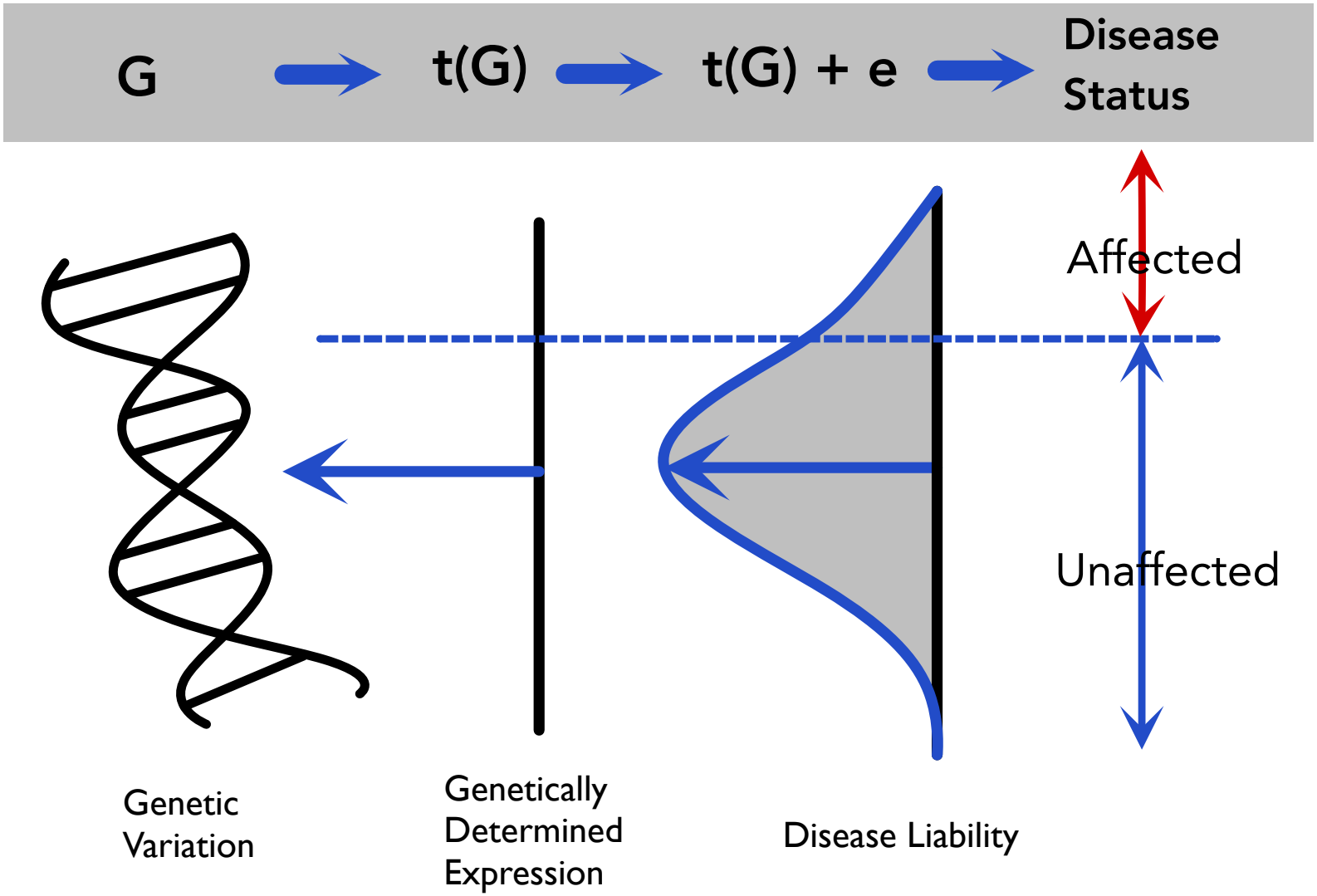
Testing the Molecular Mechanisms that Drive Complex Traits Using Transcriptome Studies

