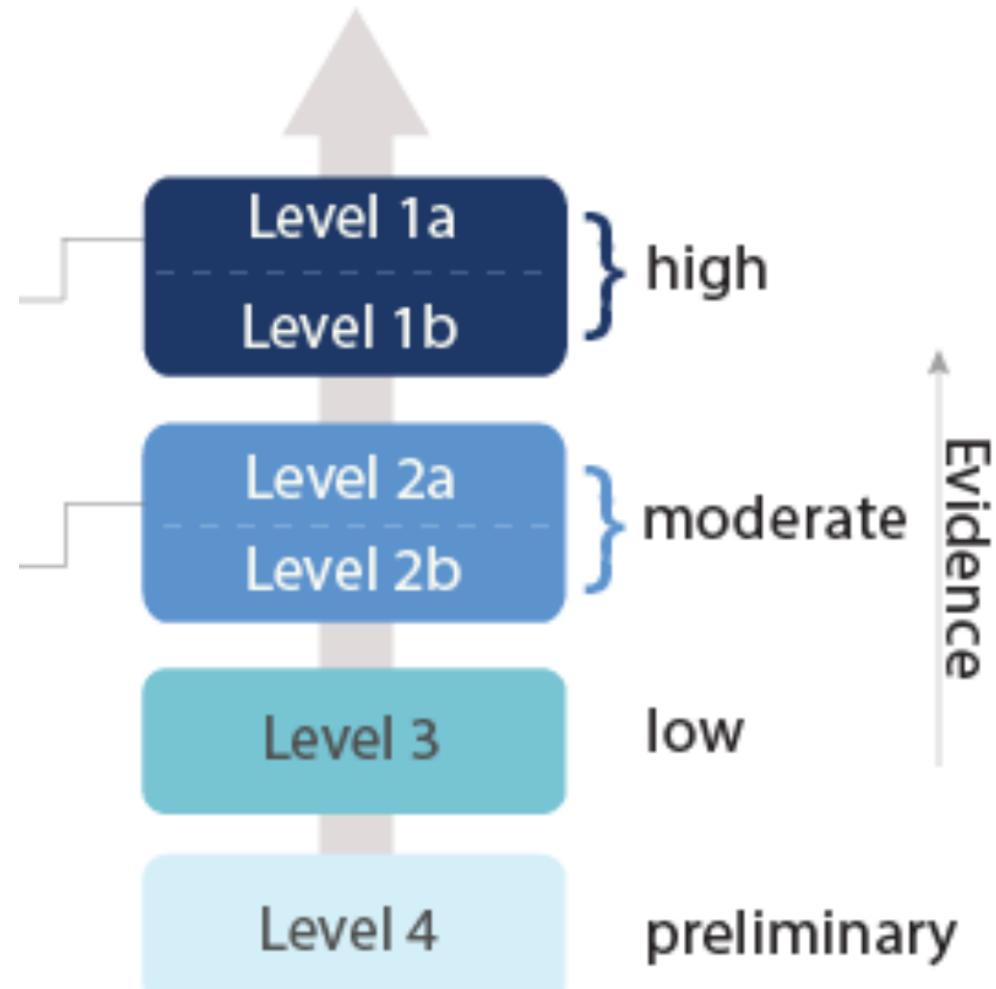
THE UNIVERSITY OF  
**CHICAGO**

# Successes and Challenges of Genome Studies

- GWAS/Sequencing
  - 10K robustly associated genetic variants
  - New insights into biology of many traits
  - Biological understanding is still lacking

# Pharmacogenomic Findings

Evidence Level	Counts	%
1a	40	3
1b	17	1
2a	96	6
2b	74	5
3	1175	76
4	145	9
Total	1547	100



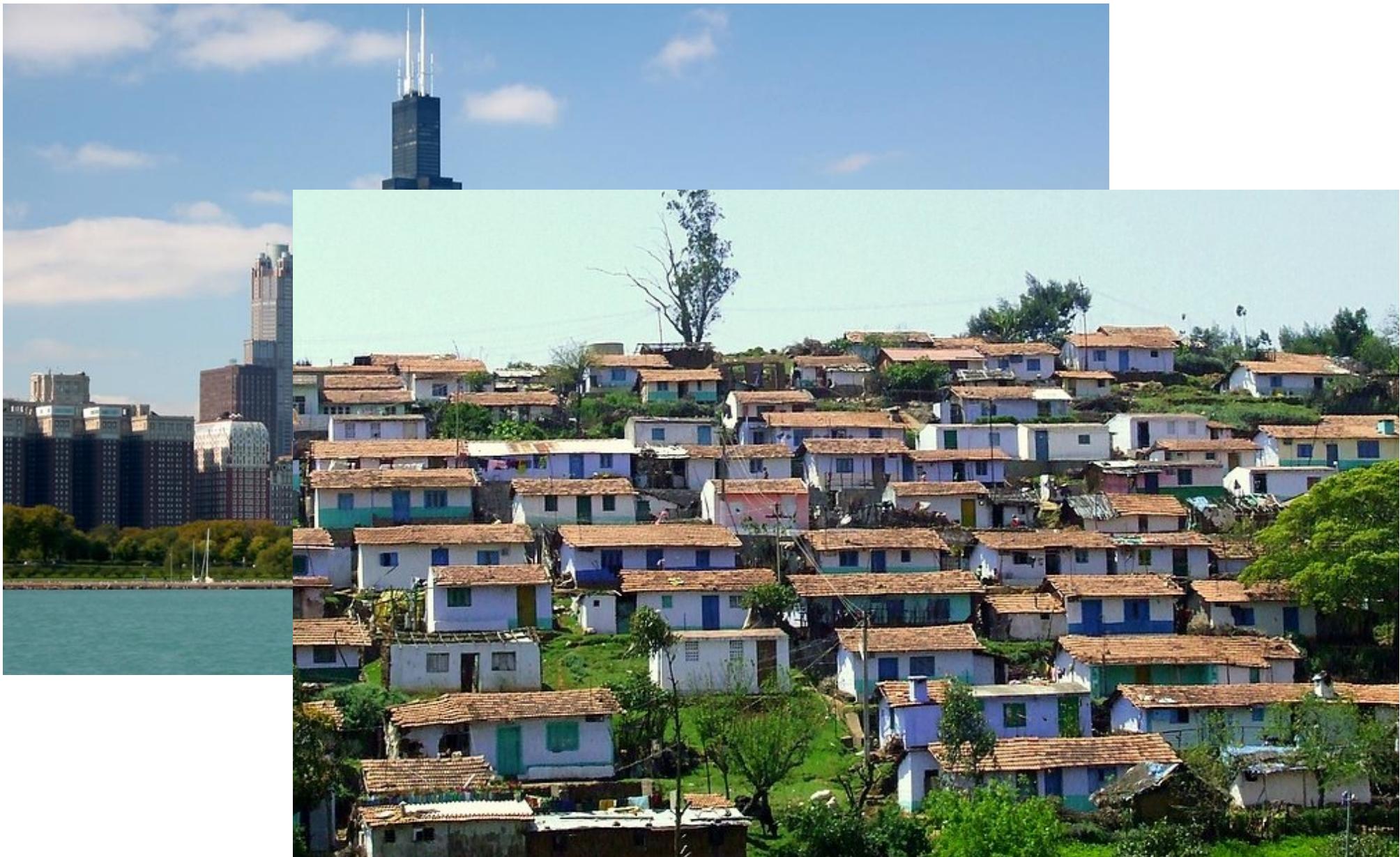
Only Level 1a finding have clinical guidelines

<https://www.pharmgkb.org/>

# Genetic Architecture of Complex Traits



# Genetic Architecture of Complex Traits



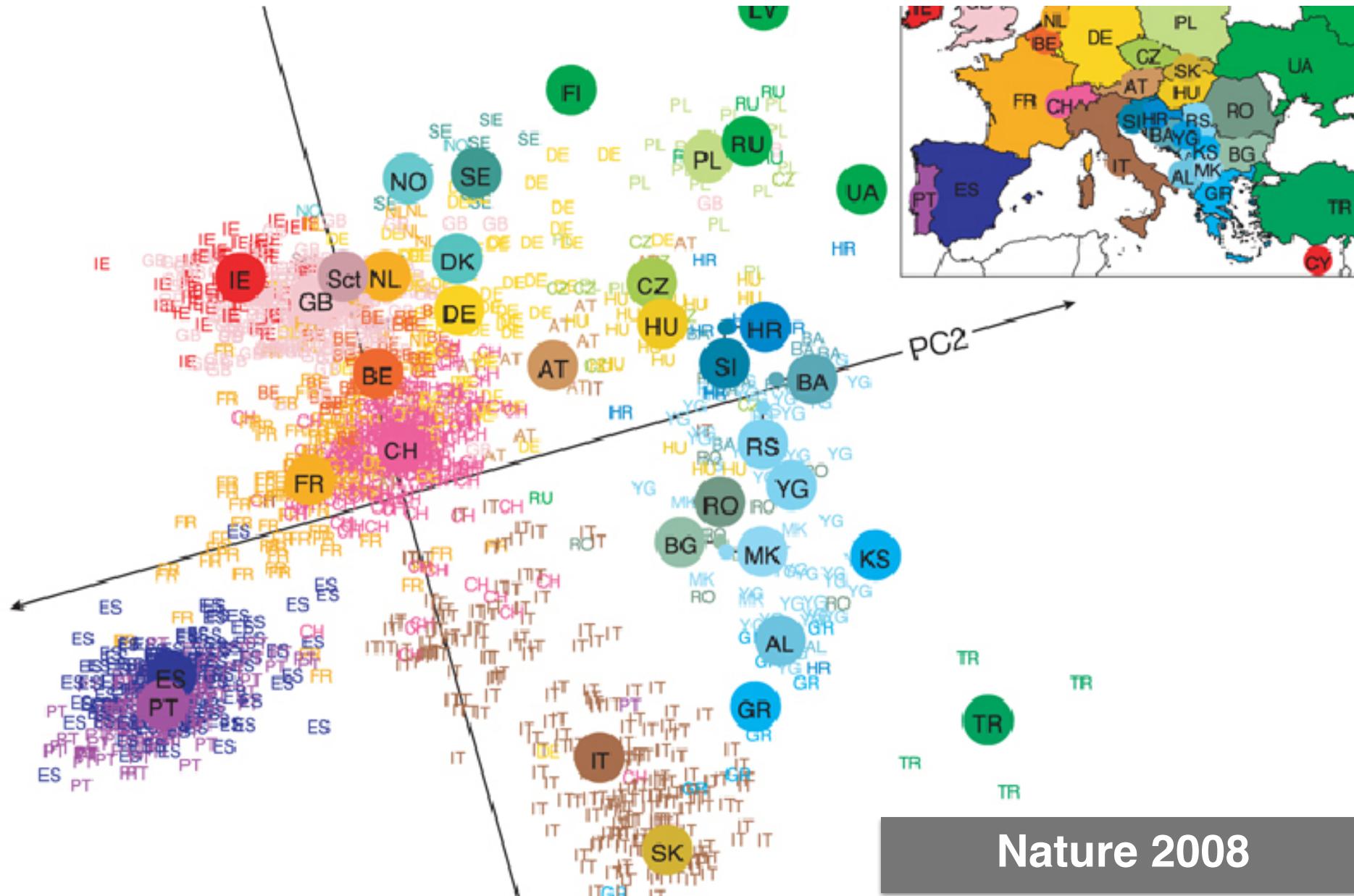
# Genetic Architecture of Complex Traits

