Clinical Utility of Pharmacogenomic Findings: Beyond Single Variants

Hae Kyung Im, PhD

THE UNIVERSITY OF CHICAGO

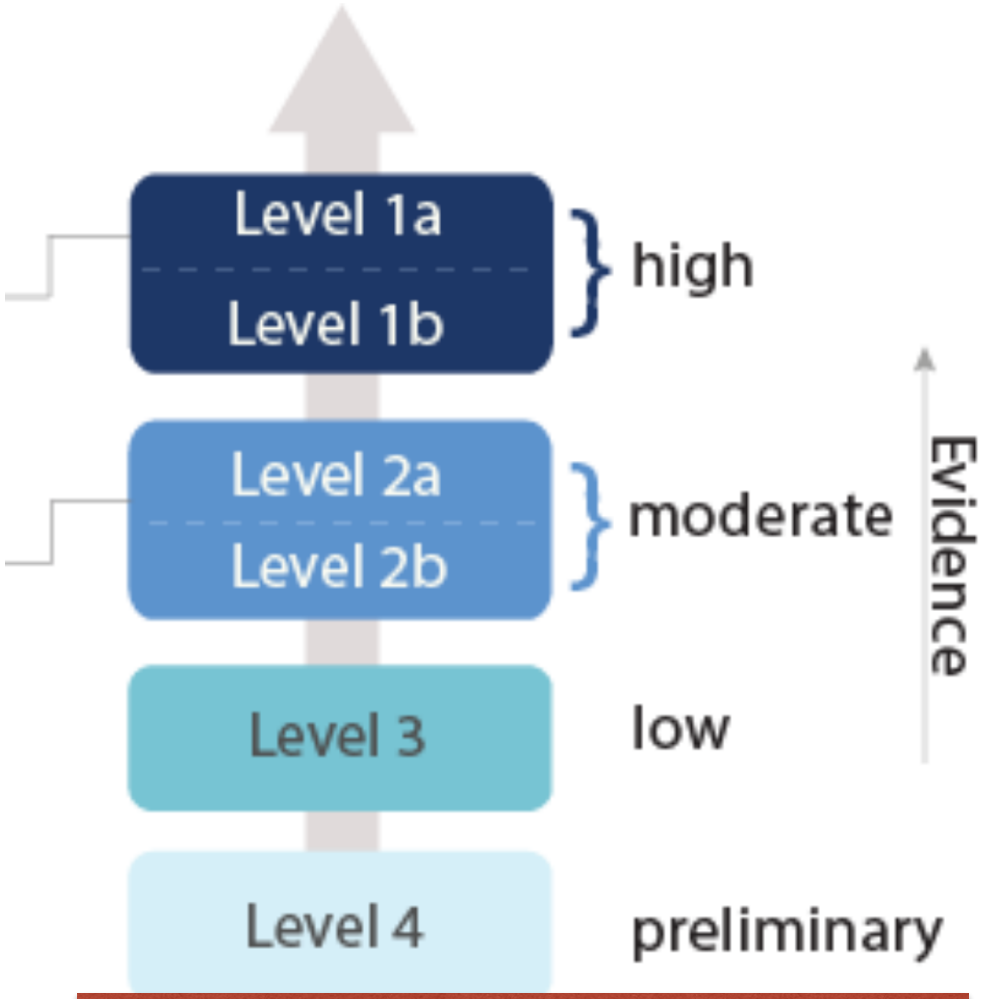
Successes and Challenges of Genome Studies

- 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

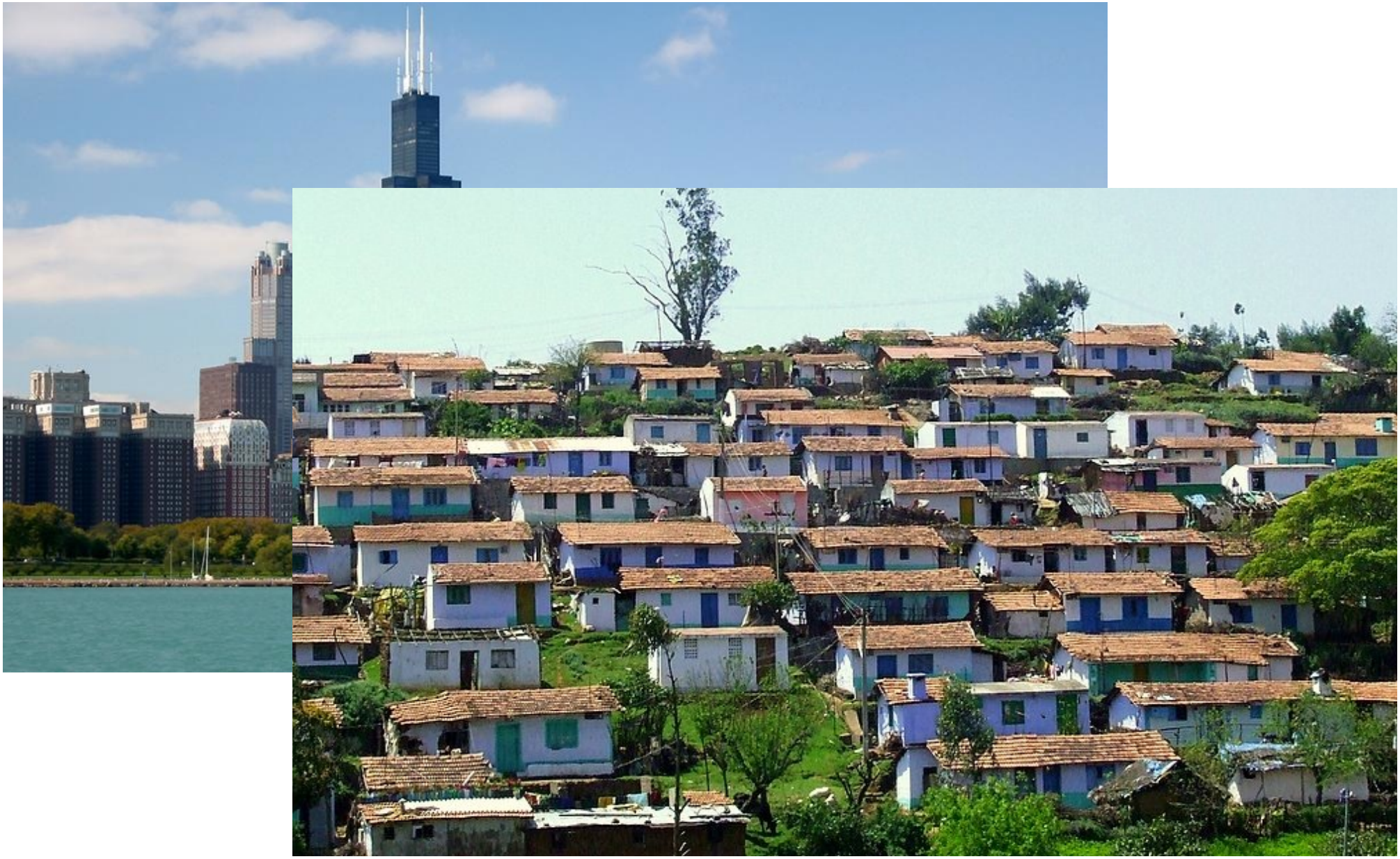
https://www.pharmgkb.org/



Only Level 1a finding have clinical guidelines

Genetic Architecture of Complex Traits

Genetic Architecture of Complex Traits

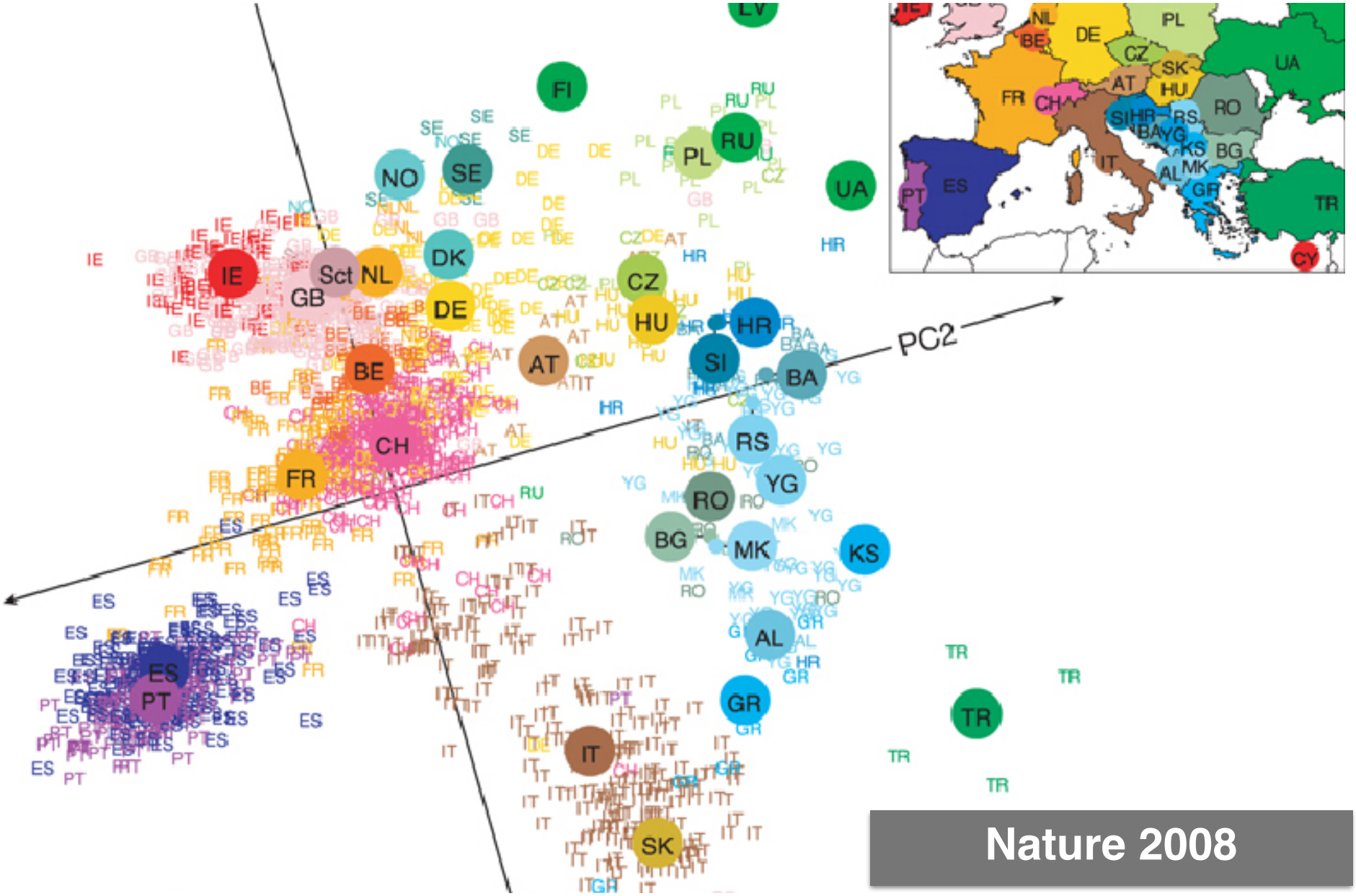


Genetic Architecture of Complex Traits

Single Variants Not Relevant for Highly Polygenic Traits

Genes mirror geography within Europe

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



Nature 2008

Prediction and Dissection to Achieve Clinical Utility

Prediction

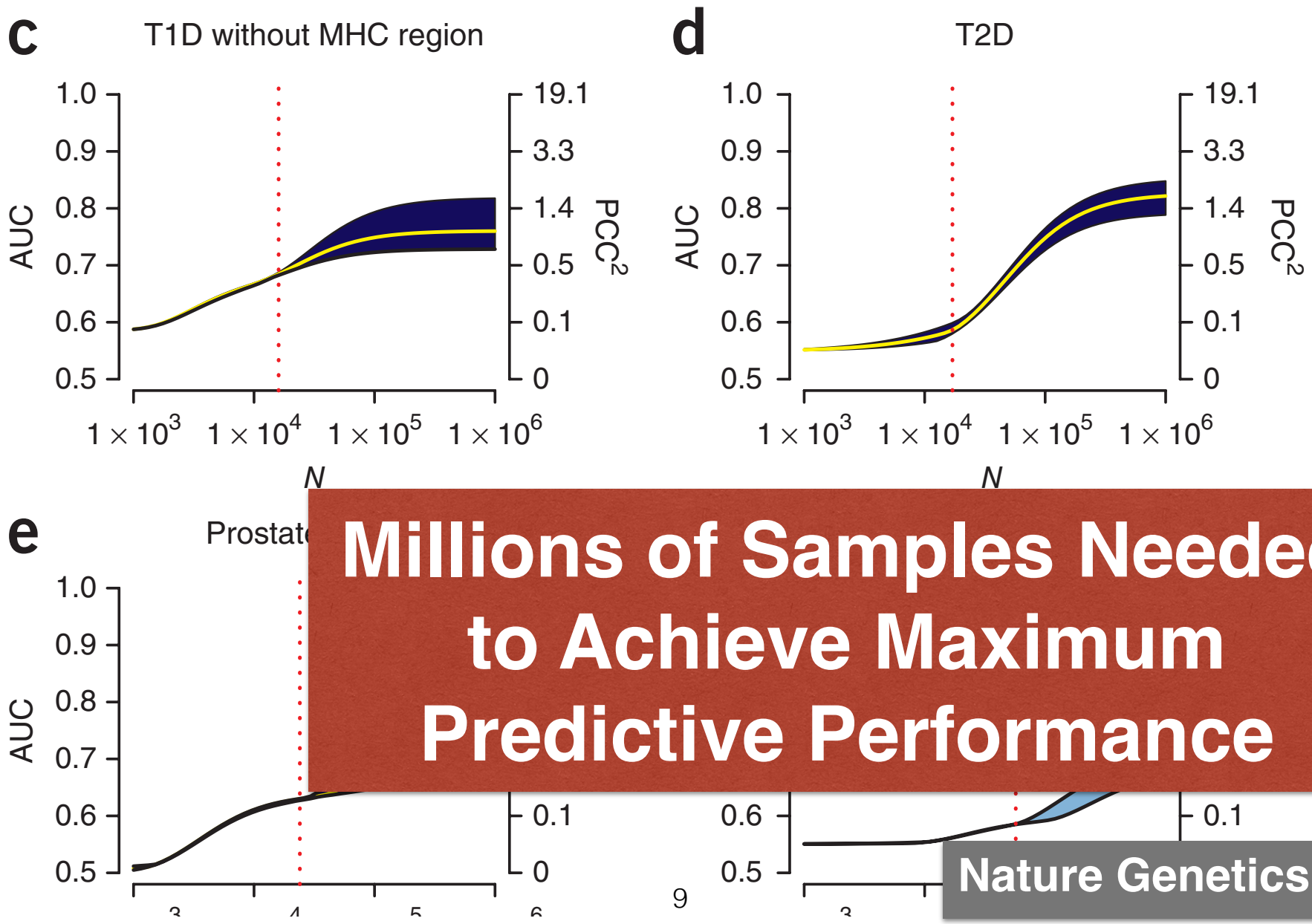
- Disease
  - risk stratification
  - intervention strategies
- Adverse events
- Efficacy of treatment

