

# The Role of Genomic Prediction in Precision Medicine

Hae Kyung Im, PhD



THE UNIVERSITY OF  
**CHICAGO**

March 6, 2015

# Overview

- Precision Medicine Initiative
- Monogenic vs. polygenic traits
- Review of prediction methods
- Poly-Omic integration: OmicKriging
- Role of regulatory variants in complex traits
- PrediXcan
- Prediction of gene expression traits

# Precision Medicine

- Obama: Precision Medicine Initiative \$215M for 2016 Budget
- Instead of “one-size fits-all-approach”
- “Right treatment, at the right time to the right person”
- Innovative approach to disease prevention and treatment based on individual differences in genes, environments, and lifestyles

<http://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative>

# Precision Medicine Implementation

## 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

# The Promise of the Human Genome Sequencing Project

- In year 2000, president Clinton announced the completion of the first draft of the human genome, which would "revolutionize the diagnosis, prevention, and treatment of most, if not all, human diseases."
- Francis Collins predicted that diagnosis of genetic diseases would be accomplished by 2010 and that treatments would start to roll out perhaps by 2015.
- Why are we not there yet?

# The Promise of the Human Genome Sequencing Project

- In year 2000, president Clinton announced the completion of the first draft of the human genome, which was expected to "revolutionize the diagnosis, prevention, and treatment of... if not all, human diseases".
- Francis Collins predicted that diagnosis of genetic diseases would be revolutionized by 2010 and that treatments would start to roll out perhaps by 2015.
- Why are we not there yet?

Prediction is hard !

# The Promise of the Human Genome Sequencing Project

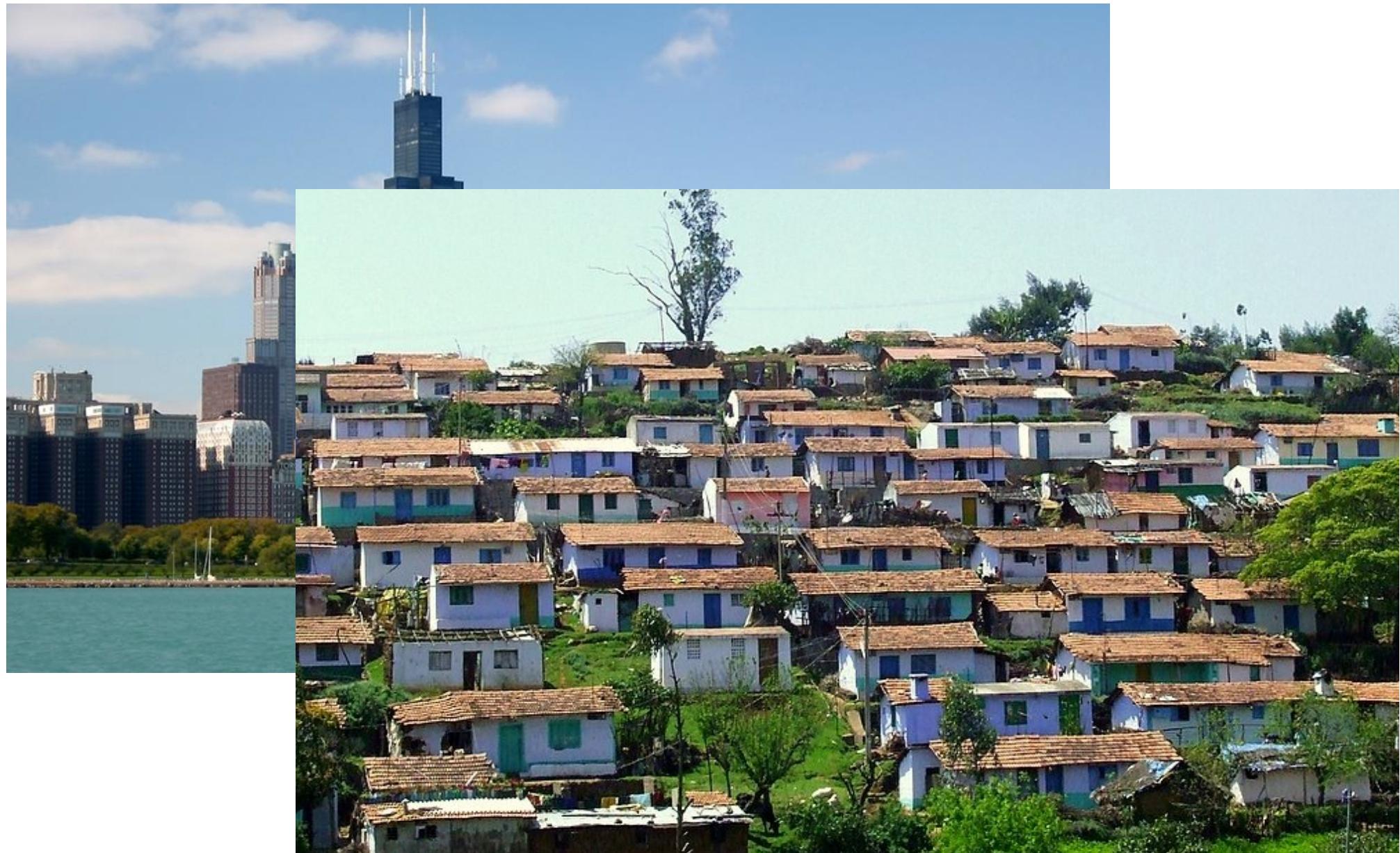
- In year 2000, president Clinton announced the completion of the first draft of the human genome. This was seen as a major step forward in the diagnosis, prevention, and treatment of human diseases.
  - Francis Collins, director of the National Institutes of Health, predicted that by 2010 there would be a revolution in the diagnosis of genetic diseases and that treatments would start to appear by 2015.
  - Why are we not there yet?
- Genetic architecture is much more complex than anticipated**

# Monogenic vs. Polygenic Architecture

# Genetic Architecture of Complex Traits



# Genetic Architecture of Complex Traits



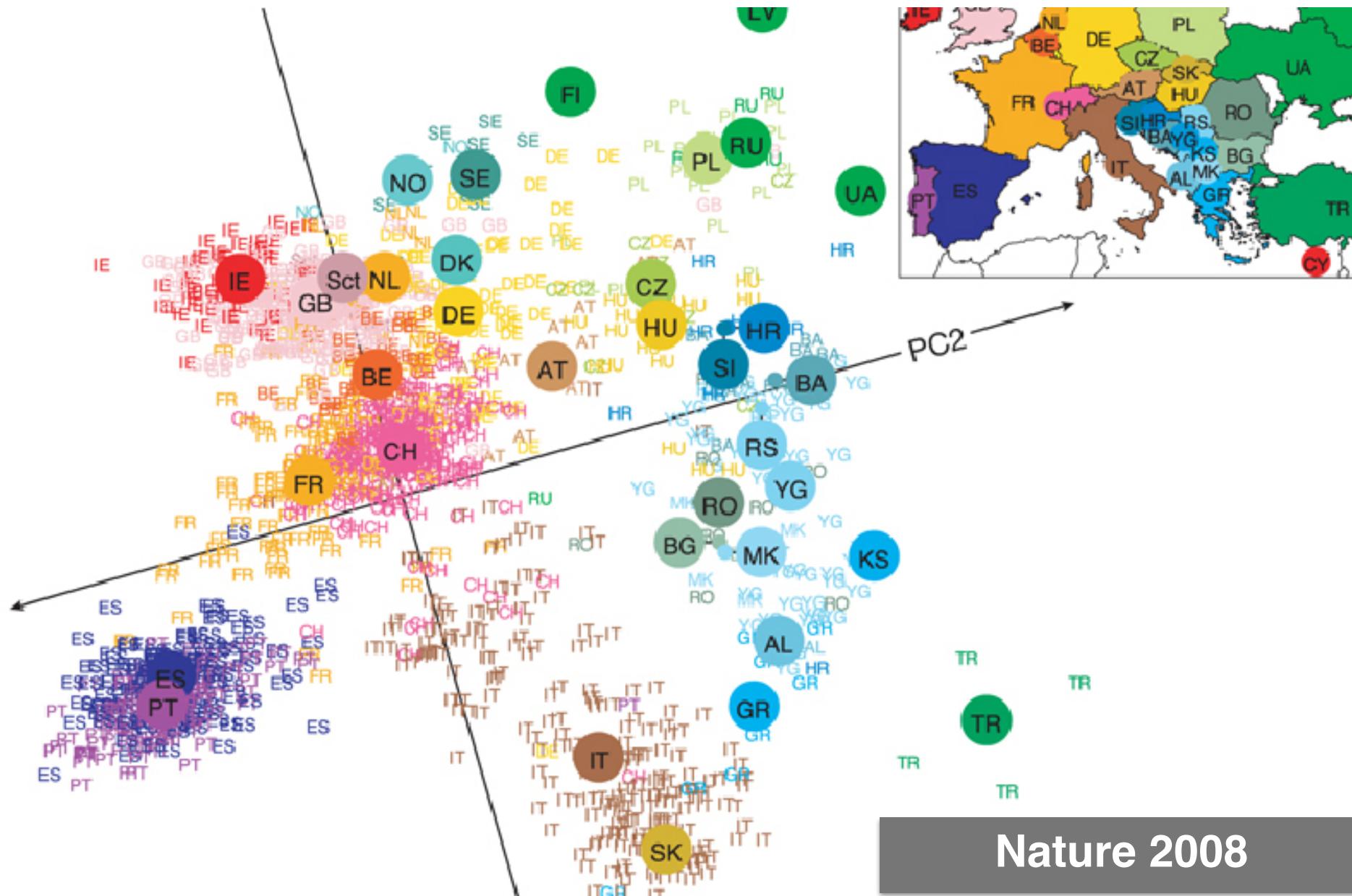
# Genetic Architecture of Complex Traits

**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>



Nature 2008

# Whole Genome Prediction Methods

# Additive Genetic Model

$$Y = \sum_{k=1}^M \beta_k X_k + \epsilon$$

Univariate  
Regression

GWAS

Penalized regression

Ridge

LASSO

Elastic  
Net

$$\|Y - X_k \beta_k\|_2$$

$$\|Y - \sum_k X_k \beta_k\|_2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2$$

# Simple Polygenic Score

## LETTERS

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

The International Schizophrenia Consortium\*

Nature 2009

$$Y = \sum_{k=1}^M \hat{\beta}_k^{\text{GWAS}} X_k$$

Univariate  
Regression



GWAS

# Best Linear Unbiased Prediction (BLUP)/Ridge

## 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>

AJHG 2011

$$Y = \sum_{k=1}^M \hat{\beta}_k^{\text{Ridge}} X_k$$

Penalized regression



Ridge

$$\|Y - \sum_k X_k \beta_k\|_2 + \lambda_2 \|\beta_2\|_2$$

# LASSO/Elastic Net Prediction

*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

*Stanford University, USA*

$$Y = \sum_{k=1}^M \hat{\beta}_k^{E-N} X_k$$

Penalized regression

LASSO

Elastic  
Net

$$\|Y - \sum_k X_k \beta_k\|_2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2$$

# Whole Genome Prediction Approaches

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>

$$Y = \sum_{k=1}^M \beta_k^L X_k + \sum_{k=1}^M \beta_k^S X_k + \epsilon$$

$$\beta_k^L \sim N(0, \sigma_L^2)$$

$$\beta_k^S \sim N(0, \sigma_S^2)$$

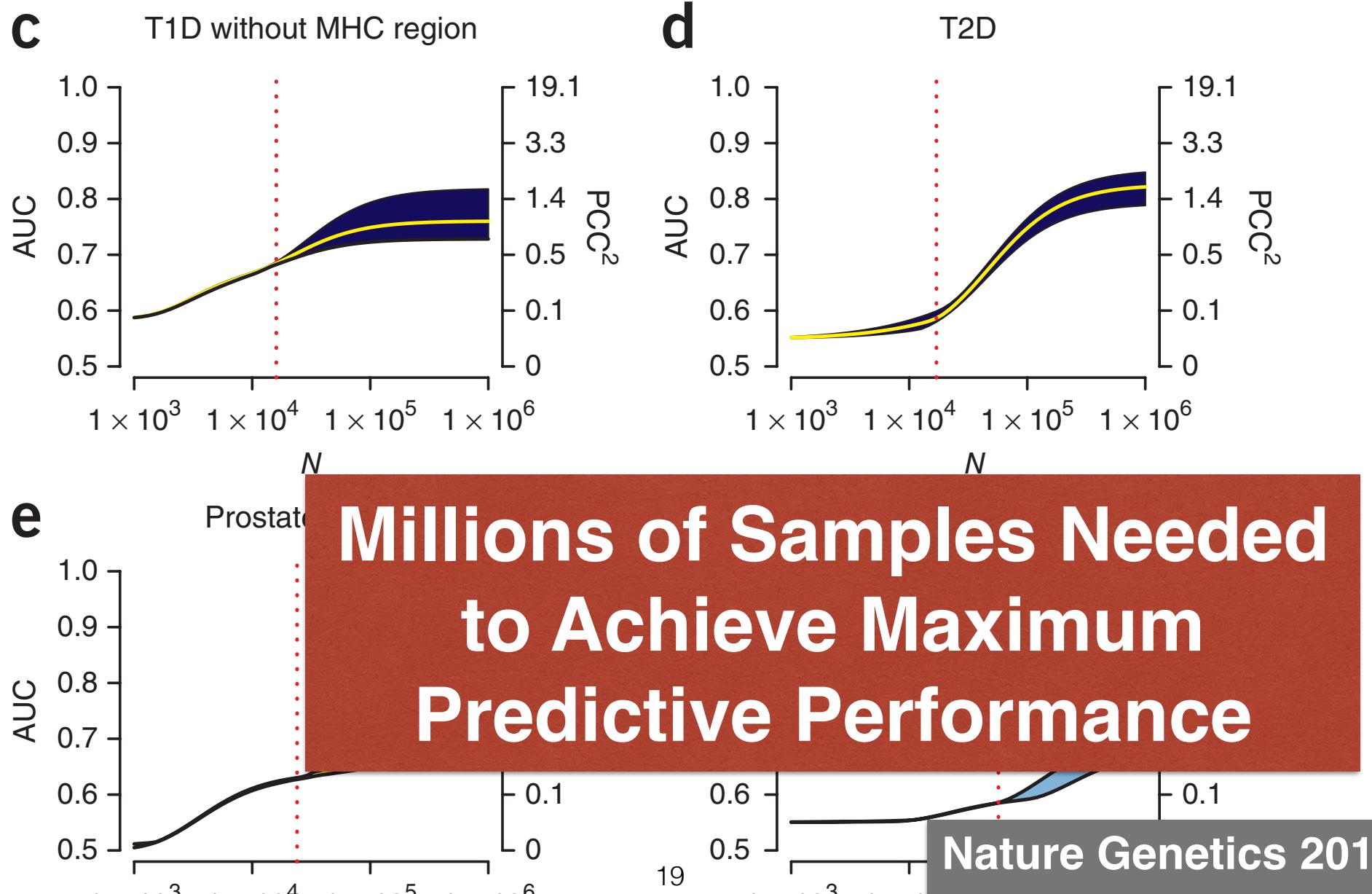
## MultiBLUP: improved SNP-based prediction for complex traits

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)