Hae Kyung Im

THE UNIVERSITY OF CHICAGO

July 21, 2014

Genetic Control of Disease Risk Through Gene Regulation



Additive Genetic Model for 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

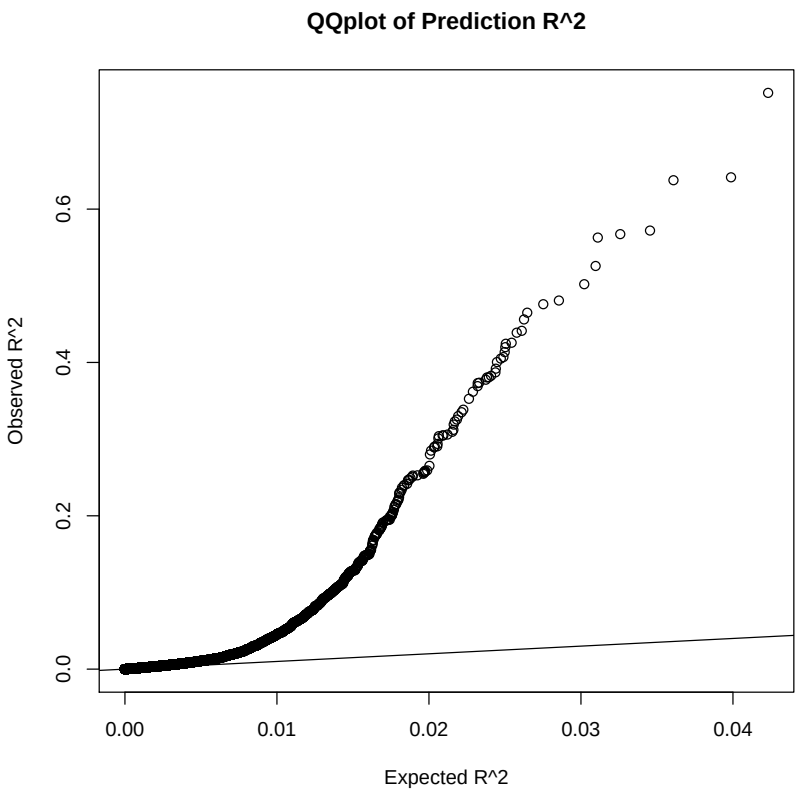
- w_k = single variant regression coefficient (Matrix eQTL output)
- w_k set to zero if p value > 0.05 for cis SNPs (1Mb TSS)
- w_k set to zero if p value $> 10^{-6}$ for trans SNPs

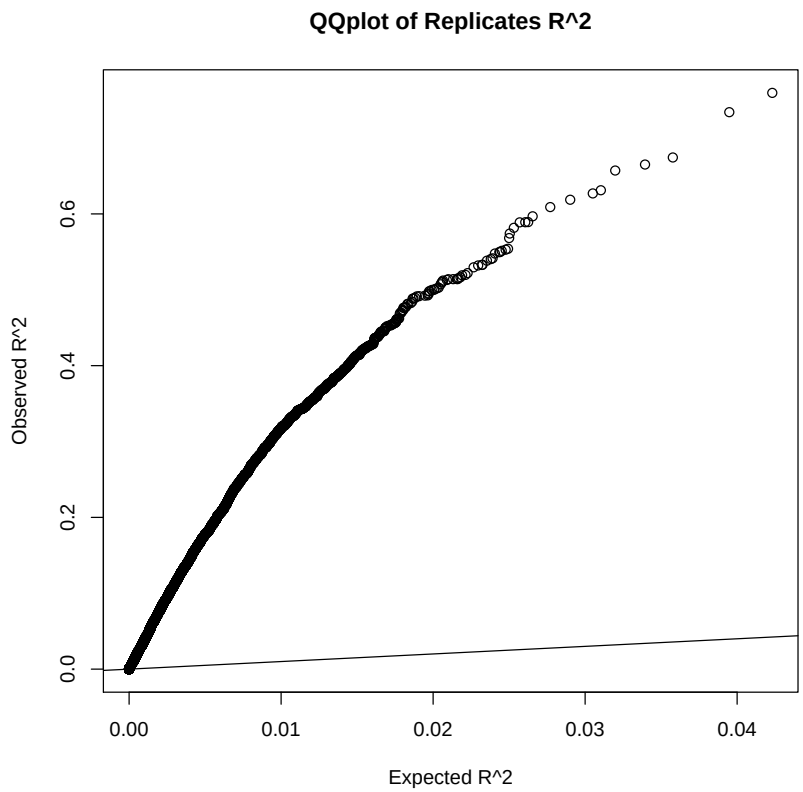
PrediXcan Flow

- Predict genetic effect on expression
- Test differential predicted expression levels between cases and controls
- For quantitative traits, perform regression
- Replicate
  - independent training set
  - independent test set
- Validate with follow up experiments