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[1], Bill Wheeler[2], Joshua Sampson[1], Patricia Hartge[1], Stephen J Chanock[1] & Ju-Hyun Park[1,3]



Millions of Samples Needed to Achieve Maximum Predictive Performance

Nature Genetics 2013

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,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

Whole Genome Prediction Approaches

RESEARCH ARTICLE

Genetic Epidemiology
OFFICIAL JOURNAL
INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY
www.geneticepi.org

Poly-Omic Prediction of Complex Traits: OmicKriging

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

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

OPEN ACCESS Freely available online

PLOS | GENETICS

Polygenic Modeling with Bayesian Sparse Linear Mixed Models

Xiang Zhou[1]*, Peter Carbonetto[1], Matthew Stephens[1,2]*

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

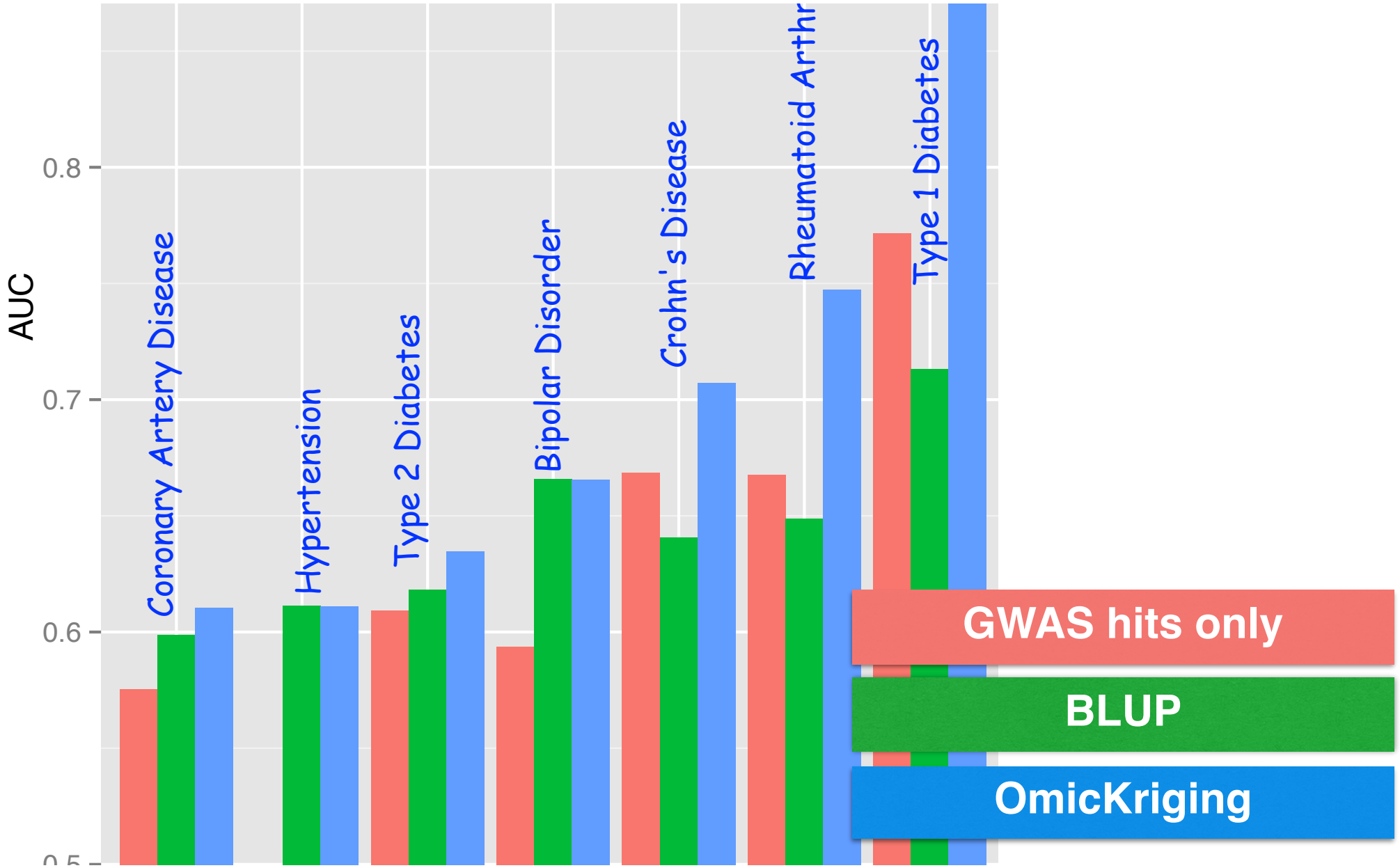
BMC Bioinformatics

SOFTWARE Open Access

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

Gad Abraham[1*], Adam Kowalczyk[1], Justin Zobel[1] and Michael Inouye[2,3]

GWAS hits vs. Whole Genome Prediction (OmicKriging)

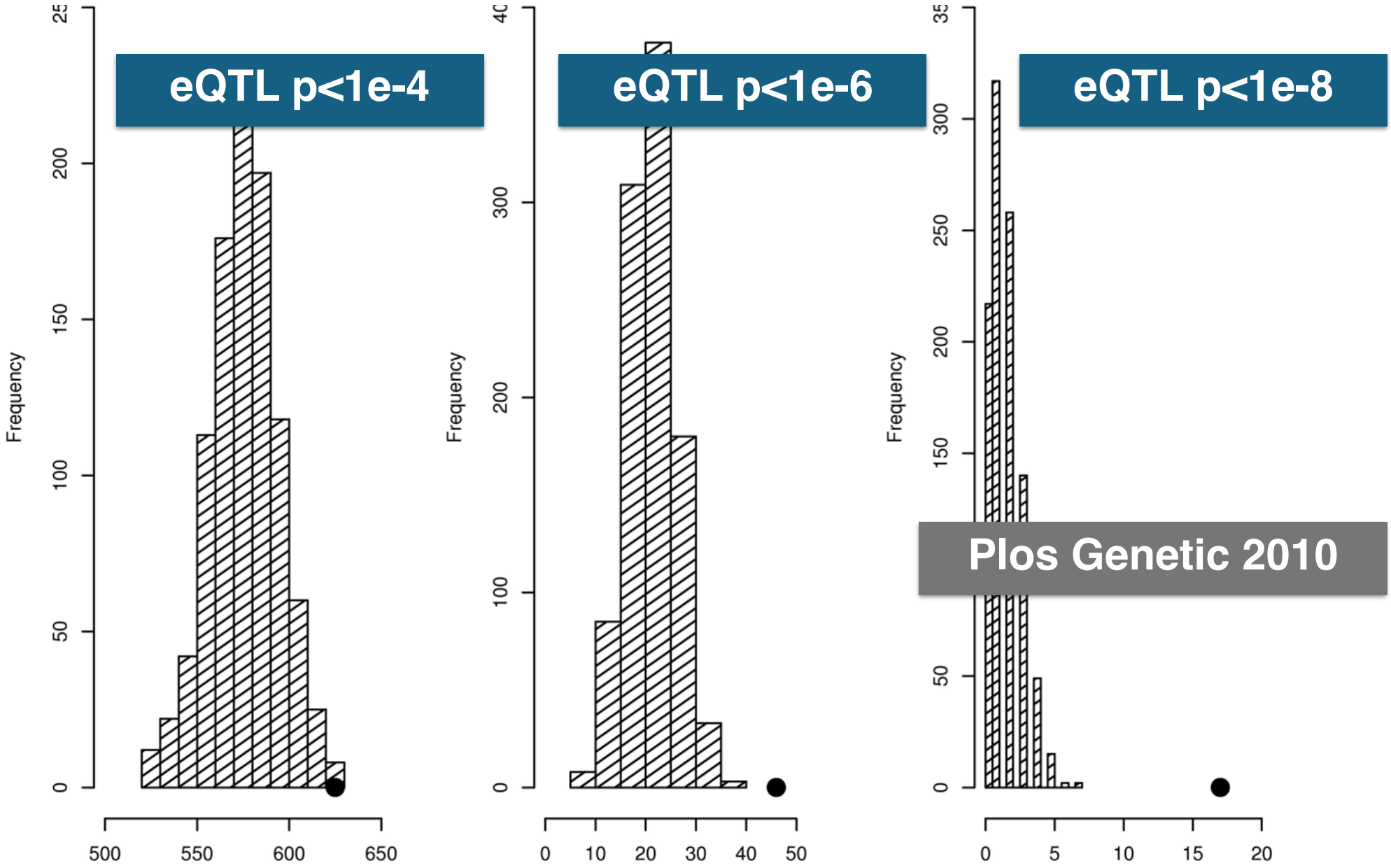


Collective Approaches to Dissect Complex Traits

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

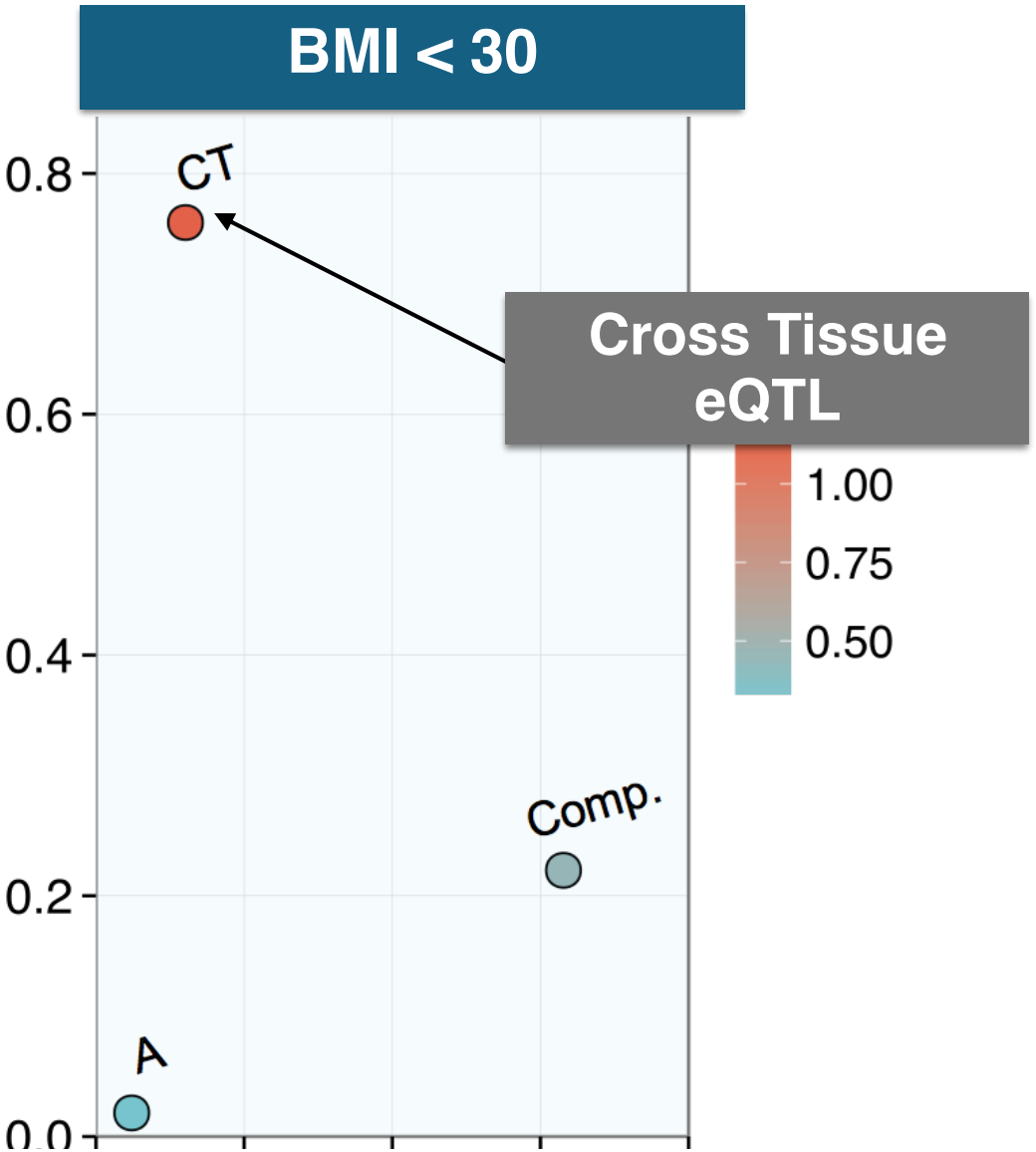
Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS