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

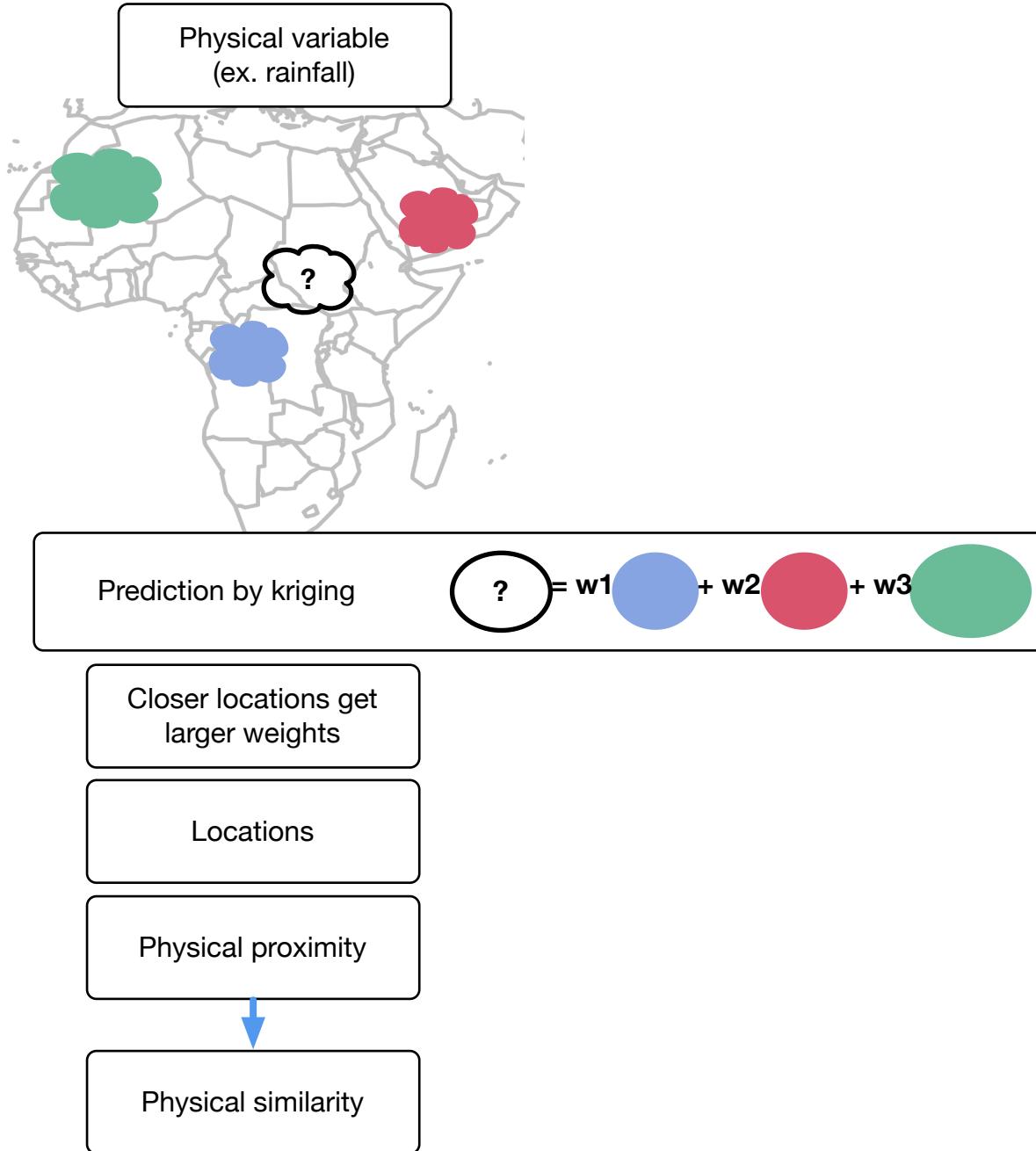
Nilanjan Chatterjee<sup>1</sup>, Bill Wheeler<sup>2</sup>, Joshua Sampson<sup>1</sup>, Patricia Hartge<sup>1</sup>, Stephen J Chanock<sup>1</sup> & Ju-Hyun Park<sup>1,3</sup>



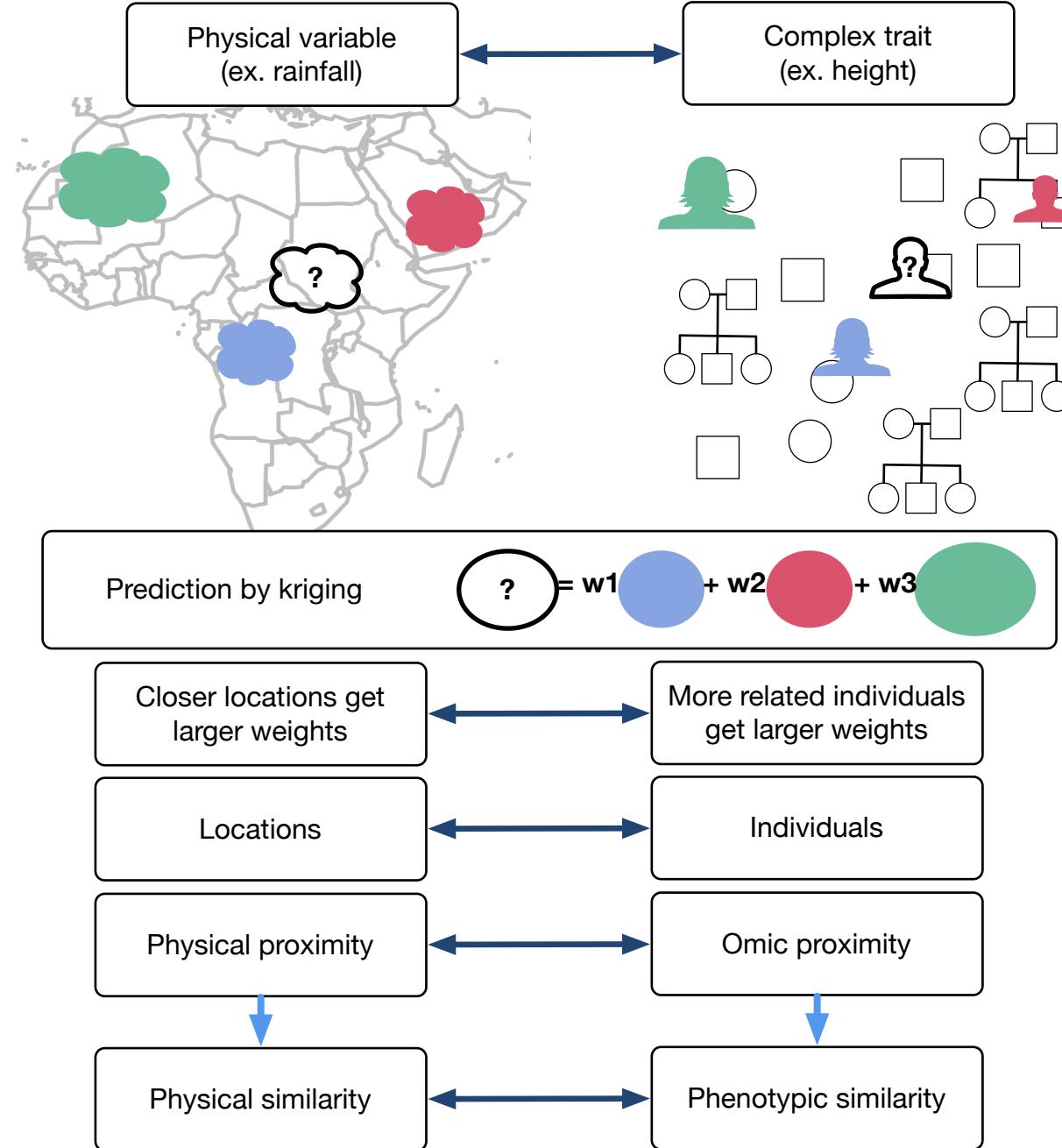
# OmicKriging: Integration of Multiple Omics Data

Genetic Epidemiology 2014

# What is Kriging?



# What is Kriging?



# Kriging

Predicted  $Y$  is the weighted average of the observations

$$\text{Prediction}(Y_{\text{new}}) = \omega_1 Y_1 + \omega_2 Y_2 + \cdots + \omega_n Y_n$$

$\omega_i$  = function(all  $n(n + 1)/2$  pairs of correlations)

Without covariates

$$\omega' = \rho' \Sigma^{-1}$$

$\rho$  the correlation between the new value and the observed values and

$\Sigma$  the correlation matrix of the observations.

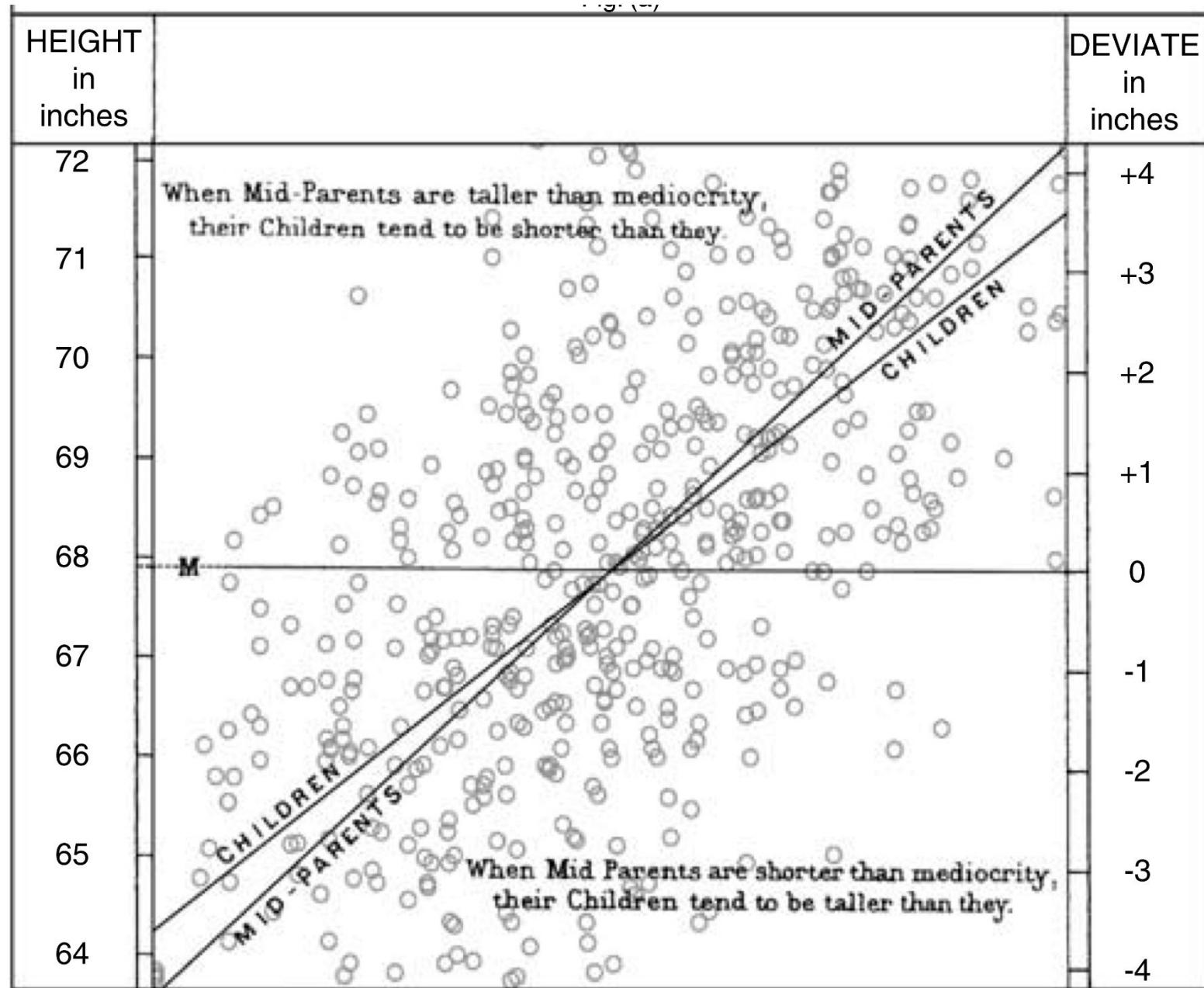
# Galton's Height Data

FAMILY HEIGHTS. from R.F.F.  
 (add 60 inches to every entry in the Table)

	Father	Mother	Sons in order of height	Daughters in order of height.
1	18.5	7.0	13.2 5.8	9.2, 9.0, 9.0
2	15.5	6.5	13.5, 12.5 2.0 3.0	5.5, 5.5
3	15.0	about 4.0	11.0 4.0	8.0
4	15.0	4.0	10.5, 8.5 4.5 6.5	7.0, 4.5, 3.0
5	15.0	-1.5	12.0, 9.0, 8.0 3.0 6.0 7.0	6.5, 2.5, 2.5
6	14.0	8.0		9.5
7	14.0	8.0	16.5, 14.0, 13.0, 13.0 2.5 10.0 4.0 1.0	10.5, 4.0
8	14.0	6.5		10.5, 8.0, 6.0
9	14.5	6.0		6.0

Hanley JA: Transmuting Women into Men. The American Statistician 2004, 58:237243.

# Galton Was Kriging with Kinship Matrix (1885)



# Kriging = BLUP (Best Linear Unbiased Prediction)

- Galton (1885): parent to offspring
- Fisher (1918) and Wright (1921): pedigree
- Formalized by Henderson (1950,1975) and Goldberger (1962)
- G-BLUP: genetic relatedness estimated using genotype
- BLUP/Kriging can be interpreted as the posterior mean of the genetic component given observations ( $Y = G + \text{error}$ )

# Kriging = BLUP (Best Linear Unbiased Prediction)

- Galton (1885): parent to offspring
- Fisher (1918) and Wright (1921): pedigree
- Formalized by Henderson (1950,1975) and Goldberger (1962)
- G-BLUP: genetic relatedness estimated using genotype
- BLUP/Kriging can be interpreted as the posterior mean of the genetic component given observations ( $Y = G + \text{error}$ )

**BLUP/Kriging translates genetic similarity  
into phenotypic prediction**

# Polyomic Model

$$Y_i = a + G_i + T_i + O_i + \epsilon_i$$

$$G_i = \sum_{l=1}^M \beta_l^G X_{il}^G \quad \text{genetic component}$$

$$T_i = \sum_{l=1}^L \beta_l^T X_{il}^T \quad \text{transcriptomic component}$$

$$O_i = \sum_{l=1}^{L'} \beta_l^O X_{il}^O \quad \text{other omic component}$$

$$(\beta_G, \beta_T, \beta_O)' \sim N(0, \Sigma_\beta)$$

# Optimal Similarity Matrix

$$Y_i = a + G_i + T_i + O_i + \epsilon_i$$

Assuming independence of  $\beta$ 's

$$\Sigma_{i,j} = \theta_G \sum_{l=1}^M X_{il}^G X_{jl}^G + \theta_T \sum_{l=1}^L X_{il}^T X_{jl}^T + \theta_O \sum_{l=1}^{L'} X_{il}^O X_{jl}^O + \theta_\epsilon \delta_{ij}$$

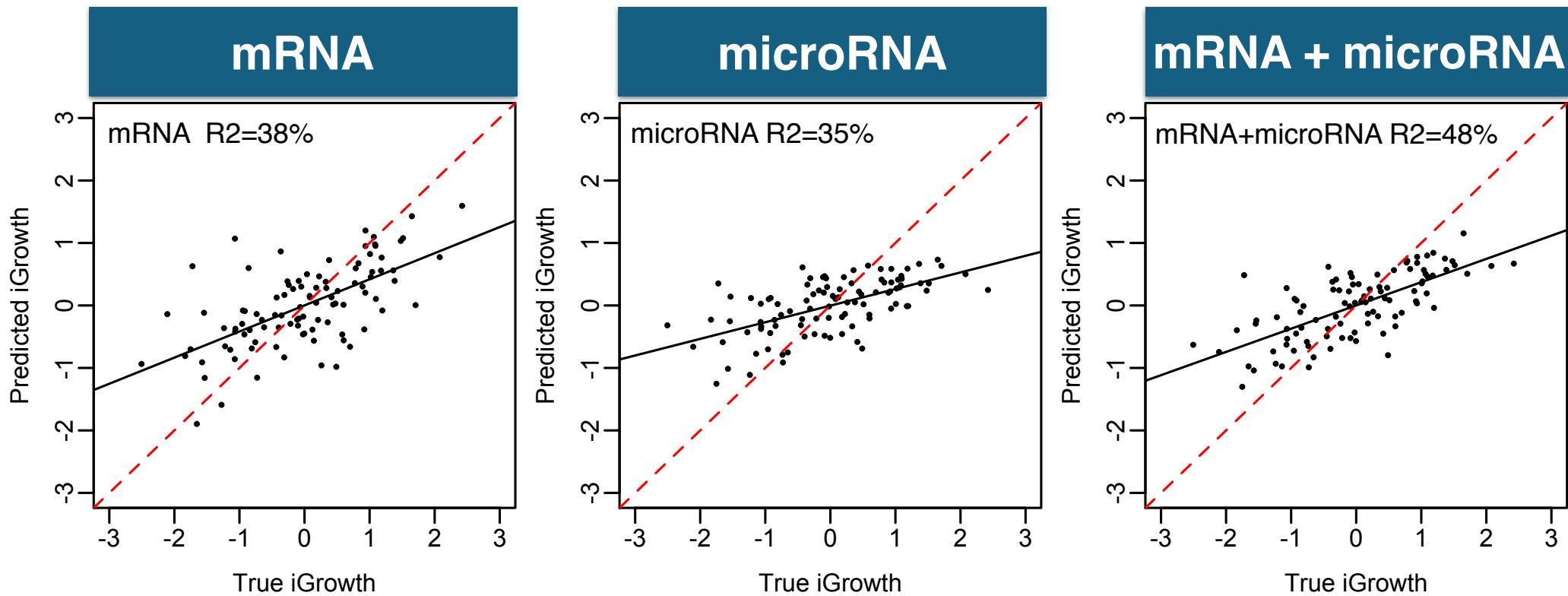
More generally

$$\begin{aligned} \Sigma_{i,j} = & \theta_G \sum_{l=1}^M X_{ik}^G X_{jk}^G + \theta_T \sum_{l=1}^L X_{ik}^T X_{jk}^T + \theta_O \sum_{k=1}^{L'} X_{ik}^O X_{jk}^O + \theta_\epsilon \delta_{ij} \\ & + \sum_{k \neq l} \text{cov}(\beta_k, \beta_l) X_{ik} X_{jl} \end{aligned}$$

# Application of OmicKriging to Cellular Growth

- Intrinsic cellular growth phenotype (Im et al 2012 PLoS Genetics)
- Genes associated with iGrowth are prognostic of survival in cancer patients
- Multiple omics data
  - 99 HapMap cell lines (CEU and YRI)
  - Genotype, mRNA, microRNA