**Single Variants Not Relevant for  
Highly Polygenic Traits**



# Genes mirror geography within Europe

John Novembre<sup>1,2</sup>, Toby Johnson<sup>4,5,6</sup>, Katarzyna Bryc<sup>7</sup>, Zoltán Kutalik<sup>4,6</sup>, Adam R. Boyko<sup>7</sup>, Adam Auton<sup>7</sup>, Amit Indap<sup>7</sup>, Karen S. King<sup>8</sup>, Sven Bergmann<sup>4,6</sup>, Matthew R. Nelson<sup>8</sup>, Matthew Stephens<sup>2,3</sup> & Carlos D. Bustamante<sup>7</sup>



# Prediction and Dissection to Achieve Clinical Utility

## Prediction

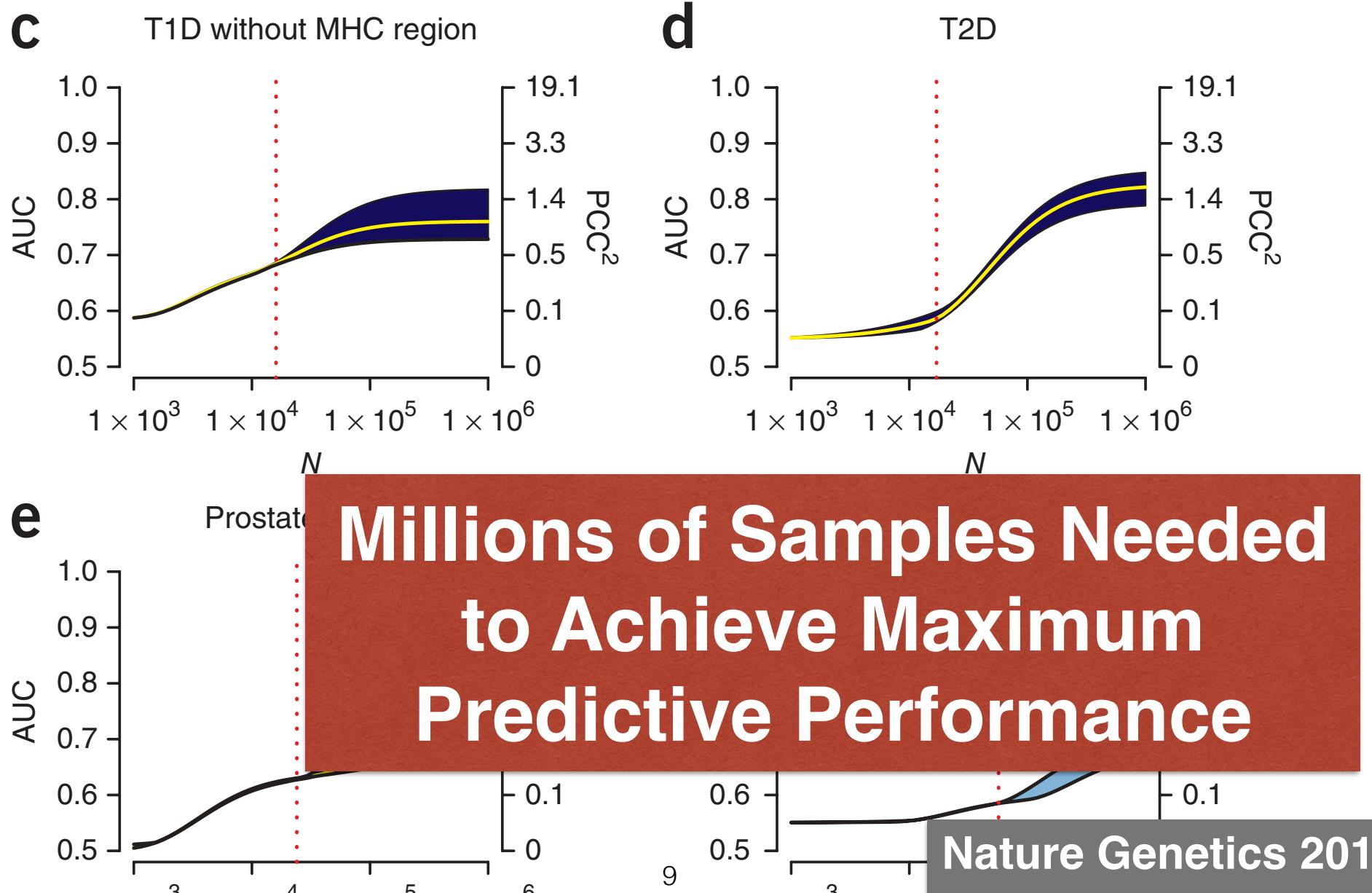
- Disease
  - risk stratification
  - intervention strategies
- Adverse events
- Efficacy of treatment

## Dissection

- Etiology of complex traits
- Mechanism by which genetic variation drives phenotypic variation
- Druggable targets

# Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies

Nilanjan Chatterjee<sup>1</sup>, Bill Wheeler<sup>2</sup>, Joshua Sampson<sup>1</sup>, Patricia Hartge<sup>1</sup>, Stephen J Chanock<sup>1</sup> & Ju-Hyun Park<sup>1,3</sup>



# Whole Genome Prediction Approaches

## LETTERS

### **Common polygenic variation contributes to risk of schizophrenia and bipolar disorder**

The International Schizophrenia Consortium\*

## REPORT

### GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,<sup>1,\*</sup> S. Hong Lee,<sup>1</sup> Michael E. Goddard,<sup>2,3</sup> and Peter M. Visscher<sup>1</sup>

# Whole Genome Prediction Approaches

RESEARCH ARTICLE

Genetic  
Epidemiology

OFFICIAL JOURNAL

INTERNATIONAL GENETIC  
EPIDEMIOLOGY SOCIETY  
[www.geneticepi.org](http://www.geneticepi.org)

## Poly-Omic Prediction of Complex Traits: OmicKriging

Heather E. Wheeler,<sup>1</sup> Keston Aquino-Michaels,<sup>2</sup> Eric R. Gamazon,<sup>2</sup> Vassily V. Trubetskoy,<sup>2</sup> M. Eileen Dolan,<sup>1</sup> R. Stephanie Huang,<sup>1</sup> Nancy J. Cox,<sup>2</sup> and Hae Kyung Im<sup>3\*</sup>

## MultiBLUP: improved SNP-based prediction for complex traits

Doug Speed and David J Balding

*Genome Res.* published online June 24, 2014  
Access the most recent version at doi:[10.1101/gr.169375.113](https://doi.org/10.1101/gr.169375.113)

OPEN  ACCESS Freely available online

 PLOS | GENETICS

## Polygenic Modeling with Bayesian Sparse Linear Mixed Models

Xiang Zhou<sup>1\*</sup>, Peter Carbonetto<sup>1</sup>, Matthew Stephens<sup>1,2\*</sup>

# Whole Genome Prediction Approaches

*J. R. Statist. Soc. B* (2005)  
**67**, Part 2, pp. 301–320

## Regularization and variable selection via the elastic net

Hui Zou and Trevor Hastie

Abraham et al. *BMC Bioinformatics* 2012, **13**:88  
<http://www.biomedcentral.com/1471-2105/13/88>



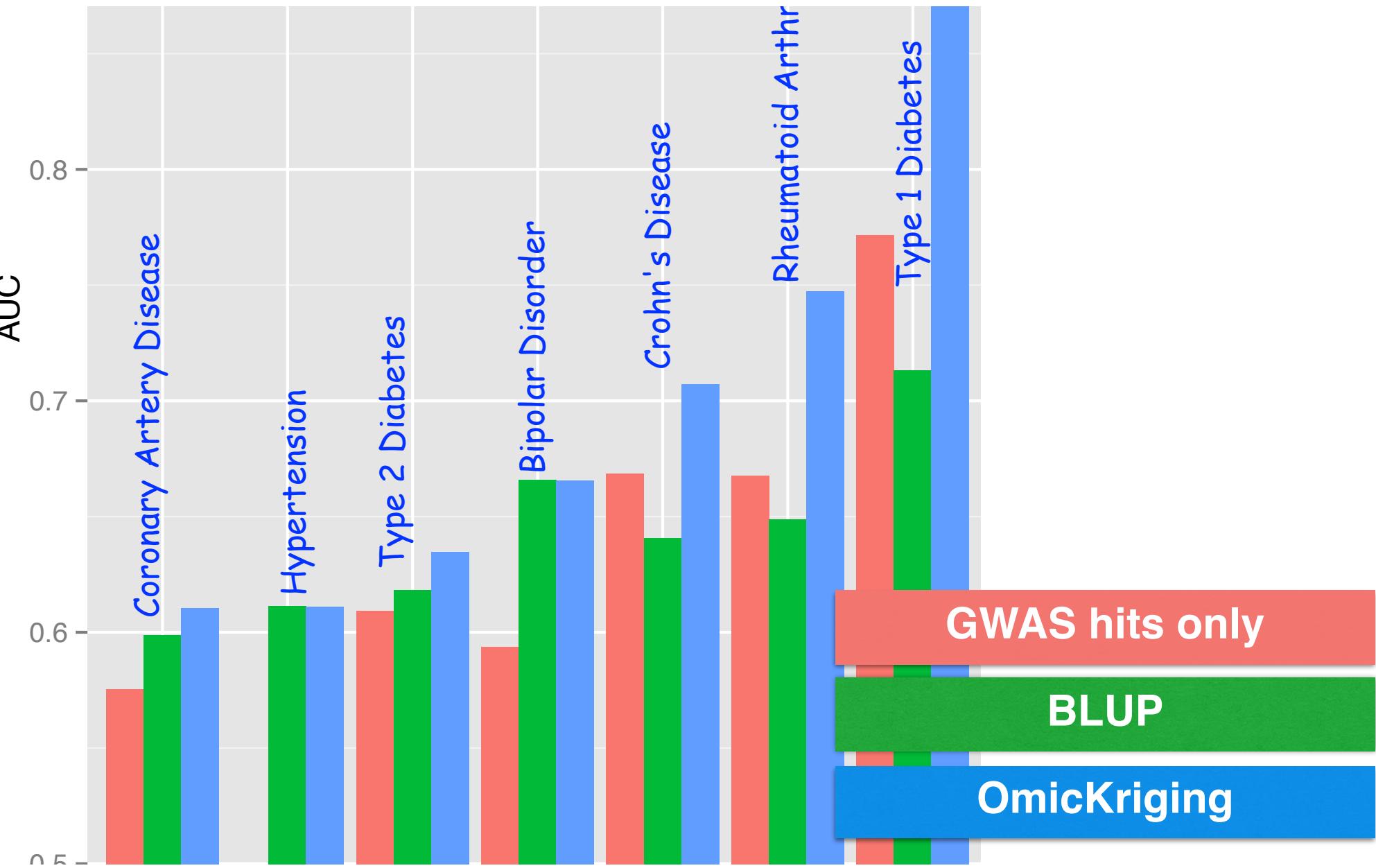
SOFTWARE

Open Access

## SparSNP: Fast and memory-efficient analysis of all SNPs for phenotype prediction

Gad Abraham<sup>1\*</sup>, Adam Kowalczyk<sup>1</sup>, Justin Zobel<sup>1</sup> and Michael Inouye<sup>2,3</sup>