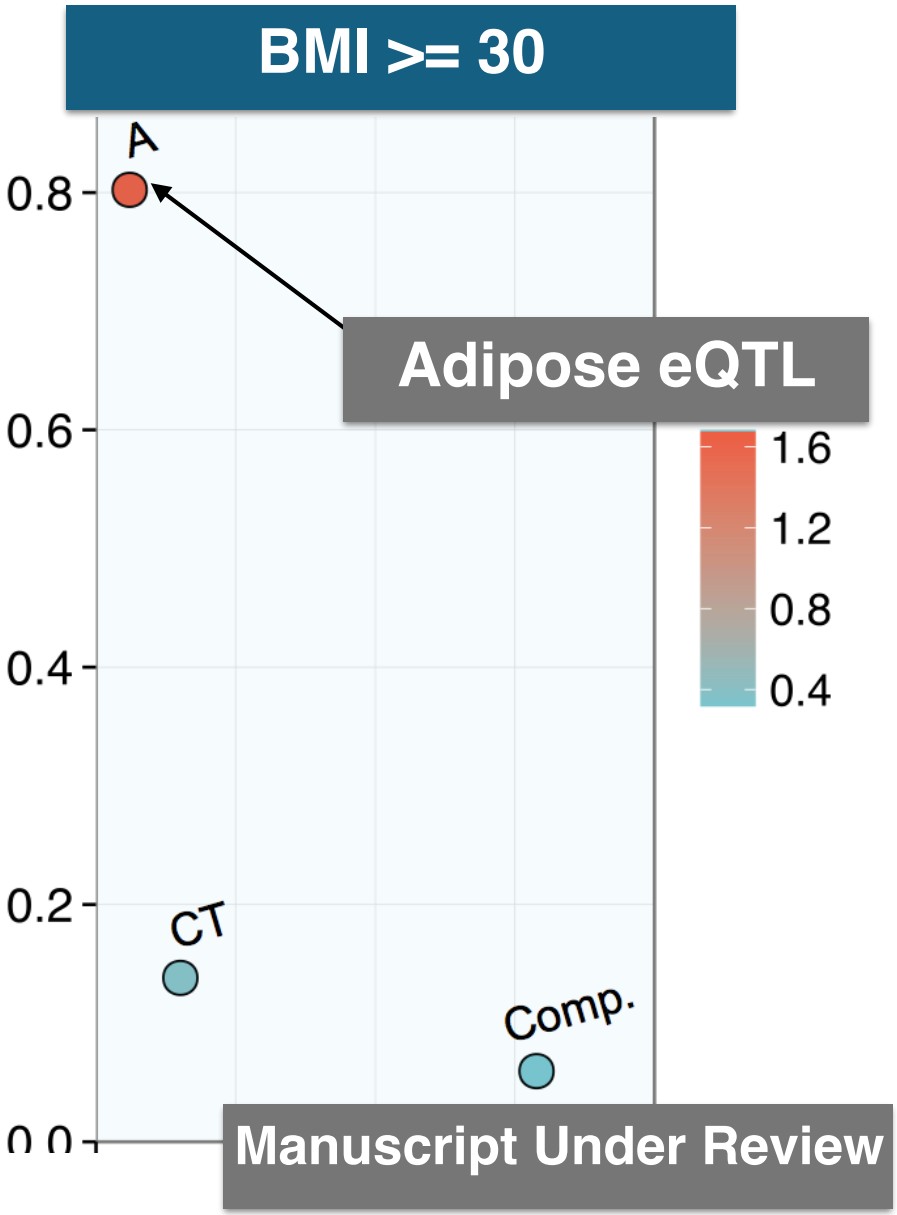
Dan L. Nicolae[1,2,3], Eric Gamazon[1], Wei Zhang[1], Shiwei Duan[1¤], M. Eileen Dolan[1,2], Nancy J. Cox[1,



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*

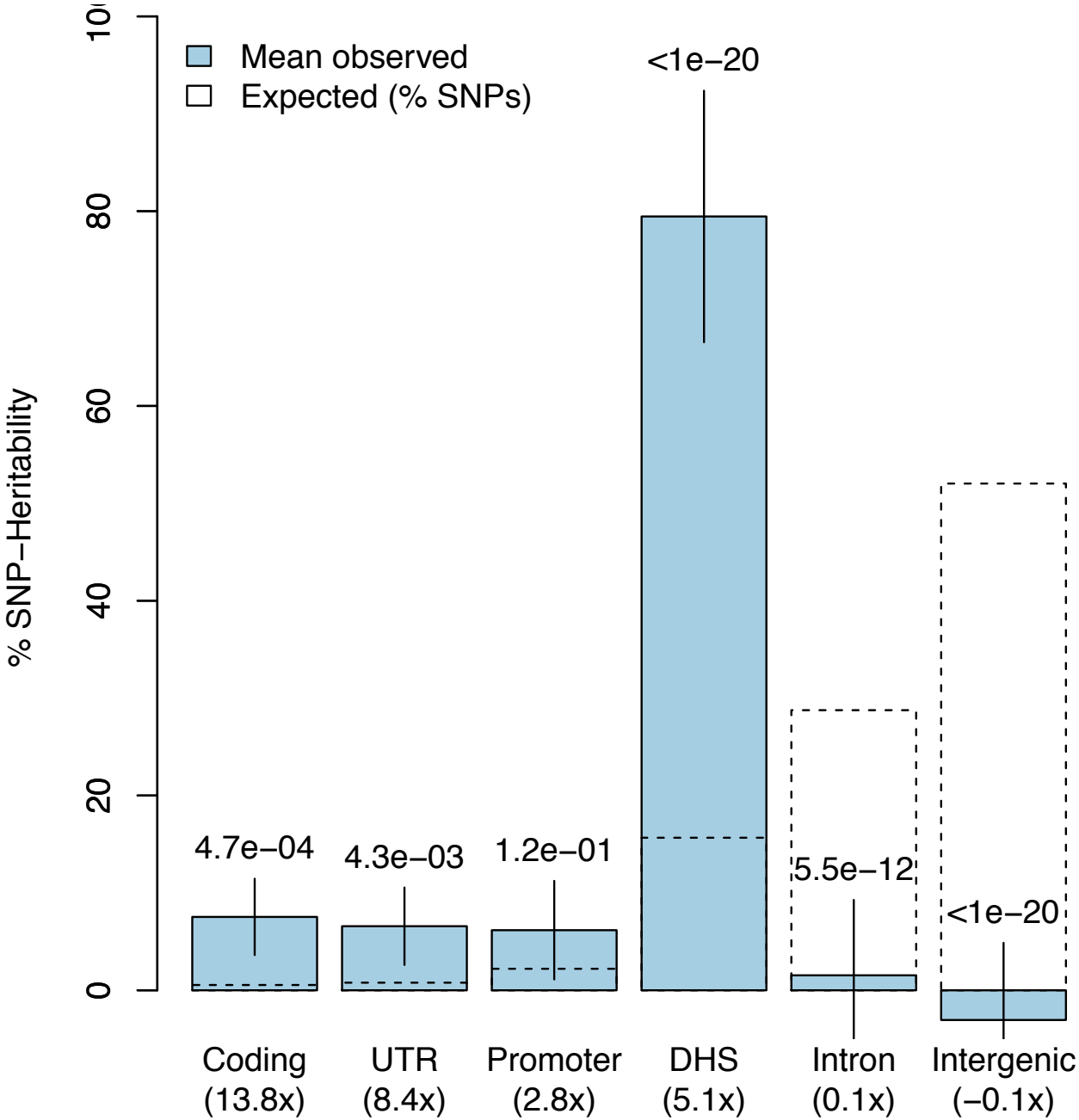




Regulatory variants explain much more heritability than coding variants across 11 common diseases