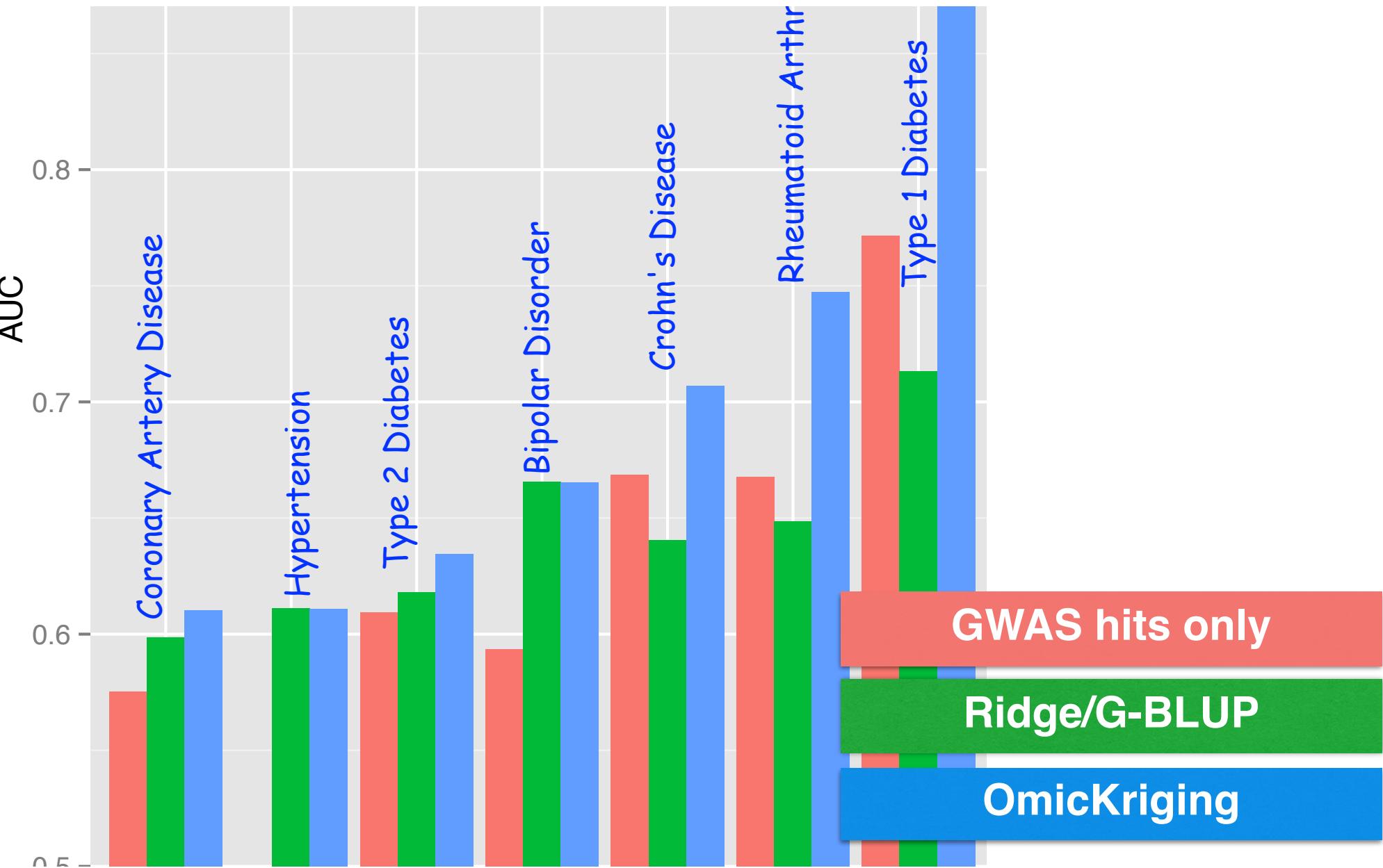
# Application of OmicKriging to Cellular Growth



# Application to Wellcome Trust Case Control Consortium

- WTCCC
- 7 disease cases and 2 control sets:
  - Coronary Artery Disease (2000)
  - Hypertension (2000)
  - Type 2 Diabetes (2000)
  - Bipolar Disorder (2000)
  - Crohn's Disease (2000)
  - Rheumatoid Arthritis (2000)
  - Type 1 Diabetes (2000)
  - 1958 Birth Cohort (1500)
  - UK National Blood Services (1500)

# GWAS hits vs. Whole Genome Prediction (OmicKriging)



# OmicKriging R Package

The screenshot shows a web browser window with the title "CRAN - Package OmicKriging". The URL in the address bar is "cran.r-project.org/web/packages/OmicKriging/index.html". The page content is as follows:

**OmicKriging: OmicKriging for Phenotypic Prediction**

This package provides functions to generate a correlation matrix from a genetic dataset and to use this matrix to predict the phenotype of an individual by using the phenotypes of the remaining individuals through kriging. Kriging is a geostatistical method for optimal prediction or best unbiased linear prediction. It consists of predicting the value of a variable at an unobserved location as a weighted sum of the variable at observed locations. Intuitively, it works as a reverse linear regression: instead of computing correlation (univariate regression coefficients are simply scaled correlation) between a dependent variable Y and independent variables X, it uses known correlation between X and Y to predict Y.

**Version:** 1.3  
**Depends:** R ( $\geq$  2.15.1), [doParallel](#)  
**Imports:** [ROCR](#), [irlba](#)  
**Published:** 2014-06-18  
**Author:** Hae Kyung Im, Heather E. Wheeler, Keston Aquino Michaels, Vassily Trubetskoy  
**Maintainer:** Hae Kyung Im <haky at uchicago.edu>  
**License:** [GPL \( \$\geq\$  3\)](#)  
**NeedsCompilation:** no  
**Materials:** [README](#)  
**CRAN checks:** [OmicKriging results](#)

**Downloads:**

**Reference manual:** [OmicKriging.pdf](#)  
**Vignettes:** [Application Tutorial: OmicKriging](#)  
**Package source:** [OmicKriging\\_1.3.tar.gz](#)  
**Windows binaries:** r-devel: [OmicKriging\\_1.3.zip](#), r-release: [OmicKriging\\_1.3.zip](#), r-oldrel: [OmicKriging\\_1.3.zip](#)  
**OS X Snow Leopard binaries:** r-release: [OmicKriging\\_1.3.tgz](#), r-oldrel: [OmicKriging\\_1.3.tgz](#)  
**OS X Mavericks binaries:** r-release: [OmicKriging\\_1.3.tgz](#)  
**Old sources:** [OmicKriging archive](#)



## 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>

<sup>1</sup>*Section of Hematology/Oncology, Department of Medicine, University of Chicago, Chicago, Illinois, United States of America;* <sup>2</sup>*Section of Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, United States of America;* <sup>3</sup>*Department of Health Studies, Lurie Children's Hospital of Chicago, Chicago, Illinois, United States of America*

Received 26 November 2013; Revised 11 March 2014; accepted revised manuscript 12 March 2014.

Published online 2 May 2014 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)). DOI 10.1002/gepi.21808

**ABSTRACT:** High-confidence prediction of complex traits such as disease risk or drug response is an ultimate goal of personalized medicine. Although genome-wide association studies have discovered thousands of well-replicated polymor-

# Summary OmicKriging

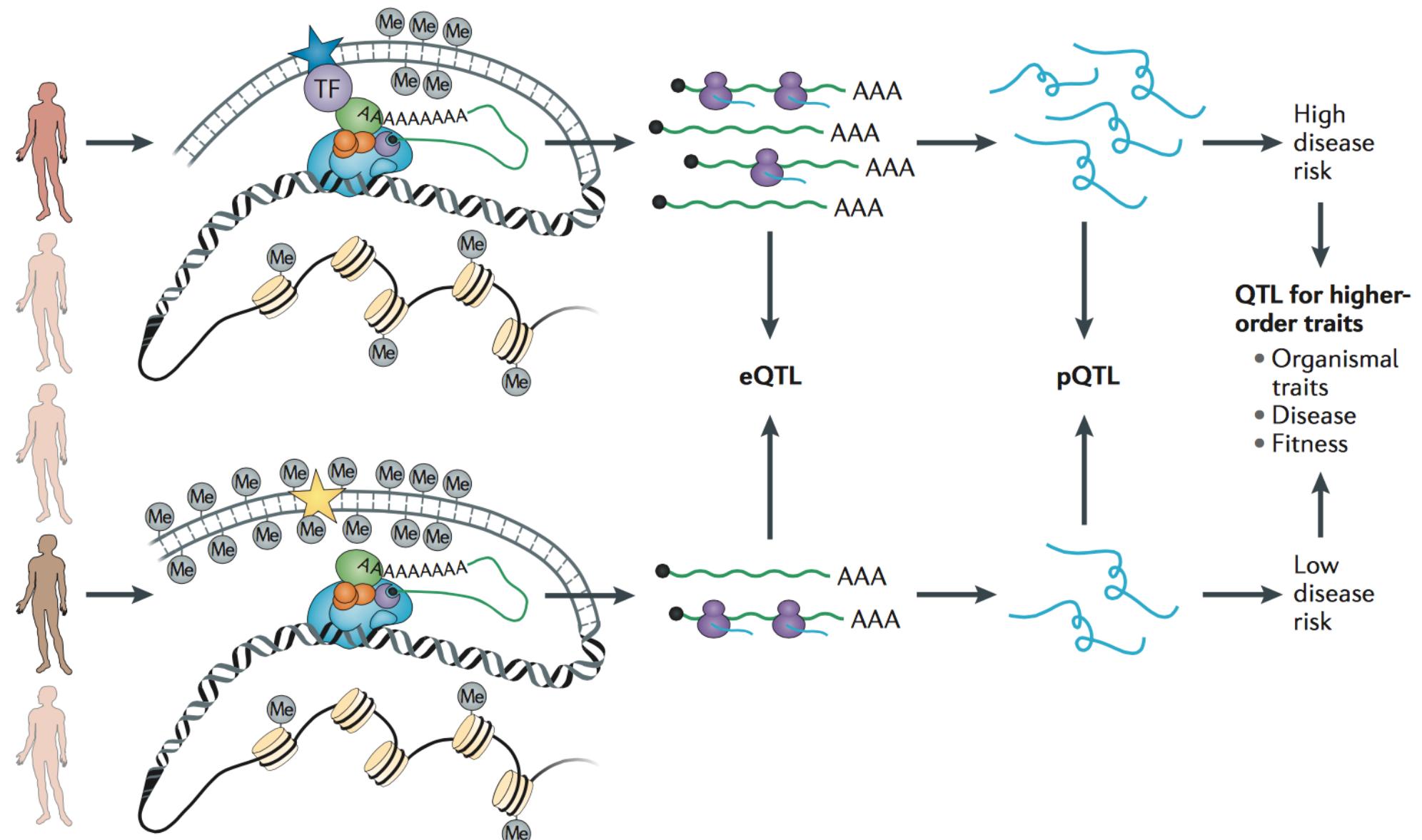
- OmicKriging is a systems approach to complex trait prediction that leverages and integrates multiple omic data
- We can attain relevant prediction even if we do not know the individual variant's contribution
- Important tool for integrating the vast amounts of data to be generated with the precision medicine initiative

# Role of Regulatory Variation in Complex Traits

# Mechanism of Genotype to Phenotype Link

- Most trait-associated SNPs are not coding
- Mechanism via regulation of gene expression levels