How Well do we Predict the Transcriptome

Trained with GTEx Whole Blood - Tested on GEUVADIS LCL





16% of predicted genes have correlation $> 10\%$

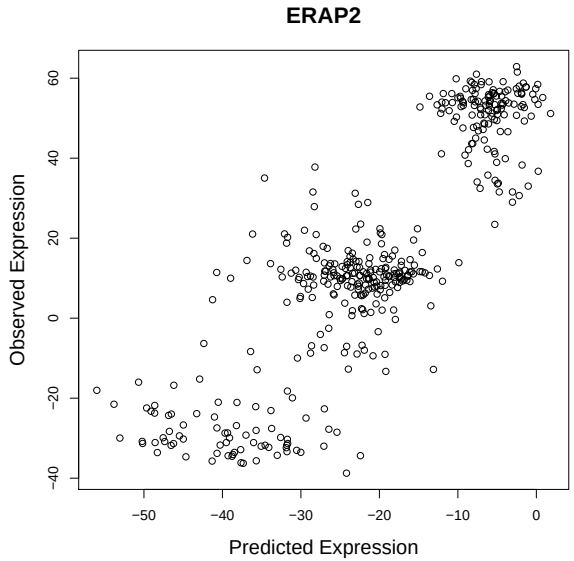
74% of genes sequenced in different labs* have correlation $> 10\%$

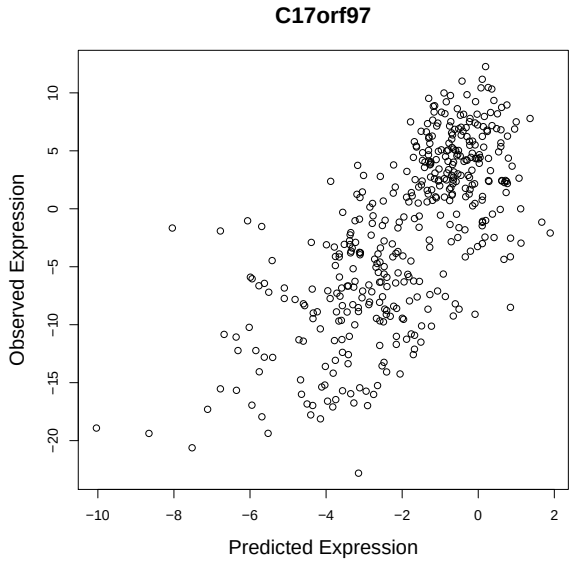
* Lappalainen et al 2013 vs. Pickrell et al 2010