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

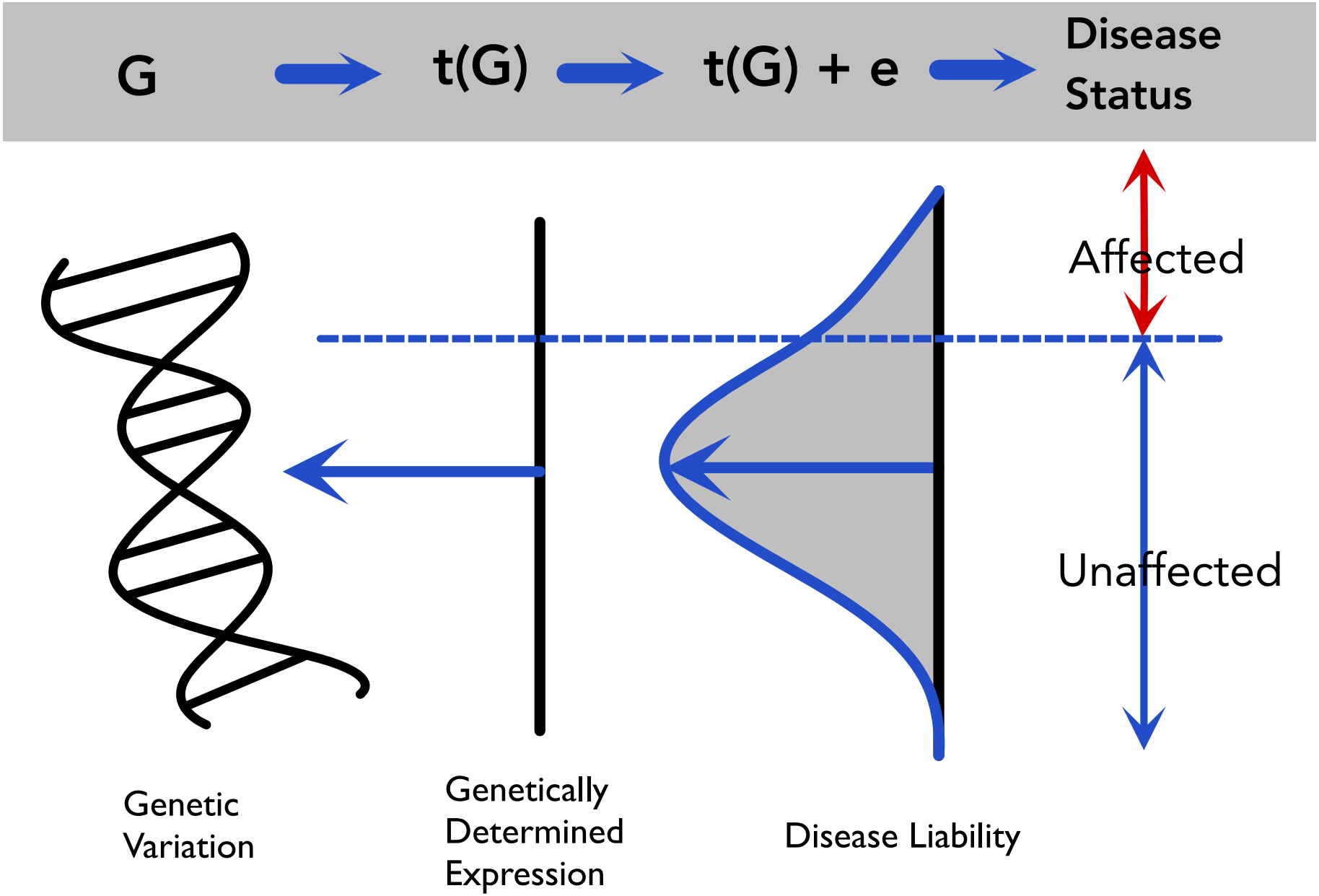
Gene Based Tests

- 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

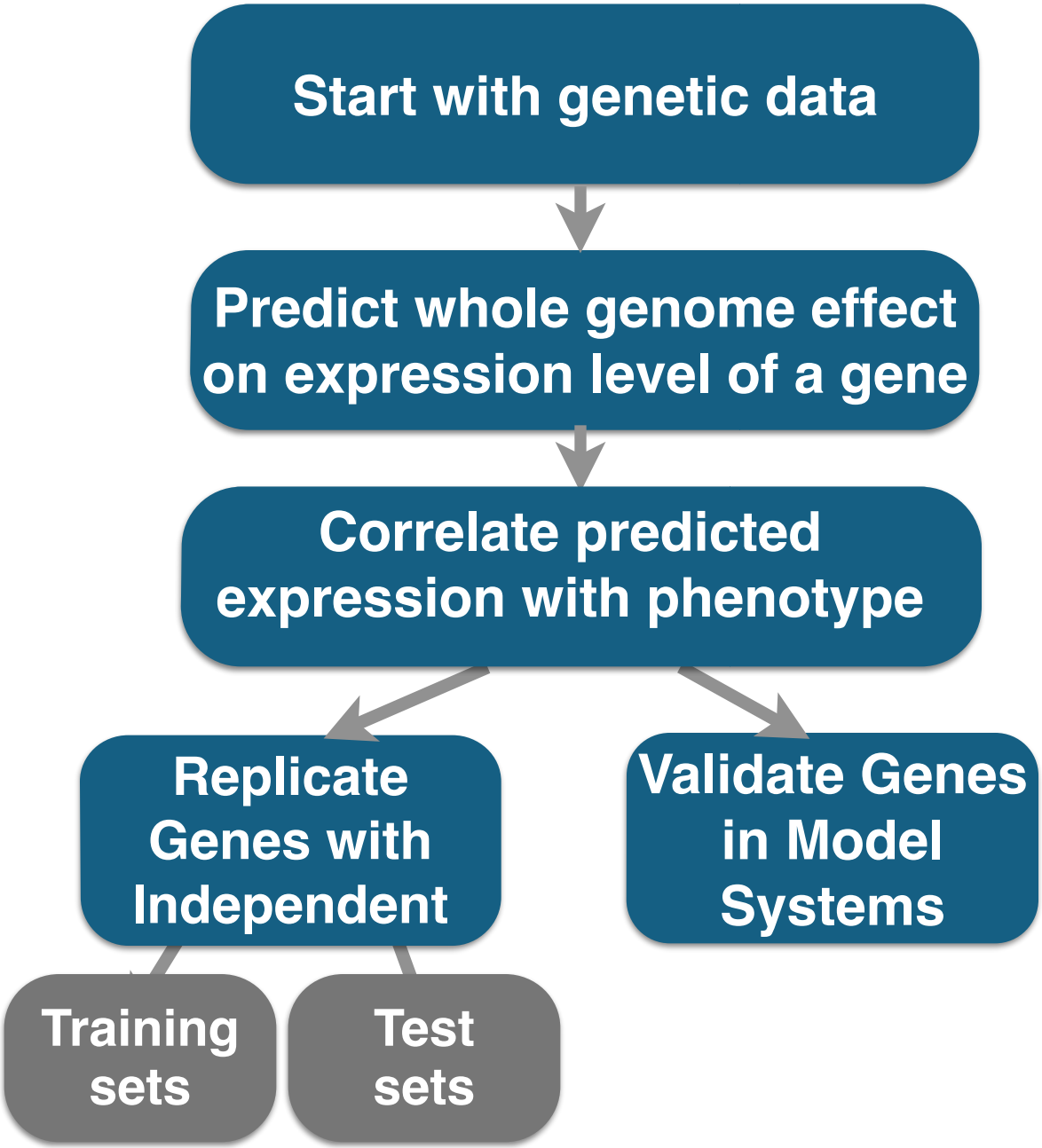
Gene Based Tests

- 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



- No transcriptome data is needed
- Only genotype and phenotype data needed
- No reverse causality
- Reduced multiple testing burden
- Mechanism is built in

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

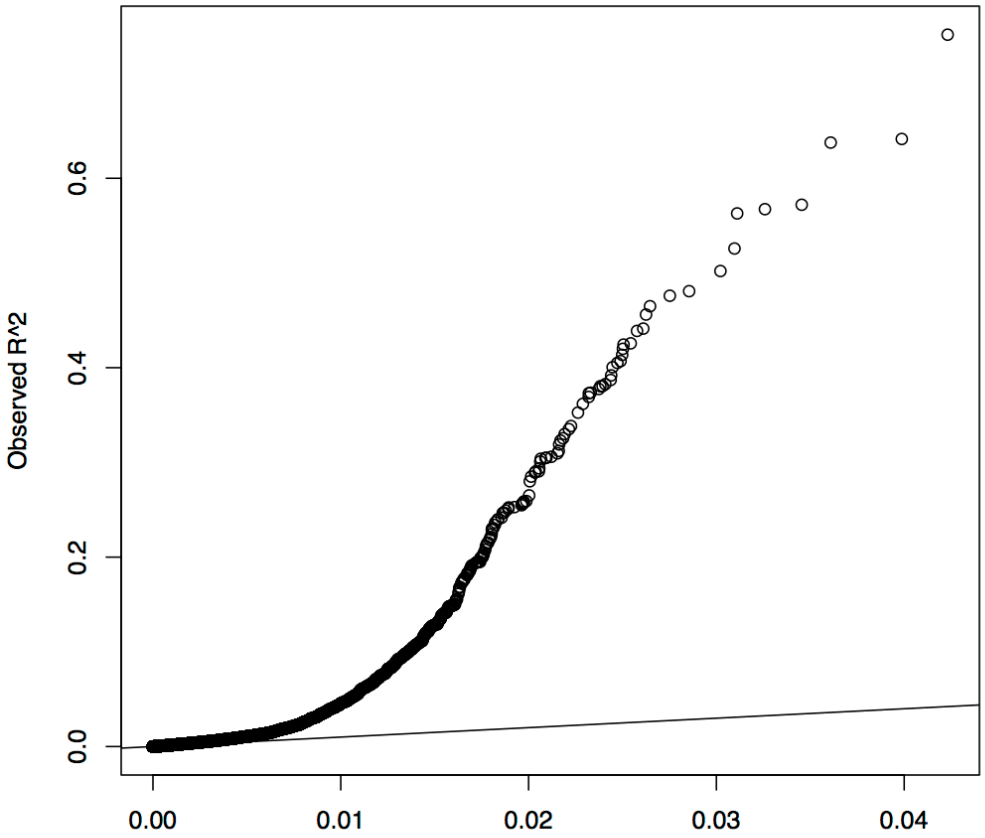
- w_k = single variant regression coefficient (Matrix eQTL output)
- w_k set to zero if p value > 0.05 for cis SNPs (1Mb TSS)
- w_k set to zero if p value $> 10^{-6}$ for trans SNPs

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)

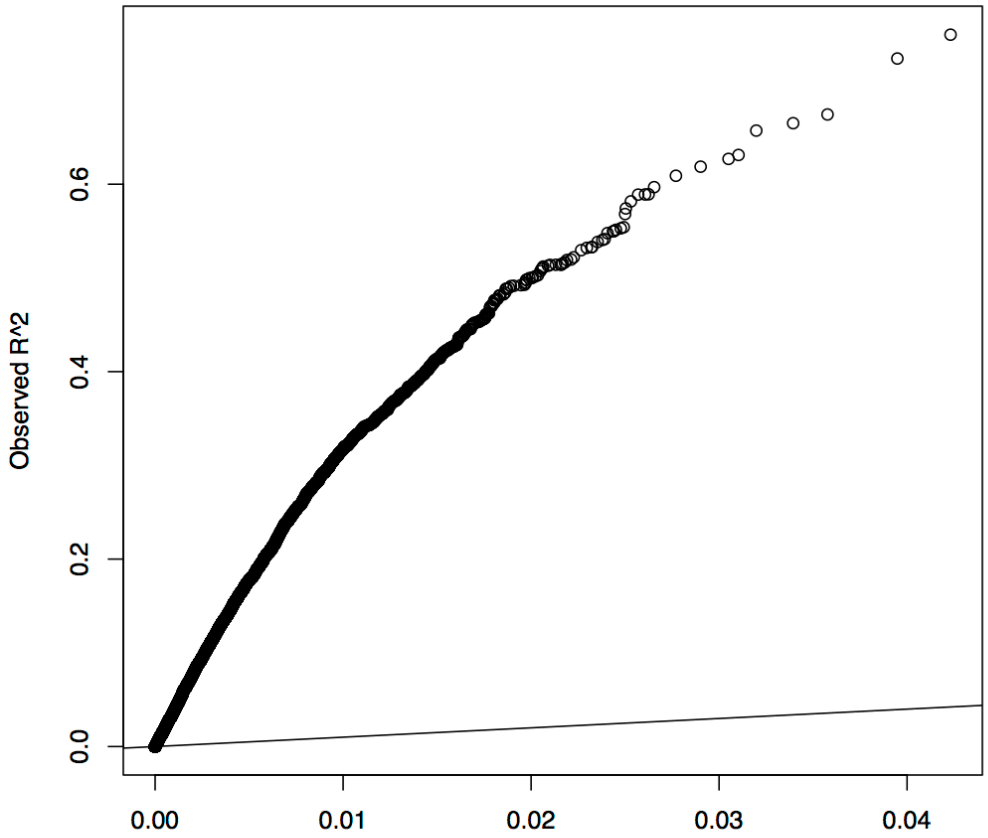
Good Prediction Performance

Prediction R^2



Training with GTEx Testing in 1K Genomes

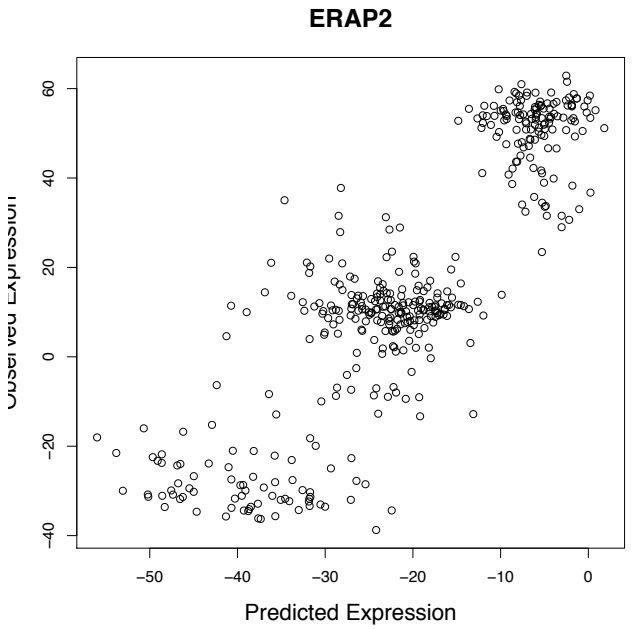
Replication R^2

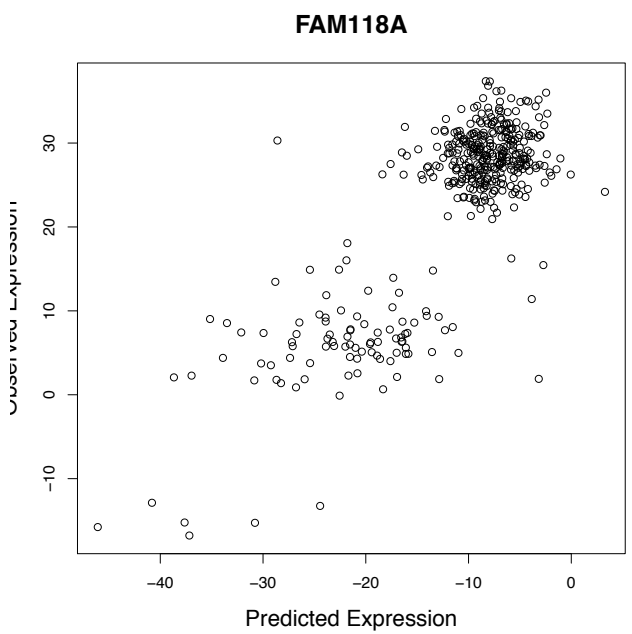


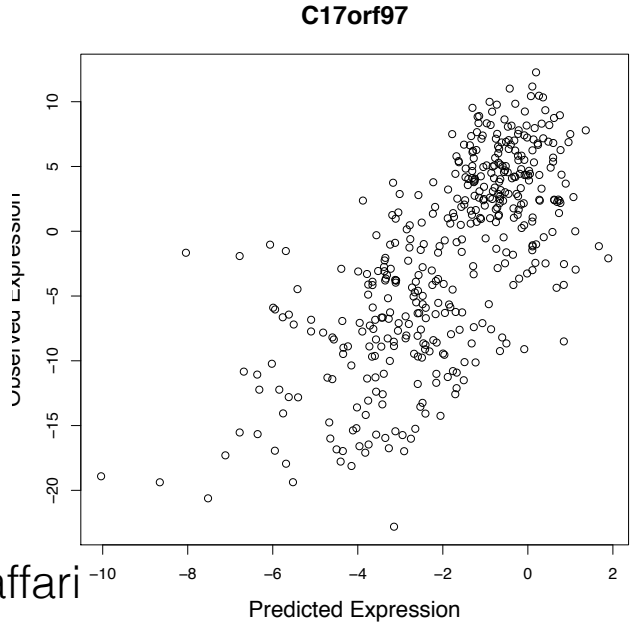
Replicate RNAseq Pickrell et al 2010 vs. 1K Genomes 2013