# GWAS hits vs. Whole Genome Prediction (OmicKriging)

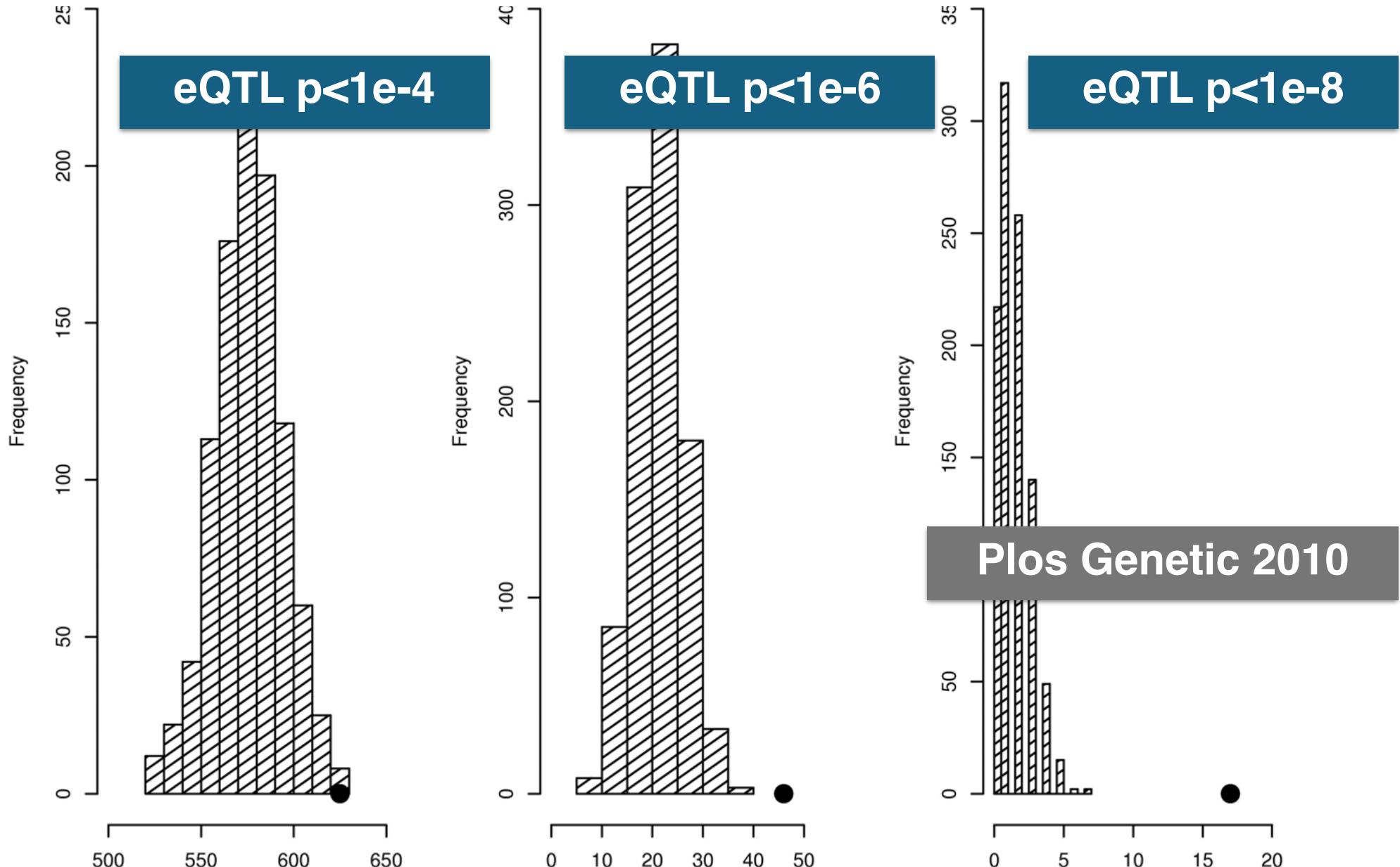


# Collective Approaches to Dissect Complex Traits

- Enrichment of functional classes
- Partitioning heritability into functional classes
- Aggregation into functional units

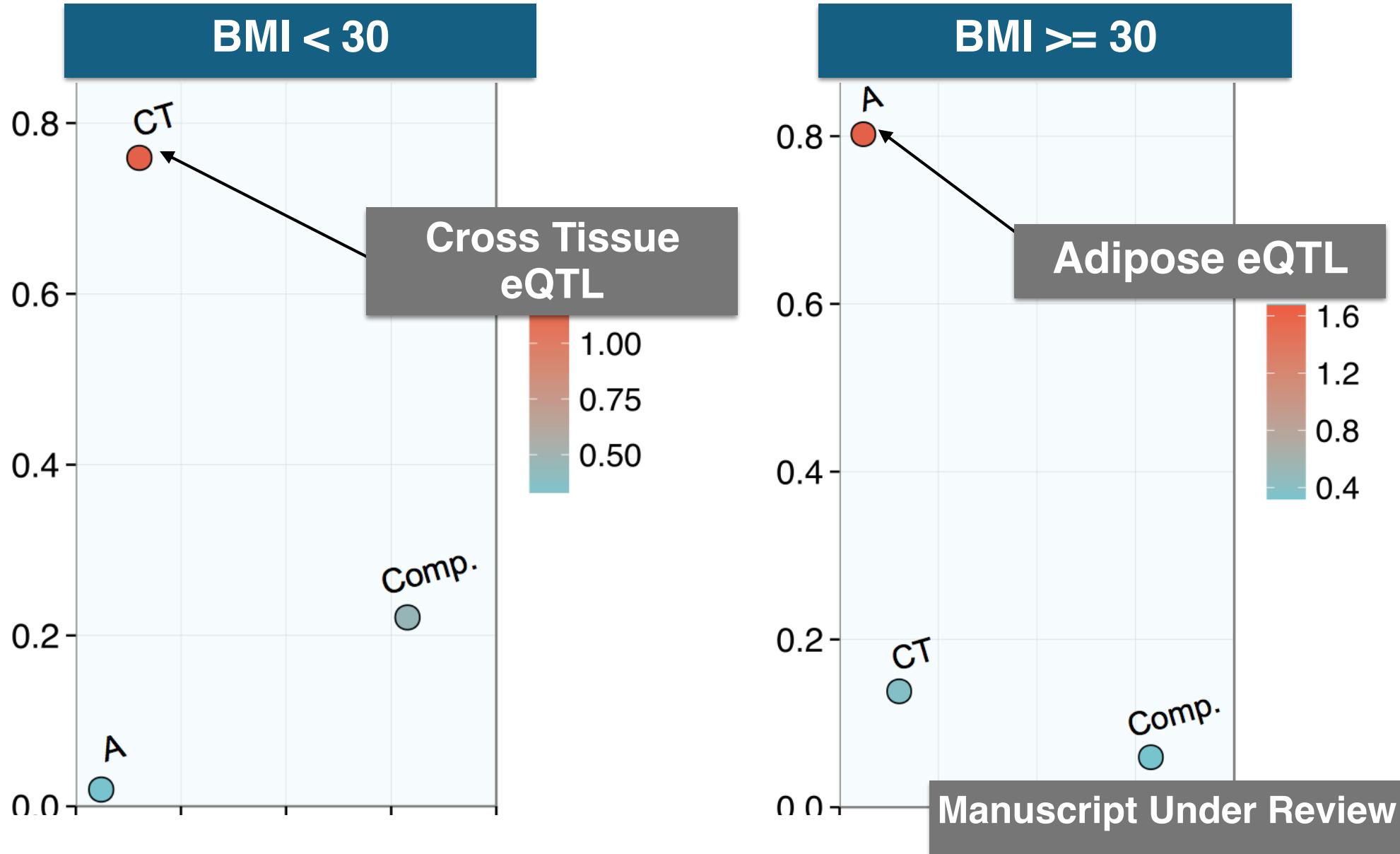
# Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS

Dan L. Nicolae<sup>1,2,3</sup>, Eric Gamazon<sup>1</sup>, Wei Zhang<sup>1</sup>, Shiwei Duan<sup>1\*</sup>, M. Eileen Dolan<sup>1,2</sup>, Nancy J. Cox<sup>1</sup>,



# Cross-tissue and tissue-specific eQTLs: locating the missing heritability of a complex trait across populations

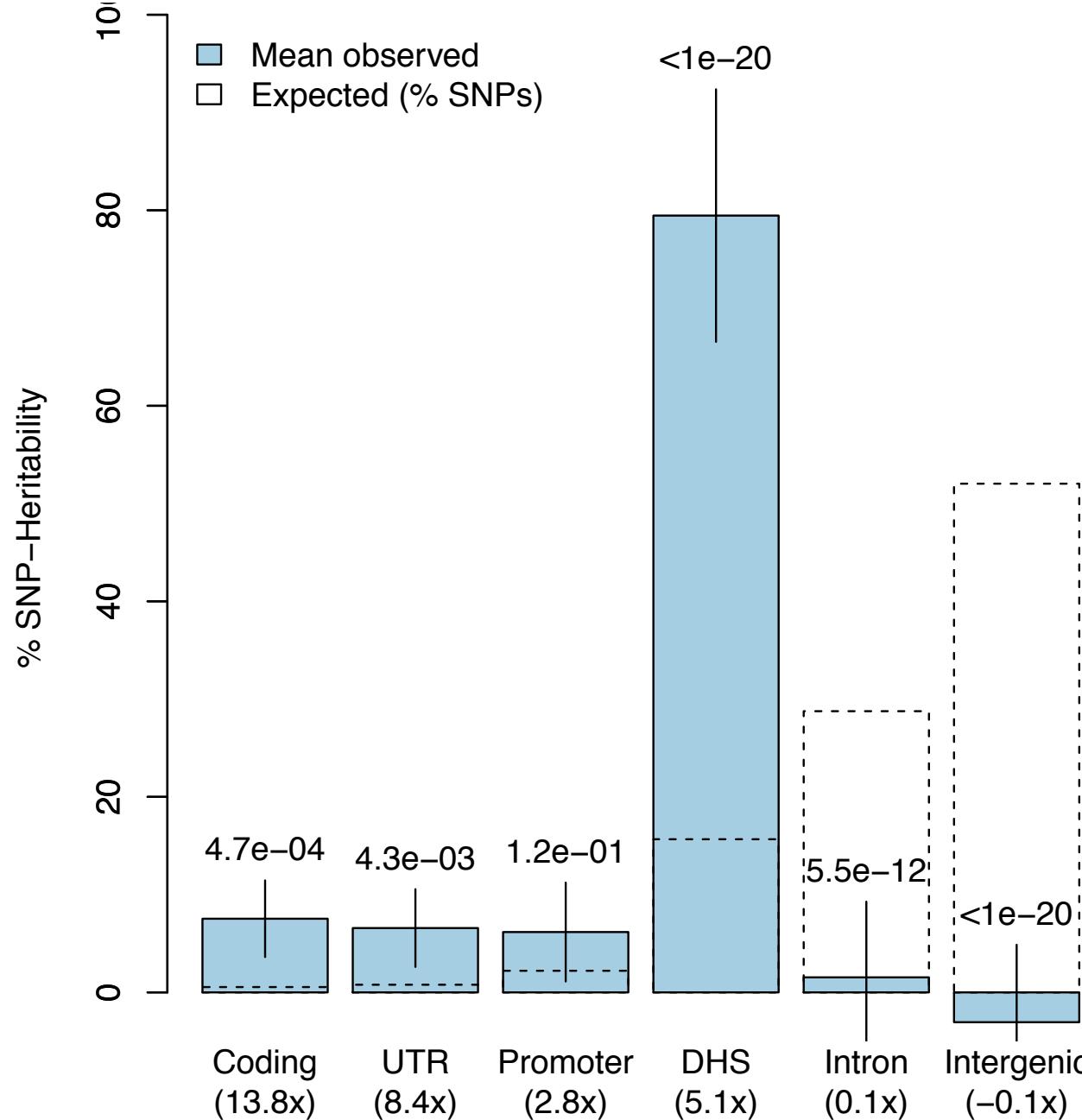
Jason M. Torres, Eric R. Gamazon, Esteban J. Parra, Jennifer E. Below, Adan Valladares-Salgado, Niels Wacher, Miguel Cruz, Craig L. Hanis, Nancy J. Cox\*



# Regulatory variants explain much more heritability than coding variants across 11 common diseases

BioRxiv 2014

Alexander Gusev, S Hong Lee, Benjamin M Neale, et al.