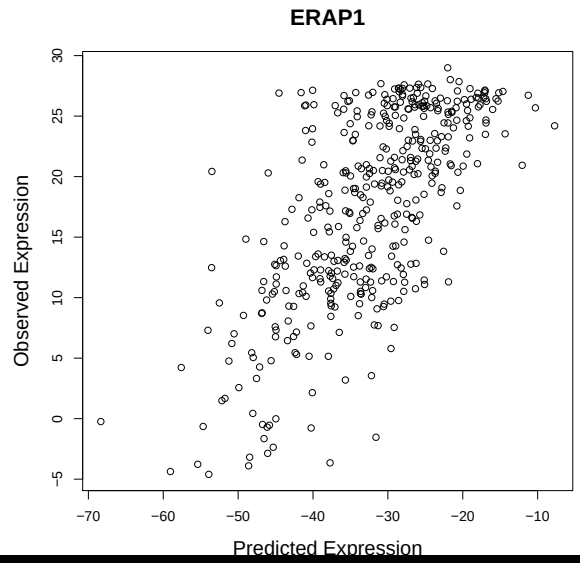
Sahar Mozaffari

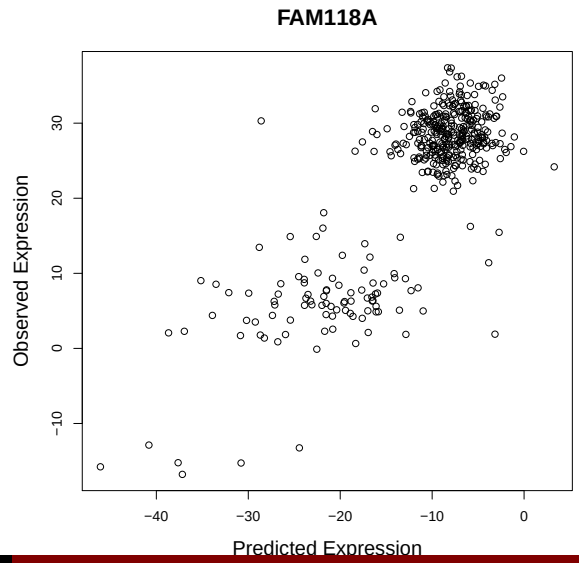
Examples of Observed vs. Predicted Expression Levels