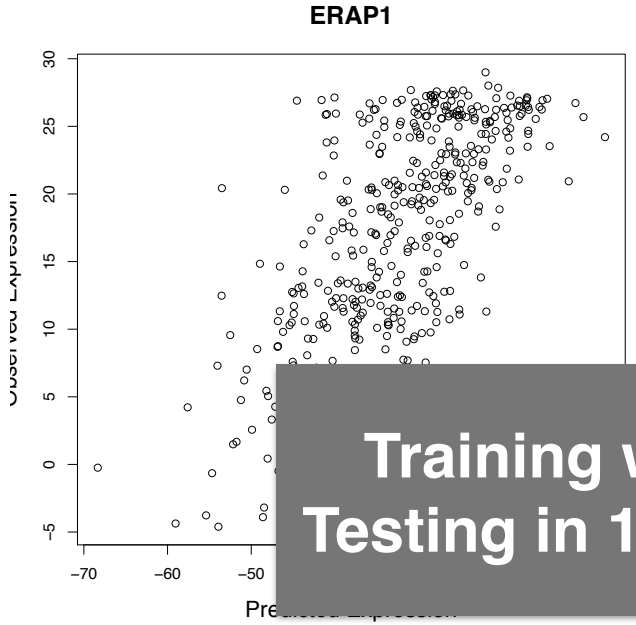
Sahar Mozaffari

Examples of Well Predicted Genes





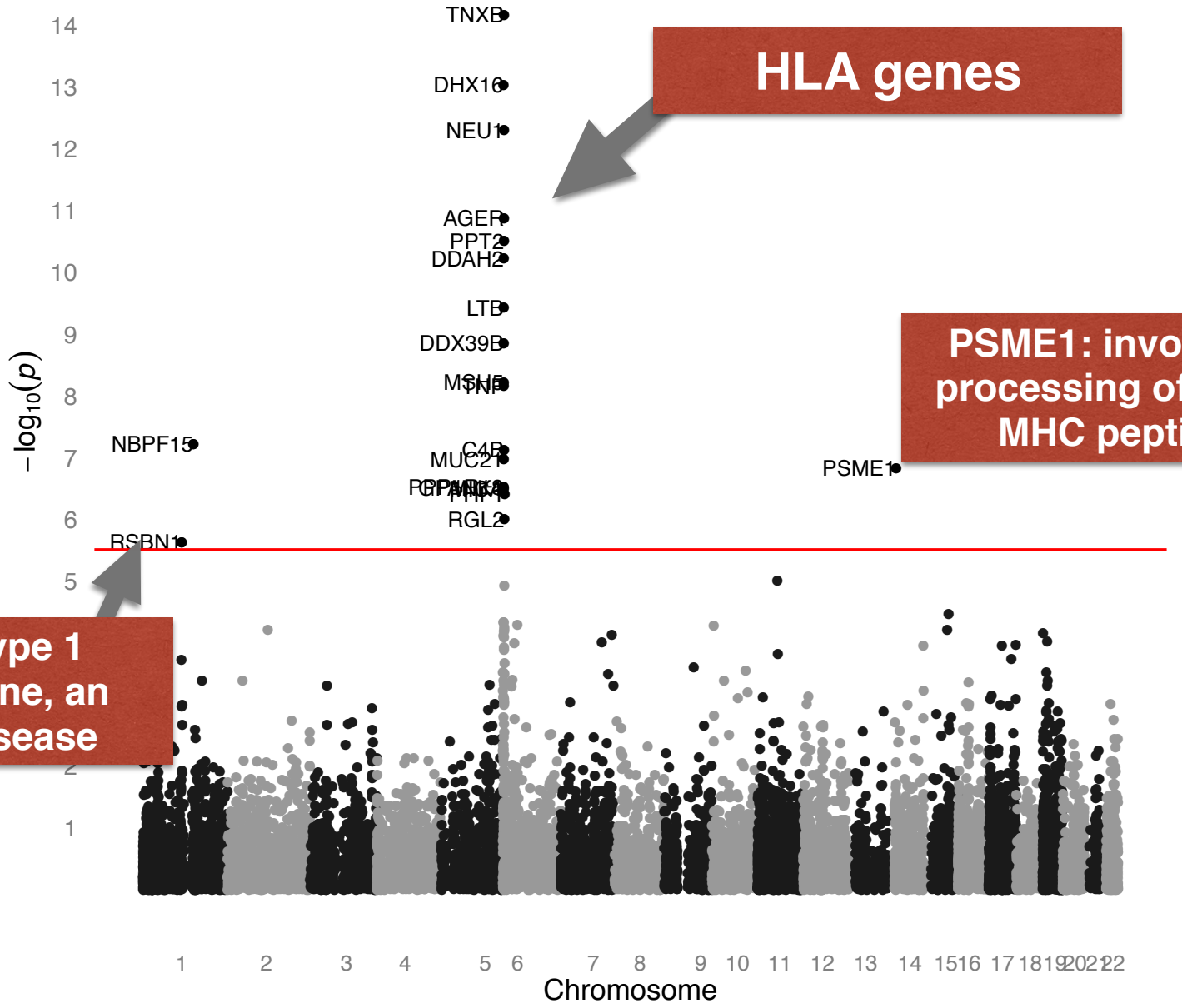




Training with GTEx
Testing in 1K Genomes

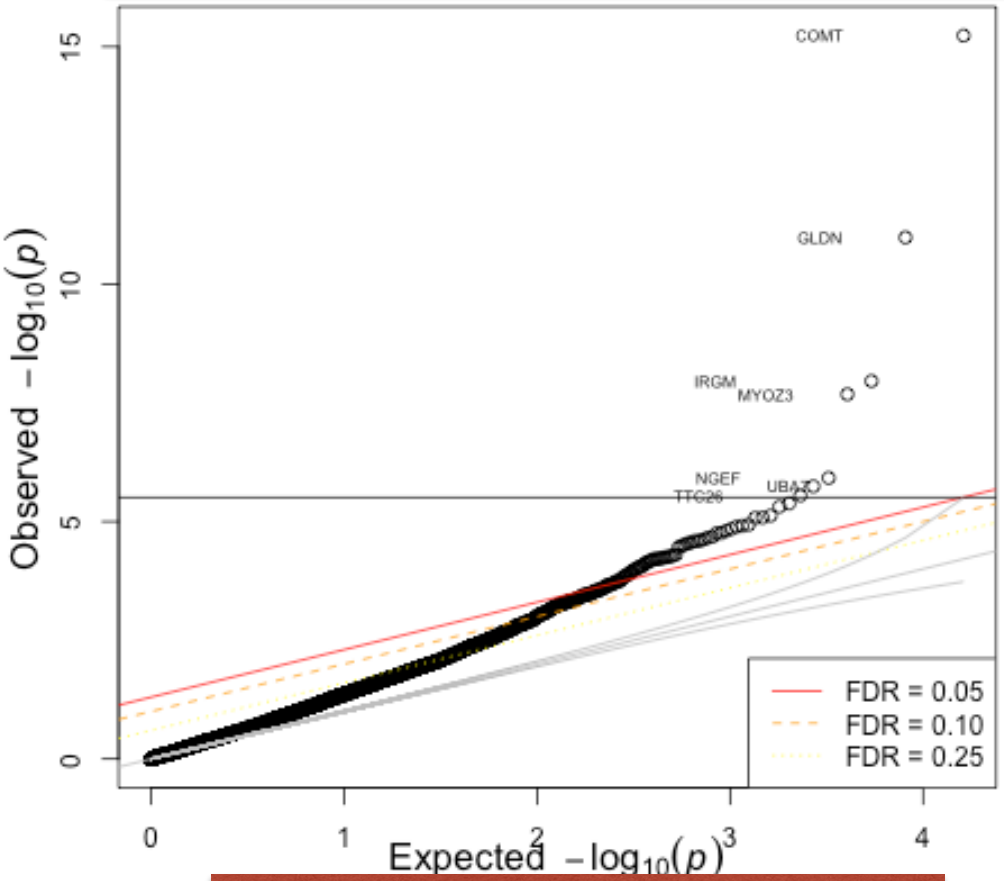
Sahar Mozaffari

Genes Associated with Rheumatoid Arthritis



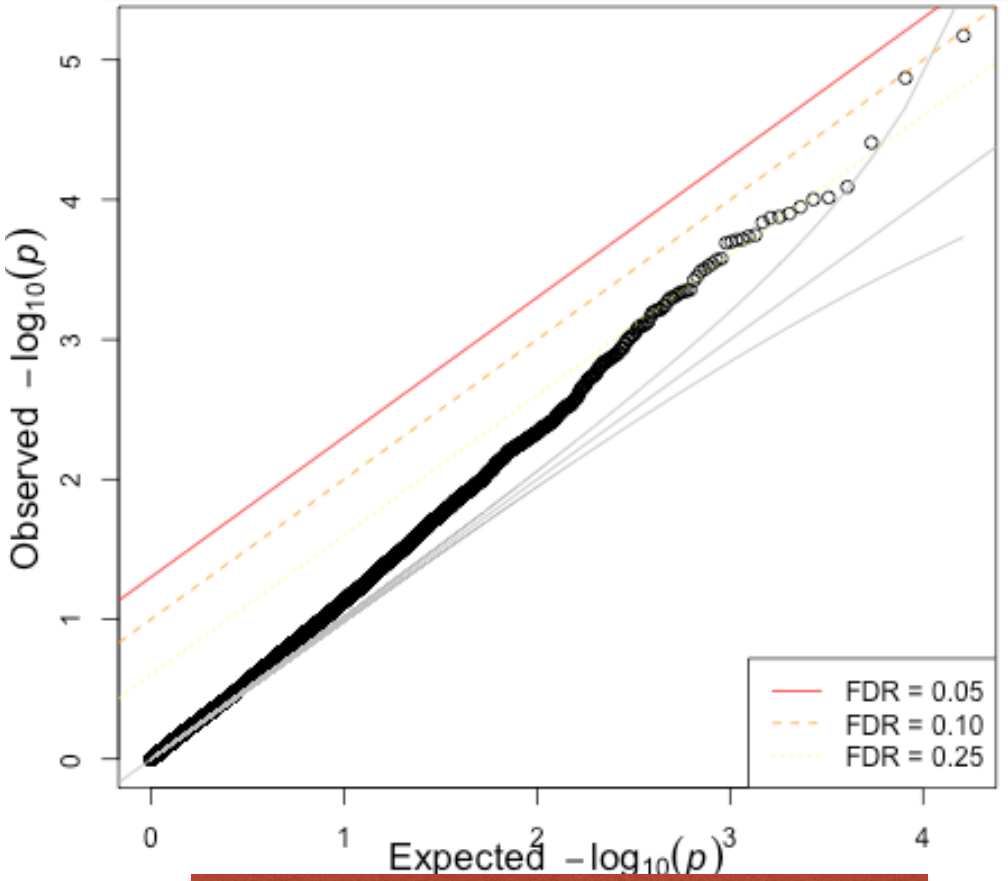
PrediXcan Results for Crohn's Disease and Hypertension