DHS: DNase  
hypersensitivity sites,  
control accessibility  
of the region thus  
levels of transcription

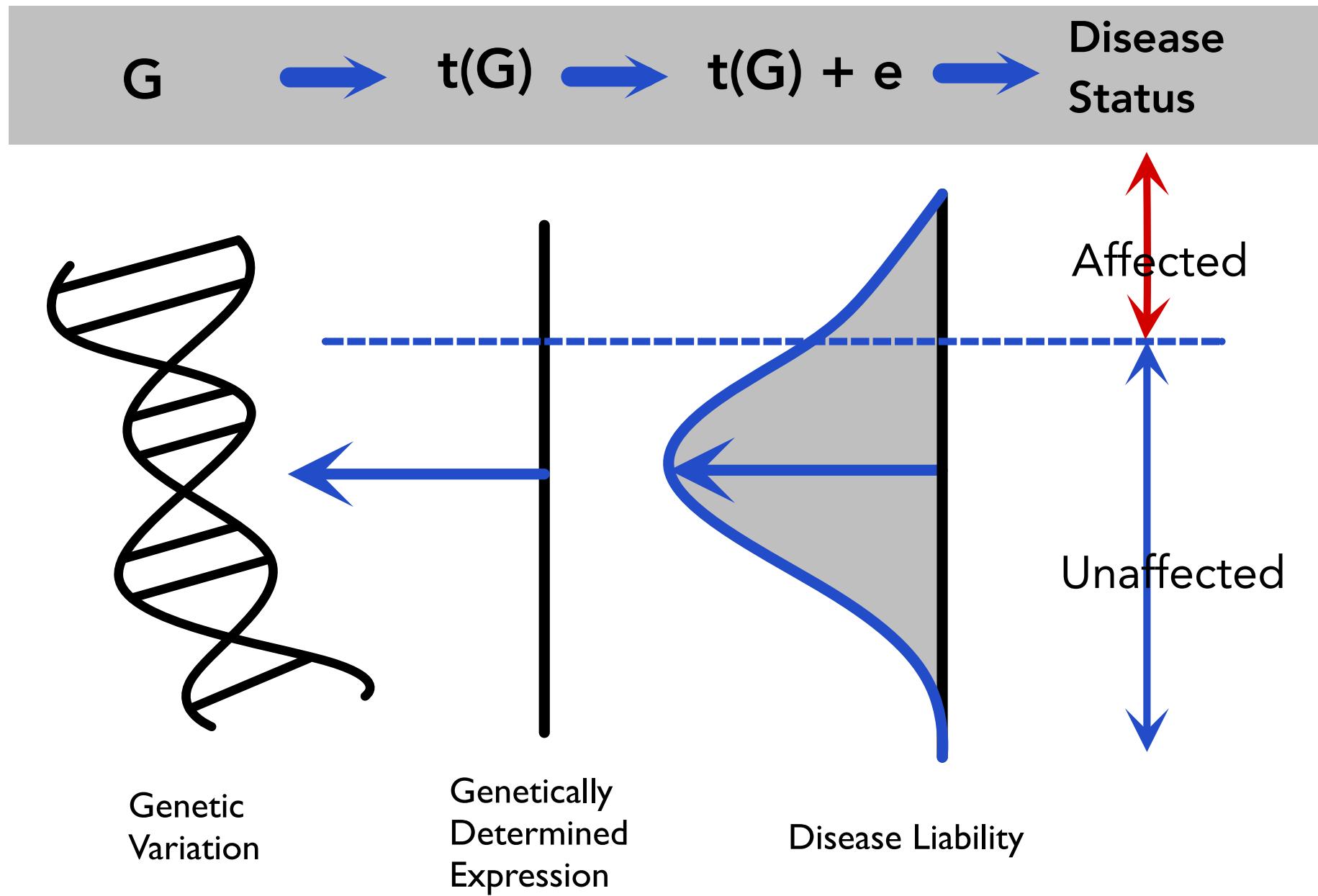
# Gene Based Tests

- Gene based association tests
  - VEGAS (Liu et al 2010 AJHG)
  - SKAT (Wu et al 2012 AJHG)
  - C-Alpha (Neale and Rivas et al 2011 Plos Genetics)
- Used extensively in whole exome studies
- Designed to address low power of rare variants

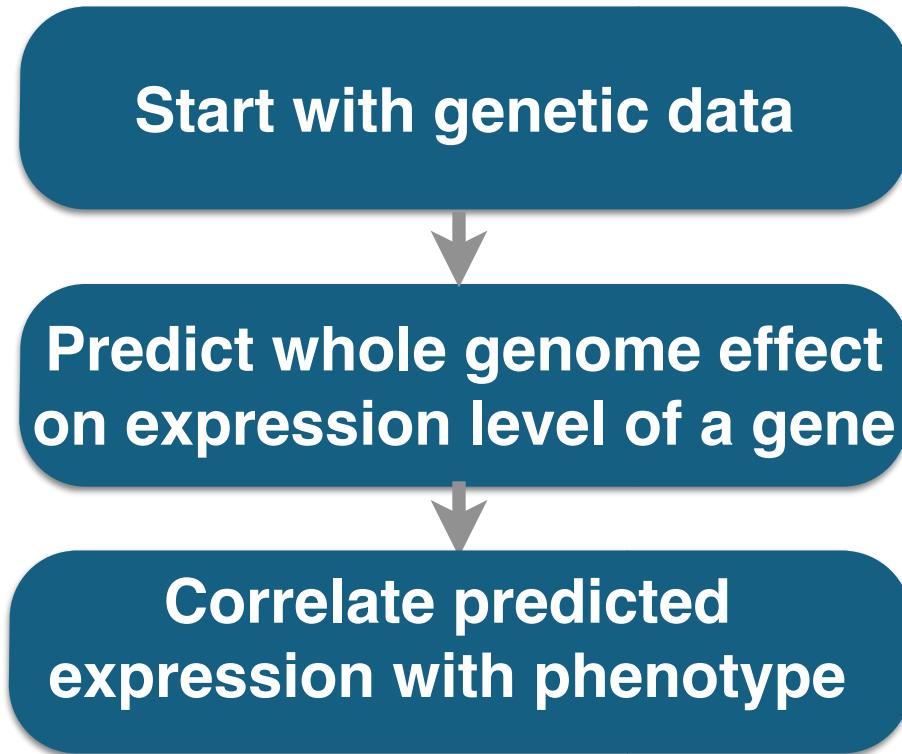
# Gene Based Tests

- Limited success of gene based tests
- More functional data needs to be integrated
- Enrichment studies indicate important role of gene regulation
- To address this issues, we propose PrediXcan
  - predict expression levels of a gene
  - correlate predicted levels with complex traits
  - scan whole genome

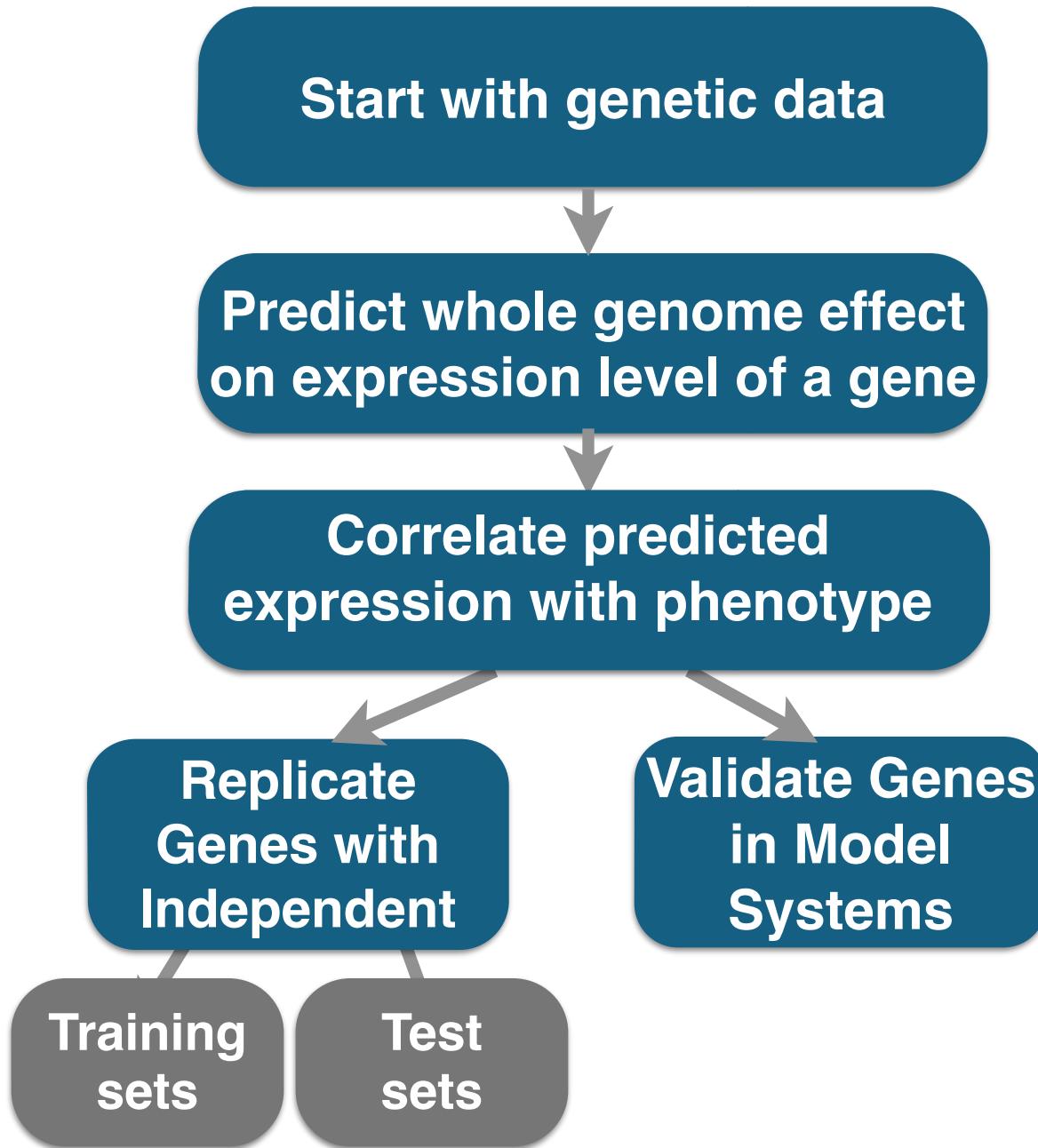
# Genetic Control of Disease Through Gene Regulation



