Leveraging GTEx to Unveil the Biology of Complex Traits

Hae Kyung Im, PhD

THE UNIVERSITY OF CHICAGO

October 5, 2015

Heritability of Gene Expression Traits

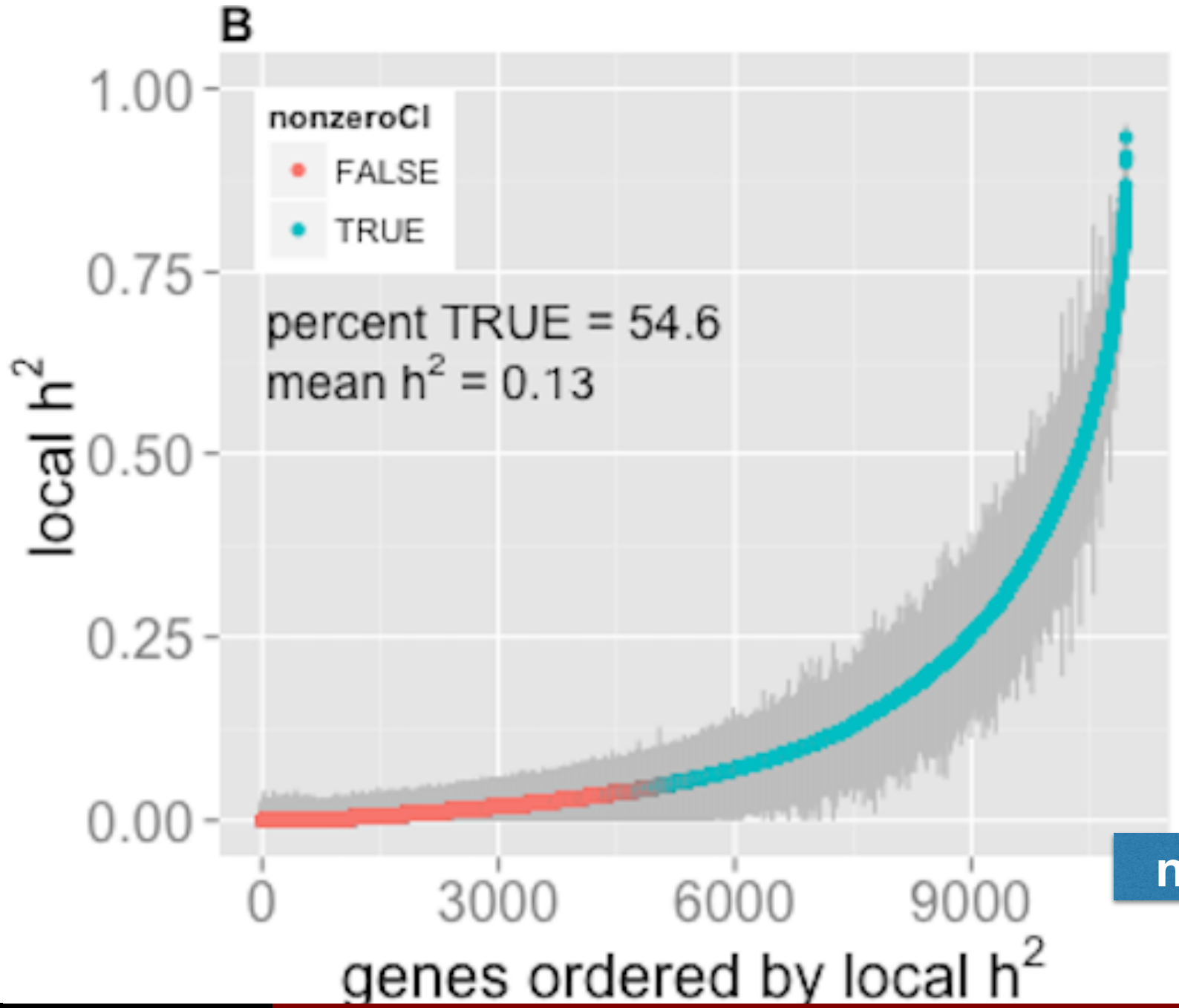
Heather Wheeler

Local + Distant Heritability Estimation

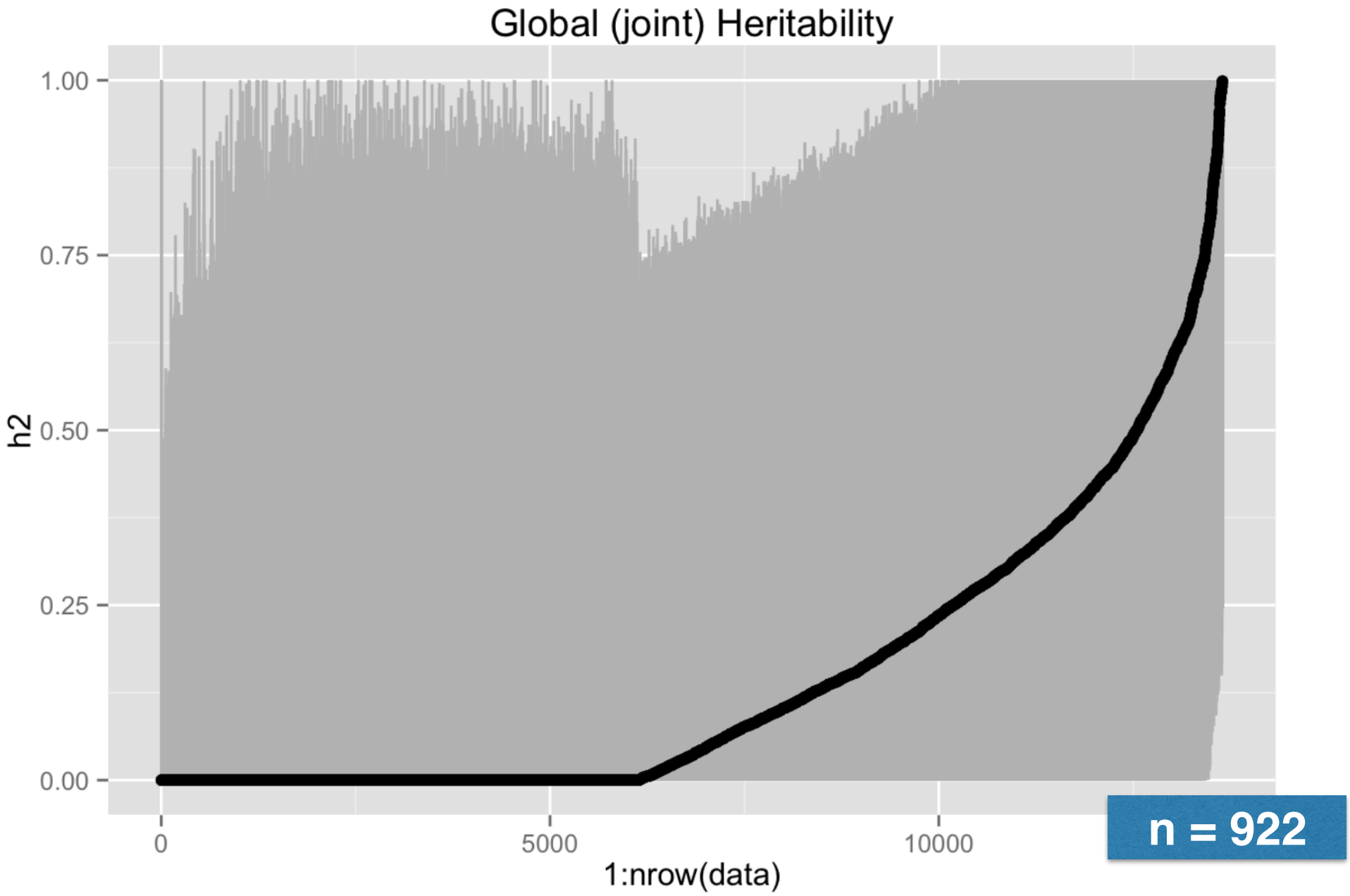
$$Y = \sum_{\text{local}} \beta_k^{\text{local}} X_k + \sum_{\text{distant}} \beta_k^{\text{distant}} X_k + \epsilon$$

Total Heritability = Local H2 + Distant H2

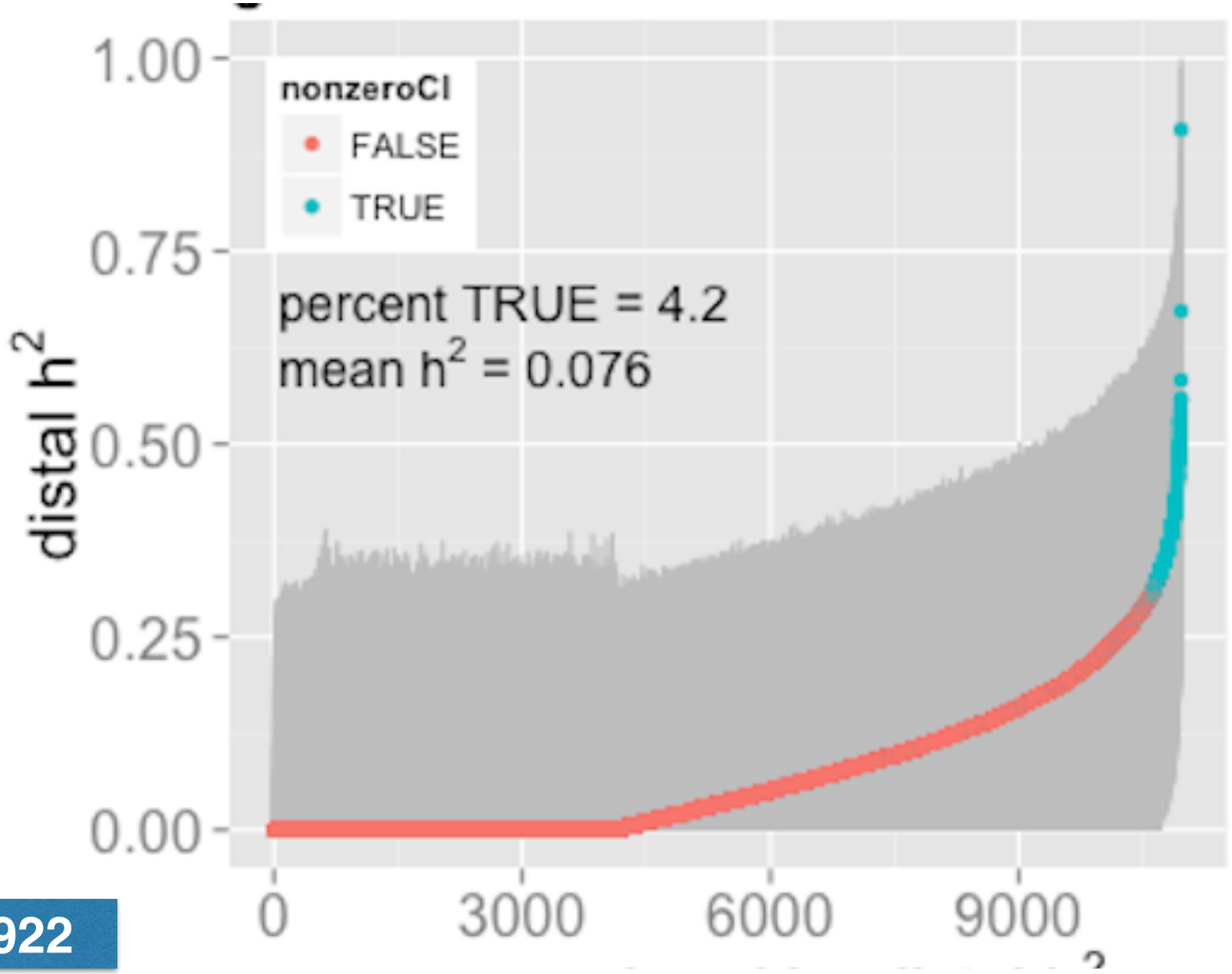
Local Heritability Well Estimated



Distant h2 Not Well Estimated (n=922)