Crohn's Disease



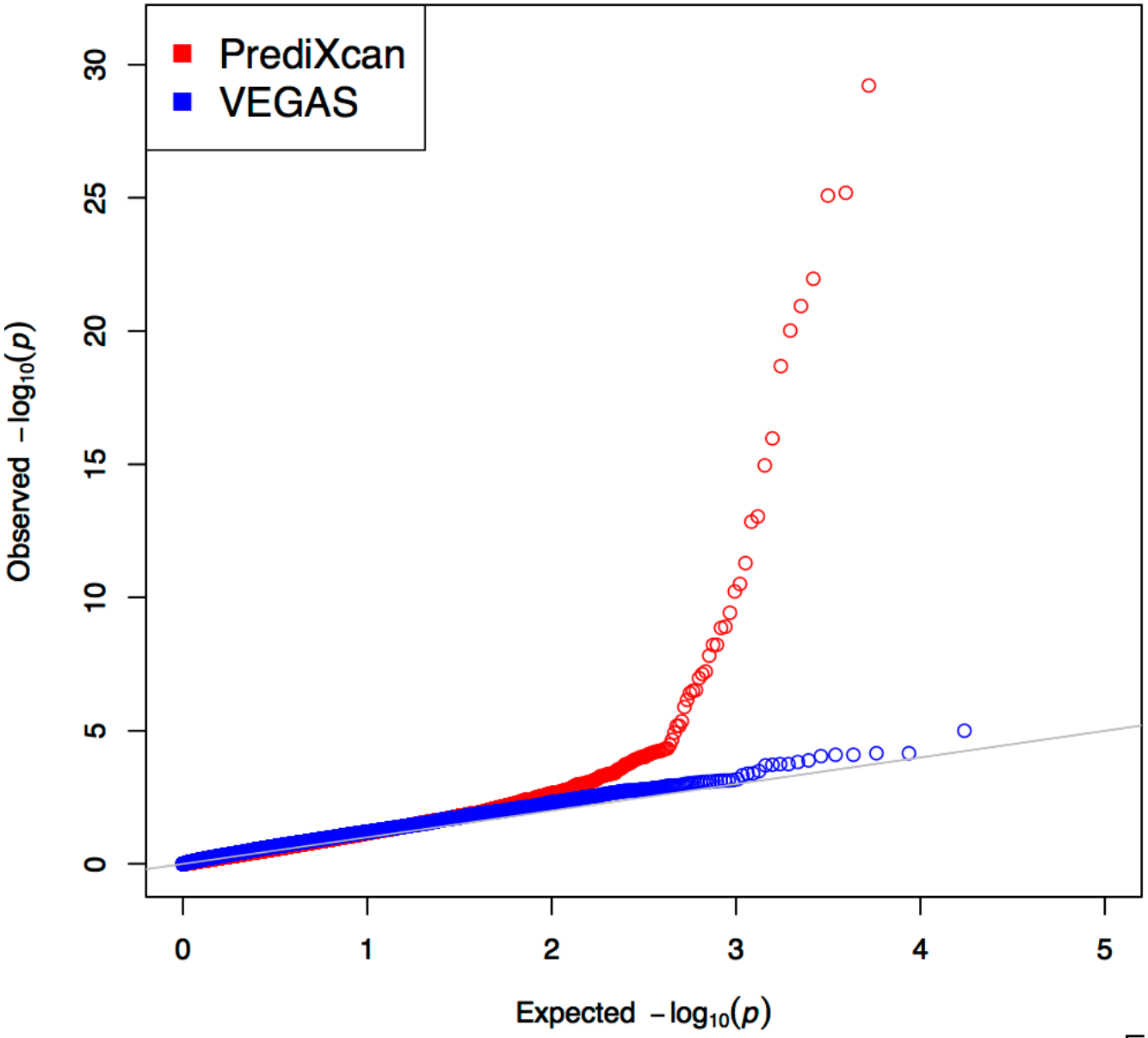
IRGM is a known Crohn's gene

Hypertension



Whole blood may not be relevant tissue

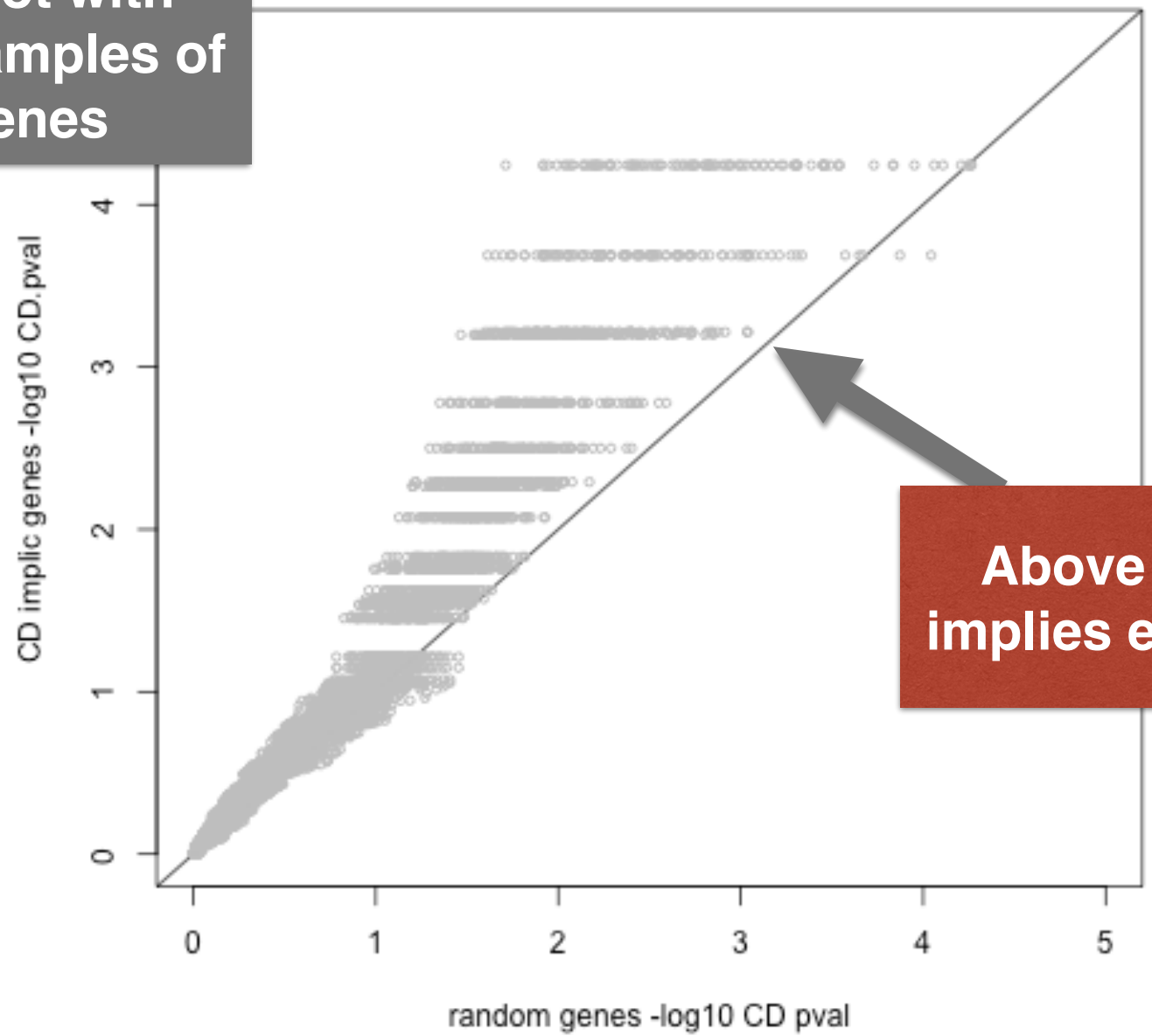
PrediXcan Outperforms VEGAS



Eric Gamazon

Enrichment of Known Crohn's Genes Among Findings

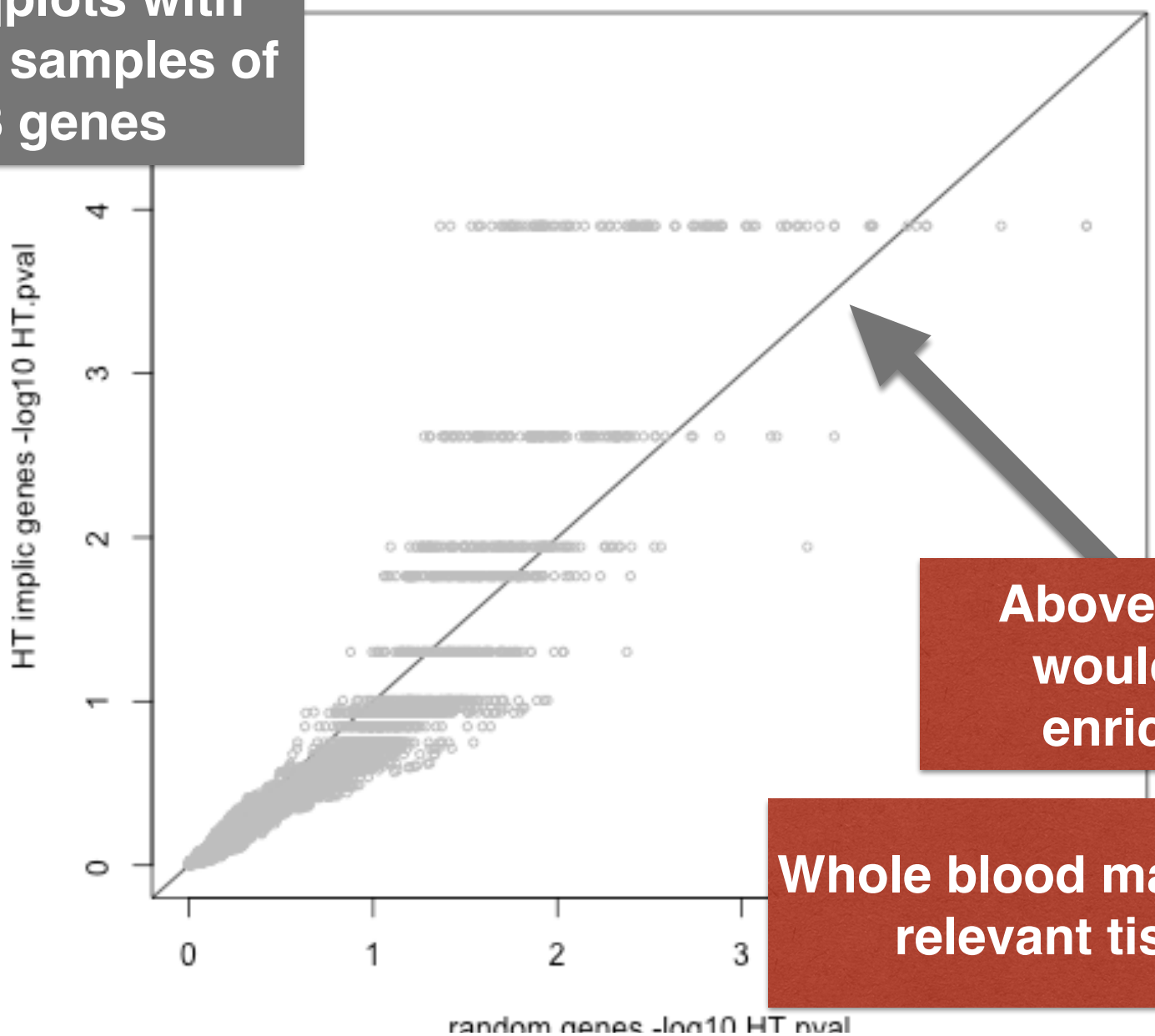
100 qqplot with random samples of 205 genes



Above this line implies enrichment

No Enrichment Among Hypertension Findings

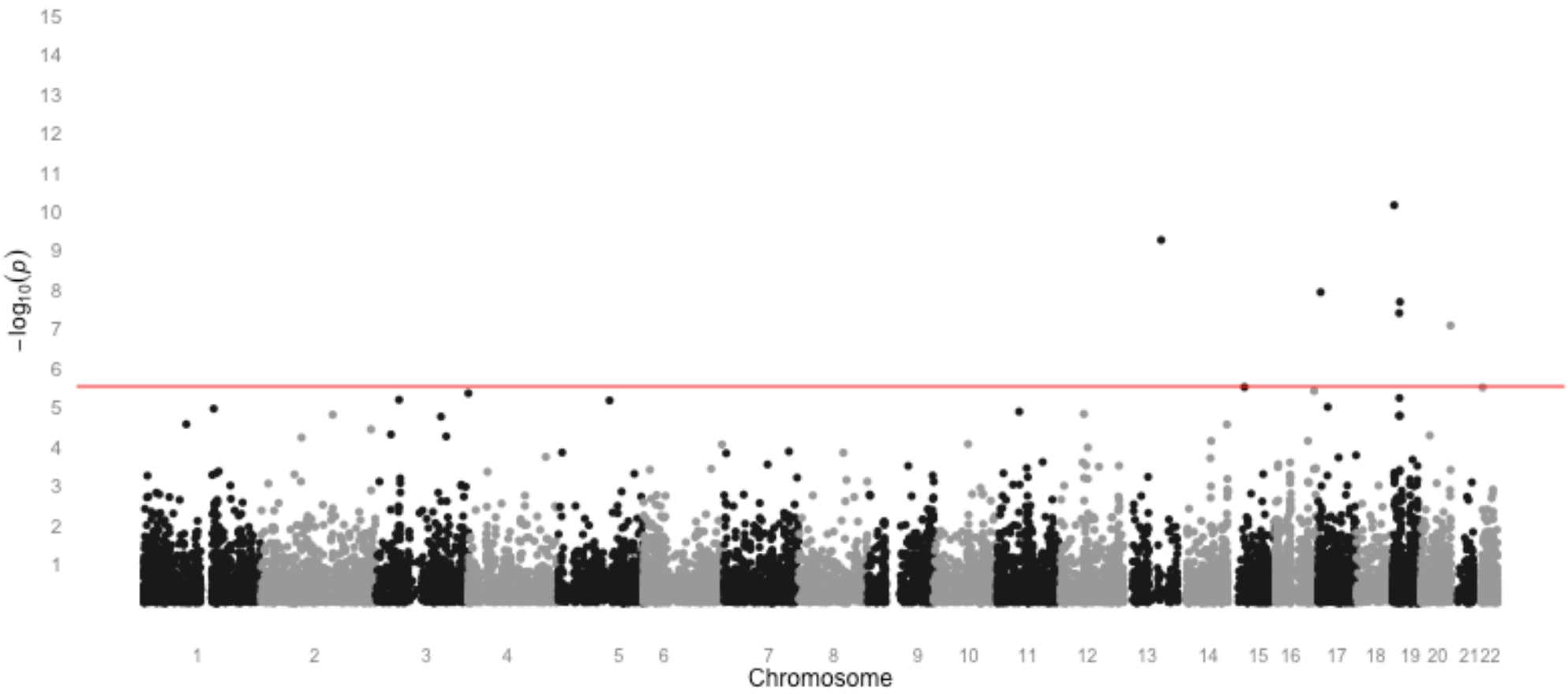
100 qqplots with random samples of 133 genes



Above this line would imply enrichment

Whole blood may not be relevant tissue

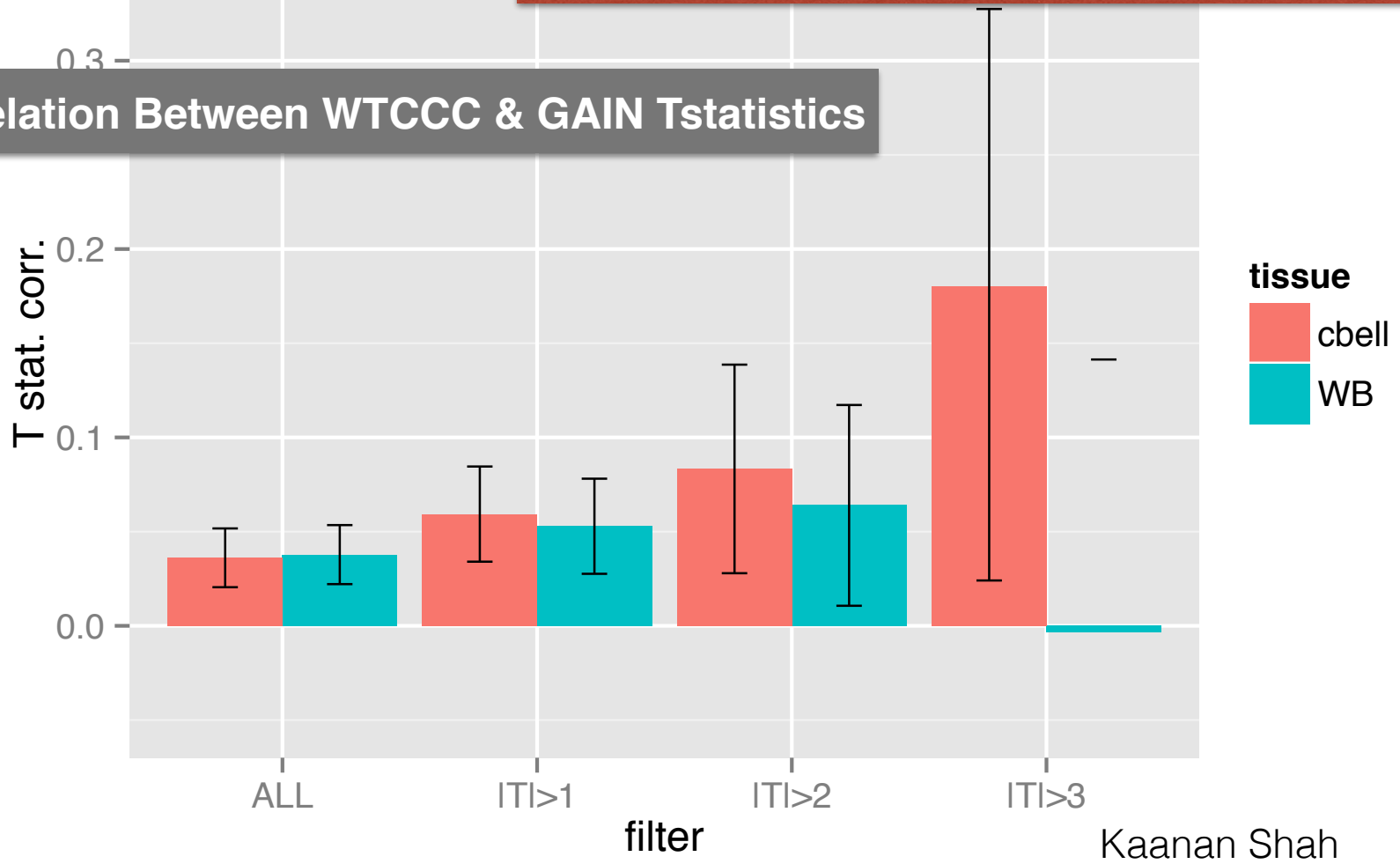
Bipolar Disorder WTCCC results



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

Bevacizumab Induced Hypertension

- Bevacizumab is a humanized monoclonal antibody that inhibits VEGF induced angiogenesis
- Hypertension is a common adverse event to bevacizumab treatment
- The incidence of hypertension with bevacizumab is 20-30%, while grade 3 or greater hypertension occurs in only 10-15% of patients.

