

Discovering Disease Susceptibility Genes Using *in-silico* Gene Expression - PrediXcan

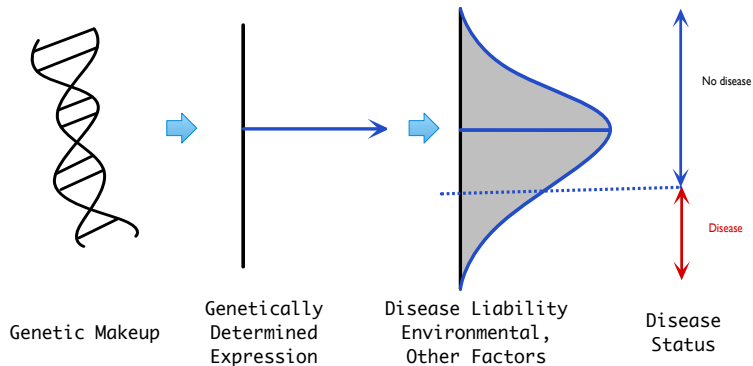
Hae Kyung Im



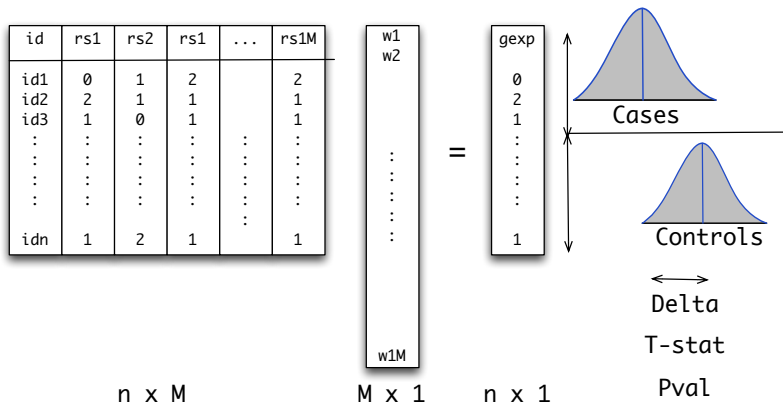
THE UNIVERSITY OF
CHICAGO

June 17, 2014

PrediXcan: Gene (Regulation) Based Association Test

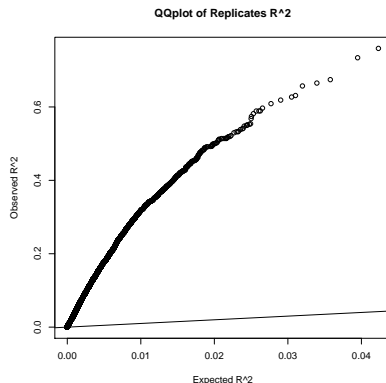
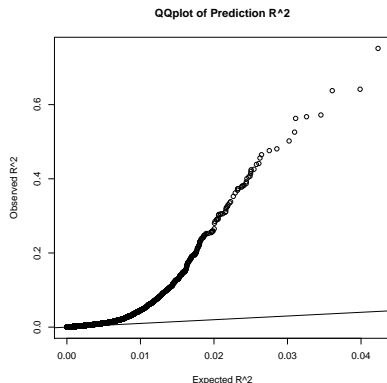