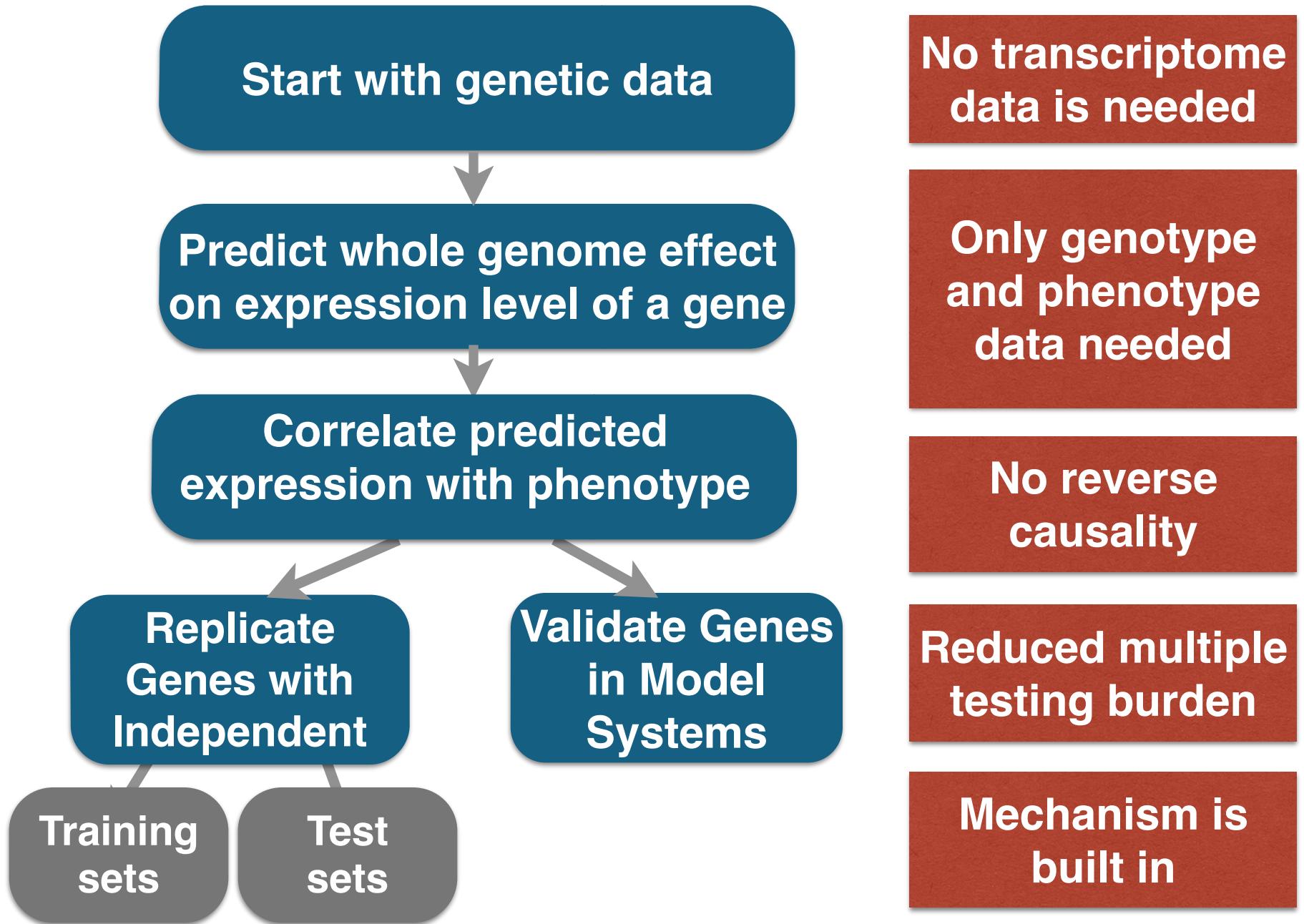
# PrediXcan Flow



# PrediXcan Flow



# PrediXcan Flow



# Additive Model for Genetic Effect Prediction

## Predicted Expression Trait

$$t_i = \sum_{k=1}^M w_k G_{ki}$$

$t_i$  is predicted effect on gene expression level for individual  $i$

$G_{ki}$  number of reference alleles for SNP  $k$  and individual  $i$

$w_k$  weight for SNP  $k$

## Simple Polygenic Model

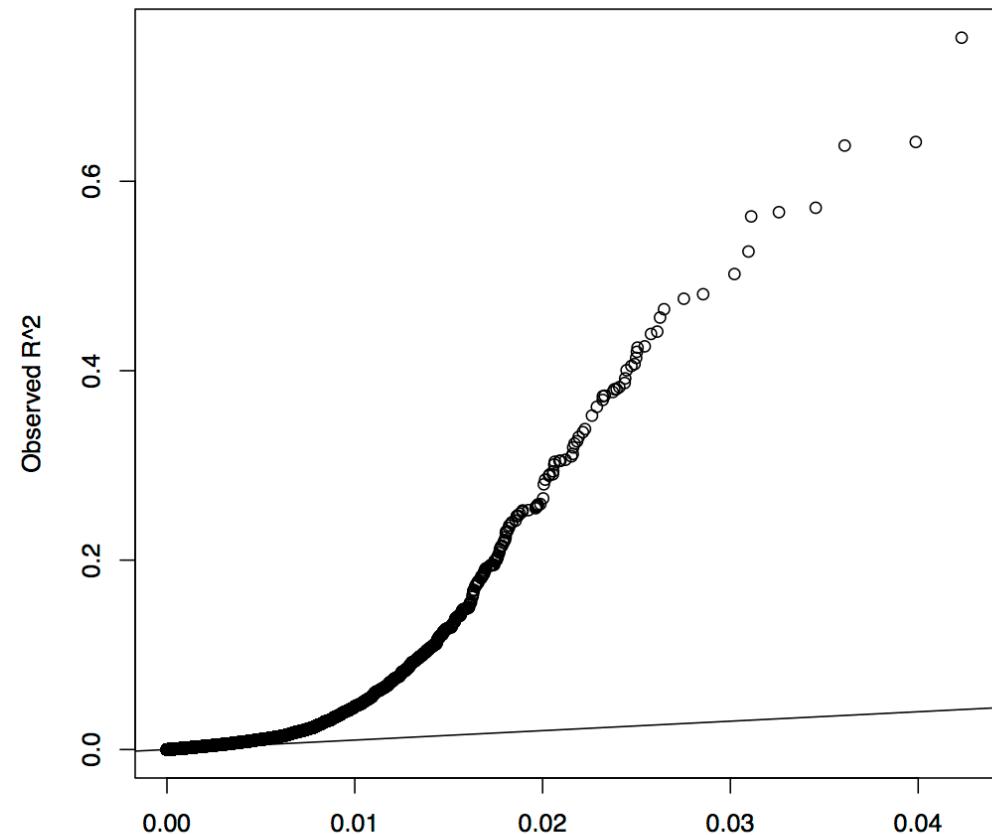
- ▶  $w_k$  = single variant regression coefficient (Matrix eQTL output)
- ▶  $w_k$  set to zero if p value > 0.05 for cis SNPs (1Mb TSS)
- ▶  $w_k$  set to zero if p value >  $10^{-6}$  for trans SNPs

# Expression Data

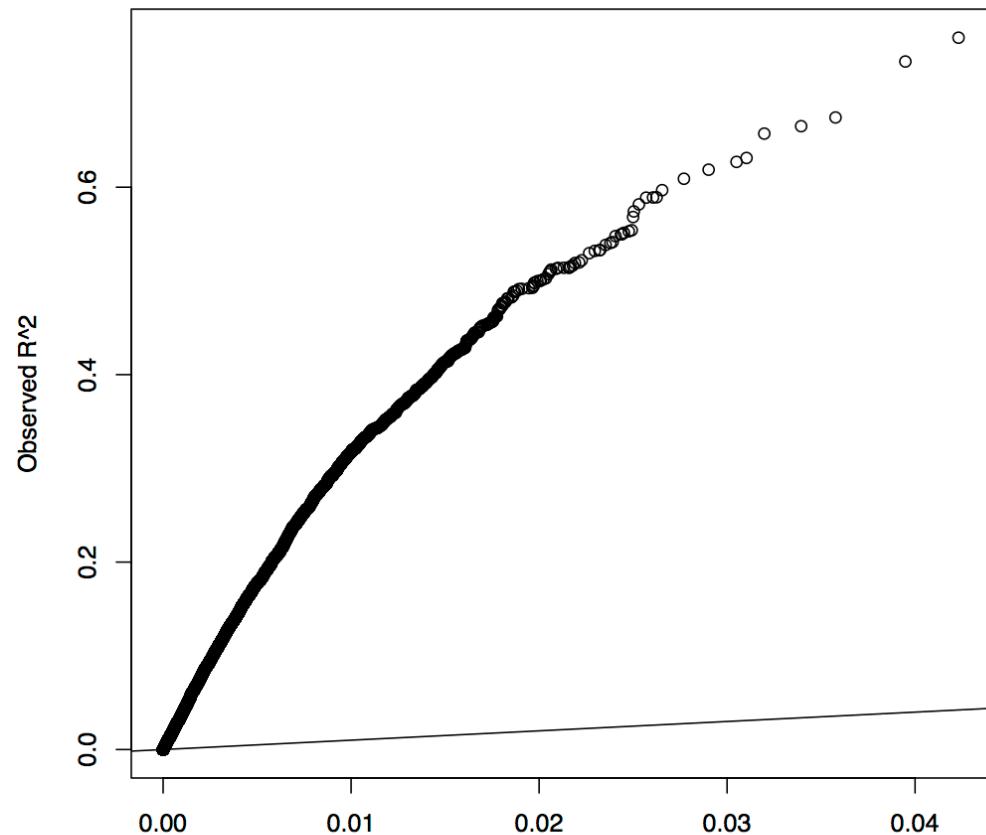
- GTEx - Genotype of Tissue Expression
  - Large scale Common Fund project
  - 900 organ donors
  - 45 tissues
  - RNAseq, whole exome seq, whole genome seq
- gEUVADIS
  - RNAseq 462 individuals from the 1000 Genomes Project
  - Cerebellum expression (Array GSE35974)