Trained with GTEx Whole Blood - Tested on GEUVADIS LCL

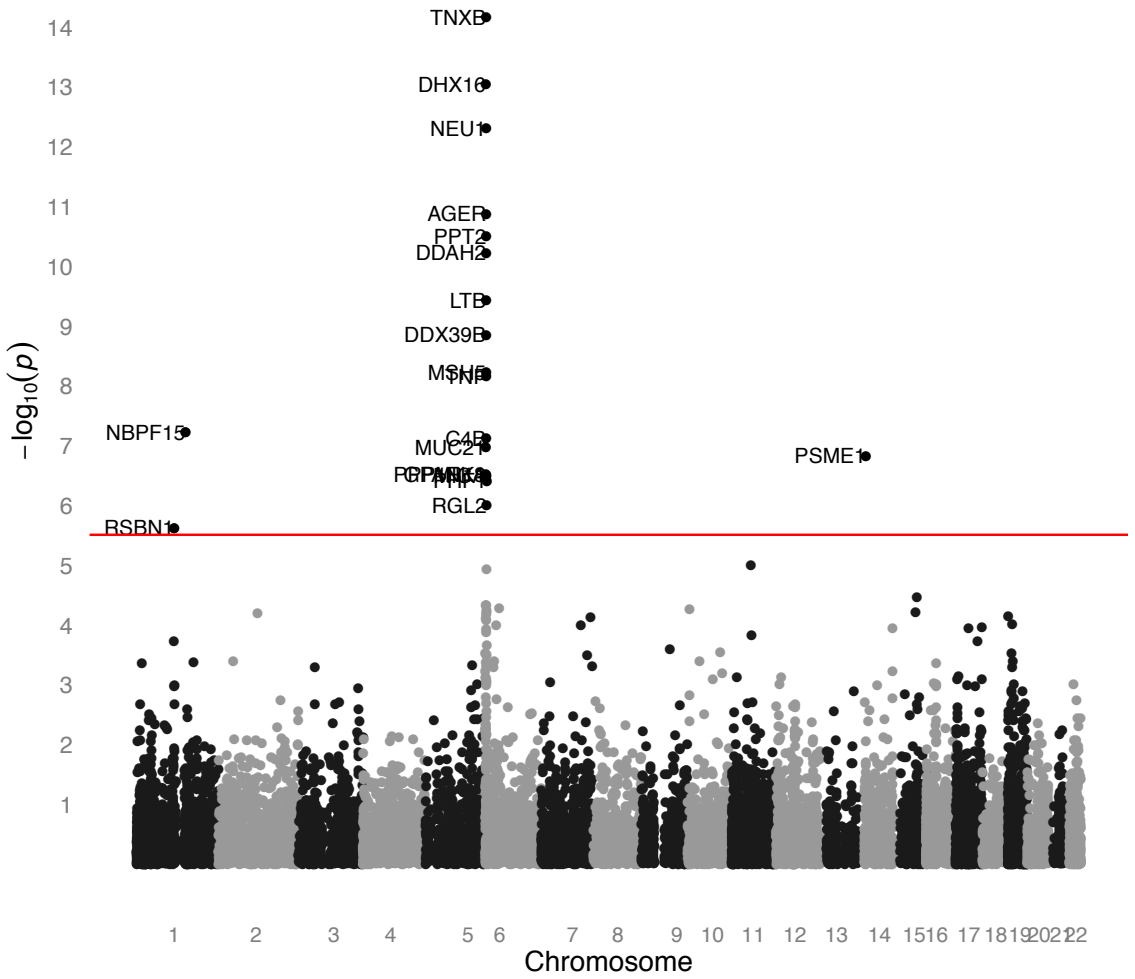






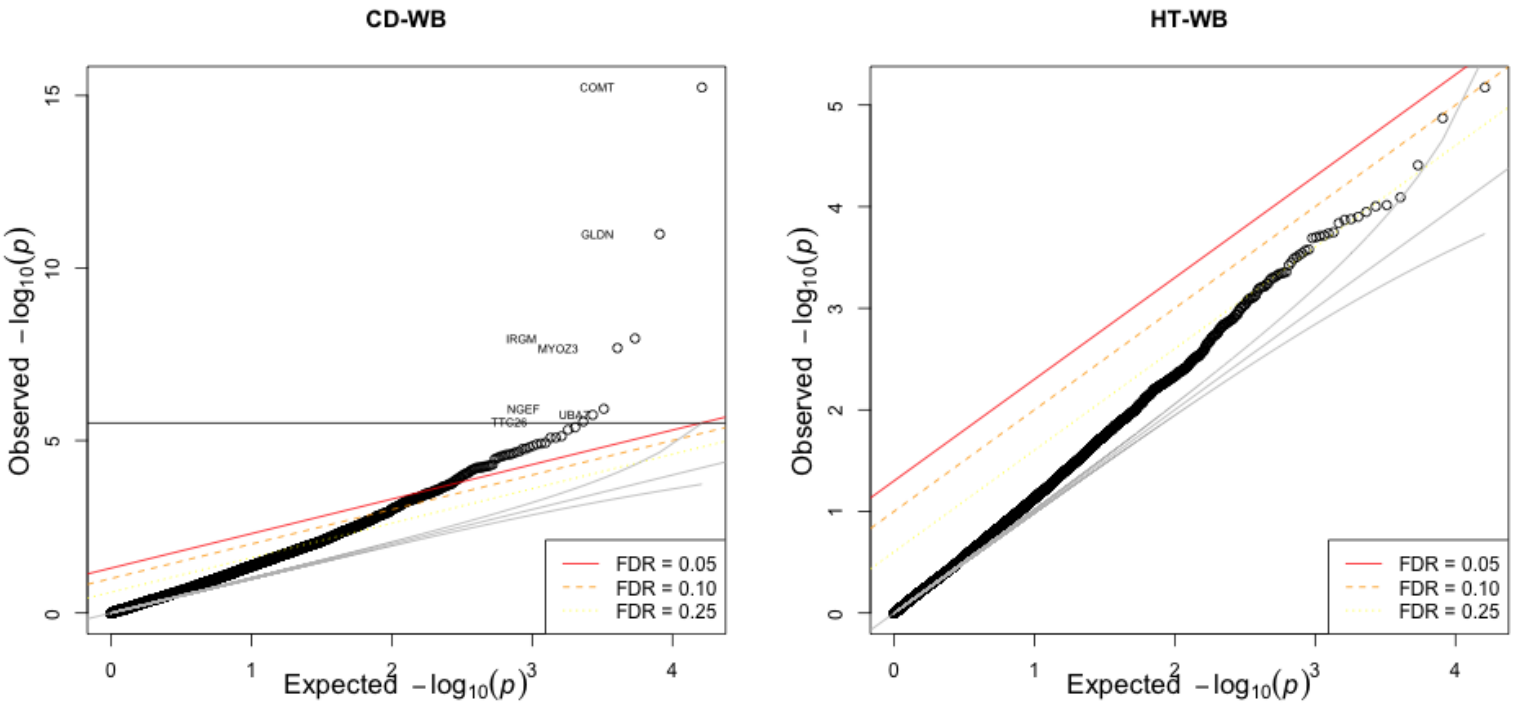


PrediXcan Results in WTCCC Rheumatoid Arthritis



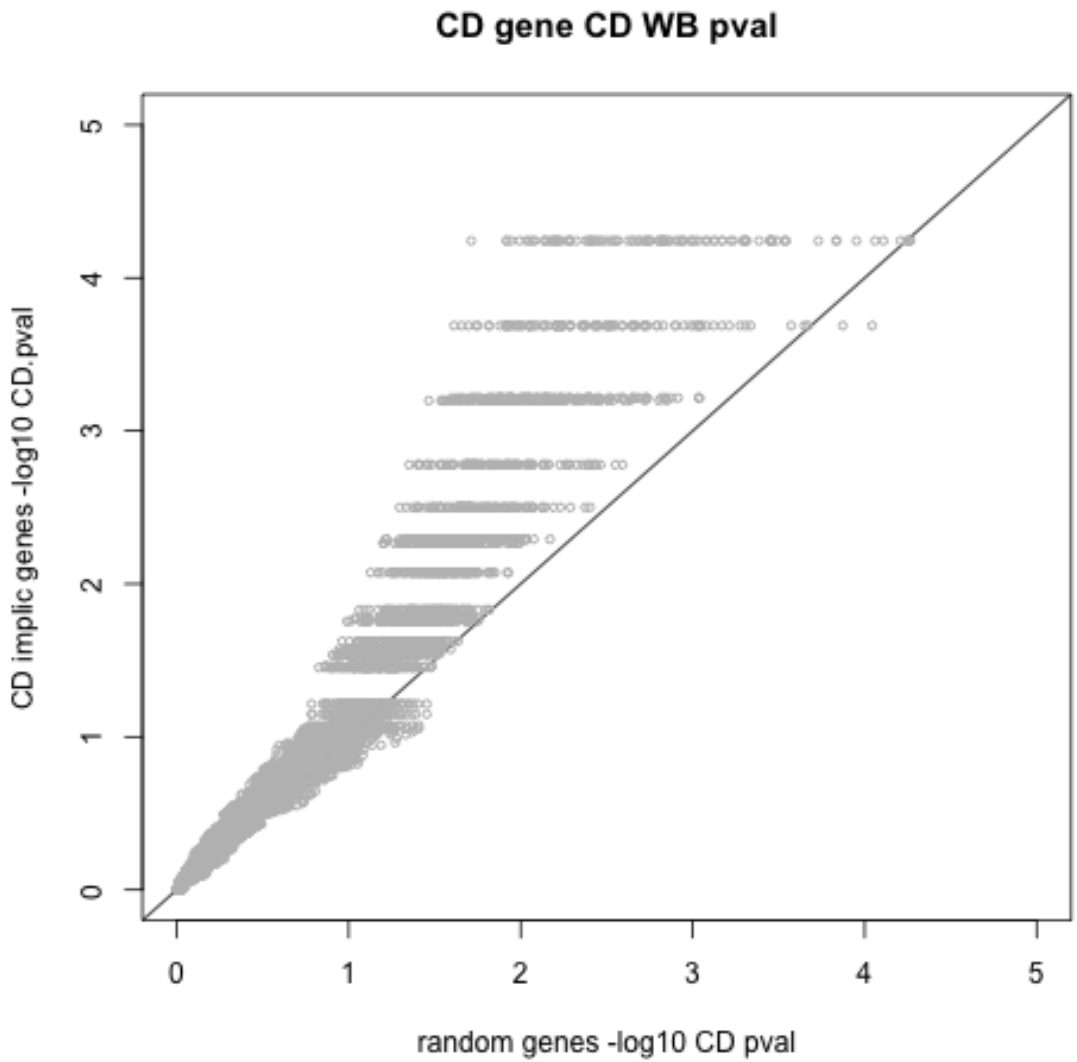
Many HLA genes as expected

PrediXcan Results in Crohn's Dis. & Hypertension