# Altered Protein Levels Influences Disease Risk

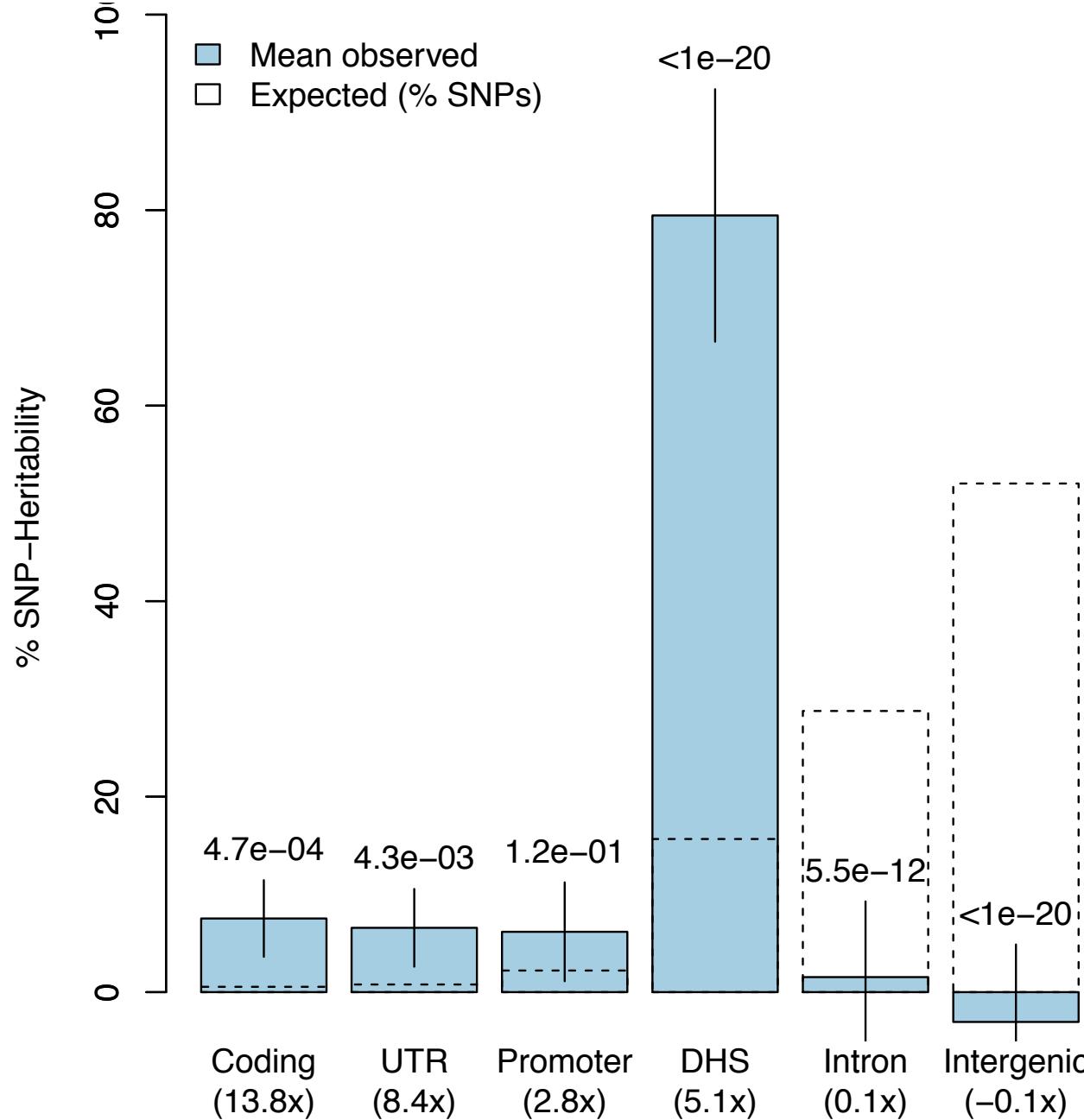


Albert & Kruglyak 2015 NGReviews

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

AJHG 2014

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



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

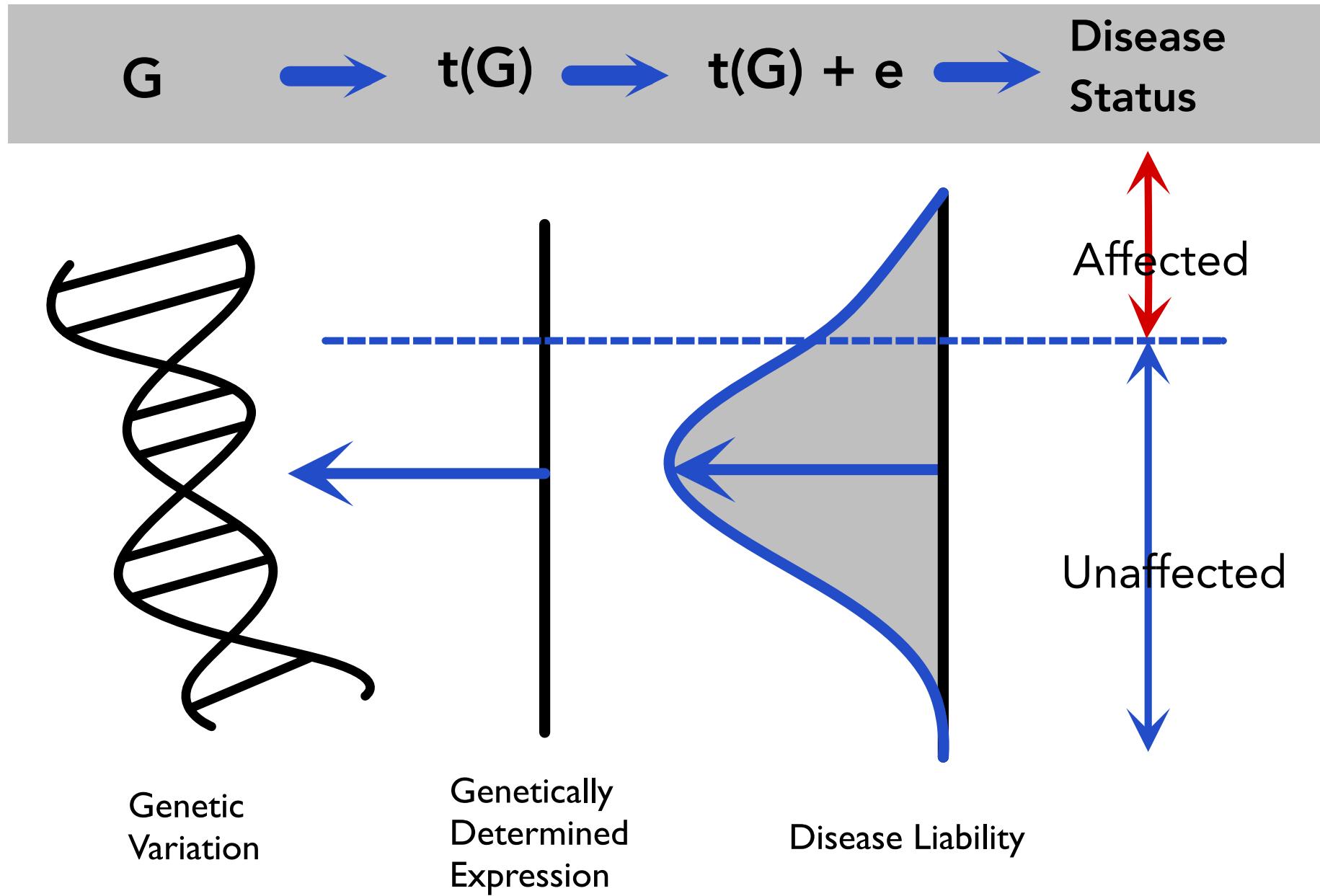
# PrediXcan

Nature Genetics - under revision

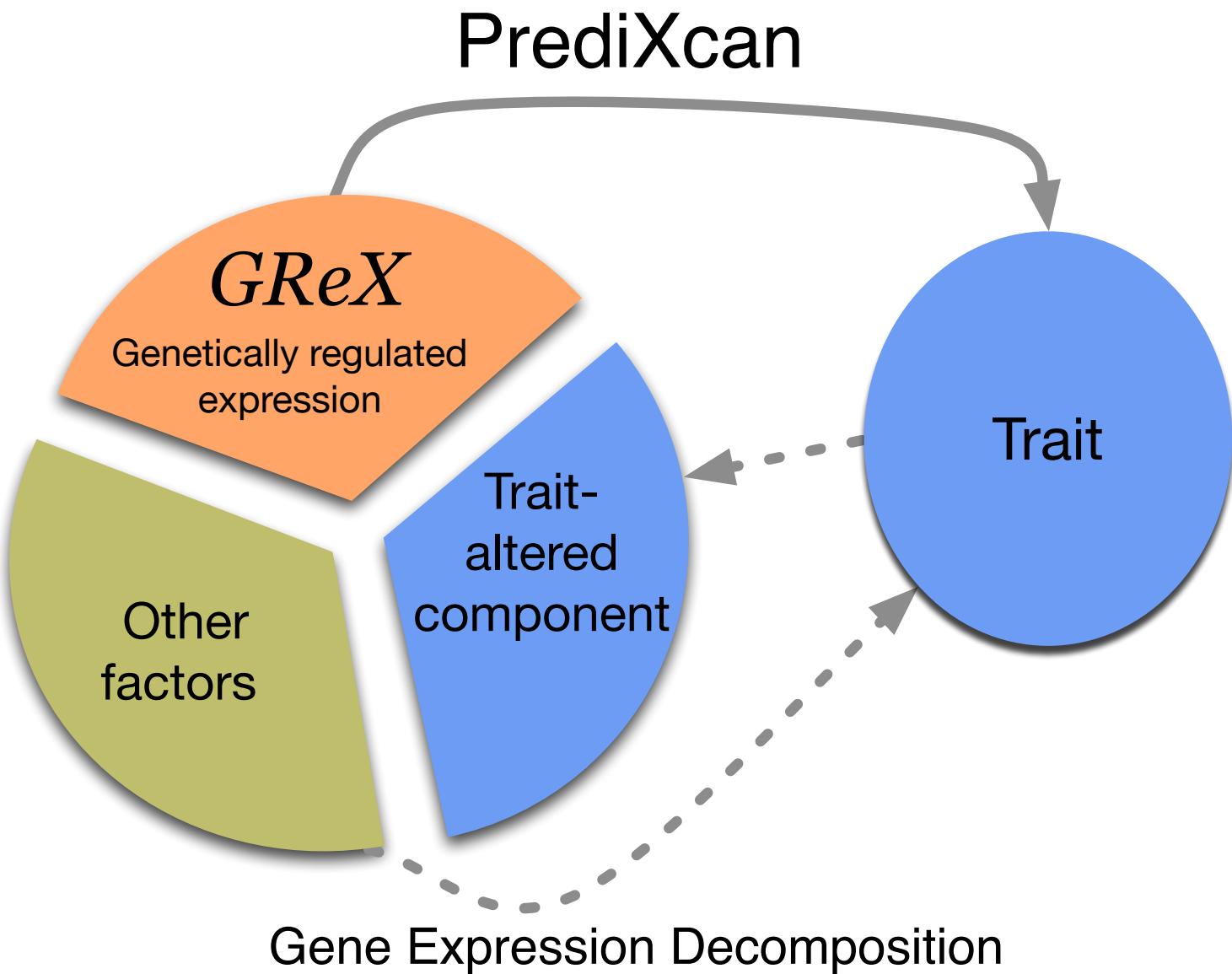
# Motivation for PrediXcan

- Lack of mechanistic understanding of most GWAS discoveries
- Large proportion of variation explained by regulatory variants
- We propose PrediXcan that tests the proposed mechanism

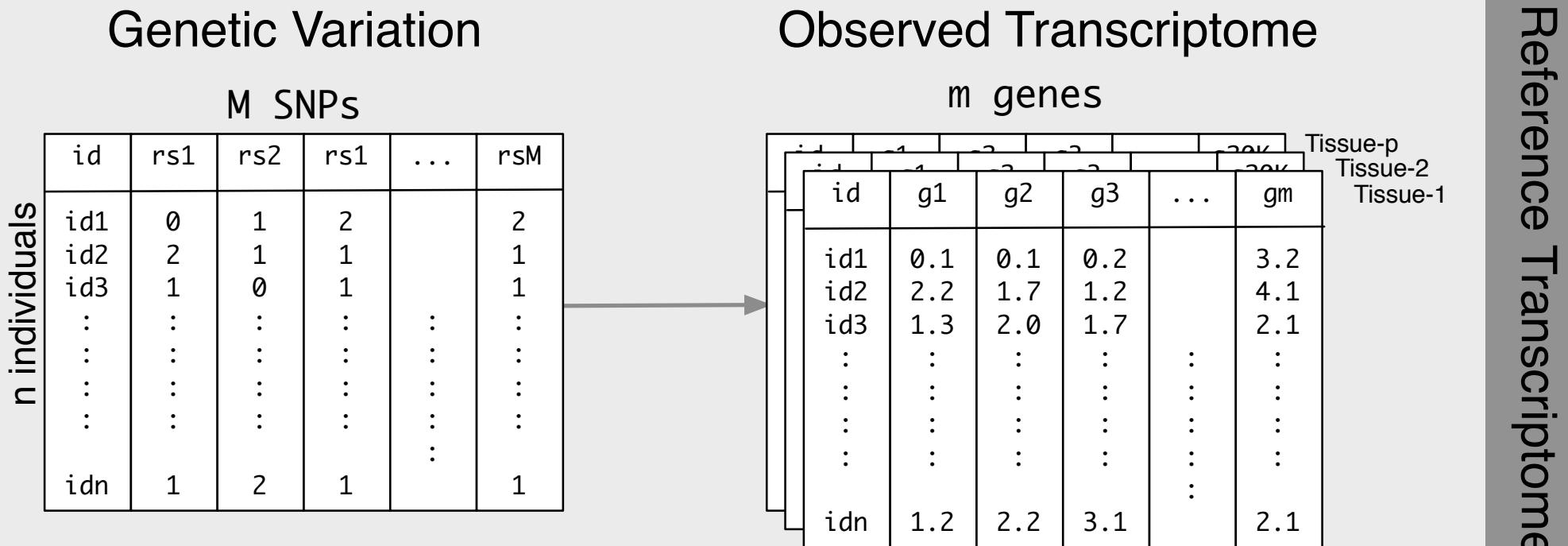
# Genetic Control of Disease Through Gene Regulation



# Mechanisms Tested by PrediXcan



# PrediXcan uses Reference Transcriptome



# PredictDB: Public Database of Weights for GReX

## PredictDB: Database of Prediction Models

		M SNPs								
		1	2	3	4	5	6	7	8	Tissue-p
		1	2	3	4	5	6	7	8	Tissue-2
		1	2	3	4	5	6	7	8	Tissue-1
		g1	w11	w12	w13					w1M
		g2	w21	w22	w23					w2M
		g3	w31	w32	w33					w3M
		:	:	:	:					:
		:	:	:	:					:
		:	:	:	:					:
		gm	wm1	wm2	wm3					wmM

Additive model of gene expression trait trained in reference transcriptome datasets

$$T = \sum_k w_k X_k + \epsilon$$

*GReX*

Weights stored in PredictDB

# PrediXcan Imputes Transcriptome & Tests Assoc.

PrediXcan on GWAS Data

Genetic Variation

M SNPs

id	rs1	rs2	rs1	...	rsM
id1	0	2	1		0
id2	1	2	2		2
id3	2	1	1		1
:	:	:	:		:
:	:	:	:		:
:	:	:	:		:
:	:	:	:		:
idn'	1	2	0		2

n<sup>'</sup> individuals

"Imputed" Transcriptome

m genes

id	g1	g2	g3	...	gm	trait
id1	0.2	0.6	0.2		3.2	0.1
id2	2.3	1.8	1.2		4.1	2.2
id3	3.3	2.2	1.7		2.1	1.3
:	:	:	:		:	:
:	:	:	:		:	:
:	:	:	:		:	:
:	:	:	:		:	:
idn'	2.2	2.0	3.1		2.1	1.2

Association  
Test

# PrediXcan: Mechanism-driven Gene-Based Test

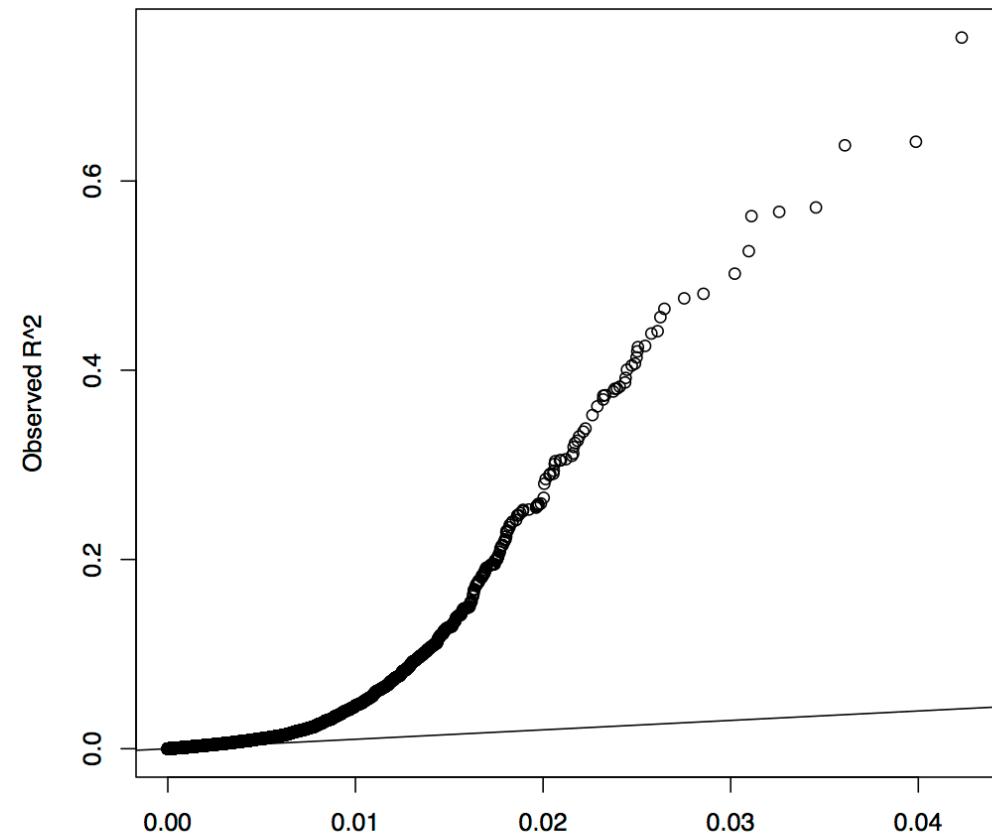
- Directly tests the molecular mechanism through which genetic variants affect phenotype
- Genes more attractive than genetic variants
  - A lot is known about their function
  - Follow up experiments can be easily devised
  - Reduced multiple testing burden
- Direction of effects
  - Positive effects: down regulation is therapeutic option
  - Negative effects: more likely to harbor deleterious rare variants
- No reverse causality issues
- Can be systematically applied to existing GWAS studies
- Tissue-specificity can be inferred

# Reference Transcriptome 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)
  - Framingham, n>5000m, whole blood
  - Depression Genes & Networks, n>900, whole blood