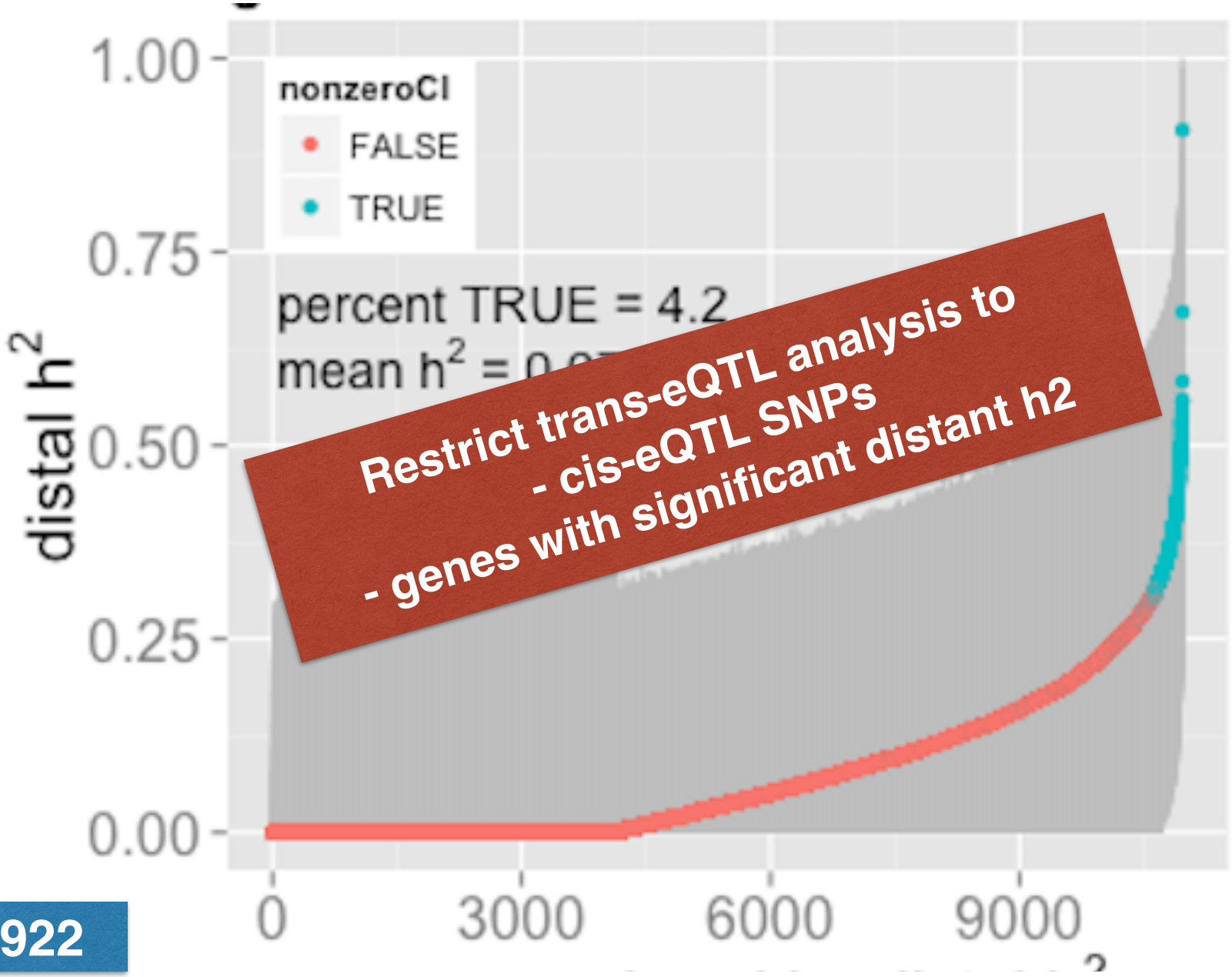


cis-eQTL Priors (FHS) Improve Distant h2 Estimates



n = 922

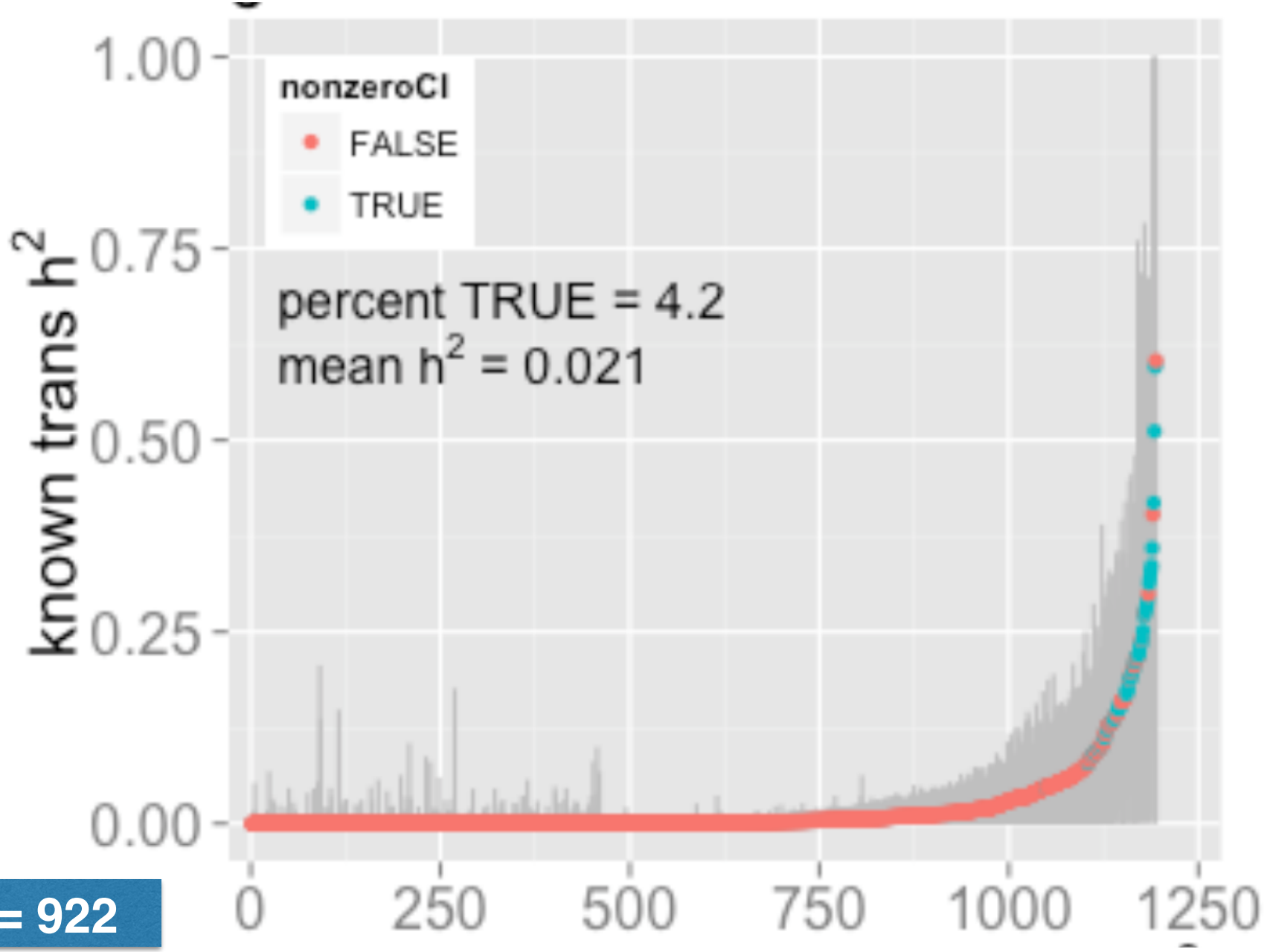
cis-eQTL Priors (FHS) Improve Distant h2 Estimates



Restrict trans-eQTL analysis to
- cis-eQTL SNPs
- genes with significant distant h2