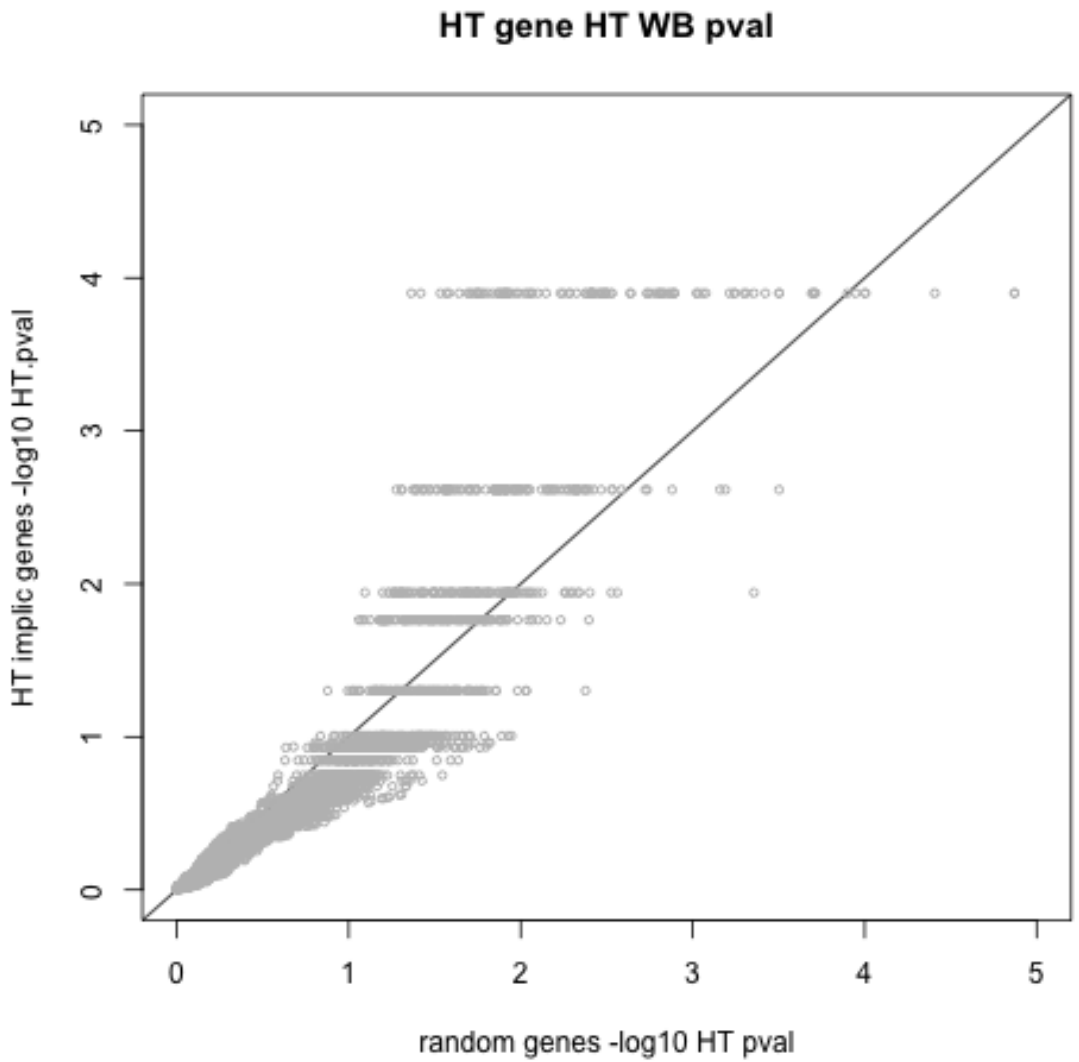


Known and novel Cronh's genes are significant
No significant Hypertension results
Whole blood may not be relevant tissue

Enrichment of Crohn's Disease Genes Among Discoveries



No Enrichment of Hypertension Genes Among Discoveries



Bipolar Disorder Replication

- GAIN (n=2000) & WTCCC Bipolar Disorder (n=5000)
- Whole Blood
- Significant genes
  - RFNG ($p_{\text{meta}} = 10^{-8}$, $p_{\text{GAIN}} = 2.5x10^{-6}$, $p_{\text{WTCCC}} = 0.00017$)

    Modulator of Notch signaling
    Implicated in neurogenesis
  - LPHN1 ($p_{\text{meta}} = 10^{-6}$, $p_{\text{GAIN}} = 0.36$, $p_{\text{WTCCC}} = 2x10^{-8}$)