# Good Prediction Performance

Prediction  $R^2$



Replication  $R^2$

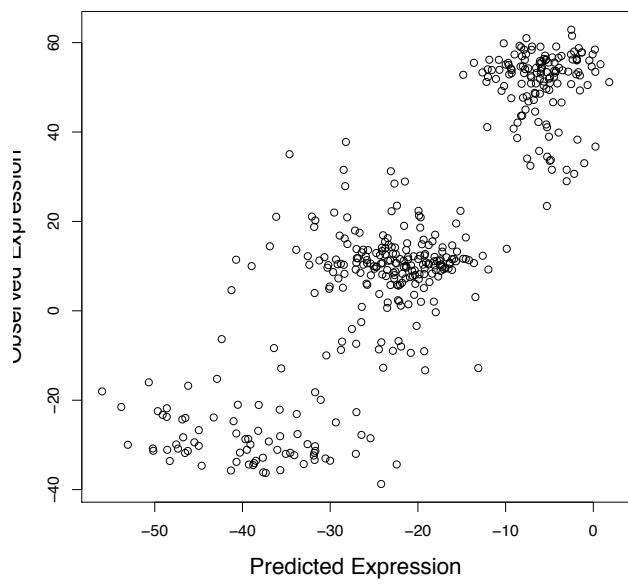


Training with GTEx  
Testing in 1K Genomes

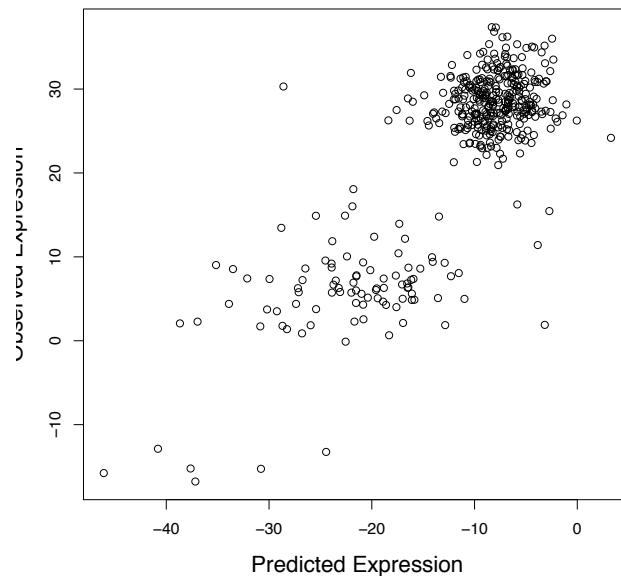
Replicate RNAseq  
Pickrell et al 2010 vs.  
1K Genomes 2013

# Examples of Well Predicted Genes

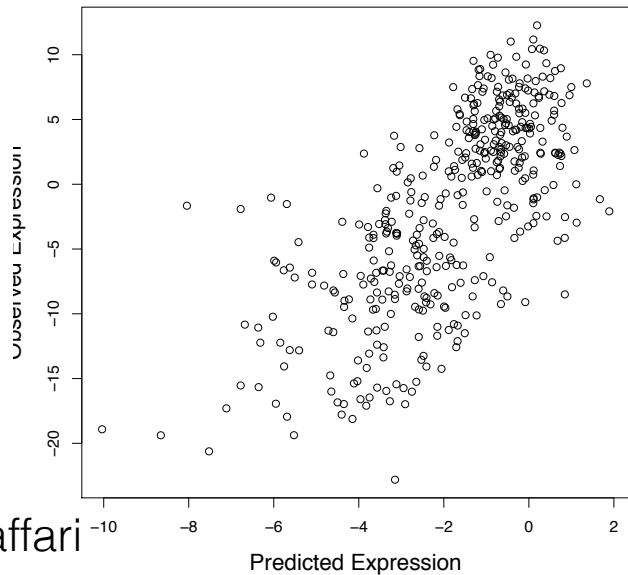
ERAP2



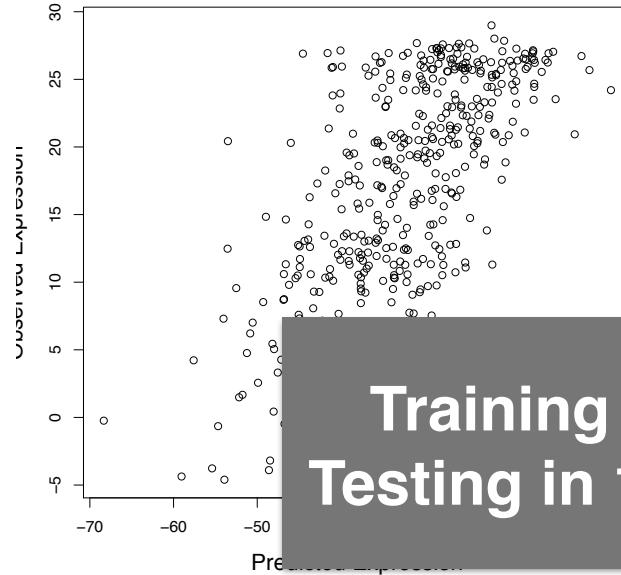
FAM118A



C17orf97

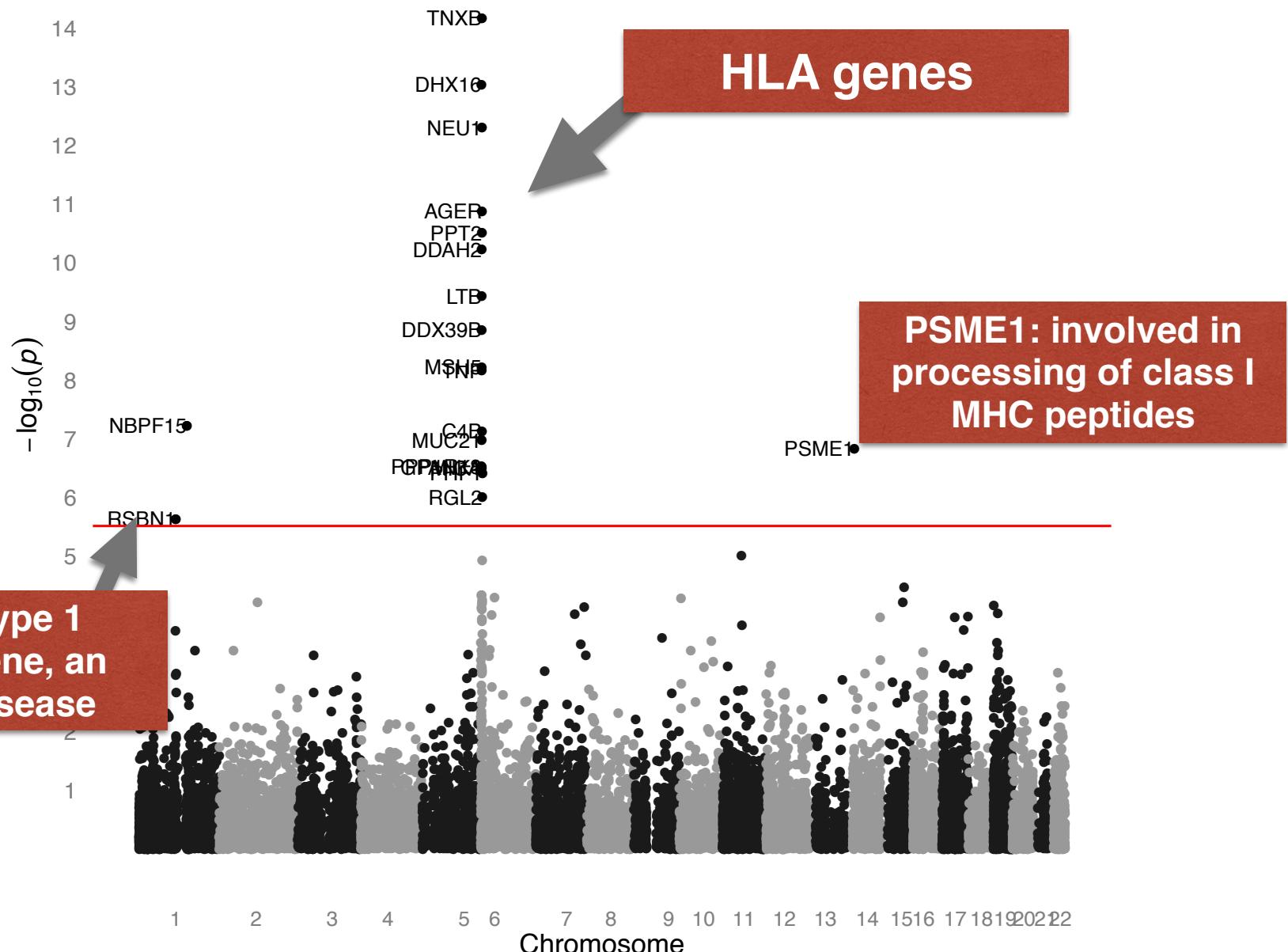


ERAP1



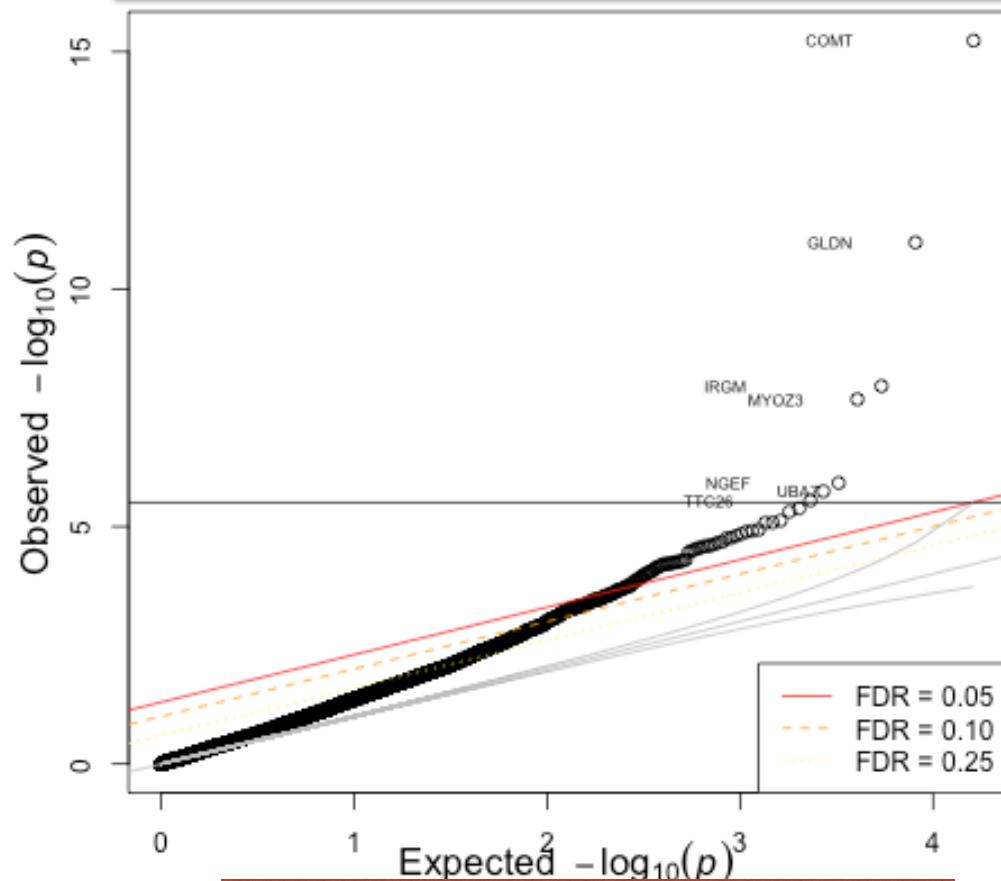
Training with GTEx  
Testing in 1K Genomes