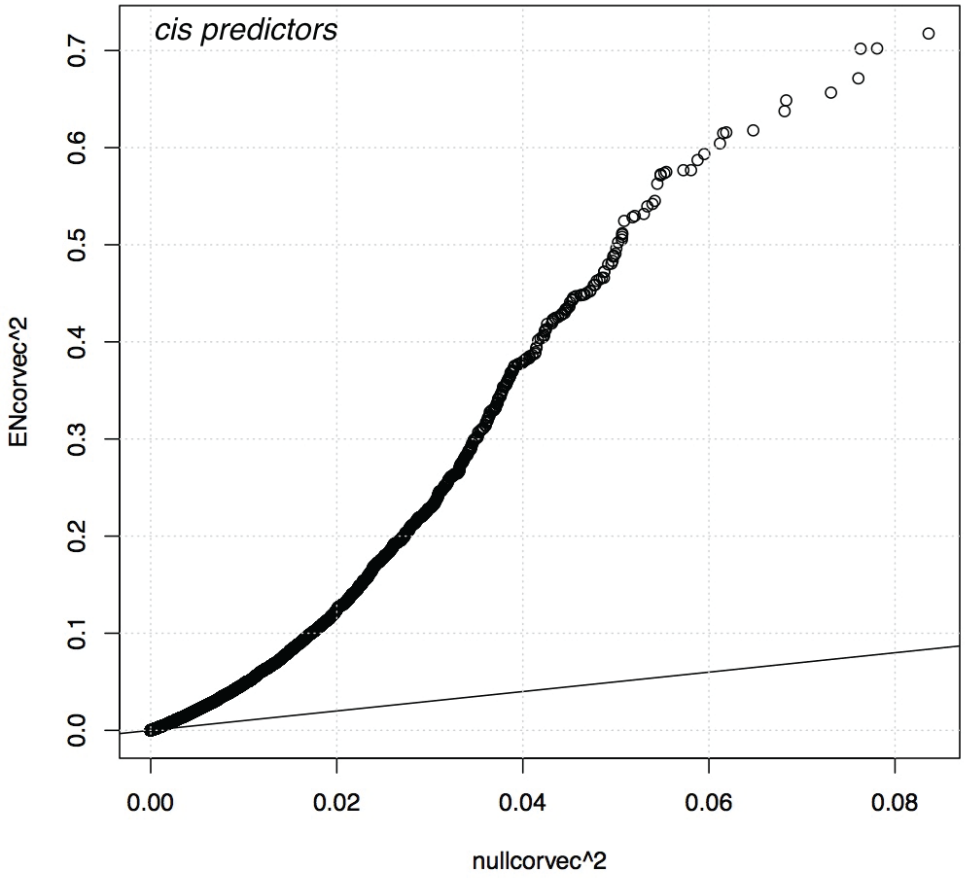
n = 922

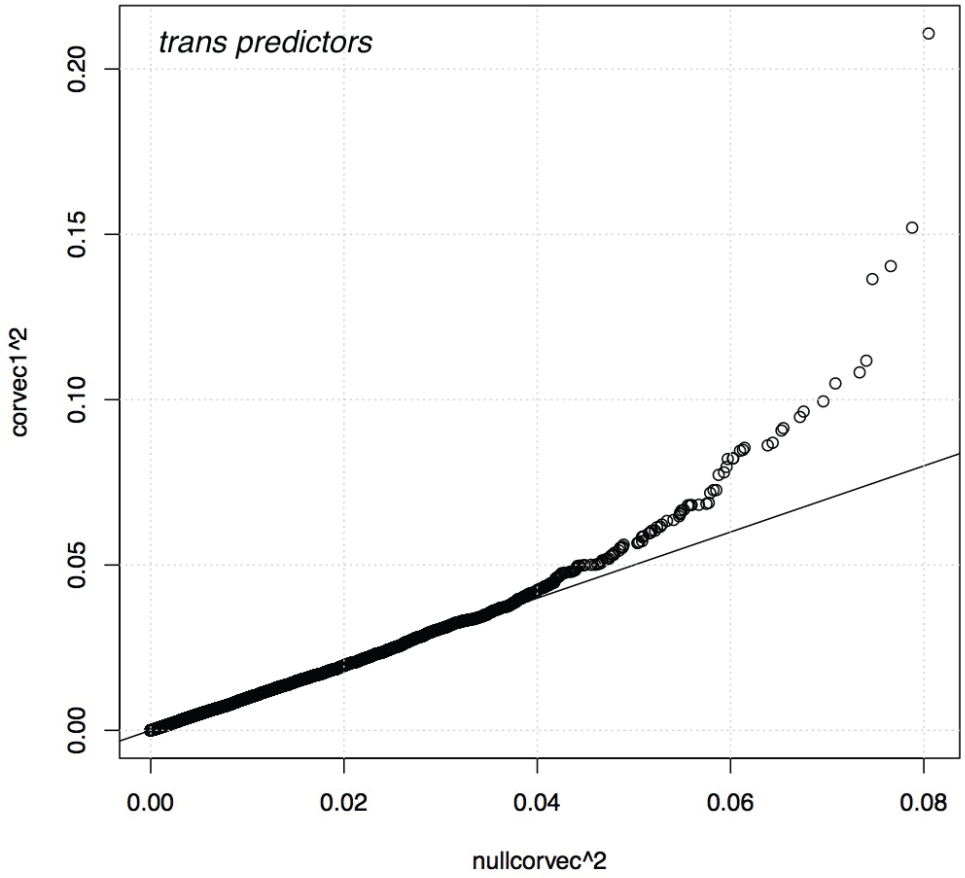
Distant eQTL Prior for Specific Genes Improves h2



n = 922

Performance of Local vs. Distant Variants





Sparsity of Gene Expression Traits

Estimating Sparse & Polygenic Components with BSLMM

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]*

1 Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America, **2** Department of Statistics, University of Chicago, Chicago, Illinois, United States of America

Abstract

Both linear mixed models (LMMs) and sparse regression models are widely used in ge
polygenic modeling in genome-wide association studies. These two approaches ma

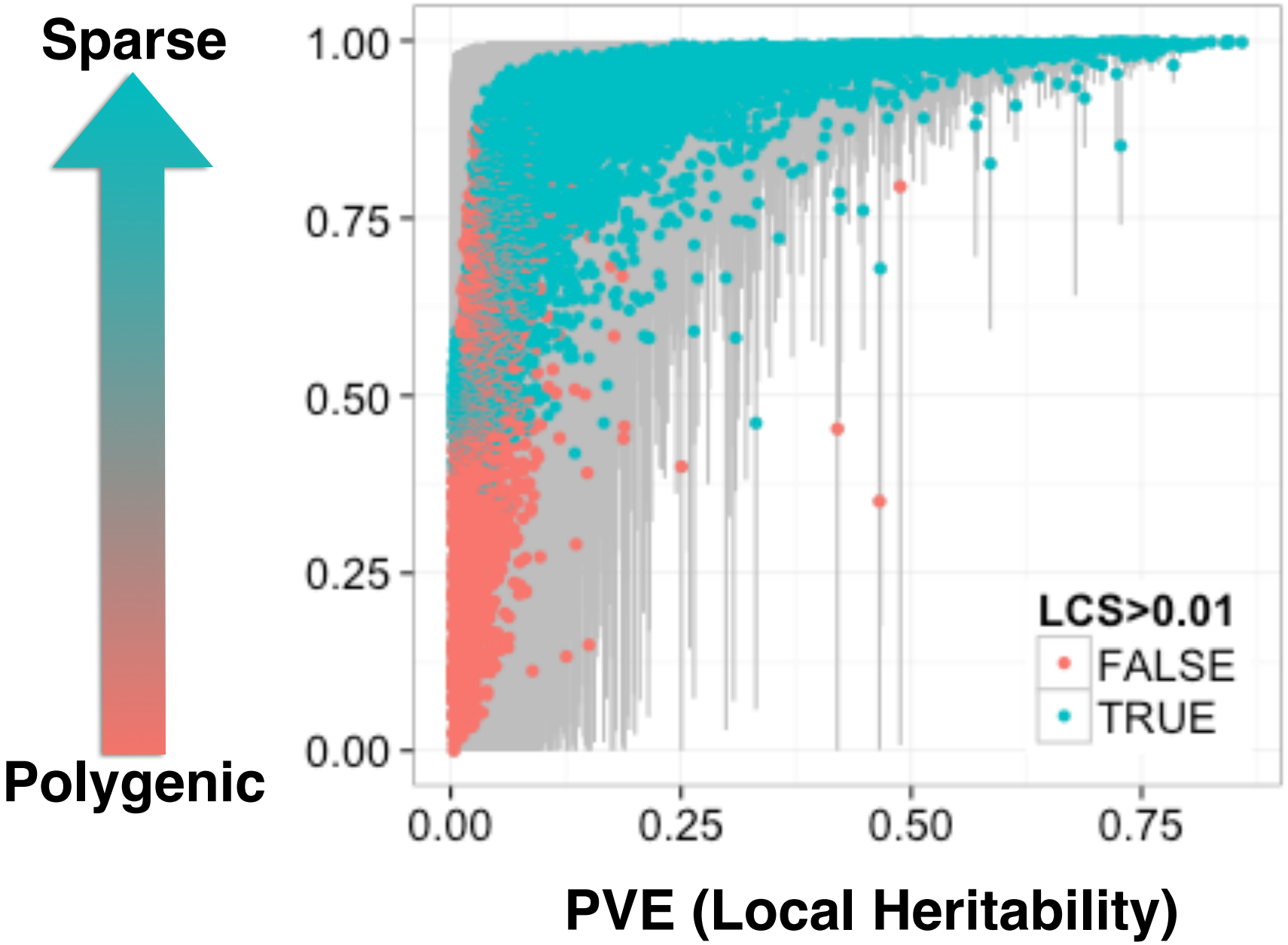
$$\mathbf{y} = 1_n\mu + \mathbf{X}\tilde{\beta} + \mathbf{u} + \boldsymbol{\varepsilon},$$

$$\mathbf{u} \sim \mathrm{MVN}_n(0, \sigma_b^2\tau^{-1}\mathbf{K}),$$

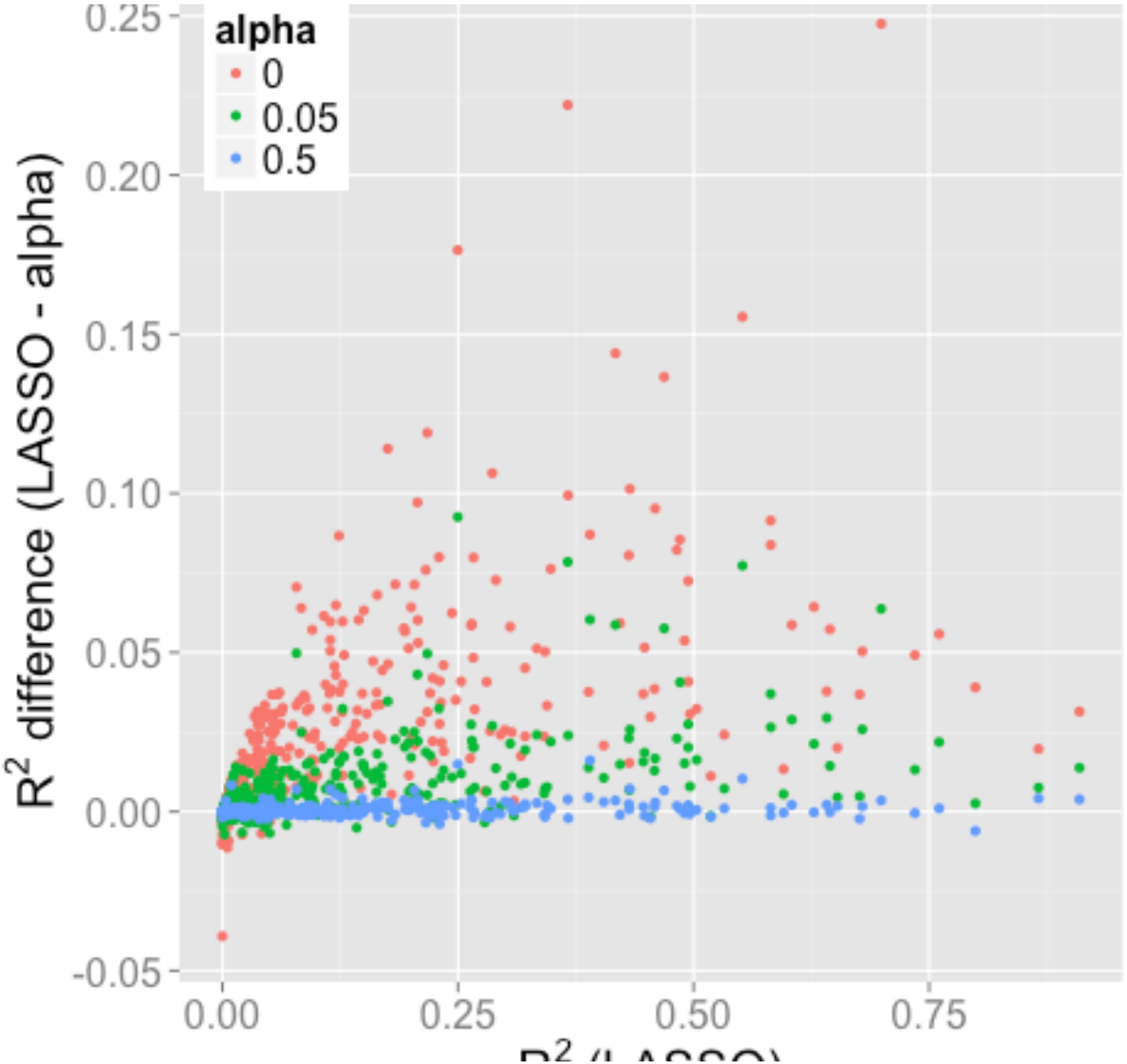
$$\boldsymbol{\varepsilon} \sim \mathrm{MVN}_n(0, \tau^{-1}\mathbf{I}_n),$$