PrediXcan Flowchart



How Well do we Predict the Transcriptome

Trained with GTEx Whole Blood - Tested on GEUVADIS LCL



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

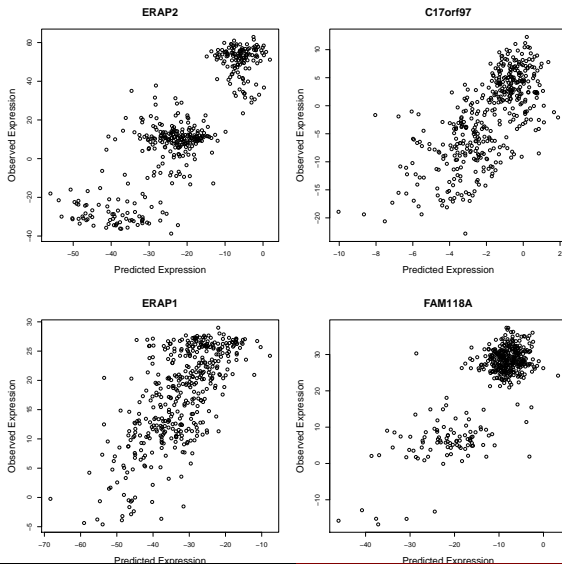
74% of genes sequenced in different labs* have correlation $> 10\%$

* Lappalainen et al 2013 vs. Pickrell et al 2010

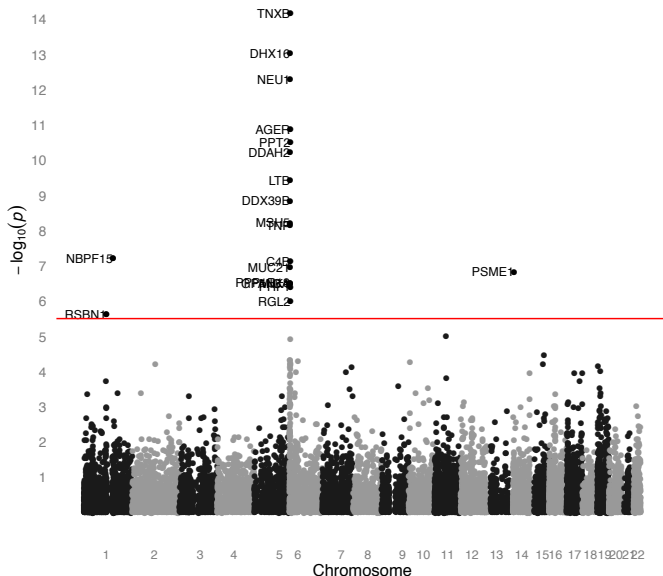
Sahar Mozaffari

Examples of Observed vs. Predicted Expression Levels

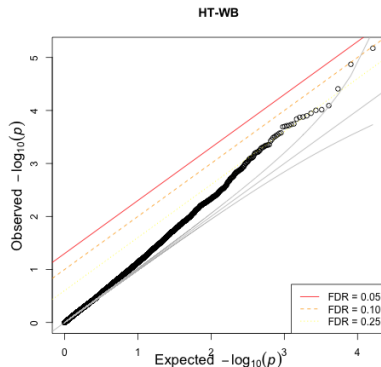
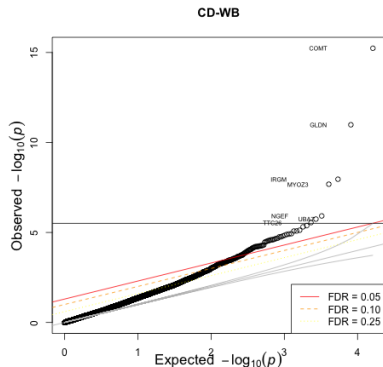
Trained with GTEx Whole Blood - Tested on GEUVADIS LCL



Differential Expression in WTCCC Rheumatoid Arthritis



Differential Expression in Crohn's Disease & Hypertension



Bipolar Disorder Replication

- ▶ GAIN (n=2000) & WTCCC Bipolar Disorder (n=5000)
- ▶ Whole Blood
- ▶ Significant genes
 - ▶ RFNG ($p_{\text{meta}} = 10^{-8}$, $p_{\text{GAIN}} = 2.5 \times 10^{-6}$, $p_{\text{WTCCC}} = 0.00017$)
Modulator of Notch signalling
Implicated in neurogenesis
 - ▶ LPHN1 ($p_{\text{meta}} = 10^{-6}$, $p_{\text{GAIN}} = 0.36$, $p_{\text{WTCCC}} = 2 \times 10^{-8}$)
Linked to brain ischemia
Receptor for TENM2 that mediates heterophilic synaptic cell-cell contact and postsynaptic specialization
Shown to recruit the neurotoxin from black widow spider venom, alpha-latrotoxin, to the synapse plasma membrane
Mouse models show behavioral phenotypes