Keston Aquino Michaels

Bevacizumab Trials

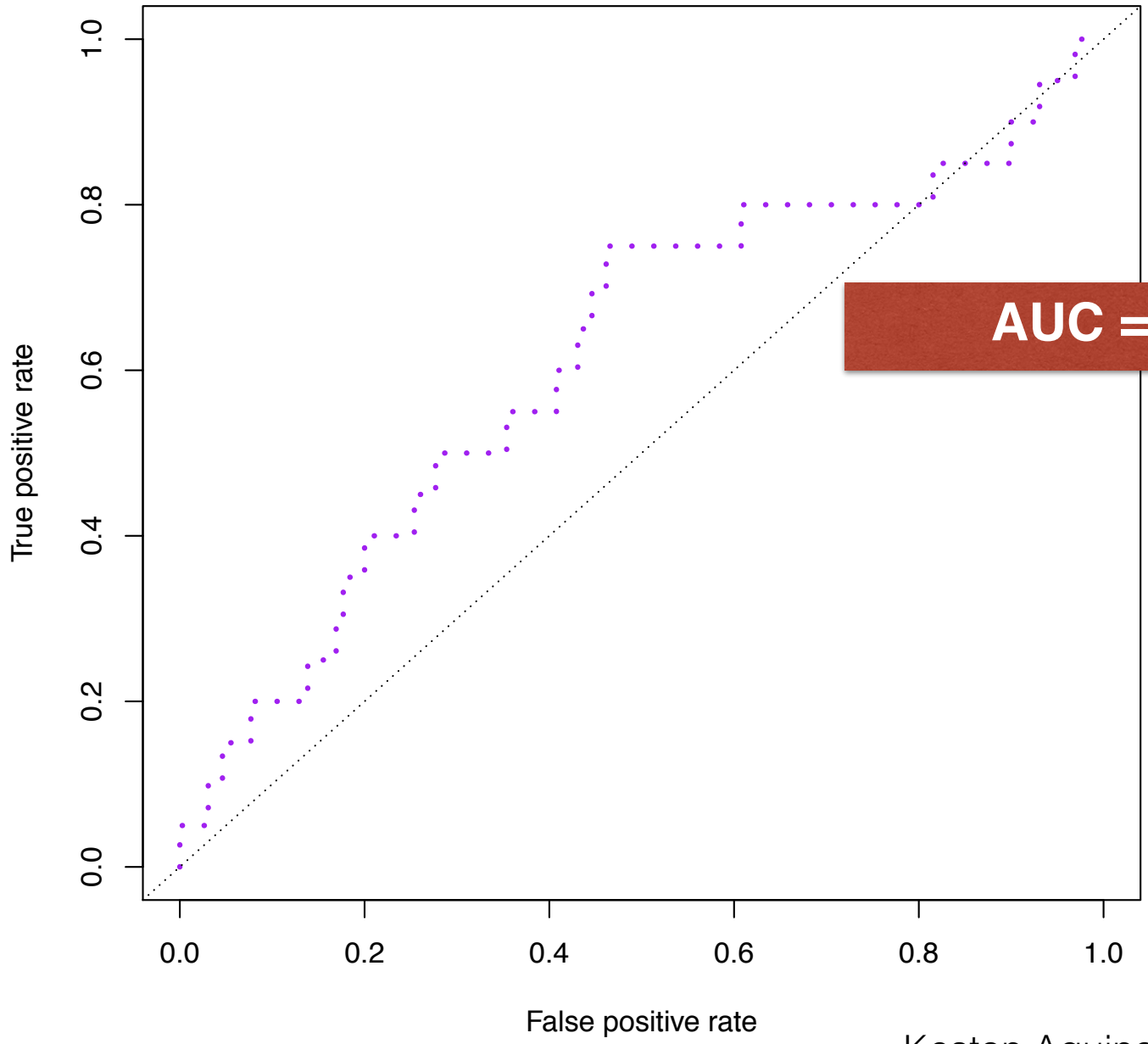
- CALGB 90401
  - a randomized double-blinded placebo controlled phase III trial comparing docetaxel and prednisone with and without bevacizumab in men with hormone refractory prostate cancer
  - n = 664 (with genotype data after QC)
  - PI: Howard McLeod
- CALGB 80303
  - a randomized phase III trial of gemcitabine plus bevacizumab versus gemcitabine plus placebo in patients with advanced pancreatic cancer
  - n = 152 (with genotype data after QC)
  - PI: Federico Innocenti

Bevacizumab Induced Hypertension

- Is primary hypertension risk score predictive of bevacizumab induced hypertension
  - Hypertension results from Cross Consortia Pleiotropy group (n~20K)
- Can we predict drug induced hypertension?
  - 90401 training set
  - 80303 test set
- Dissection of Hypertension

Keston Aquino Michaels & Heather Wheeler

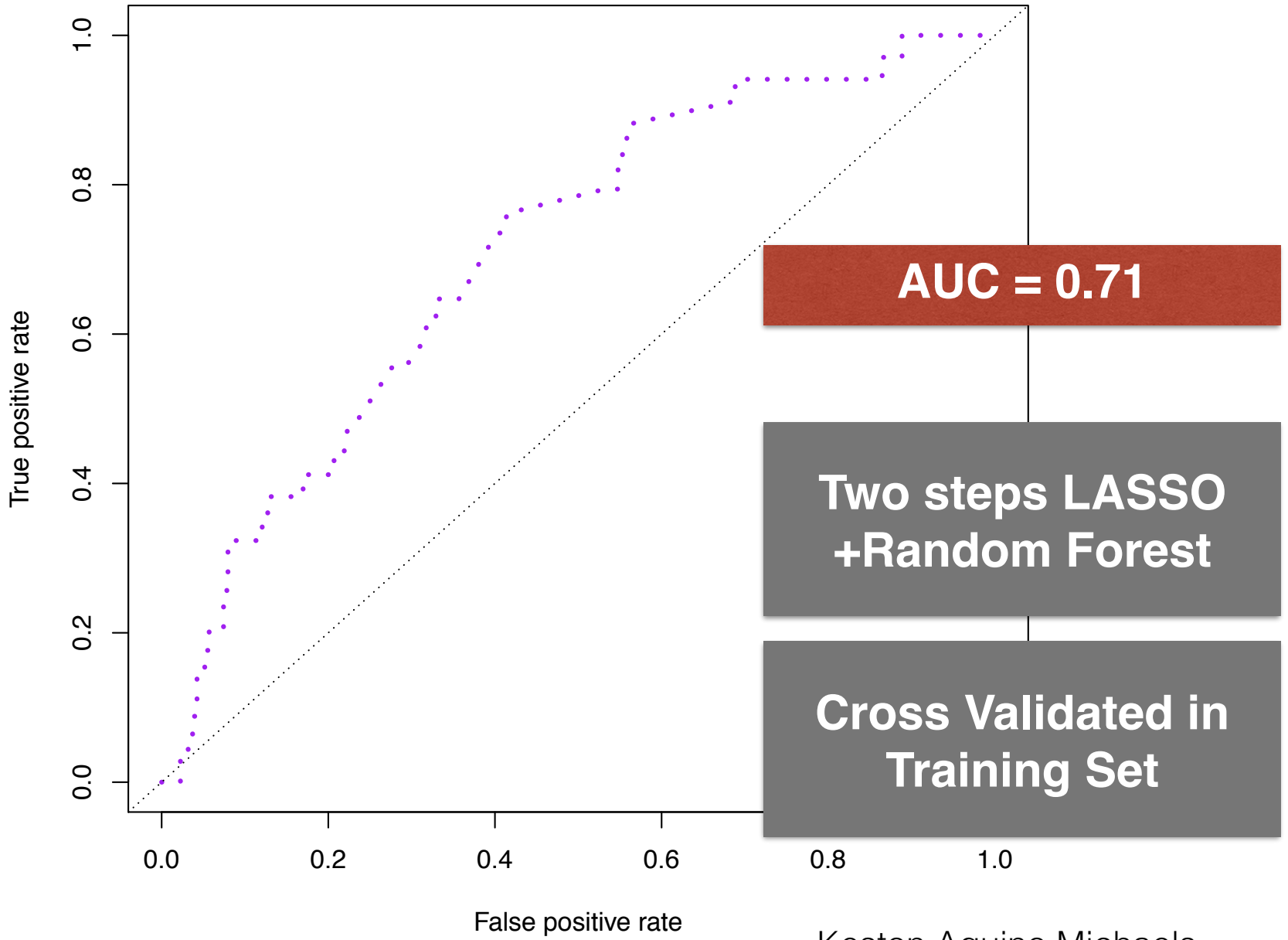
Primary Hypertension score Predicts Bev-induced HT