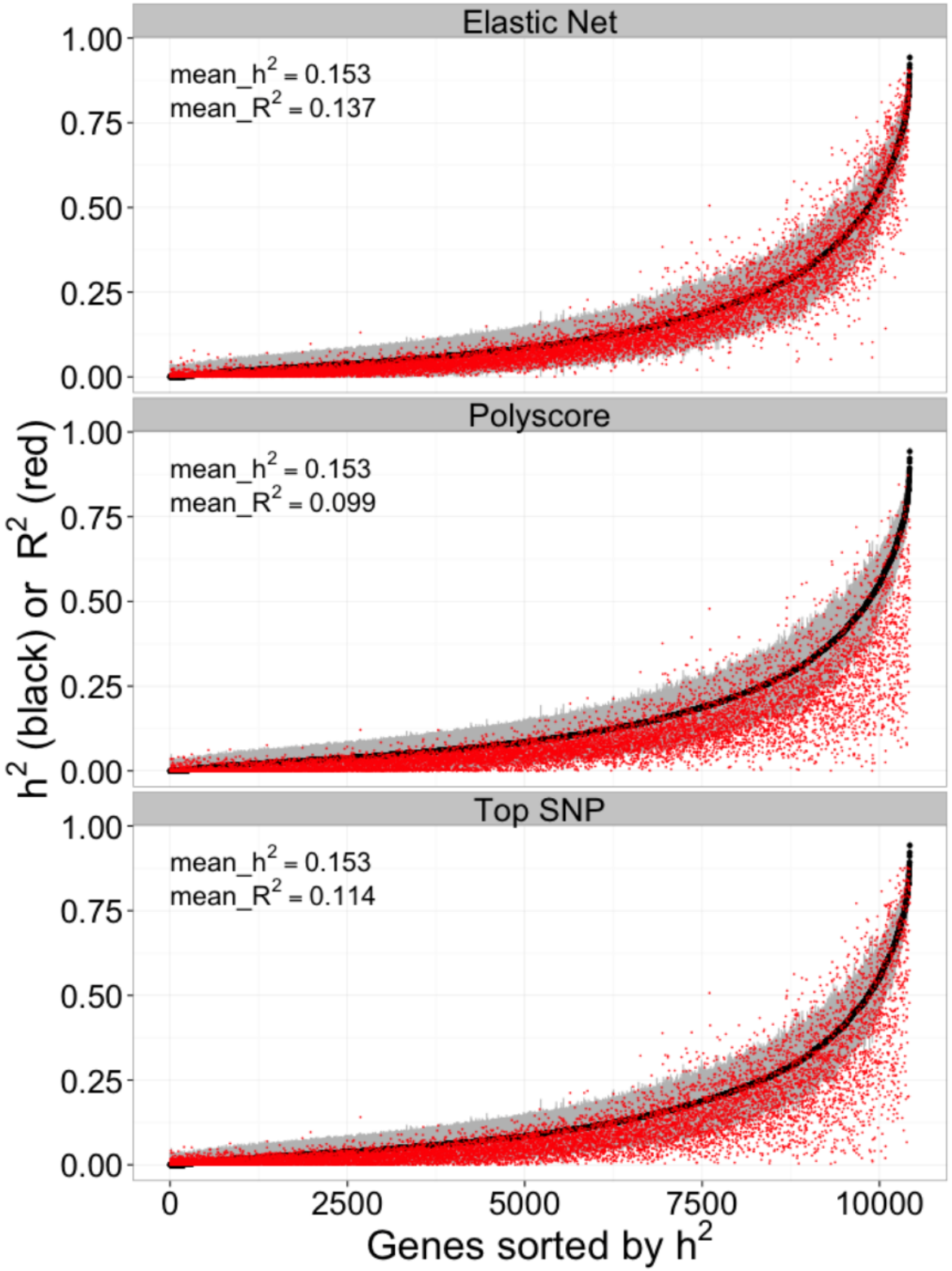
$$\tilde{\beta}_i \sim \pi \mathrm{N}(0, \sigma_a^2\tau^{-1}) + (1-\pi)\delta_0,$$