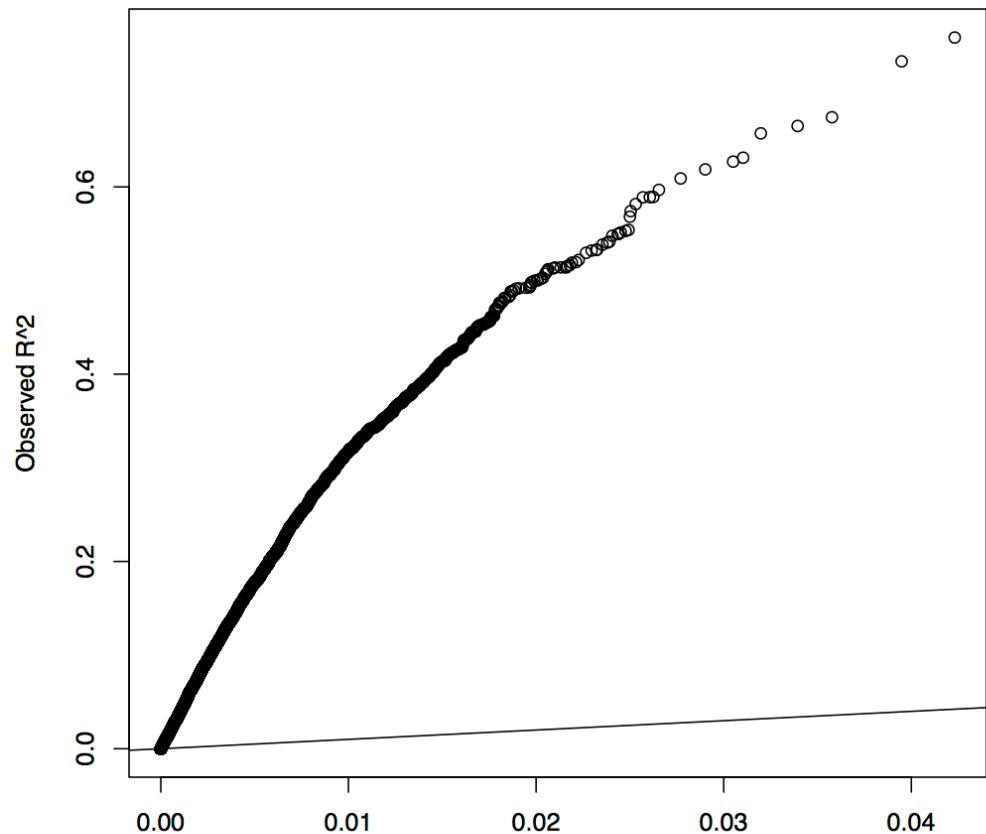
# Good Prediction Performance

Prediction R<sup>2</sup>



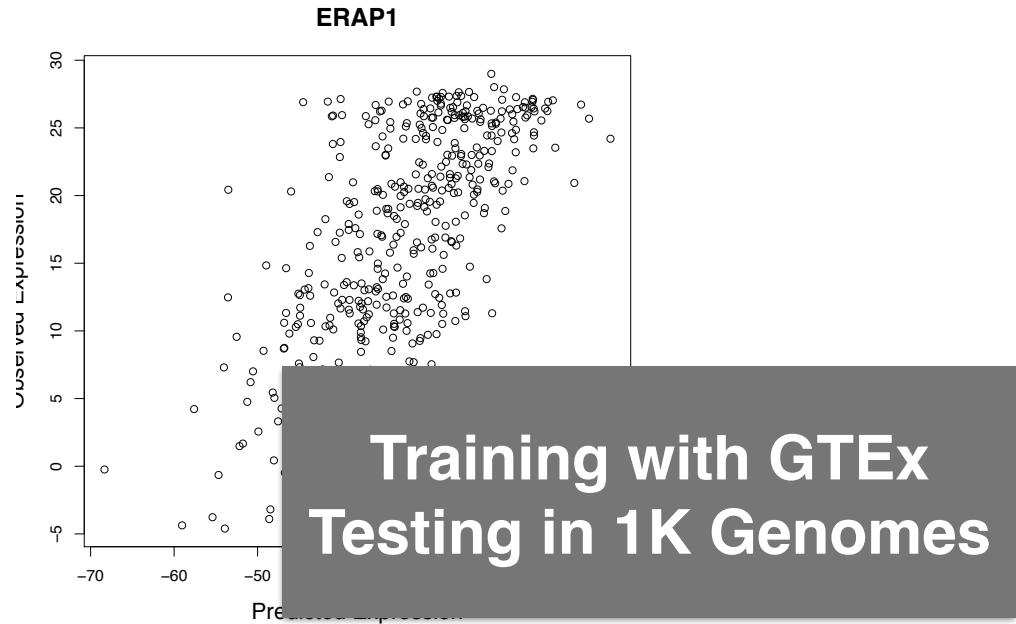
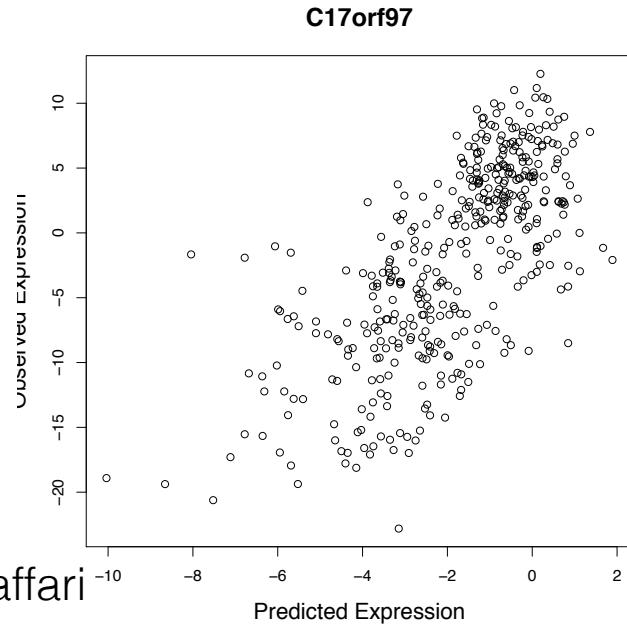
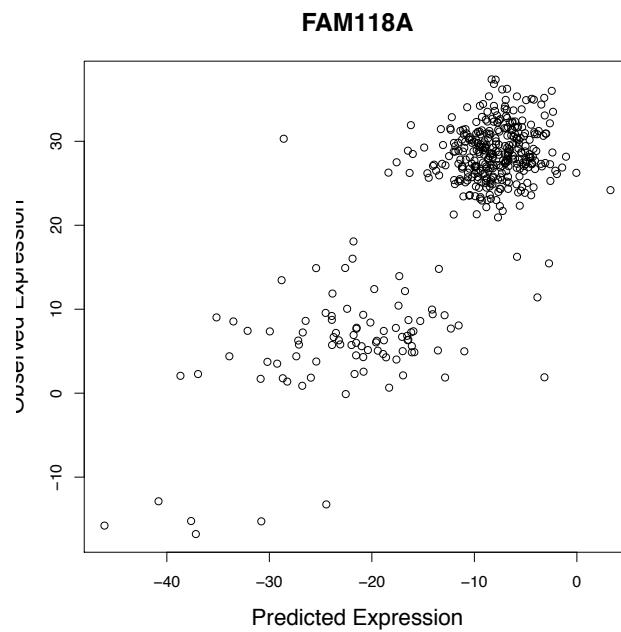
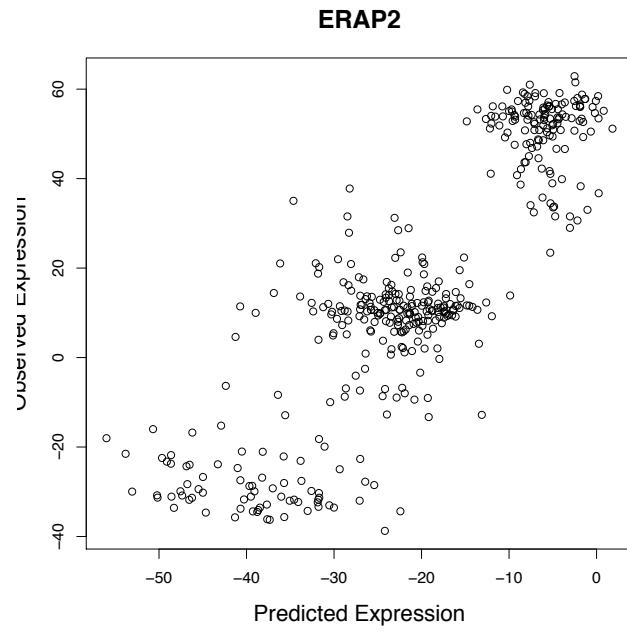
Training with GTEx  
Testing in 1K Genomes

Replication R<sup>2</sup>

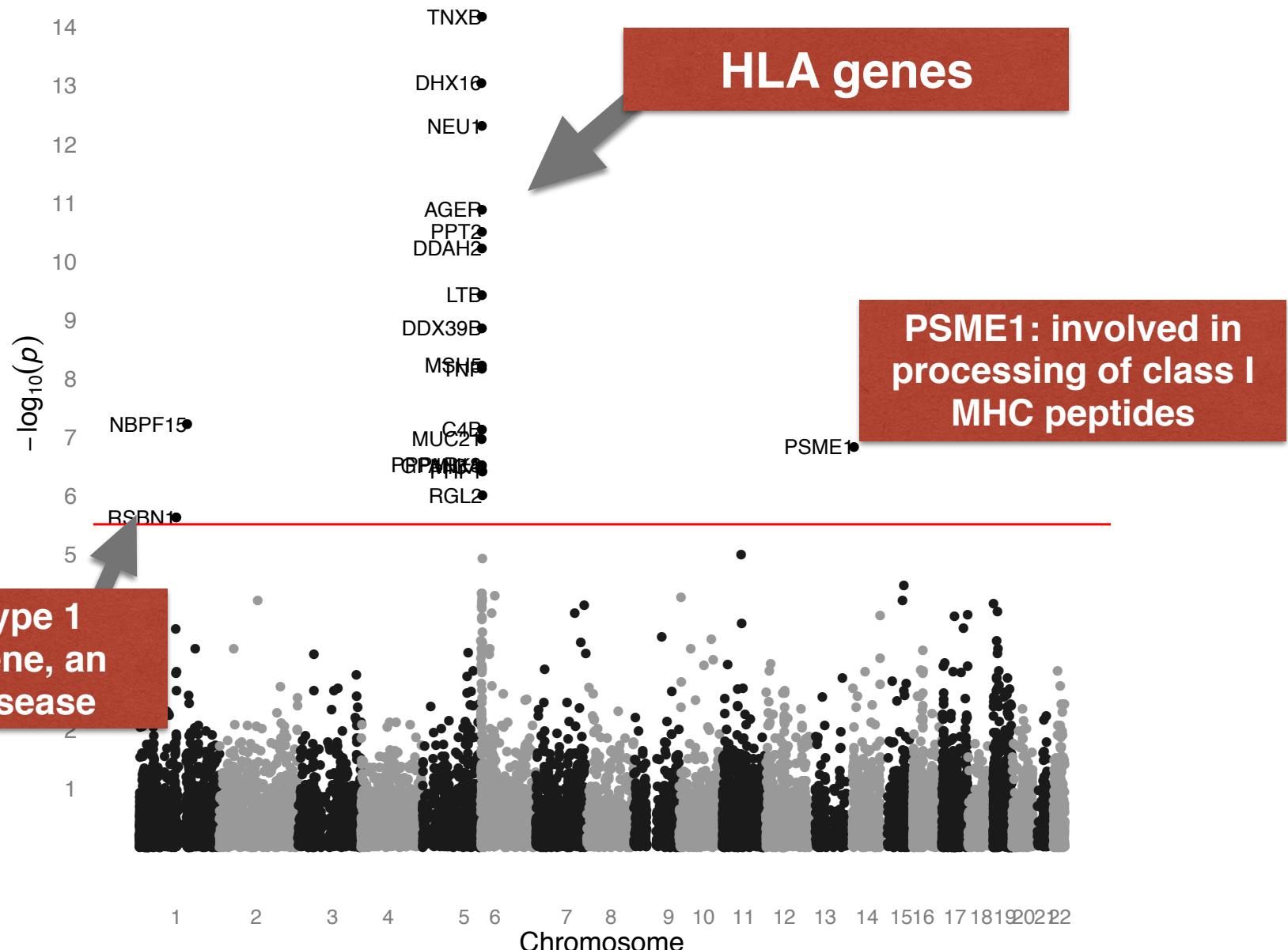


Replicate RNAseq  
Pickrell et al 2010 vs.  
1K Genomes 2013

# Examples of Well Predicted Genes

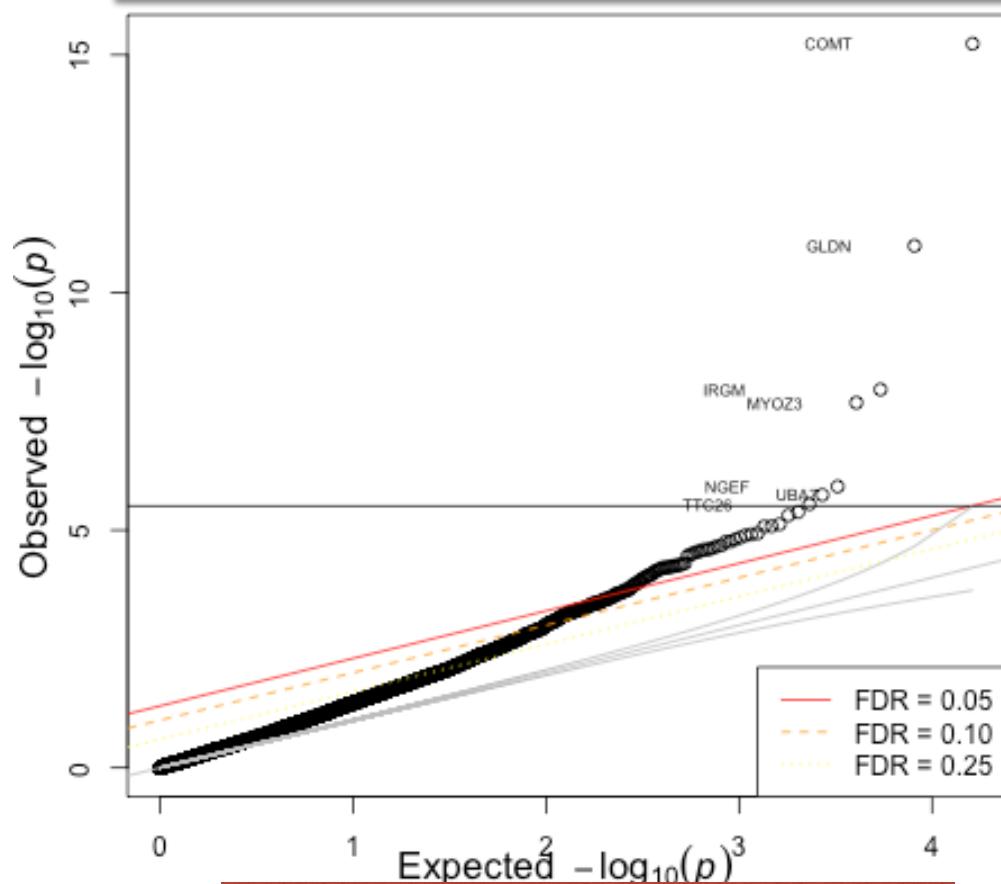


# Genes Associated with Rheumatoid Arthritis



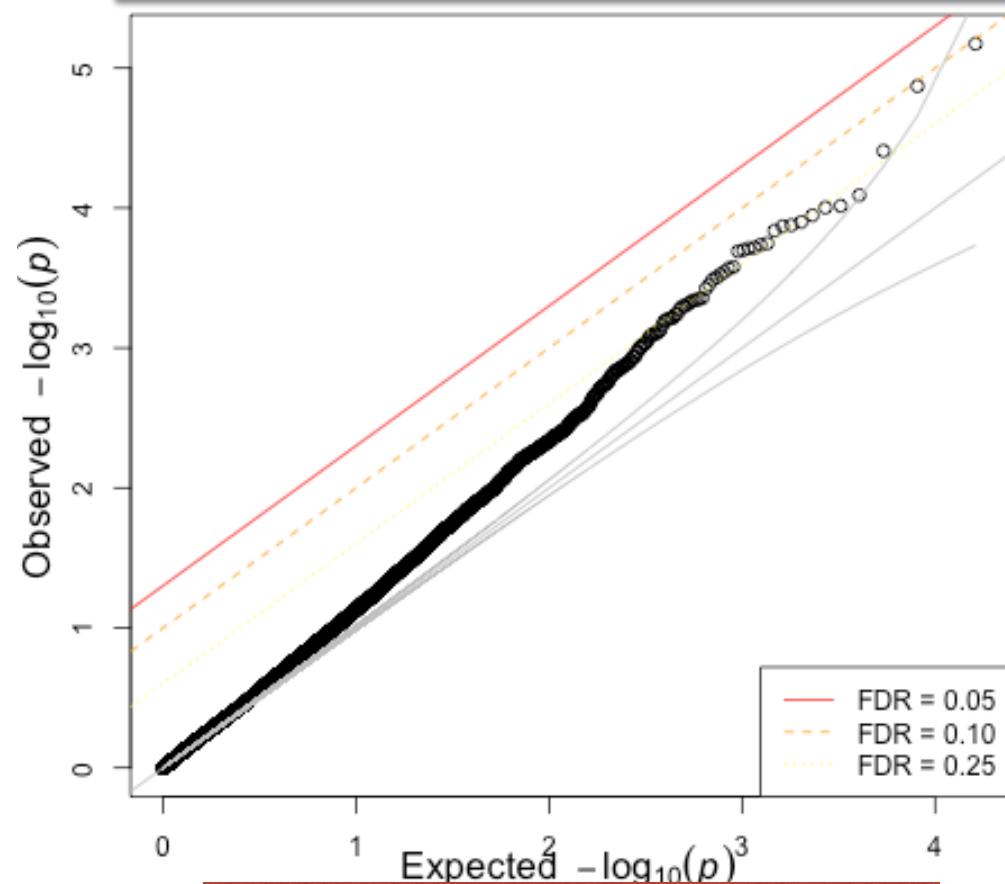
# PrediXcan Results for Crohn's Disease and Hypertension