AUC = 0.62

Keston Aquino Michaels

Bev-Hypertension Predicted Within Study



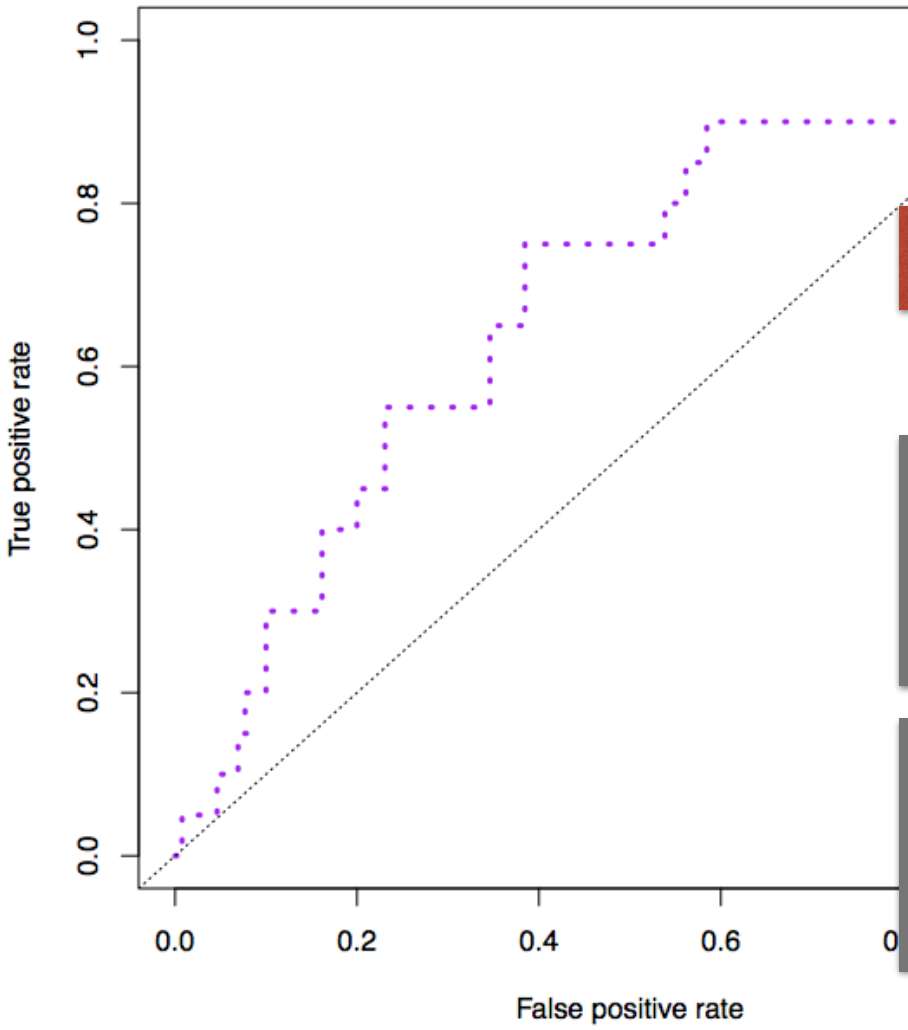
AUC = 0.71

Two steps LASSO +Random Forest

Cross Validated in Training Set

Keston Aquino Michaels

Bev-Hypertension Predicted in Independent Study



AUC = 0.68

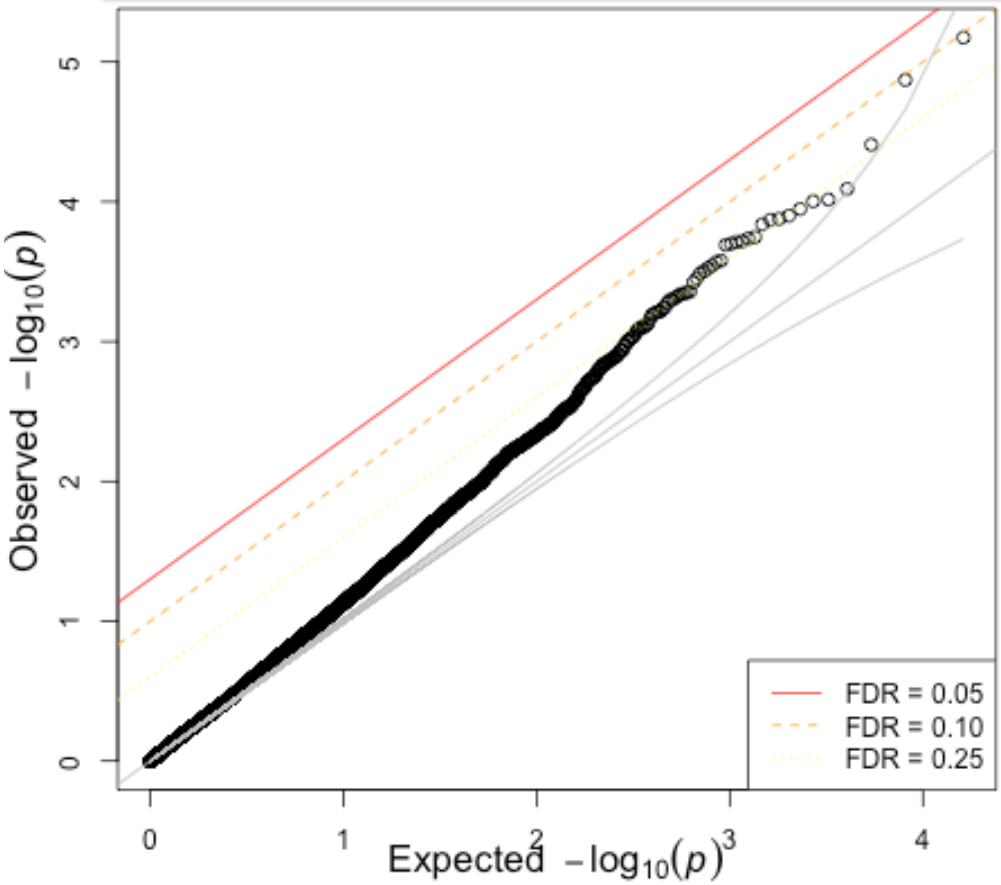
Two steps LASSO +Random Forest

Validated in Independent Set AUC down to 0.68 from 0.71

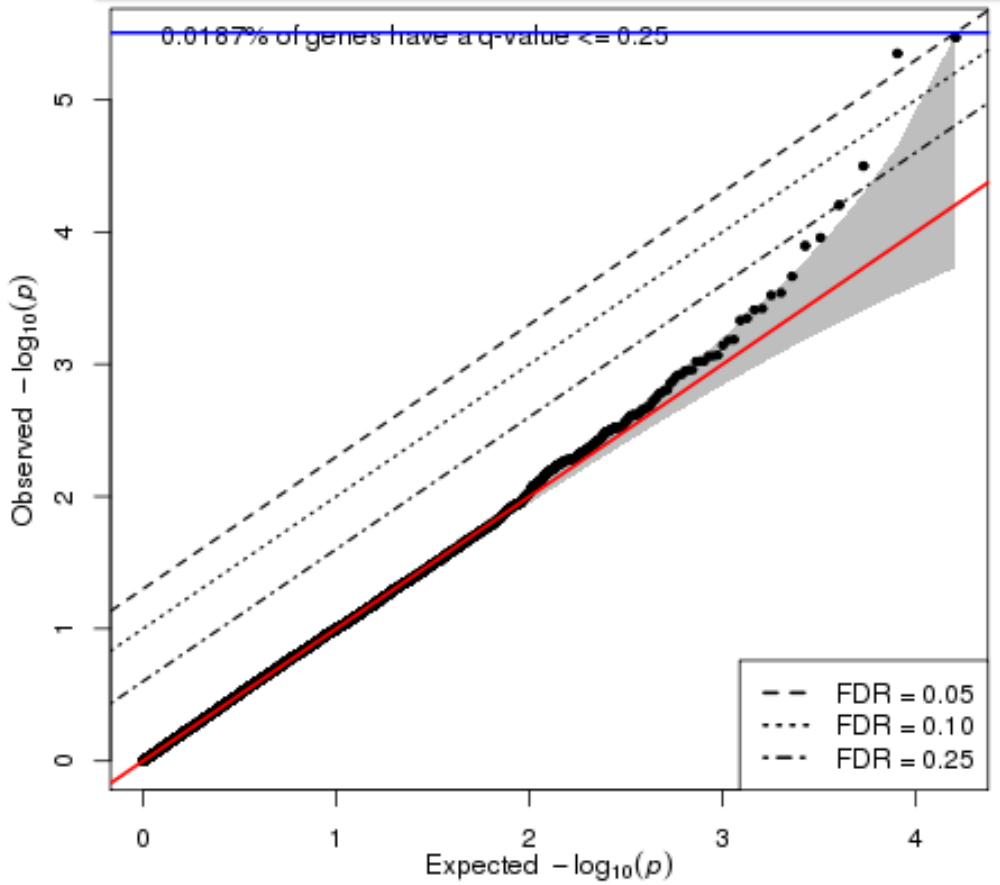
Keston Aquino Michaels

PrediXcan Results Primary and Bev-Hypertension

WTCCC Hypertension



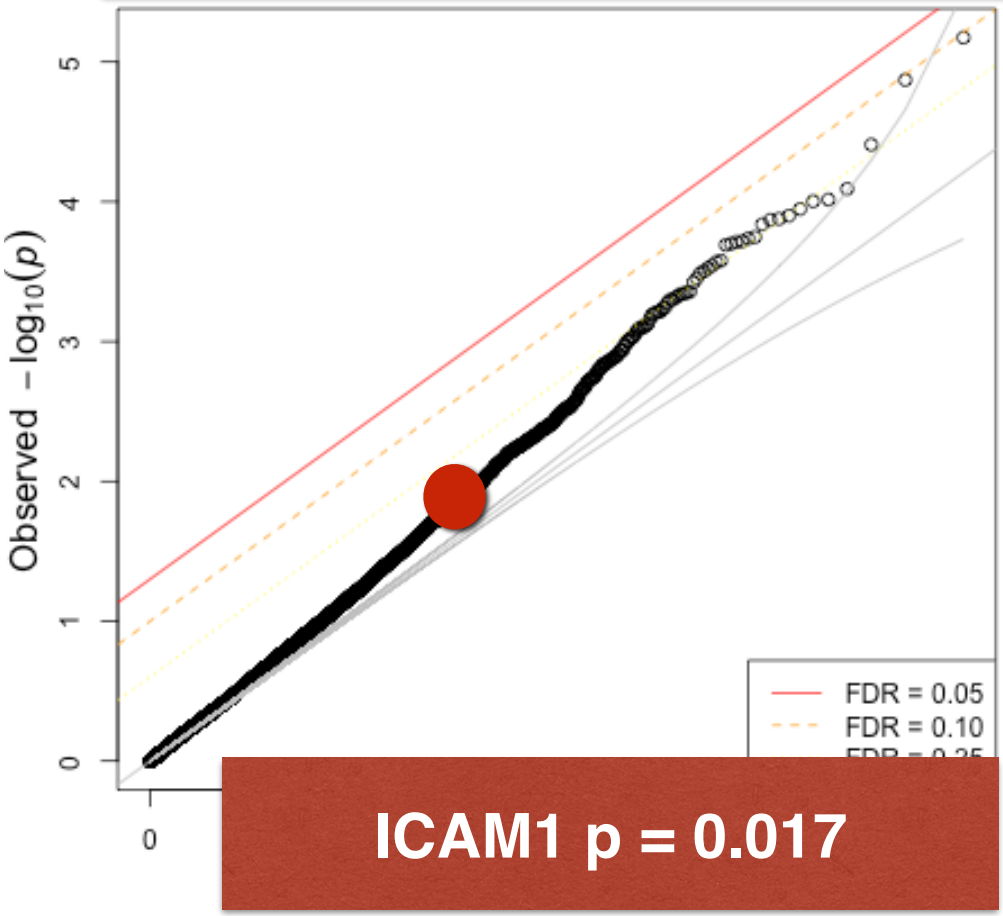
Bevacizumab Hypertension



Heather Wheeler

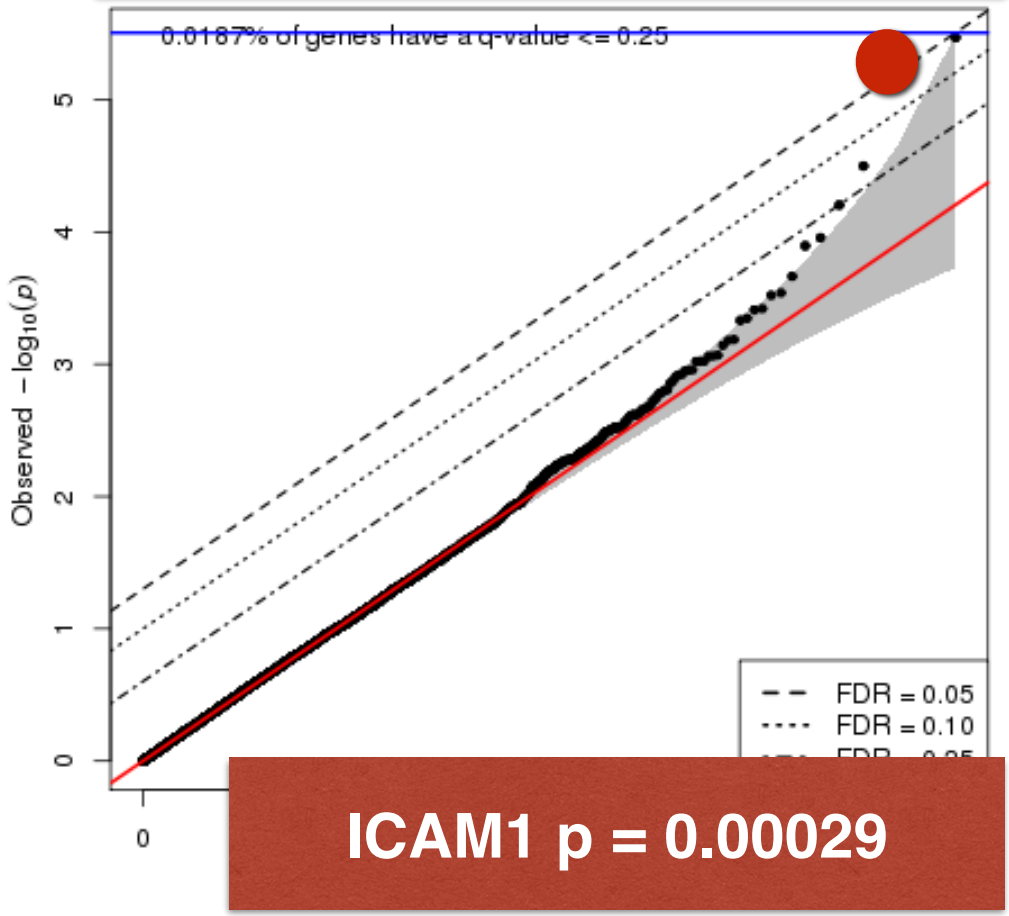
PrediXcan Results Primary and Bev-Hypertension

WTCCC Hypertension



ICAM1 p = 0.017

Bevacizumab Hypertension



ICAM1 p = 0.00029

Heather Wheeler

Hypertension and ICAM1

- Genetically predicted expression levels of ICAM1 was associated with
  - Primary hypertension WTCCC
  - Bevacizumab induced hypertension CALGB 90401
- Genetically predicted serum levels of ICAM1 was associated with
  - Bev induced hypertension CALGB 90401
- Increased levels of ICAM1 were associated with blood pressure in induced hypertension mice model

Summary

- Most single variant findings have limited clinical utility
- Whole genome approaches to prediction improves utility
- Aggregation, partitioning and enrichment studies improves dissection
- Bevacizumab induced hypertension example
  - primary hypertension results help in predicting drug induced hypertension
  - succesfully predicted bevacizumab induced hypertension in indenpendet study
- PrediXcan: novel gene based test that test mechanism yielded promising findings
- Need more samples and better methods

Acknowledgements

Contributors

- Keston Aquino Michaels
- Heather Wheeler
- Nancy Cox
- Eileen Dolan
- Kaanan P. Shah
- Sahar Mozaffari
- Eric Gamazon
- Kouros Owzar
- CALGB 80303 Federico Innocenti/Mark Ratain/Yusuke Nakamura
- CALGB 90401 Howard McLeod
- CALGB 40101 Deanna Kroetz
- GTEx Consortium

Data sources

- Cross Consortia Pleiotropy XCP summary results
- WTCCC

Funding

- HKI was funded in part by Uchicago CTSA NCI K12CA139160
- University of Chicago Diabetes Research and Training Center: P60 DK20595, P30 DK020595
- Genotype of Tissue Expression GTEx R01 MH090937 and R01 MH101820
- Pharmacogenomics of Anticancer Agents PAAR UO1GM61393
- Pharmacogenomics Research Network (PGRN) Statistical Analysis Resource (P-STAR) U19 HL065962
- Conte Center grant P50MH094267

Lipid Markers AUC

Manickam et al 2011 J Clinical Lipidology