# Genes Associated with Rheumatoid Arthritis



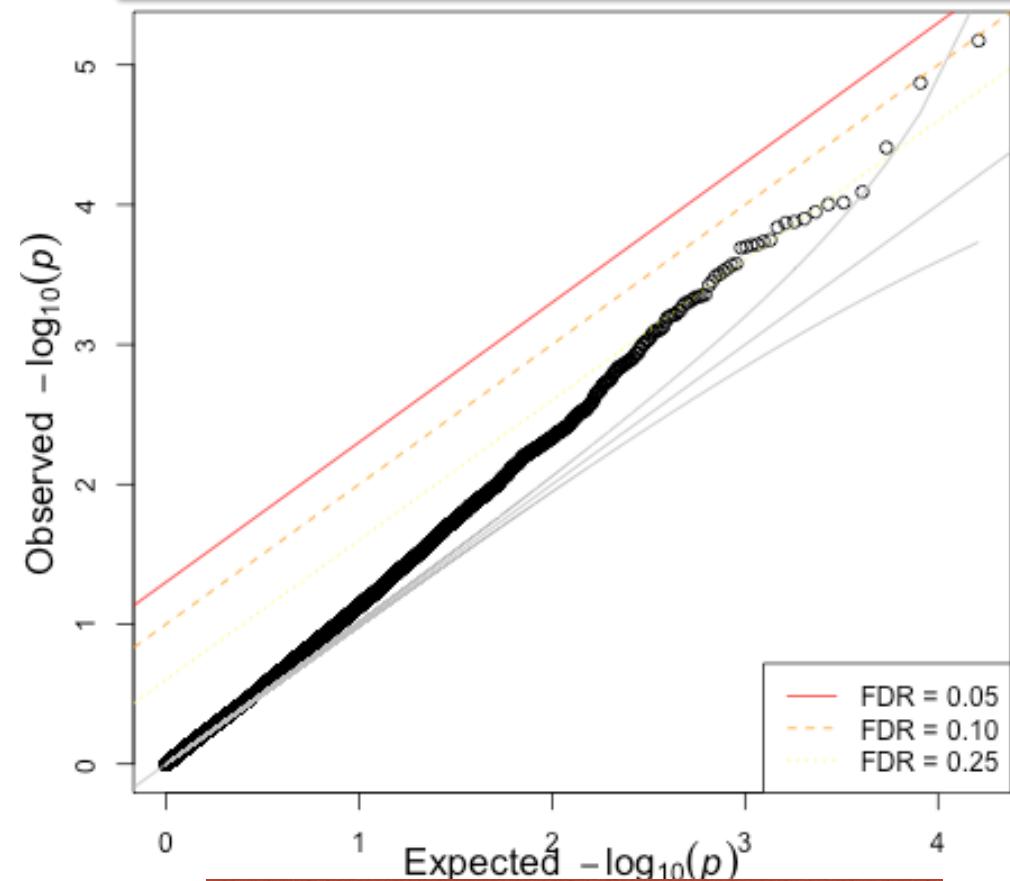
# PrediXcan Results for Crohn's Disease and Hypertension