Highly Heritable Genes are Sparse



Sparse Models Outperform Polygenic Models in Prediction



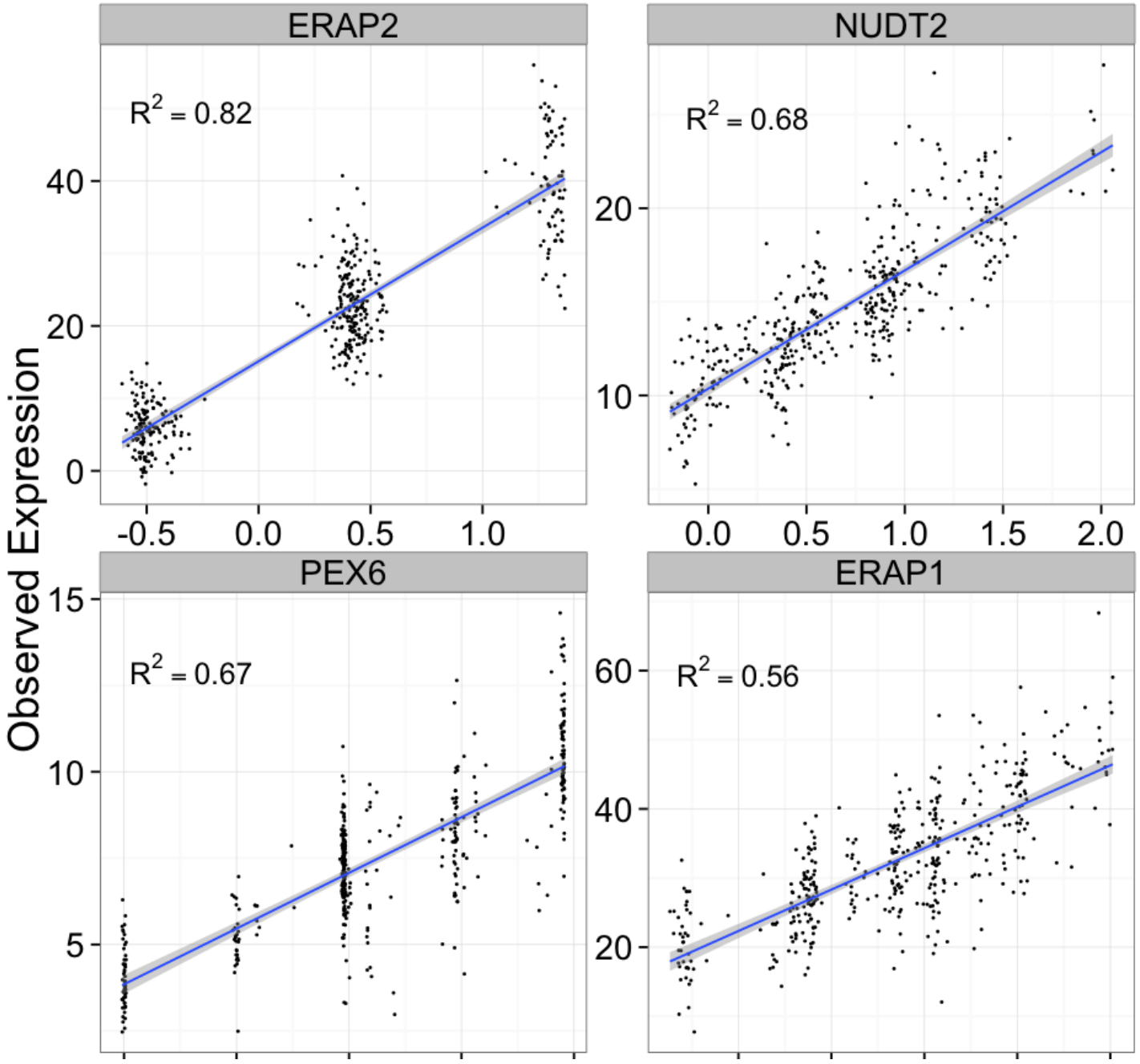


Sparse Models Outperform Polygenic Models

Polygenic architecture not supported

Using top eQTL suboptimal

Examples of well predicted genes

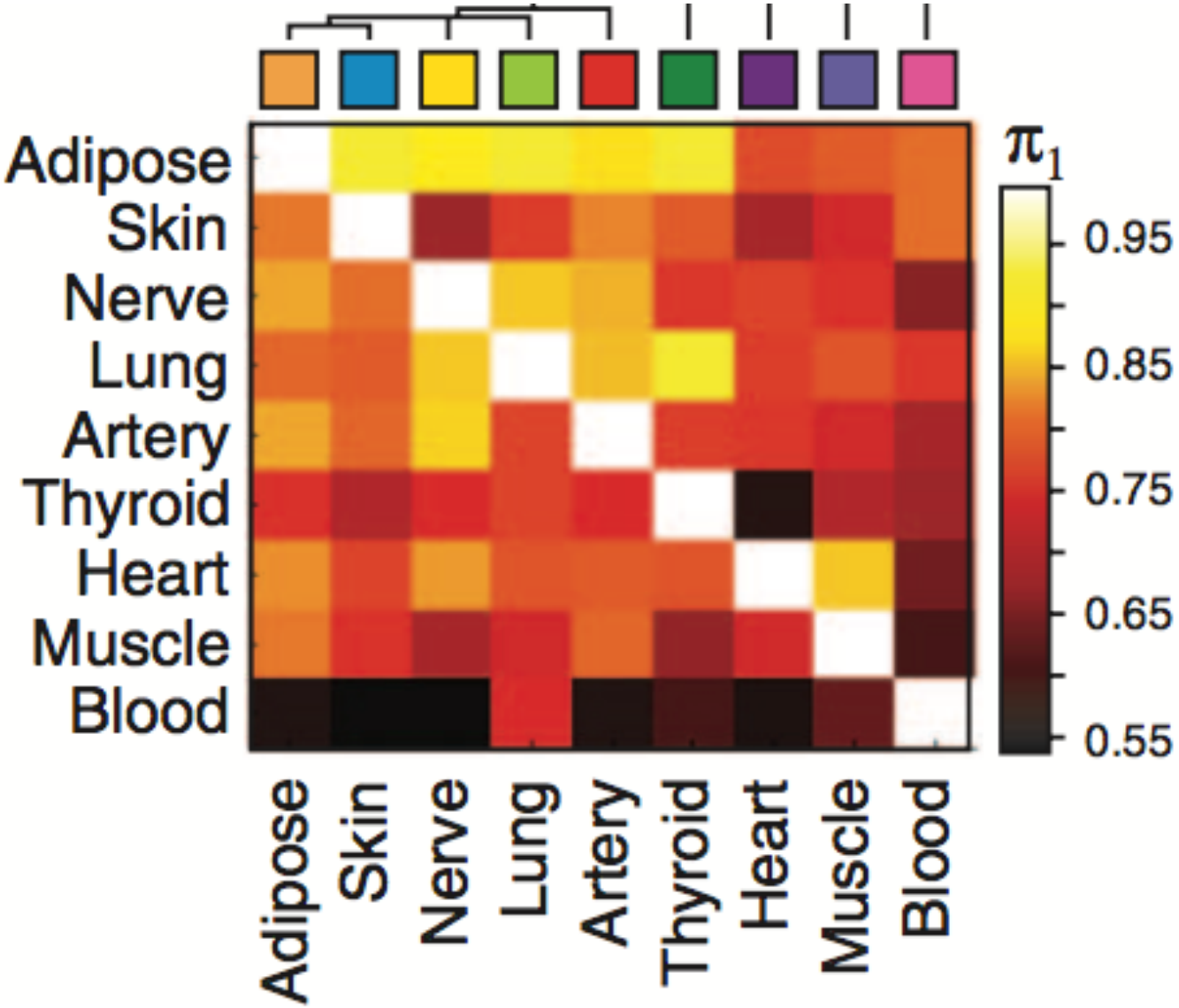


ERAP2: one causal variant

PEX6: two causal variants

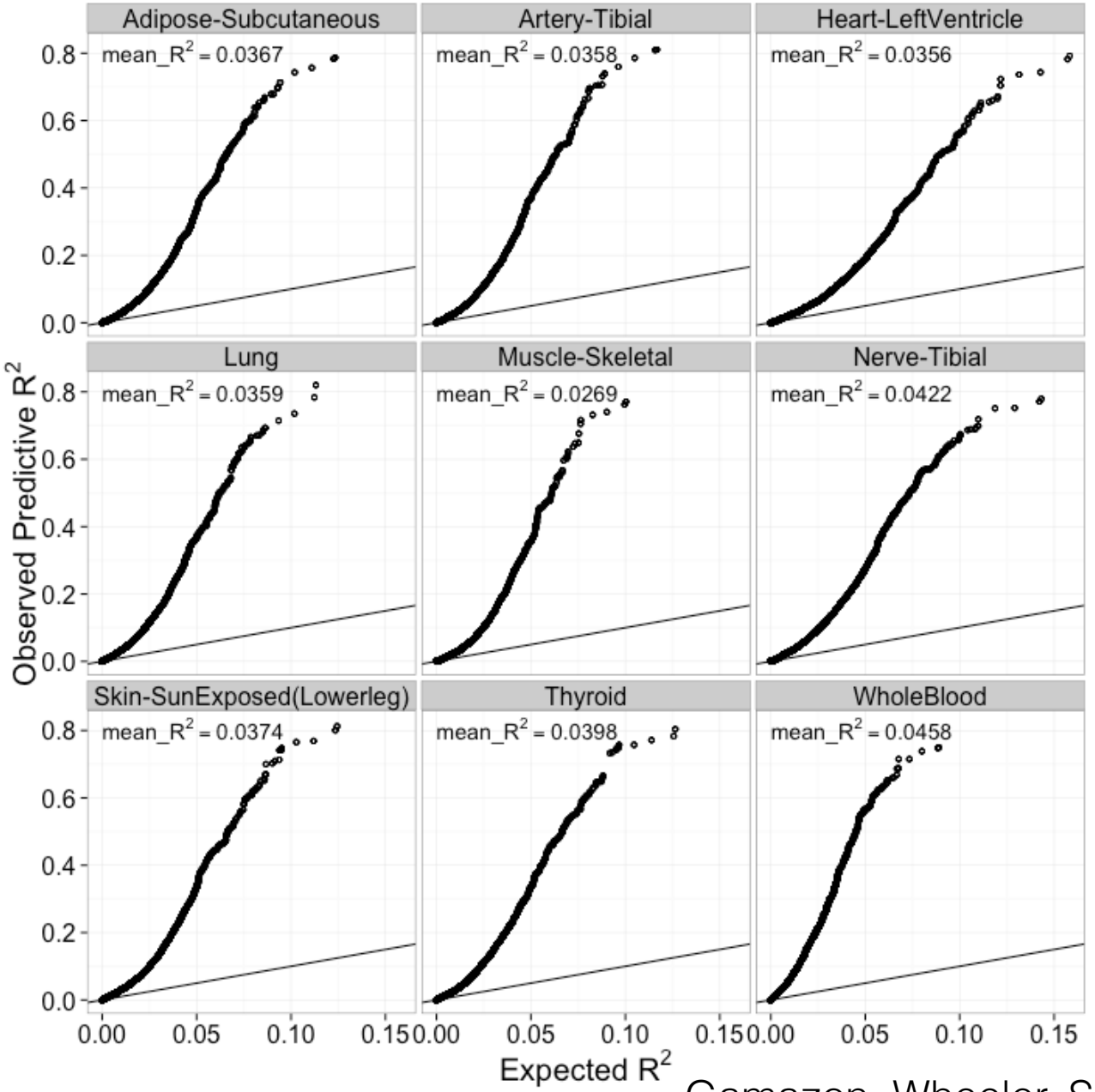
Cross Tissue and Tissue Specific Architecture