## Crohn's Disease



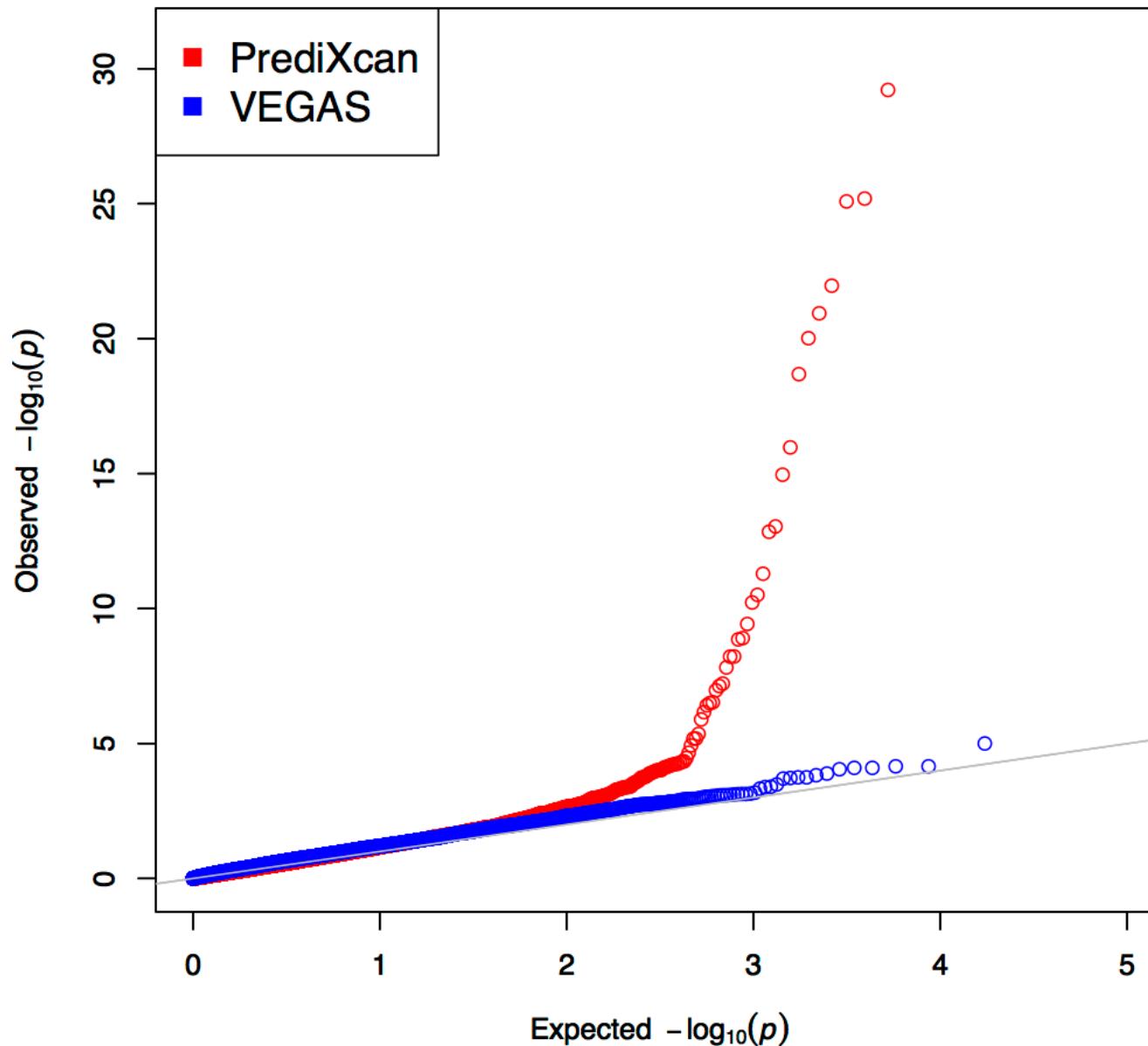
IRGM is a known  
Crohn's gene

## Hypertension



Whole blood may not be  
relevant tissue

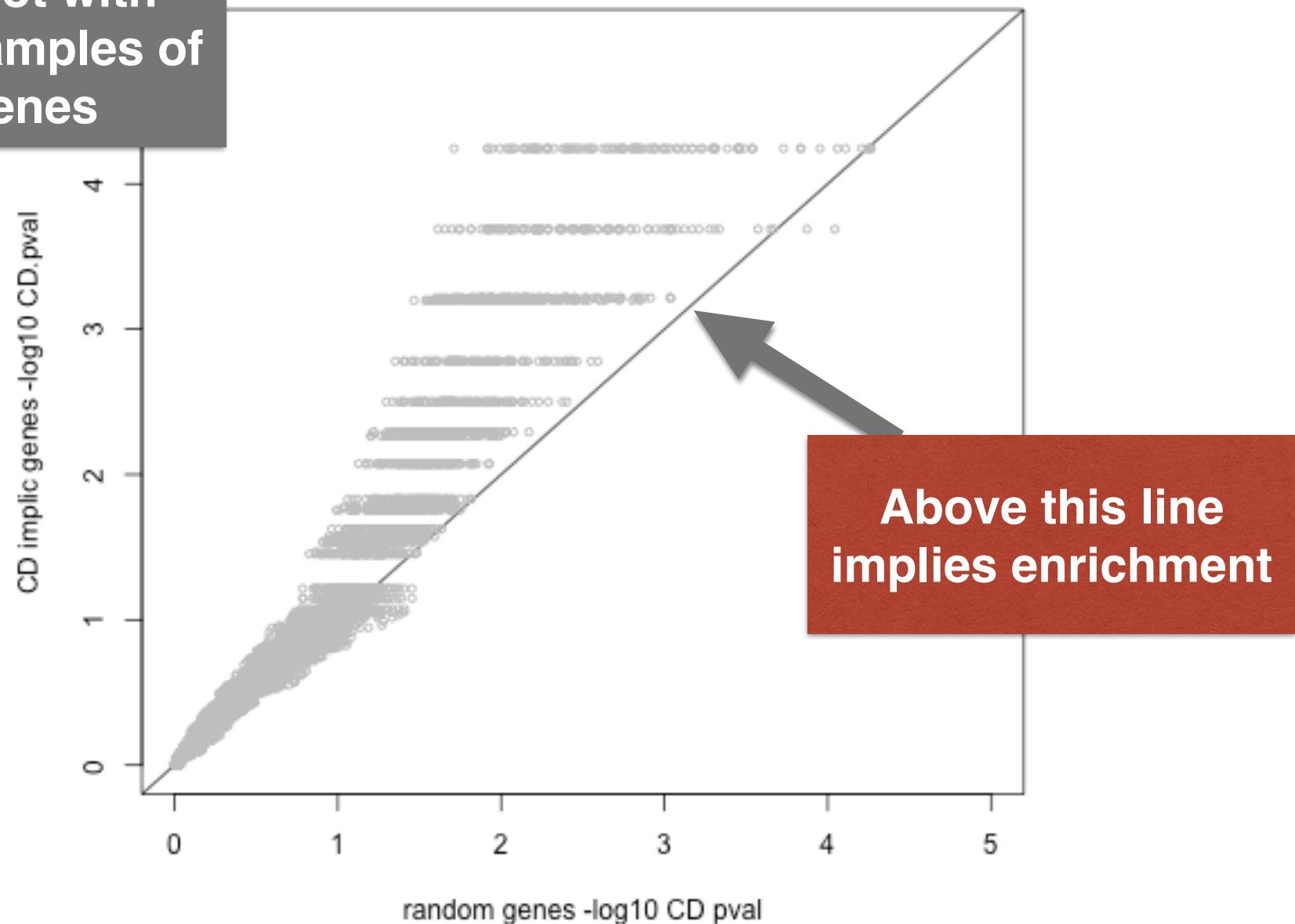
# PrediXcan Outperforms VEGAS



Eric Gamazon

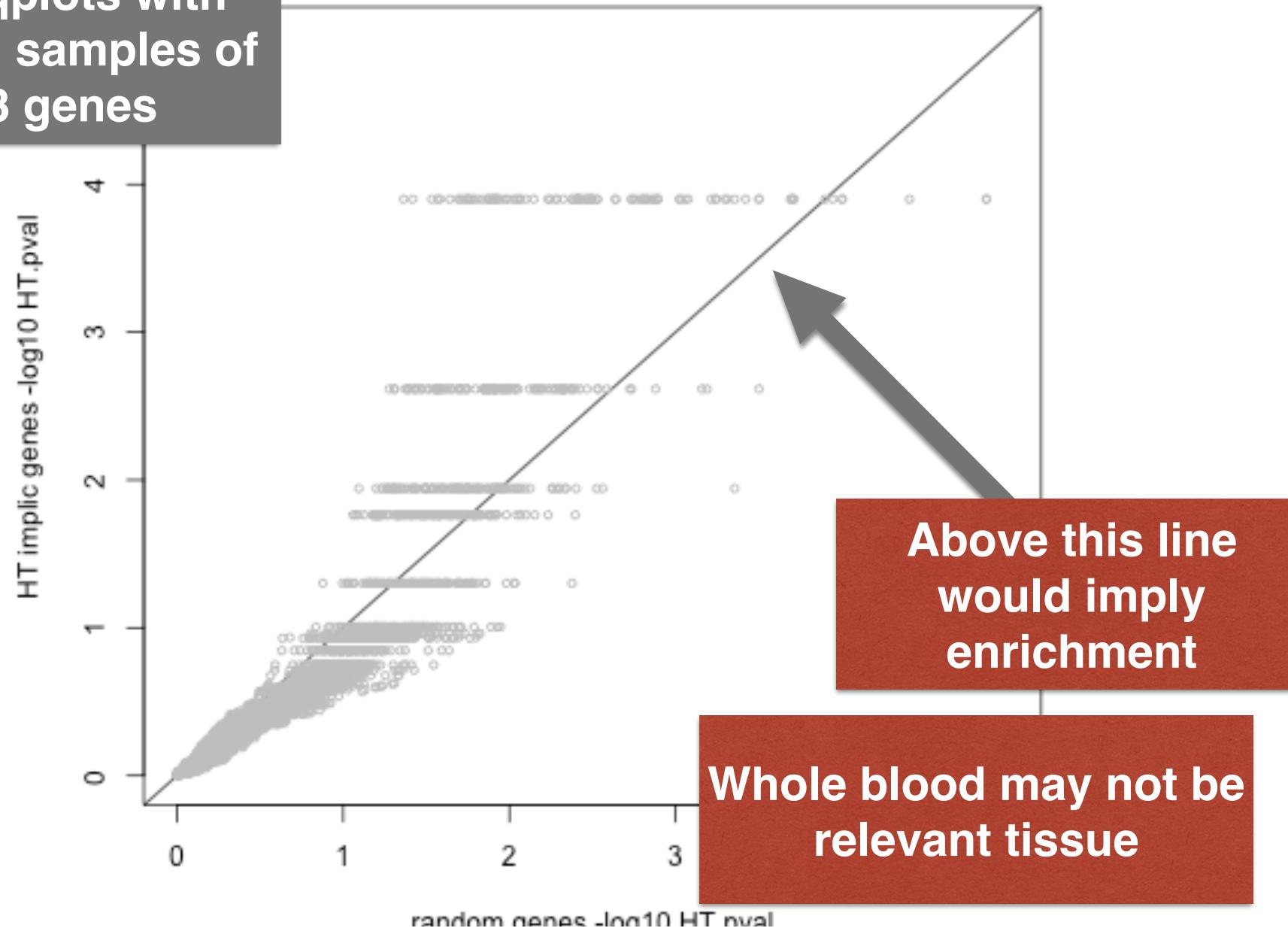
# Enrichment of Known Crohn's Genes Among Findings

100 qqplot with  
random samples of  
205 genes



# No Enrichment Among Hypertension Findings

100 qqplots with  
random samples of  
133 genes



# 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

# Prediction of Gene Expression Traits

# Genetic Architecture to Improve Prediction

- Local and distant regulation (heritability)
- Sparsity/Polygenicity
- This information guides us to improve prediction, i.e. estimates of GReX

# Local/Distant Heritability Estimation

- Gene expression trait model

$$Y = \sum_{\text{local}} \beta_k^{\text{local}} X_k + \sum_{\text{distant}} \beta_k^{\text{distant}} X_k + \epsilon$$

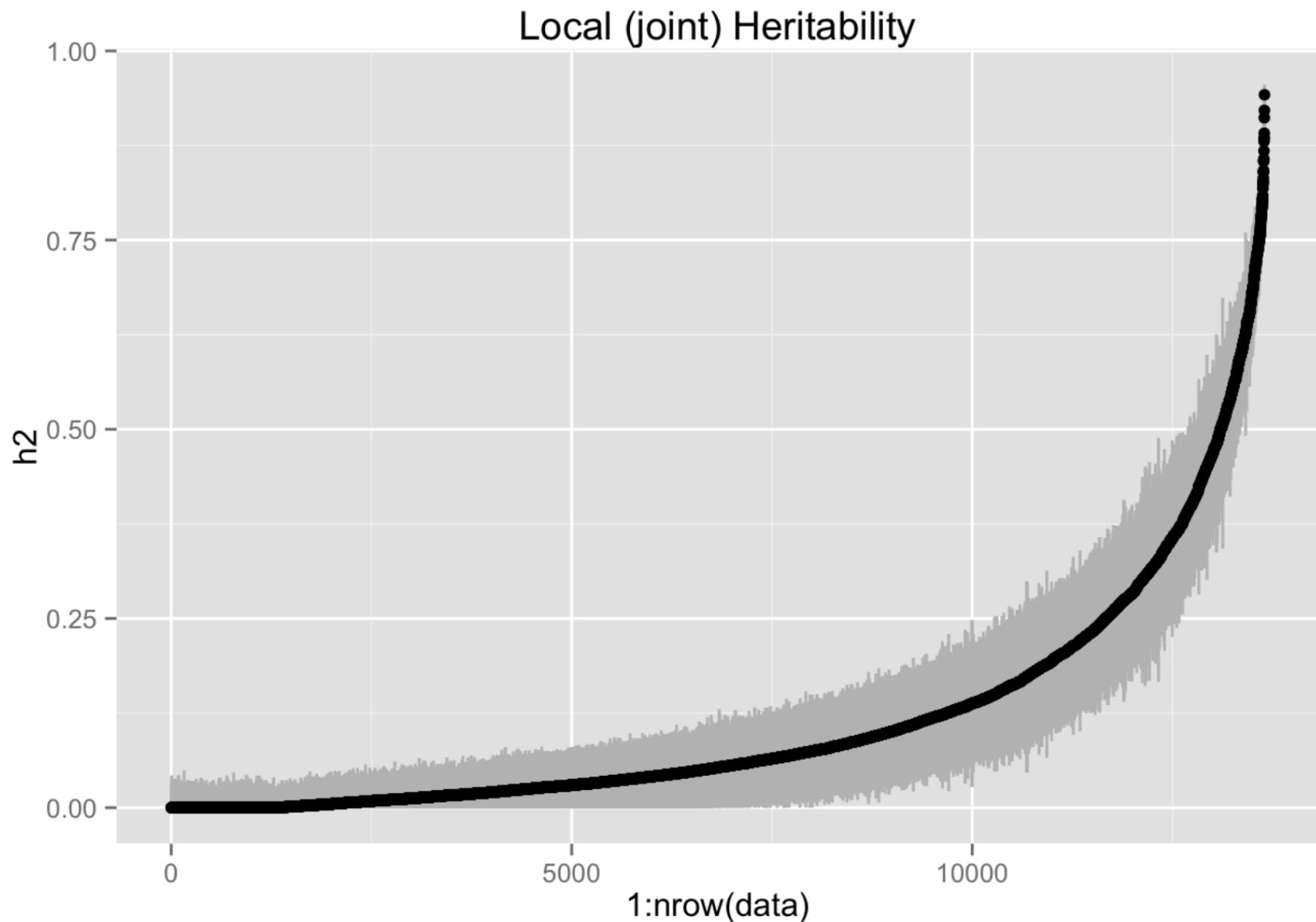
- REML to estimation of local and distant contributions jointly
- Covariance of local component: GRM using SNPs nearby
- Covariance of distant component: GRM using distant SNPs
- We use GCTA as REML calculator

**Total Heritability = Local H<sup>2</sup> + Distant H<sup>2</sup>**

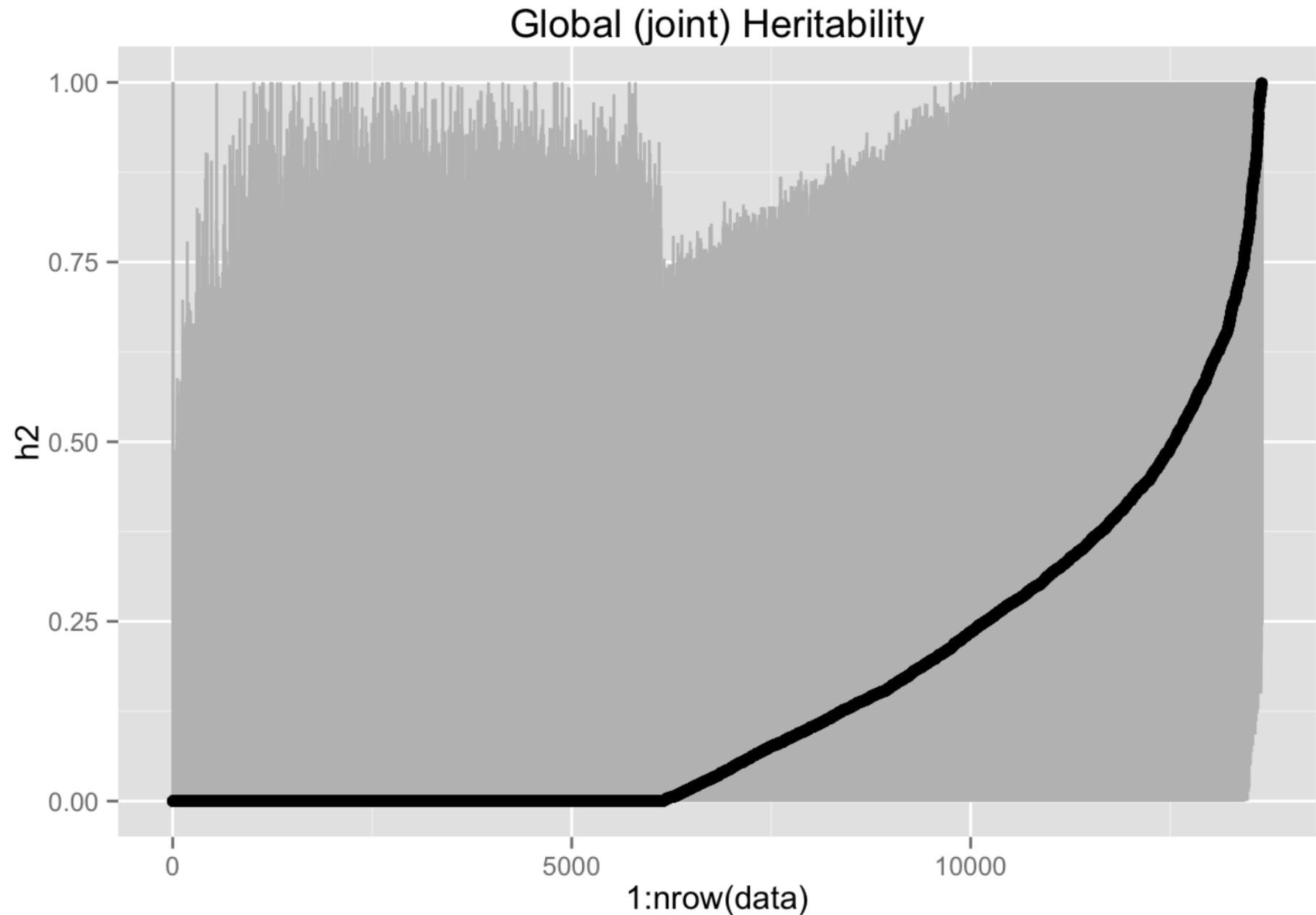
# Whole Blood Expression Data: DGN

- Battle et al. “Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals.” *Genome Research* 2014, 24(1):14-24
- Whole blood from Depression Genes and Networks study
- n = 922
- RNA-seq