## Crohn's Disease



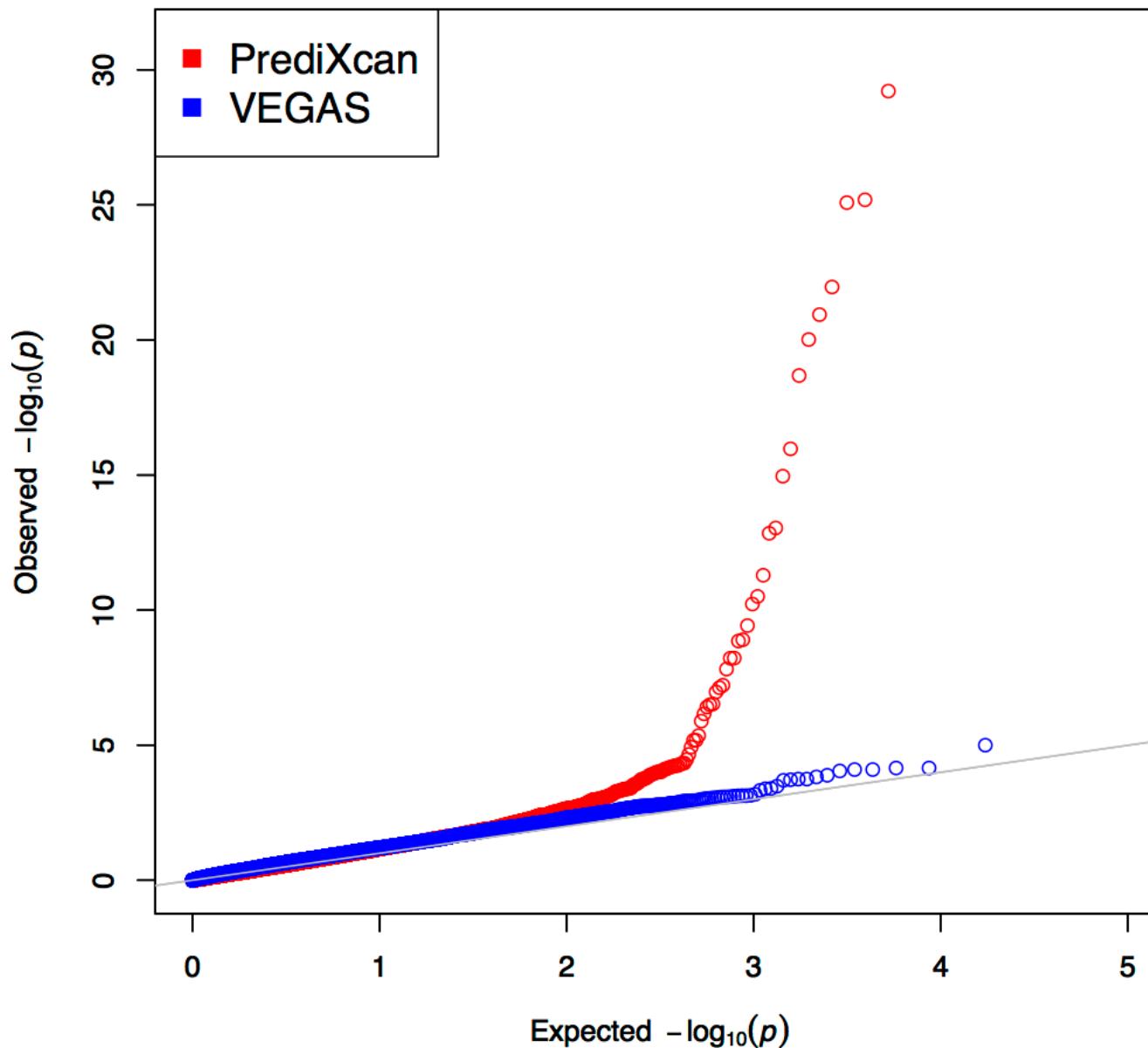
IRGM is a known  
Crohn's gene

## Hypertension



Whole blood may not be  
relevant tissue

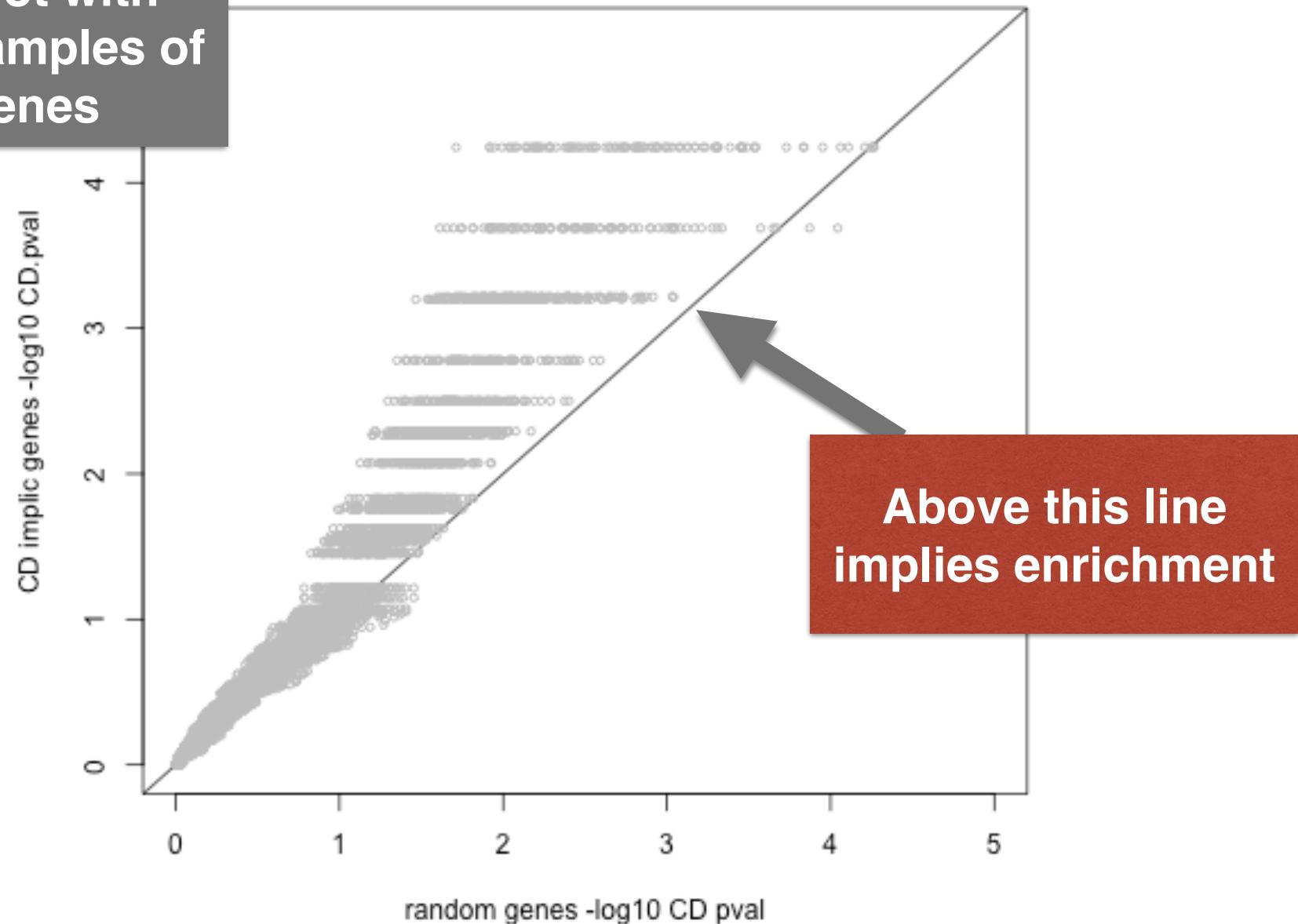
# PrediXcan Outperforms VEGAS



Eric Gamazon

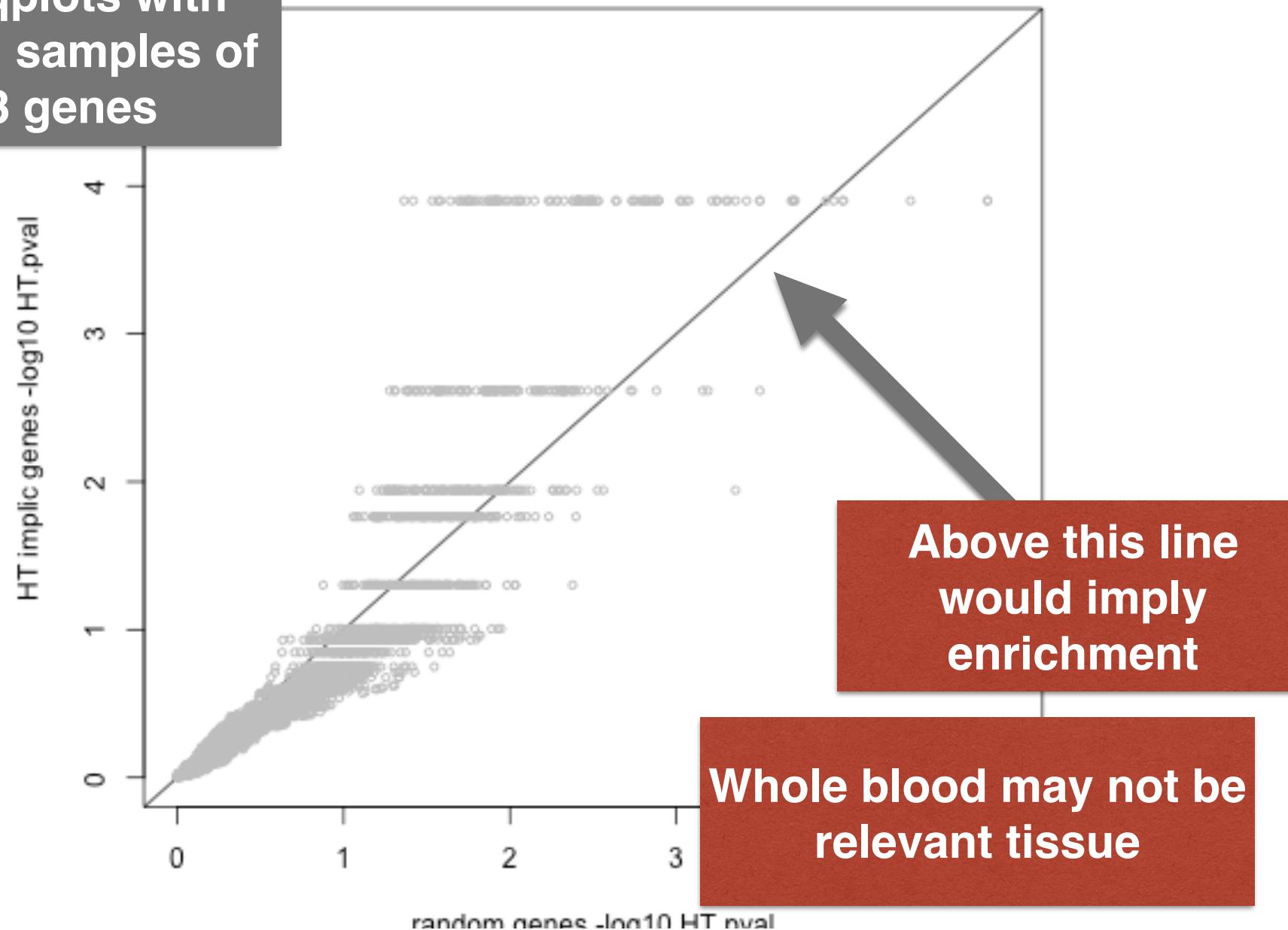
# Enrichment of Known Crohn's Genes Among Findings

100 qqplot with  
random samples of  
205 genes

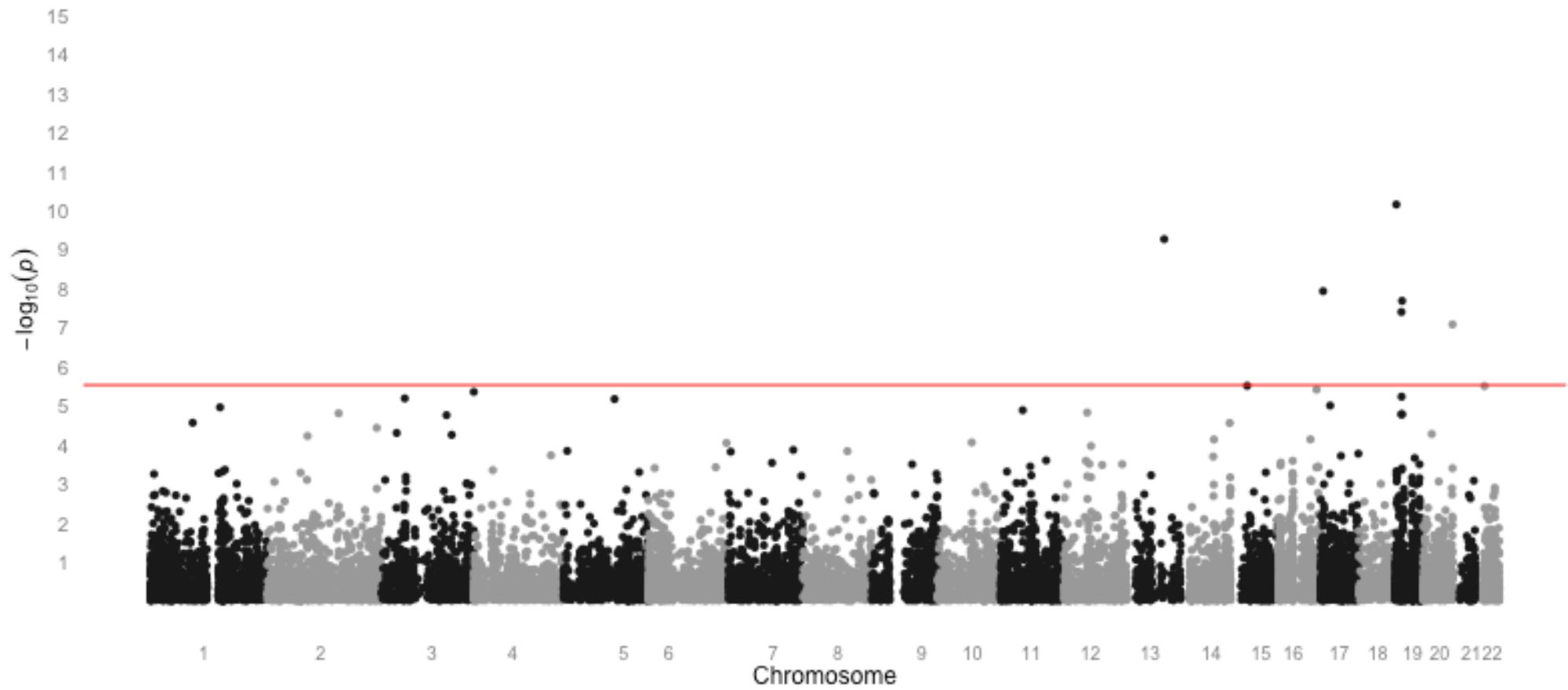


# No Enrichment Among Hypertension Findings

100 qqplots with  
random samples of  
133 genes



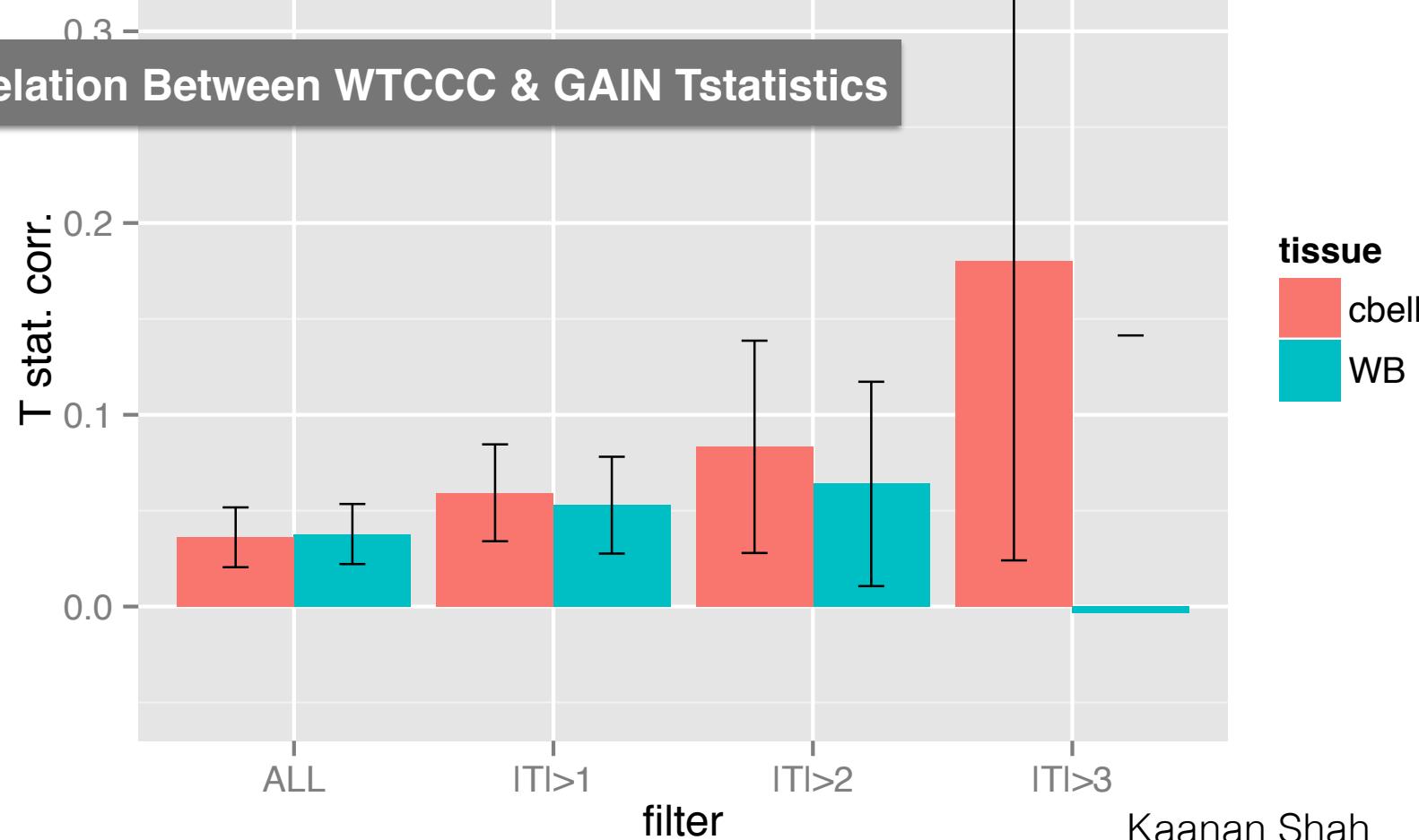
# Bipolar Disorder WTCCC results



# Significant Concordance Between Independent Bipolar Studies

Higher correlation for cerebellum based predictions than whole blood based

Correlation Between WTCCC & GAIN Tstatistics



Kaanan Shah

# PrediXcan: a Gene Discovery Approach

- PrediXcan is a powerful gene based association test
- It directly tests the molecular mechanism through which genetic variants affect phenotype
- Reduced multiple testing burden compared to single variant approach
- Unlike other gene based tests, it provides direction of effects
- Advantages relative to gene expression studies
  - Applicable to any GWAS datasets  
gene expression levels are predicted from genotype data
  - No reverse causality  
disease status does not affect germline DNA
  - Multiple Tissues can be evaluated  
tissue expressions are only needed to build prediction models

# Challenges of Pharmacogenomic Studies

- Smaller sample size
- Even more important to integrate prior data
- Integrate other functional data
- Heritability estimates are harder
  - Limited family data
  - Usually samples greater than 1K are needed for GCTA