eQTLs Are Highly Shared Across Tissues



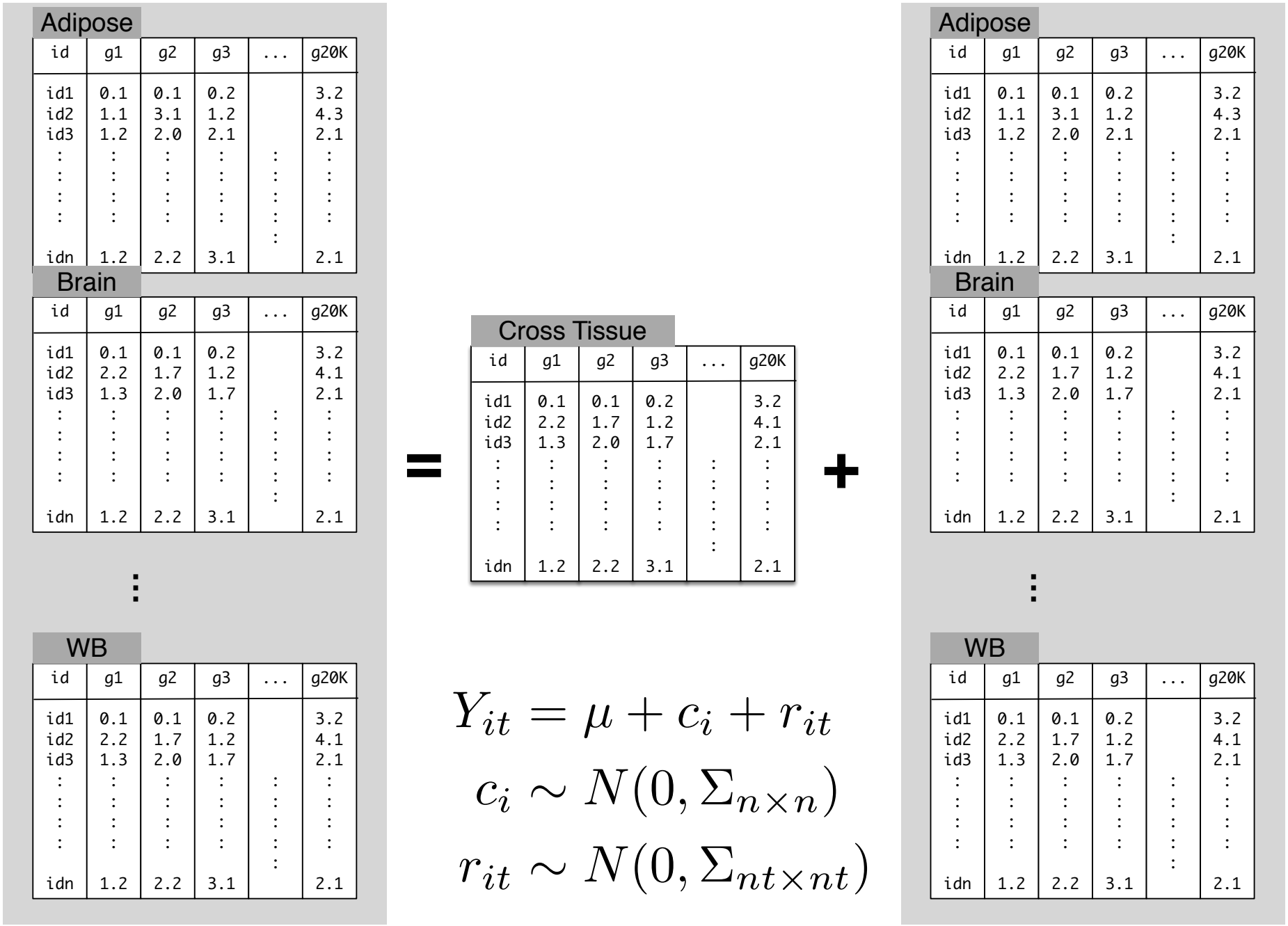
GTEx Consortium, Science 2015

Predicted Expression Performs Well Across Tissues



Gamazon, Wheeler, Shah et al NG 2015

Splitting Tissue Specific and Cross Tissue Components



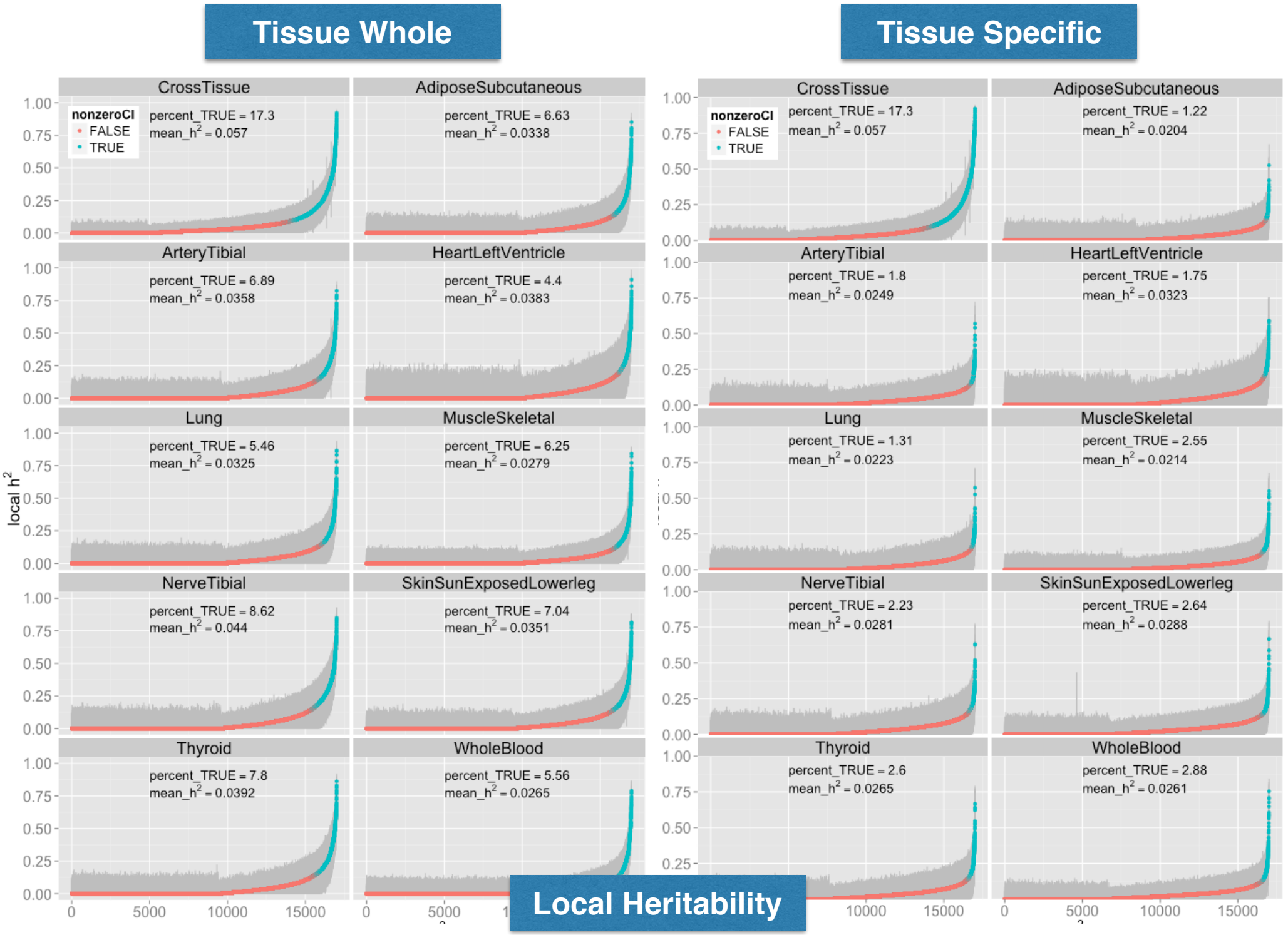
$$Y_{it} = \mu + c_i + r_{it}$$

$$c_i \sim N(0, \Sigma_{n \times n})$$

$$r_{it} \sim N(0, \Sigma_{nt \times nt})$$

Cross-Tissue Less Noisy & Larger Sample Size

Tissue	n
Cross-Tissue	**450**
Muscle-Skeletal	361
WholeBlood	339
Skin-SunExposed(Lowerleg)	303
Adipose-Subcutaneous	298
Artery-Tibial	285
Lung	279
Thyroid	279
Nerve-Tibial	256
Heart-LeftVentricle	190


Tissue Whole
Tissue Specific
CrossTissue
AdiposeSubcutaneous
ArteryTibial
HeartLeftVentricle
Lung
MuscleSkeletal
NerveTibial
SkinSunExposedLowerleg
Thyroid
WholeBlood
nonzeroCI
FALSE
TRUE
percent_TRUE = 17.3
mean_h2 = 0.057
percent_TRUE = 6.63
mean_h2 = 0.0338
percent_TRUE = 6.89
mean_h2 = 0.0358
percent_TRUE = 4.4
mean_h2 = 0.0383
percent_TRUE = 5.46
mean_h2 = 0.0325
percent_TRUE = 6.25
mean_h2 = 0.0279
percent_TRUE = 8.62
mean_h2 = 0.044
percent_TRUE = 7.04
mean_h2 = 0.0351
percent_TRUE = 7.8
mean_h2 = 0.0392
percent_TRUE = 5.56
mean_h2 = 0.0265
percent_TRUE = 17.3
mean_h2 = 0.057
percent_TRUE = 1.22
mean_h2 = 0.0204
percent_TRUE = 1.8
mean_h2 = 0.0249
percent_TRUE = 1.75
mean_h2 = 0.0323
percent_TRUE = 1.31
mean_h2 = 0.0223
percent_TRUE = 2.55
mean_h2 = 0.0214
percent_TRUE = 2.23
mean_h2 = 0.0281
percent_TRUE = 2.64
mean_h2 = 0.0288
percent_TRUE = 2.6
mean_h2 = 0.0265
percent_TRUE = 2.88
mean_h2 = 0.0261
local h2
0.00
0.25
0.50
0.75
1.00
0
5000
10000
15000
Local Heritability