# Local Heritability Can Be Well Estimated



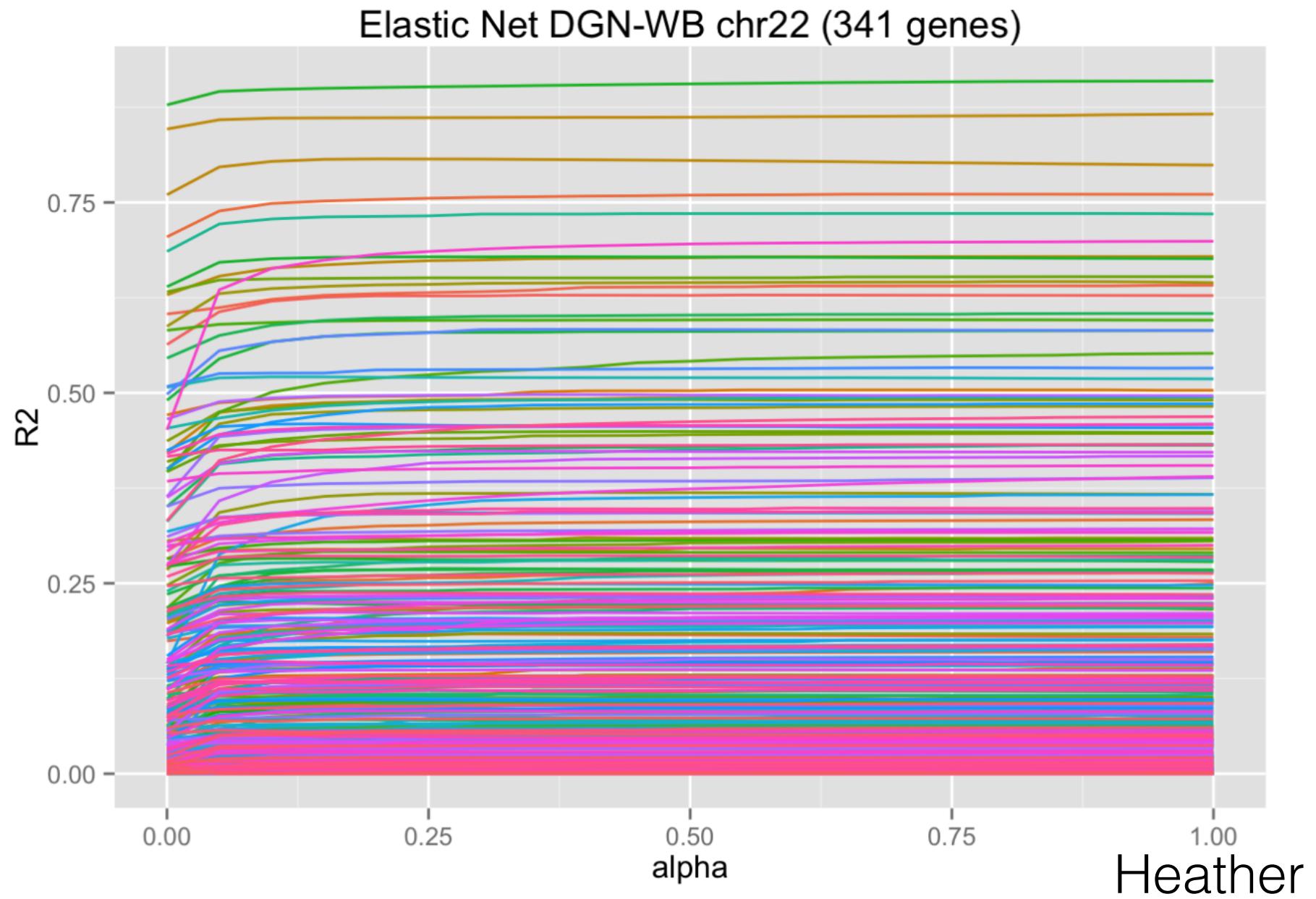
# Distant Heritability Not Reliable



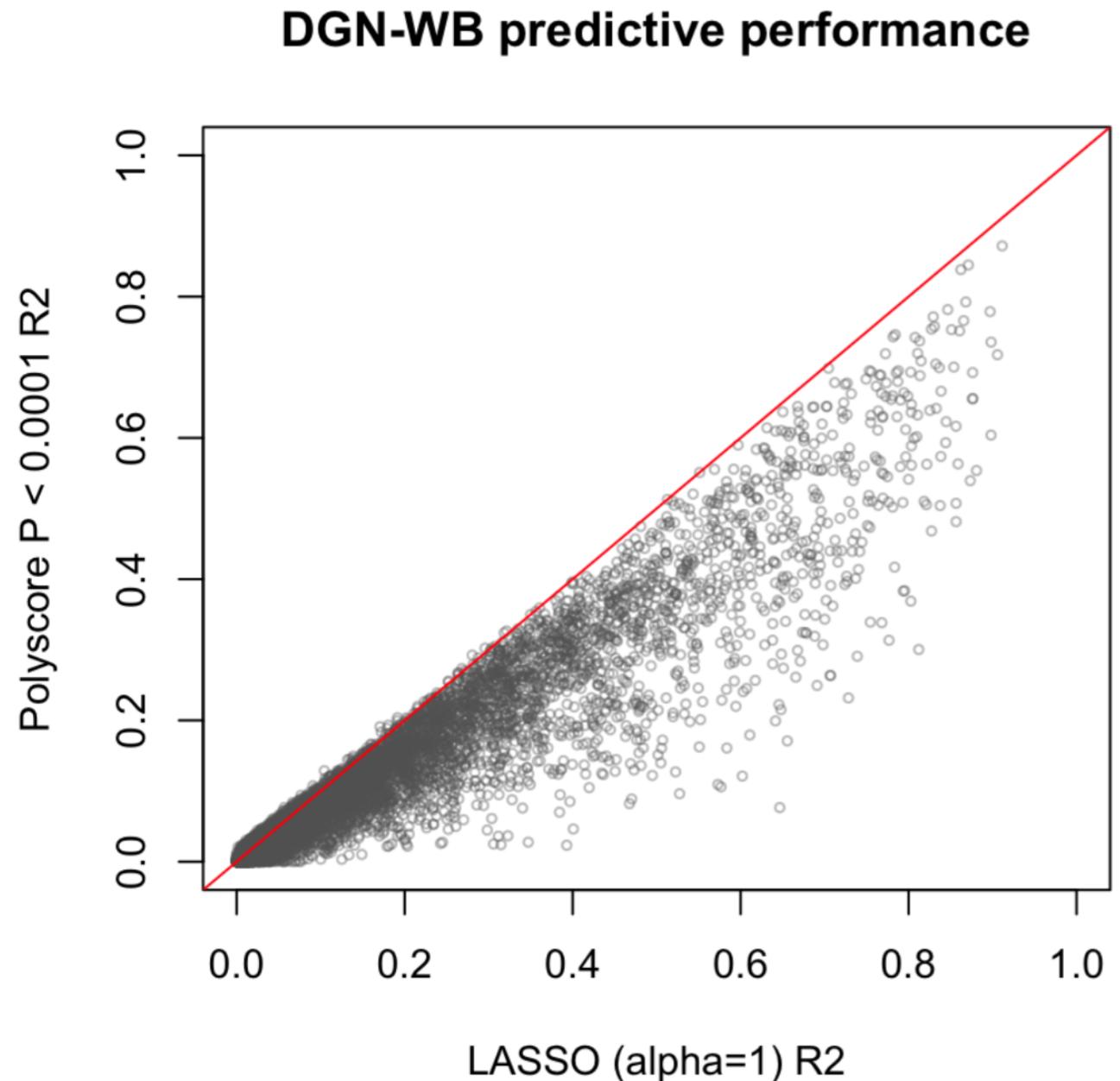
# Proportion of LASSO to Ridge as Measure of Sparsity

- Only local component can be assessed
- LASSO performs slightly better than E-N 0.50 in cross validated R<sup>2</sup>

# Performance vs sparsity



# E-N & LASSO Outperform Polygenic Score



# Whole Blood DGN (n=922) + 38 GTEx Tissue Models

PredictDB

E-N.0.5

Adipose-Subcutaneous\_0.5.db

AdrenalGland\_0.5.db

Artery-Aorta\_0.5.db

Artery-Coronary\_0.5.db

Brain-Anterior...4)\_0.5.db

Brain-Caudate...)\_0.5.db

Brain-Cerebellar Hemisphere\_0.5.db

Brain-Cerebellum\_0.5.db

Brain-Cortex\_0.5.db

Brain-FrontalCortex\_0.5.db

Brain-Hippocampus\_0.5.db

Brain-Hypothalamus\_0.5.db

Brain-NucleusAccumbens\_0.5.db

Brain-Putamen\_0.5.db

Breast-MammaryTissue\_0.5.db

Cells-EBV-transformedLymphocytes\_0.5.db

Cells-TransformedK562\_0.5.db

Colon-Sigmoid\_0.5.db

Colon-Transverse\_0.5.db

DGN-WB\_0.5.db

Esophagus-Gastroesophageal Junction\_0.5.db

Esophagus-Mucosa\_0.5.db

Esophagus-Muscularis\_0.5.db

Heart-AtrialAppendage\_0.5.db

Heart-LeftVentricle\_0.5.db

Liver\_0.5.db

Lung\_0.5.db

Nerve-Tibial\_0.5.db

Ovary\_0.5.db

Pancreas\_0.5.db

Pituitary\_0.5.db

Skin-NotSunExposed\_0.5.db

Skin-SunExposed\_0.5.db

SmallIntestine-TerminalIleum\_0.5.db

Spleen\_0.5.db

Stomach\_0.5.db

Testis\_0.5.db

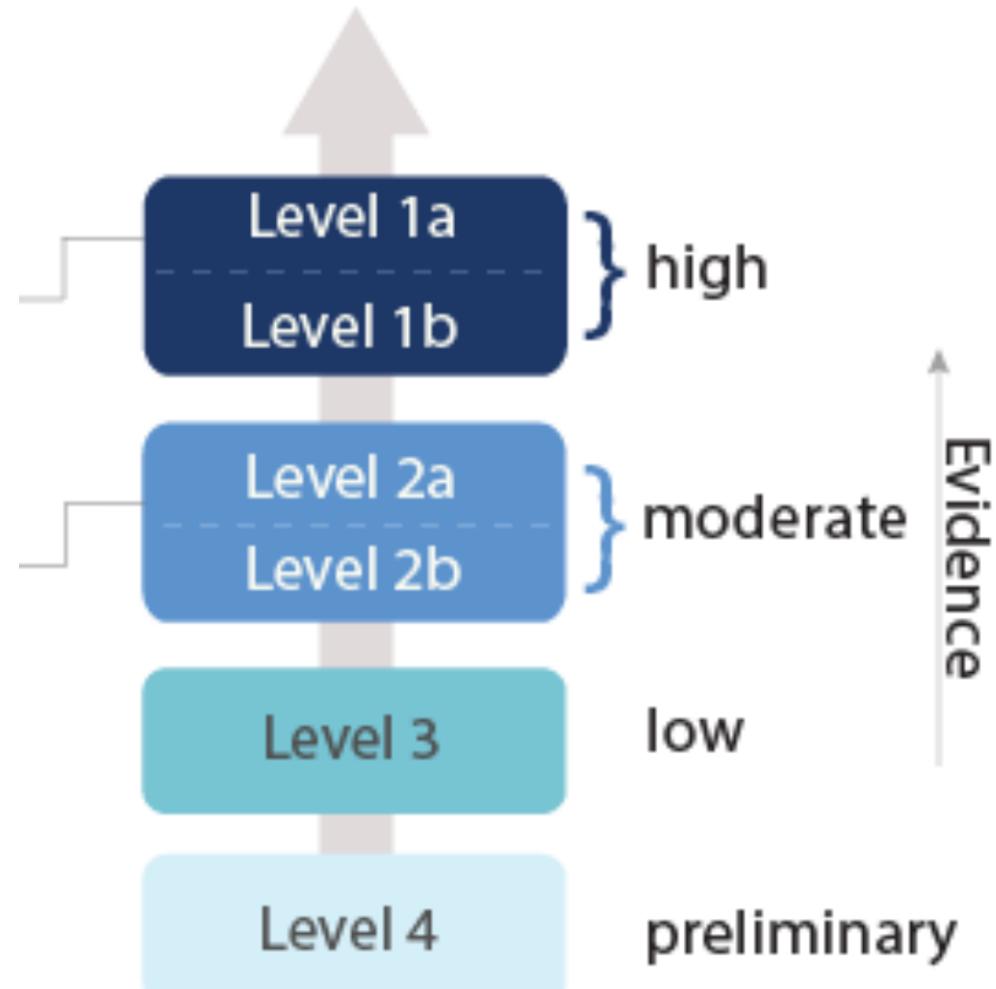
Thyroid\_0.5.db

Whole.Blood\_0.5.db

# Challenges in Pharmacogenomic Predictions

# 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 findings have clinical guidelines

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

# 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

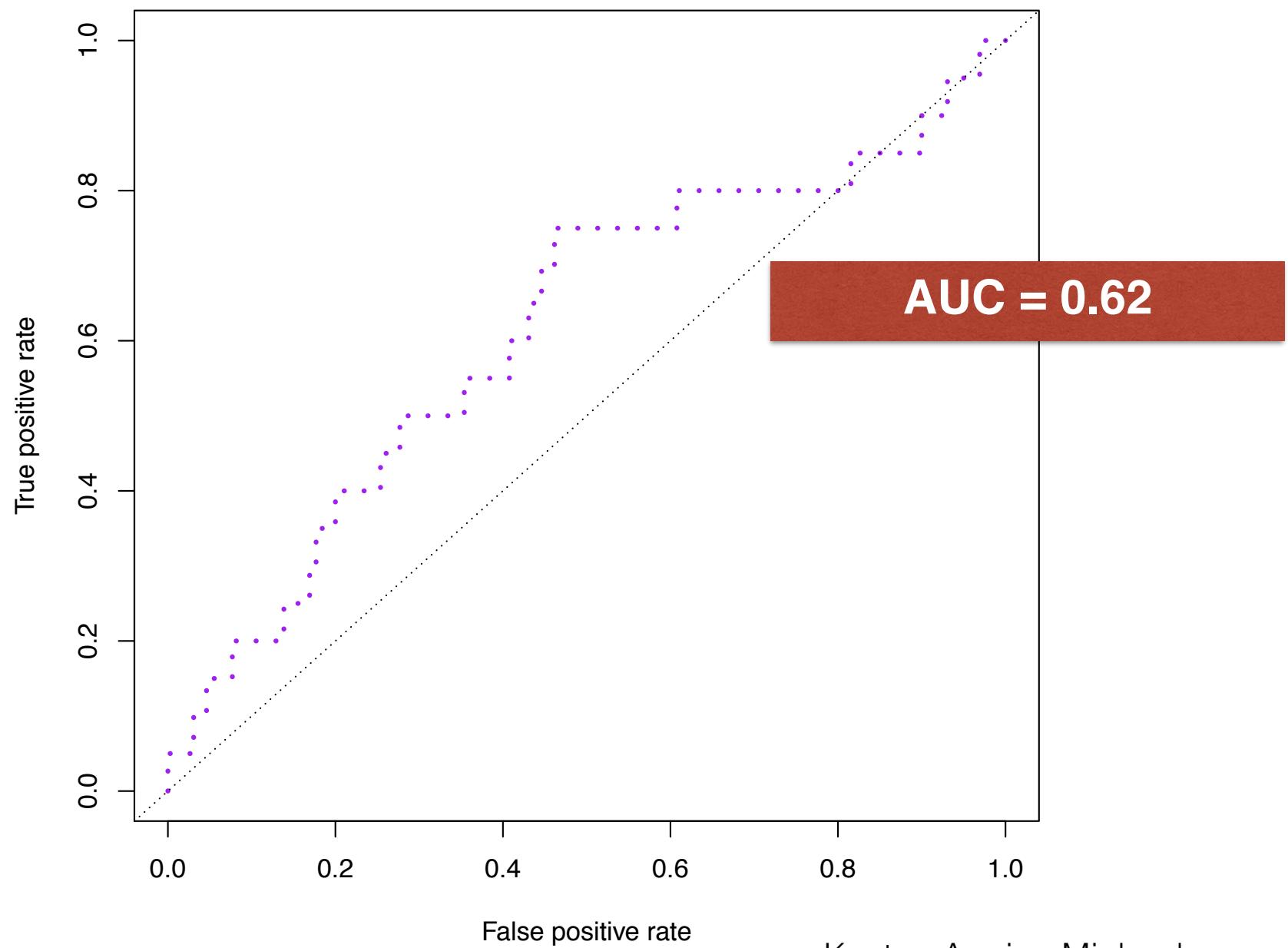
- CALGB 90401
  - a randomized double-blinded placebo controlled phase III trial comparing docetaxel and prednisone with and without bevacizumab in men with hormone refractory prostate cancer
  - n = 664 (with genotype data after QC)
  - PI: Howard McLeod
- CALGB 80303
  - a randomized phase III trial 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