Kaanan P. Shah

Acknowledgement

Nancy J. Cox
Eric R. Gamazon
Kaanan P. Shah
Sahar Mozaffari
Heather E. Wheeler
Jason M. Torres
Keston Aquino-Michaels

GTE_x Consortium

Funding Sources

UC CTSA K12 grant
NCI K12CA139160

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

GTE_x
R01 MH090937 and R01 MH101820

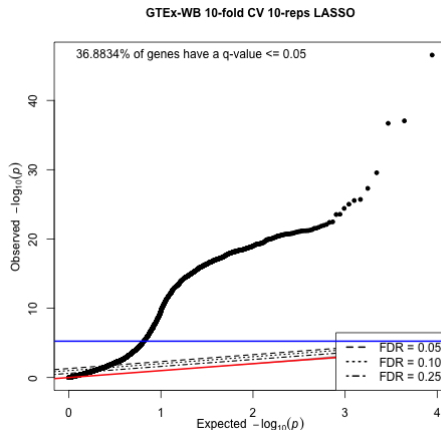
PAAR
NIH/NIGMS U01GM61393

Conte Center grant
P50MH094267

DHX16 PPP1R18 MUC21 MICA DDX39B TNF LTB PRRC2A
GPANK1 C6orf25 DDAH2 MSH5 NEU1 CFB SKIV2L DOM3Z
C4B TNXB PPT2 RNF5 AGER RGL2 PHF1

How Well do we Predict the Transcriptome

Cross Validated LASSO within GTEx Whole Blood



~ 5500 genes with correlation > 0.10

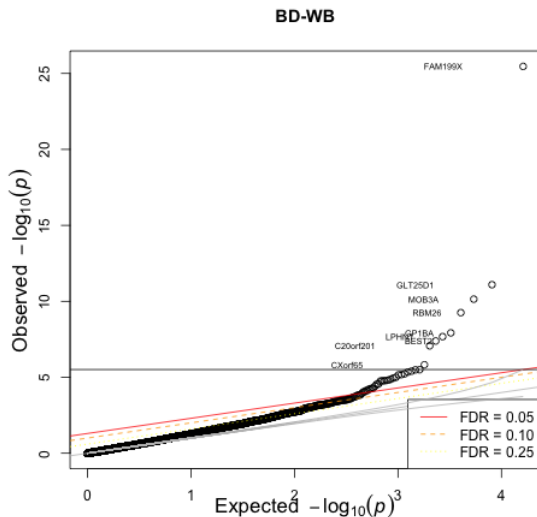
Heather E Wheeler

Number of eQTL's selected by LASSO

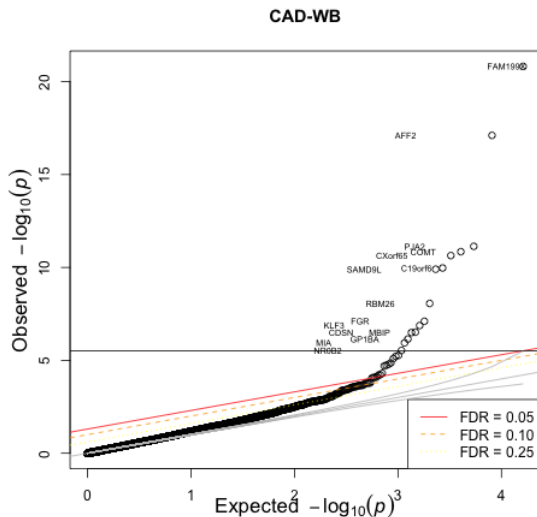
0	1	2	3	4	5	6	7	8	9+
10542	3092	1632	845	574	421	283	260	202	1489