Tissue Whole

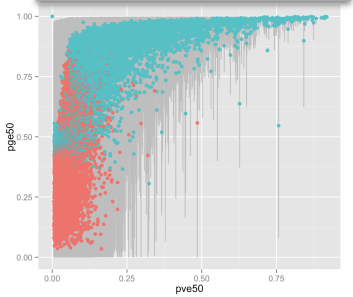
Tissue Specific

Distant Heritability with FHS cis-eQTL subset

Tissue Whole

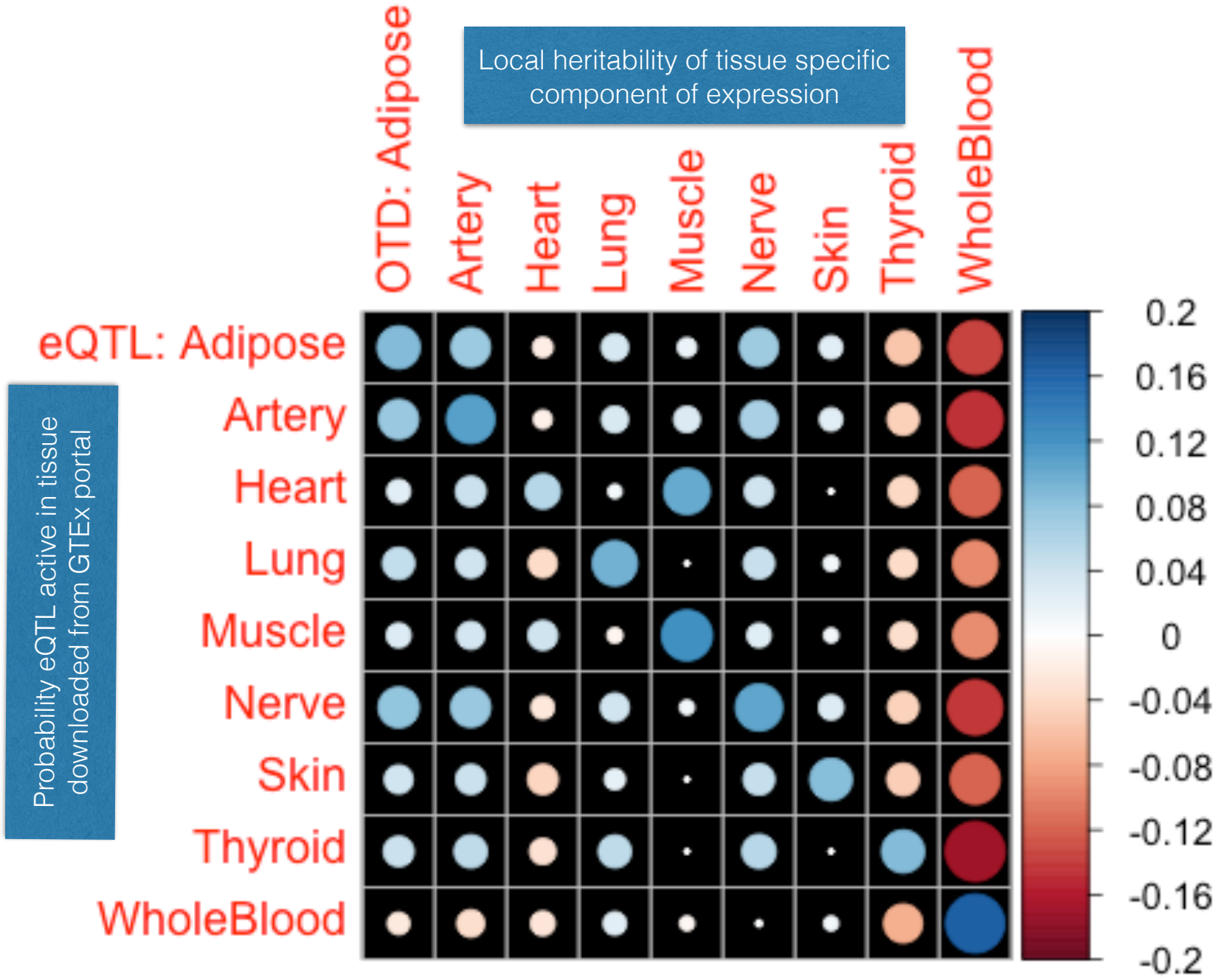
Tissue Specific

Cross Tissue

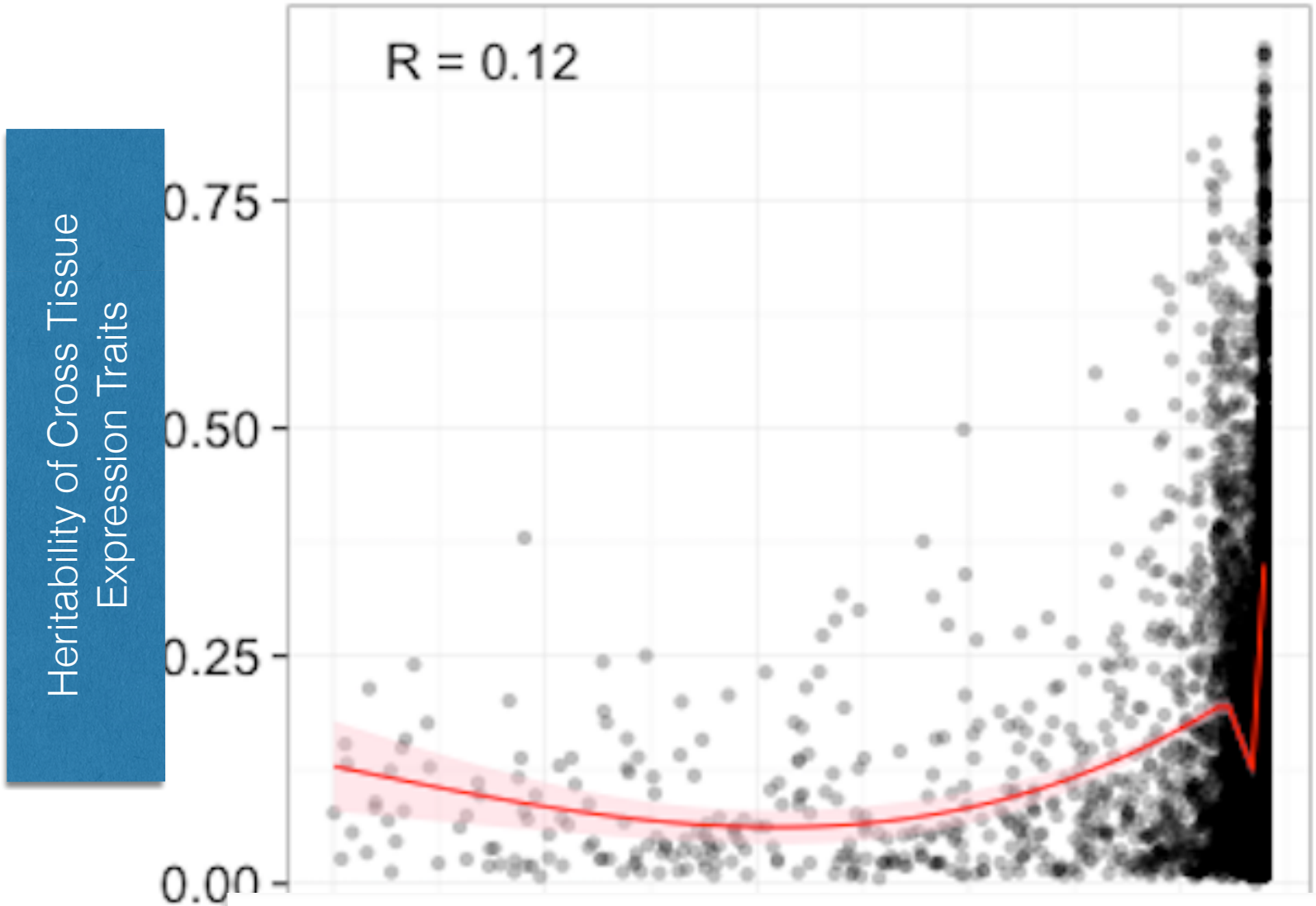


Sparsity

Correlation bw Tissue Specific h2 vs Prob eGene in Tissue

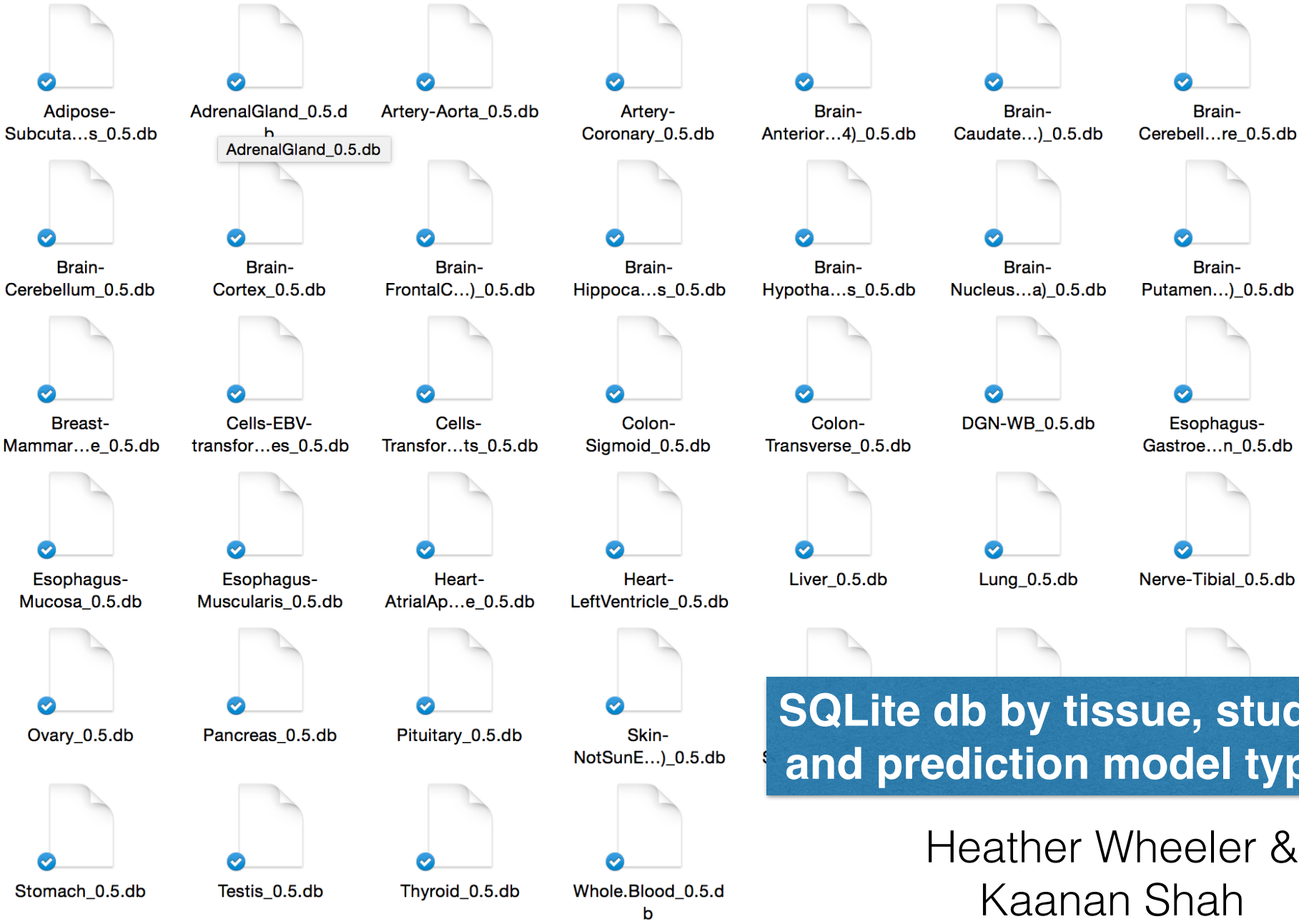


Tissue Independence Correlates with Cross Tissue h2



Measure of Tissue Independence $S_g = -\sum_t p_{t,g} \log p_{t,g}$

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



SQLite db by tissue, study, and prediction model type

Heather Wheeler &
Kaanan Shah

Summary

- Quantified local and distant heritability of gene expression traits
- Local heritability is well estimated
- Distant heritability can only be estimated using strong functional priors to reduce number of genetic markers
- For heritable genes architecture is sparse rather than polygenic
- Orthogonal Tissue Decomposition
  - Defined new phenotypes cross tissue & orthogonal tissue specific expressions
  - Found consistency with multi-tissue

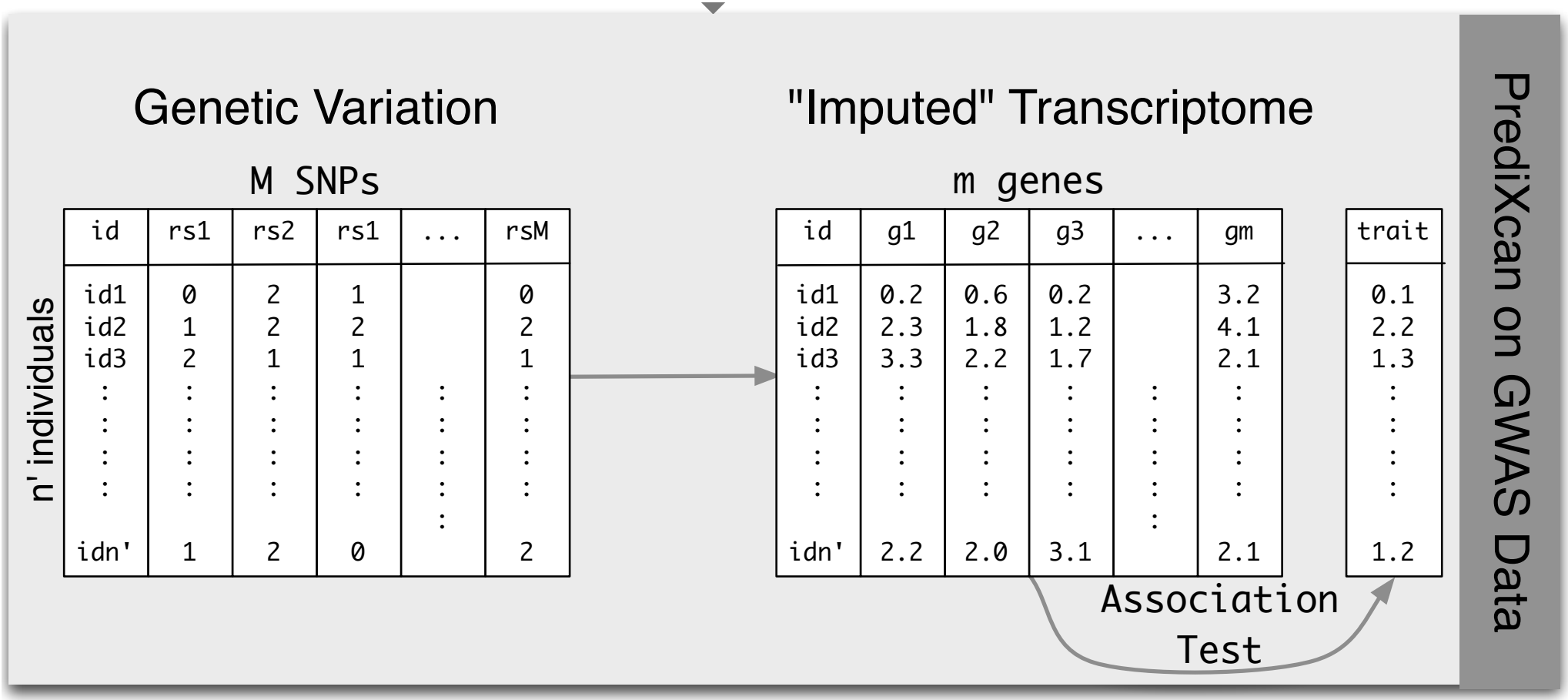
ASHG - Invited Session #13
Wednesday 11-1pm Room 316
“Secure, Efficient, and Scalable Computational Genetics via Summary Statistics”

MetaXcan

a summary statistics based gene-level association test

Alvaro Barbeira

PrediXcan Imputes Transcriptome & Tests Assoc.