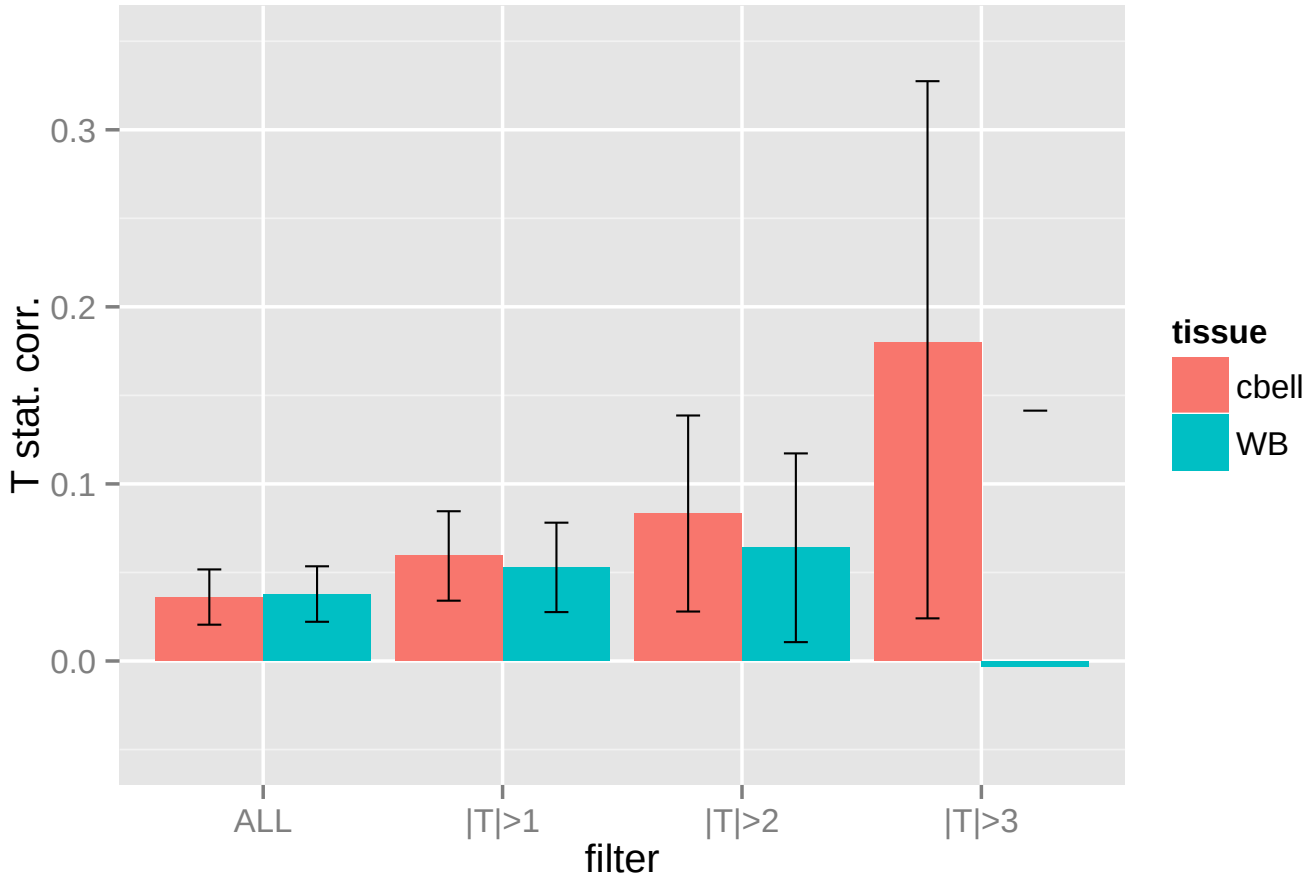
    Receptor for TENM2 that mediates heterophilic synaptic cell-cell contact and postsynaptic specialization
    Candidate gene for mental disorder based on mouse model phenotypes

Kaanan P. Shah

Concordance between GAIN and WTCCC

** Significant Correlation in T statistics

** Higher correlation for cerebellum than whole blood based results



PrediXcan: 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

Conclusions

- PrediXcan is a promising novel gene discovery approach
- Application to the WTCCC data recapitulates known genes and identifies many novel genome-wide significant ones
- Bipolar Disorder genes replicated in independent datasets

Acknowledgement

Contributors

Nancy J. Cox
Eric R. Gamazon
Heather E. Wheeler
Kaanan P. Shah
Sahar Mozaffari
Dan L. Nicolae
Keston Aquino-Michaels

GTEx Consortium

Funding Sources

UC CTSA K12 grant
NCI K12CA139160

UC DRTC
University of Chicago Diabetes Research and Training Center; P60 DK20595, P30 DK020595

GTEx
R01 MH090937 and R01 MH101820

PAAR
NIH/NIGMS UO1GM61393

Conte Center grant
P50MH094267