Heather E Wheeler

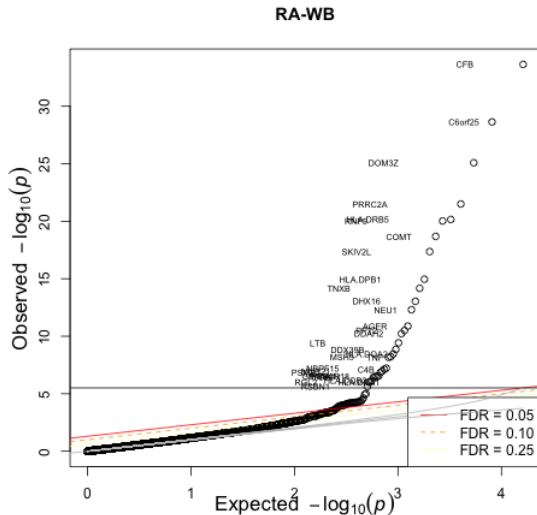
Bipolar Disorder vs. Predicted WB Expression



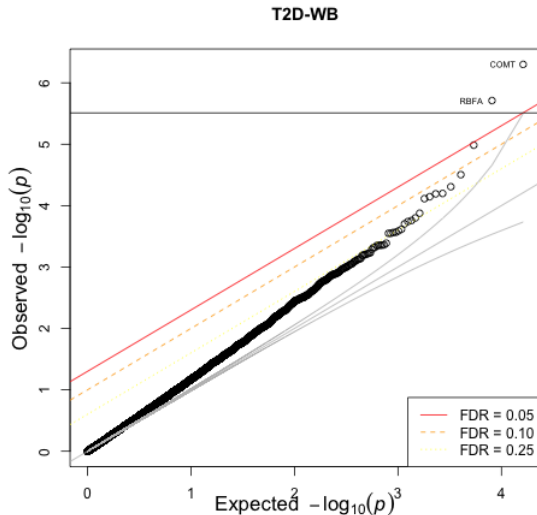
Coronary Artery Disease vs. Predicted WB Expression



Rheumatoid Arthritis vs. Predicted WB Expression



Type 2 Diabetes vs. Predicted WB Expression



- ▶ It directly tests whether genetic variants affect traits through gene expression regulation
- ▶ Built-in mechanism
- ▶ Applicable to any GWAS datasets
gene expression levels are computed from genotype data
- ▶ No reverse causality
Disease status cannot affect in silico levels thus ruling out reverse causality
- ▶ Reduced multiple testing burden
- ▶ Direction of effects
 - ▶ Positive: good candidates for drug target knocking down is easier
 - ▶ Negative: likely to find enrichment of rare variants in cases

PrediXcan: Training Datasets for Transcriptome Prediction

- ▶ GTE_x
 - ▶ Tissues
whole blood (n=185), tibial nerve (n=98), muscle (n=138)
84% Europeans, 14% African Americans
 - ▶ RNAseq
- ▶ Bipolar Disorder
 - ▶ Brain - cerebellum (n=138, Europeans)
 - ▶ Affymetrix Human Gene 1.0 ST
- ▶ Framingham
 - ▶ White blood cell (n=5000, European Americans)
 - ▶ Affymetrix GeneChip Exon Array
- ▶ GEUVADIS/1000 Genomes
 - ▶ LCL (n=373 Europeans, n=89 Africans)
 - ▶ RNAseq

- ▶ PrediXcan is a powerful gene based association test
- ▶ Directly tests the hypothesis that genetic variants affect disease status through the regulation of genes
- ▶ Application to the WTCCC data recapitulates known genes and identifies many novel genome-wide significant ones
- ▶ Bipolar Disorder genes replicated on independent dataset
- ▶ Promising peripheral neuropathy gene was identified