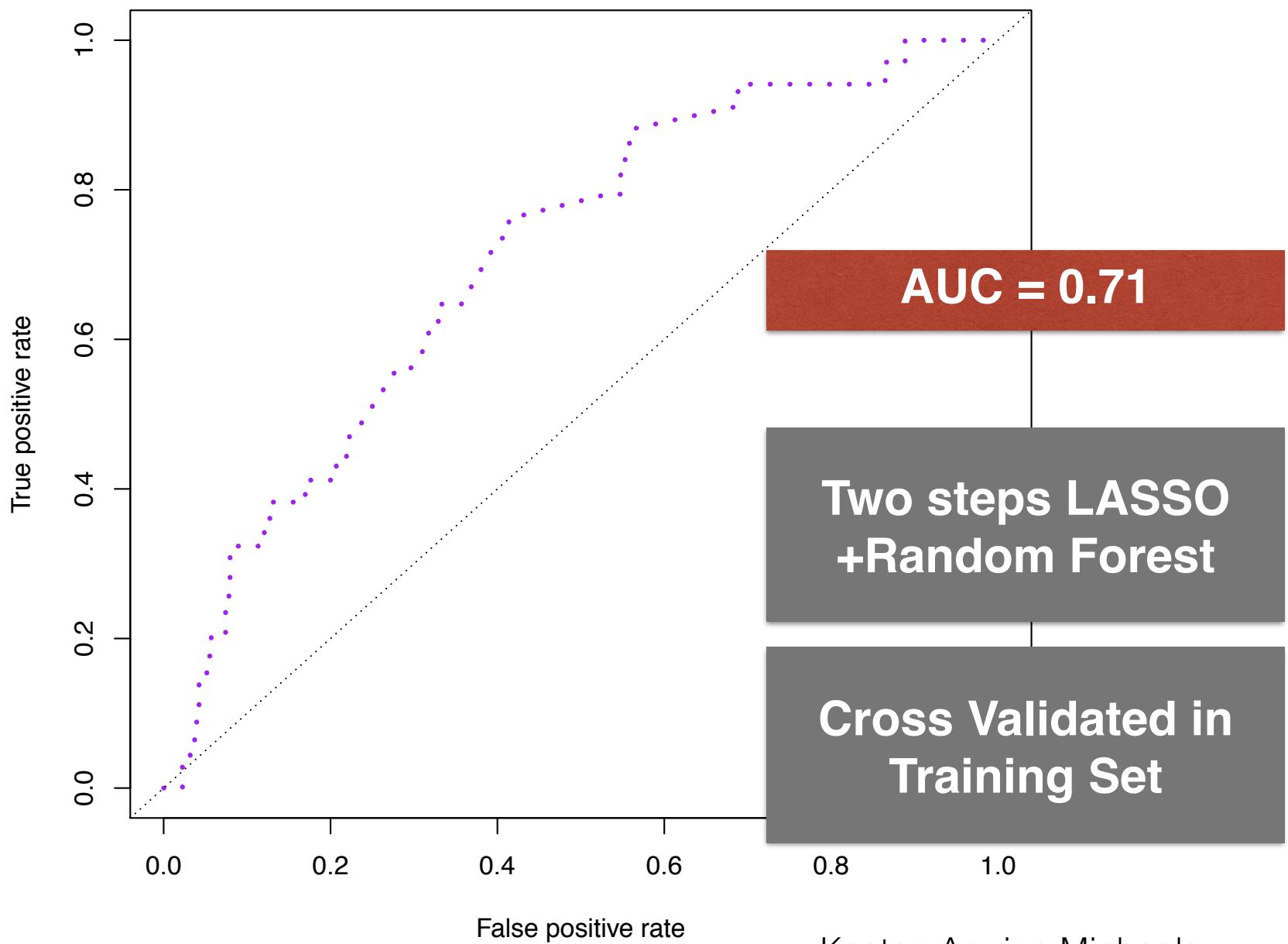
Keston Aquino Michaels & Heather Wheeler

# Primary Hypertension Score Predicts Bev-induced HT



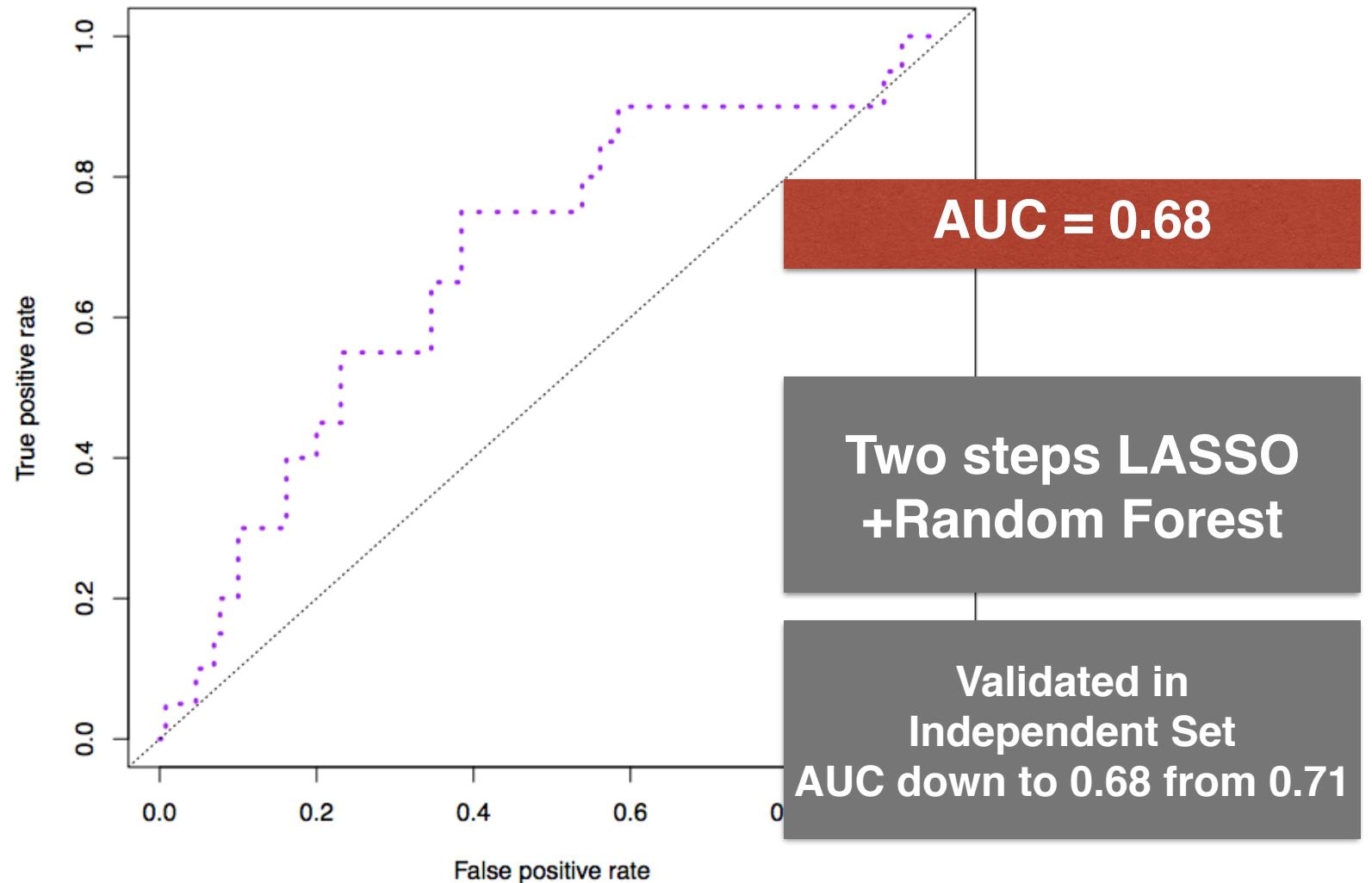
Keston Aquino Michaels

# Bev-Hypertension Predicted Within Study



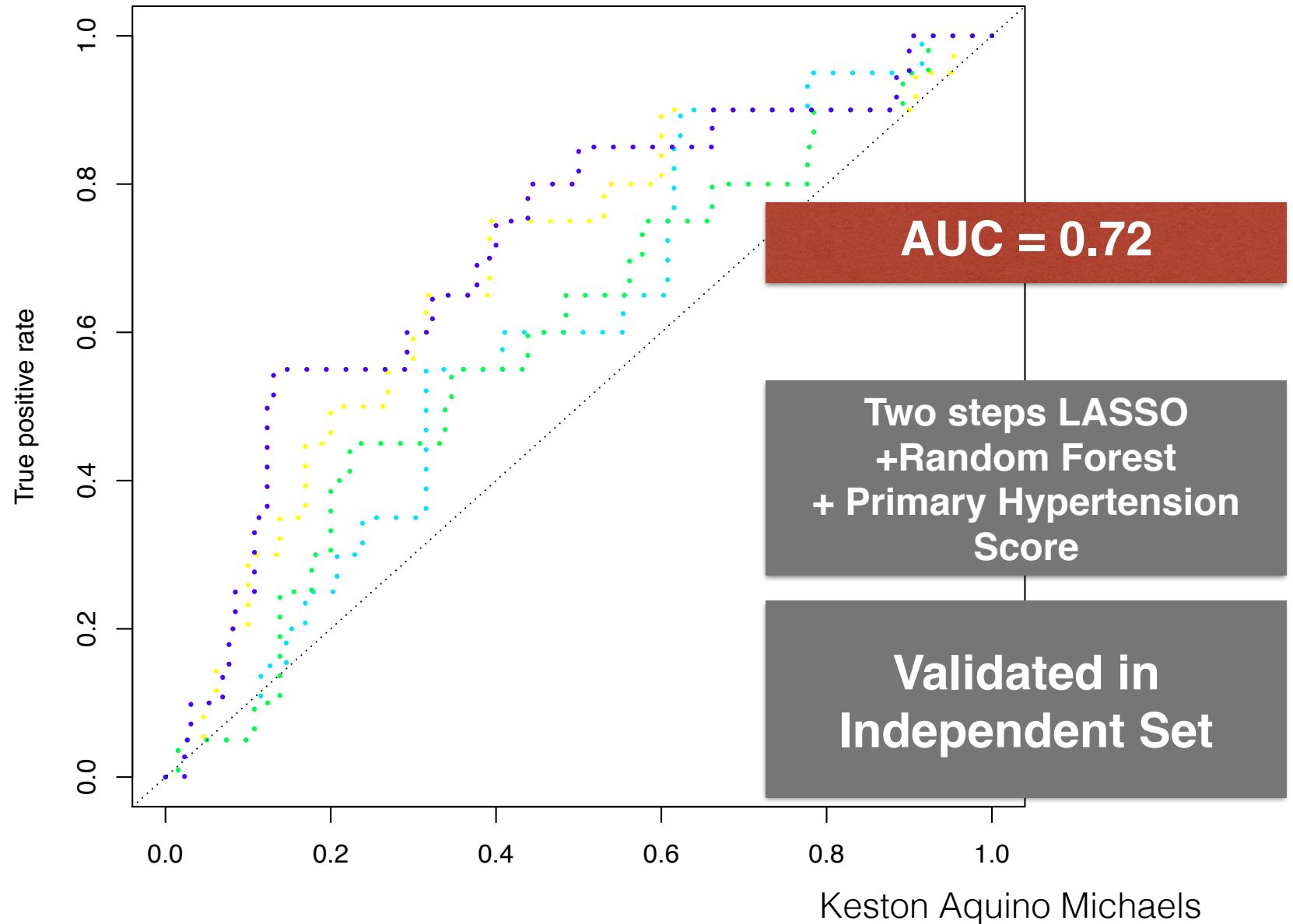
Keston Aquino Michaels

# Bev-Hypertension Predicted in Independent Study



Keston Aquino Michaels

# Bev-Hypertension Predicted in Independent Study



# Summary Pharmacogenomics

- Most single variant findings have limited clinical utility
- Whole genome approaches to prediction improves utility
- Bevacizumab induced hypertension example
  - primary hypertension results help in predicting drug induced hypertension
  - successfully predicted bevacizumab induced hypertension in independent study
  - combining primary + bevacizumab induced HT leads to improved prediction

# Summary

- Shift from monogenic to polygenic paradigm
- Systems approach to genomics
  - Most single variant findings have limited clinical utility
  - Whole genome approaches to prediction improves utility
- Larger sample sizes will be needed, 1 Million+
- OmicKriging: prediction method that integrates heterogeneous sources of data well suited for data from the Precision Medicine Initiative
- Large role of regulation variants in complex traits
- PrediXcan: novel gene based test that test mechanism
- Prediction of gene expression traits

# Conclusion

- recognizing the complexity of the genetic architecture and mechanisms of genetic control,
- collecting deep phenotype data from large number of individuals,
- broadly sharing data and results, and
- integrating multiple sources of data
- using mechanism-driven tests

We will achieve the promise of precision medicine

# Thank You!

## Contributors

- Heather Wheeler
- Nancy Cox
- Eric Gamazon
- Keston Aquino Michaels
- Sahar Mozaffari
- Kaanan P. Shah
- Nicholas Knoblauch
- Vassily Trubetskoy
- GTEx Consortium

## Data sources

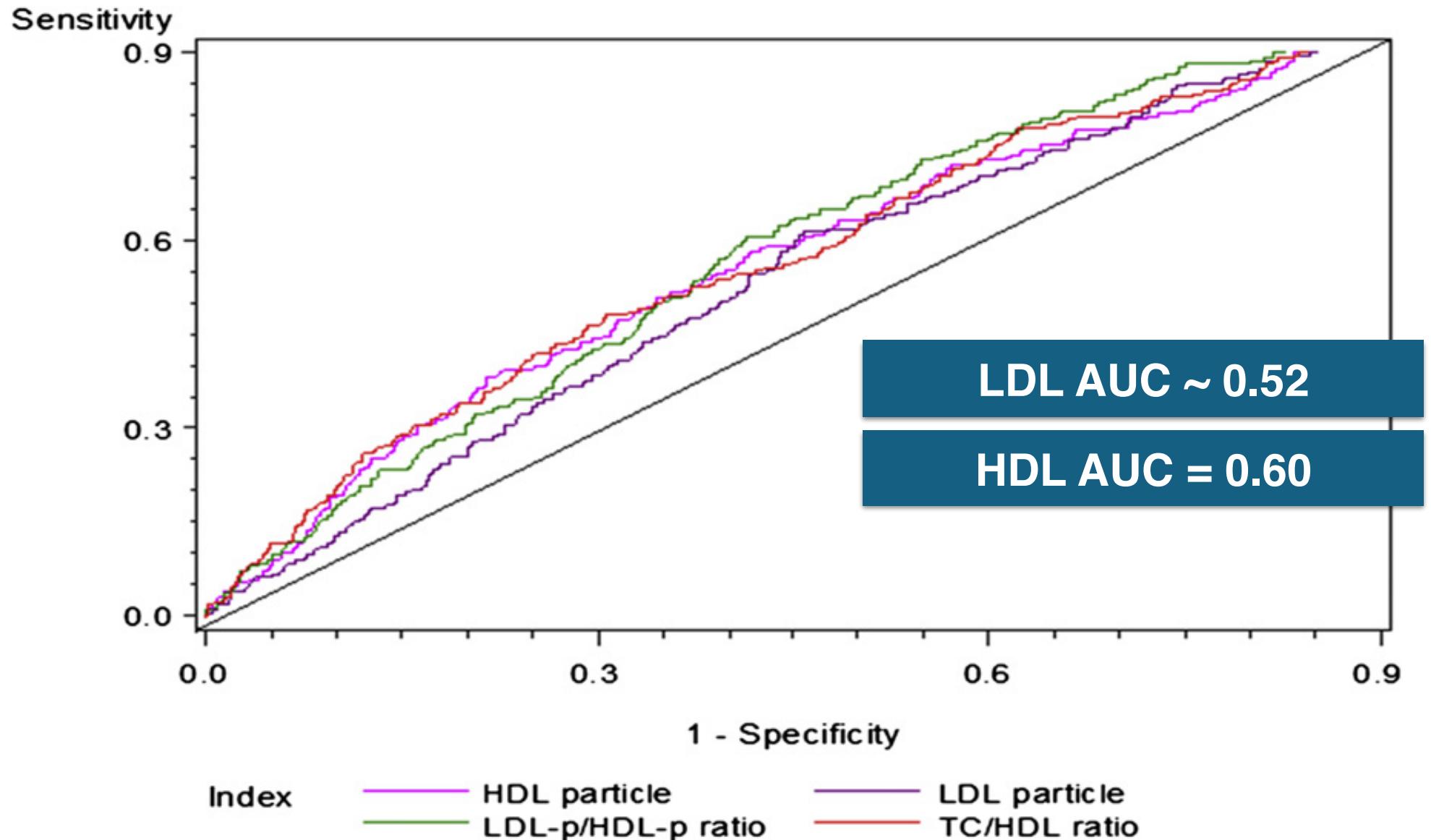
- WTCCC
- GAINS/Bipolar Disorder
- GoKinD
- Disease Genes & Networks

## 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
- P50DA037844 - Integrated GWAS of complex behavioral and gene expression traits in outbred rats
- 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



# 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>,