MetaXcan Formula

$$\hat{\gamma}_g = \frac{\mathrm{Cov}(T_g, Y)}{\hat{\sigma}_g^2}$$

$$= \frac{\mathrm{Cov}(\sum_{l \in \mathrm{Model}_g} w_{lg} X_l, Y)}{\hat{\sigma}_g^2}$$

$$= \sum_{l \in \mathrm{Model}_g} \frac{w_{lg}\mathrm{Cov}(X_l, Y)}{\hat{\sigma}_g^2}$$

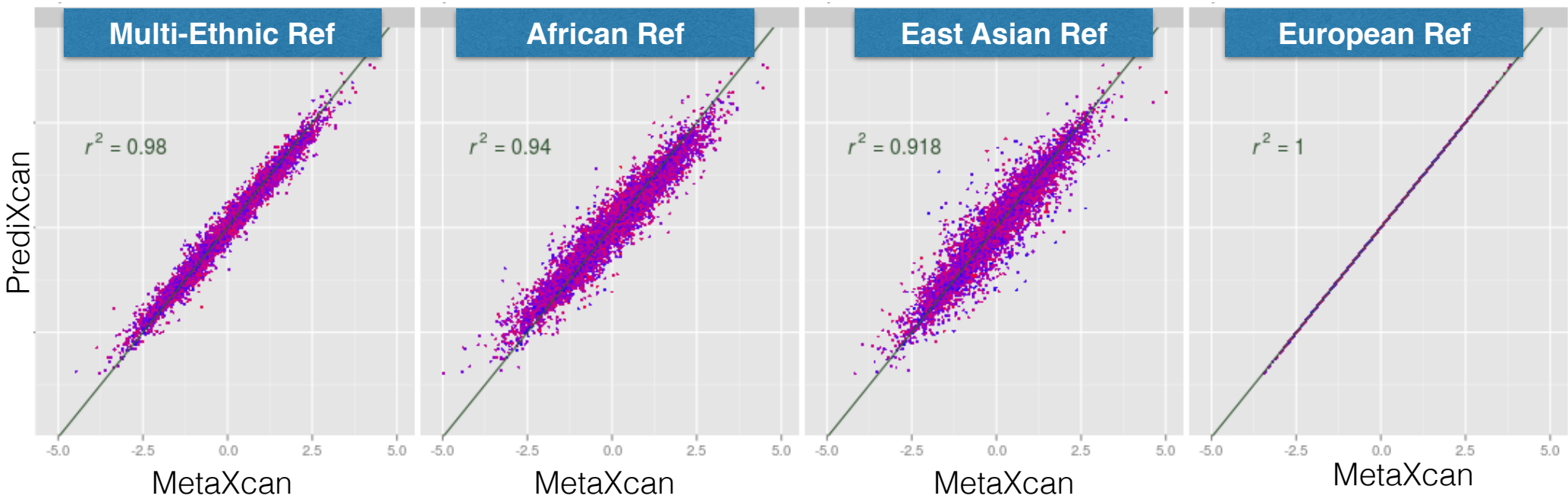
$$= \sum_{l \in \mathrm{Model}_g} \frac{w_{lg}\hat{\beta}_l \sigma_l^2}{\hat{\sigma}_g^2}$$

MetaXcan Formula

$$Z_g = \sum_{l \in \text{Model}_g} w_{lg} \frac{\sigma_l}{\hat{\sigma}_g} \frac{\hat{\beta}_l}{\text{se}(\hat{\beta}_l)} \sqrt{\frac{1 - R_l^2}{1 - R_g^2}}$$

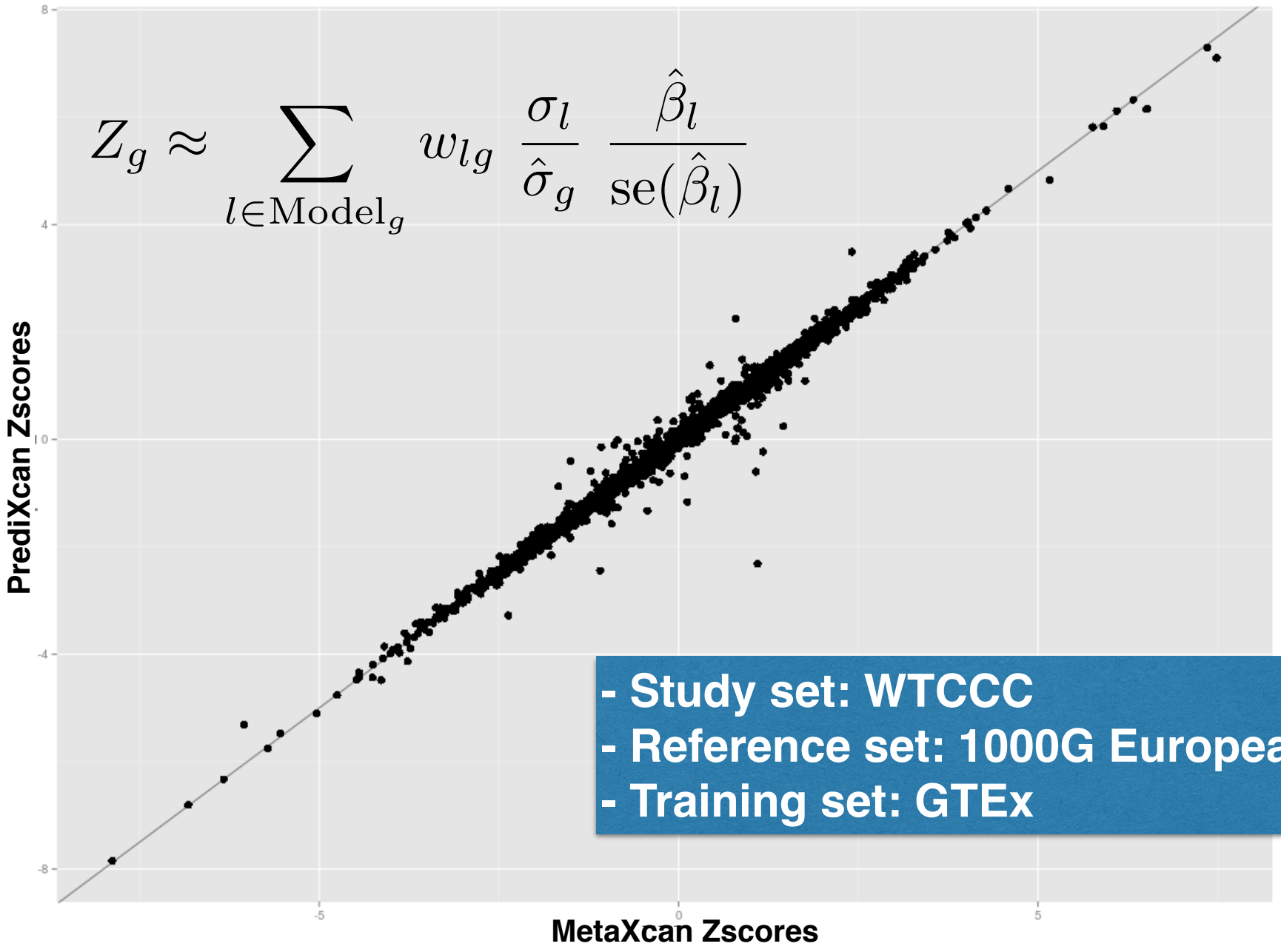
Robustness to Reference Population Differences

PrediXcan vs. MetaXcan
Study set: European
Training set: GTEx



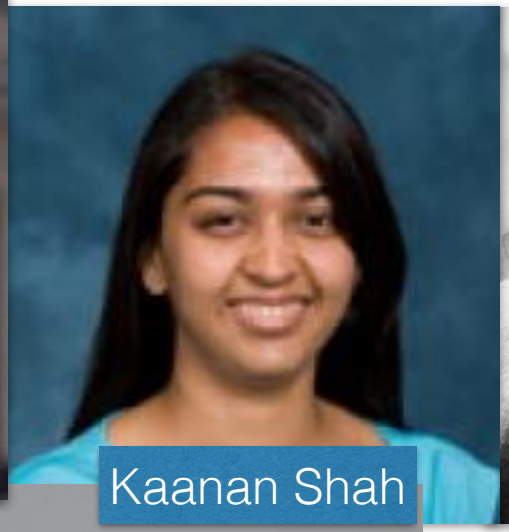
WTCCC T1D PrediXcan vs MetaXcan

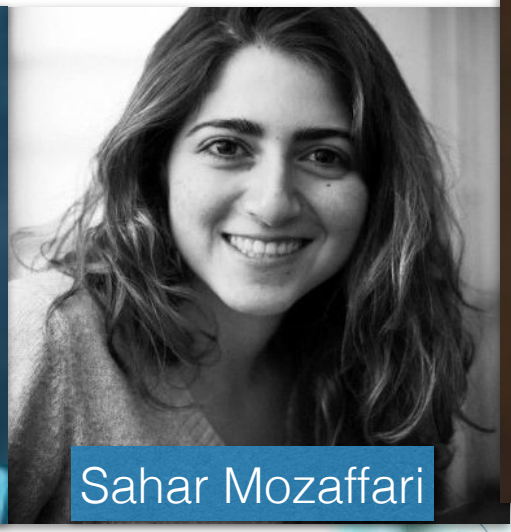
$$Z_g \approx \sum_{l \in \text{Model}_g} w_{lg} \frac{\sigma_l}{\hat{\sigma}_g} \frac{\hat{\beta}_l}{\text{se}(\hat{\beta}_l)}$$



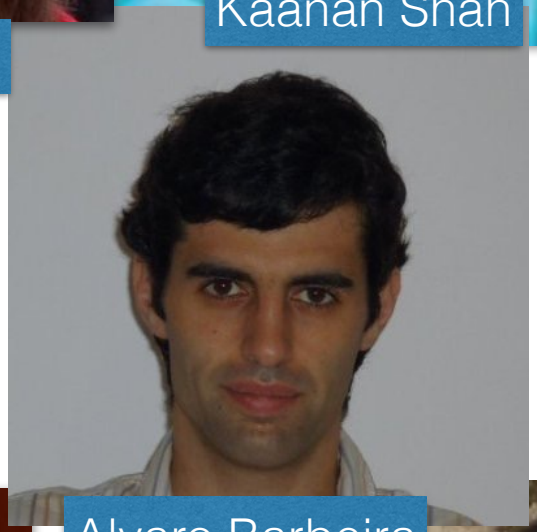
- Study set: WTCCC
- Reference set: 1000G European
- Training set: GTEx

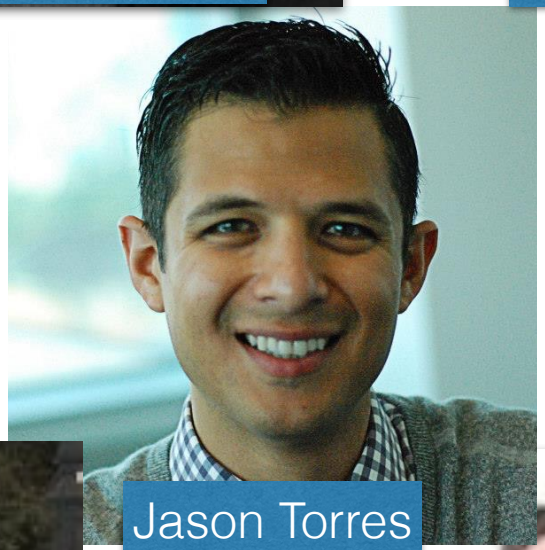

Heather Wheeler


Kaanan Shah


Sahar Mozaffari


Nancy Cox


Alvaro Barbeira


Jason Torres

Eric Gamazon


Dan Nicolae


Keston Aquino-Michaels

Thank You

Contributors

- **Heather Wheeler**
- **Alvaro Barbeira**
- Nancy Cox
- Kaanan P. Shah
- Eric Gamazon
- Dan Nicolae
- Keston Aquino Michaels
- Sahar Mozaffari
- Nicholas Knoblauch
- GTEx Consortium
- Josh Denny, Robert Carrol, Anne Eyler

Data sources

- WTCCC
- NIMH: Depression Genes & Networks

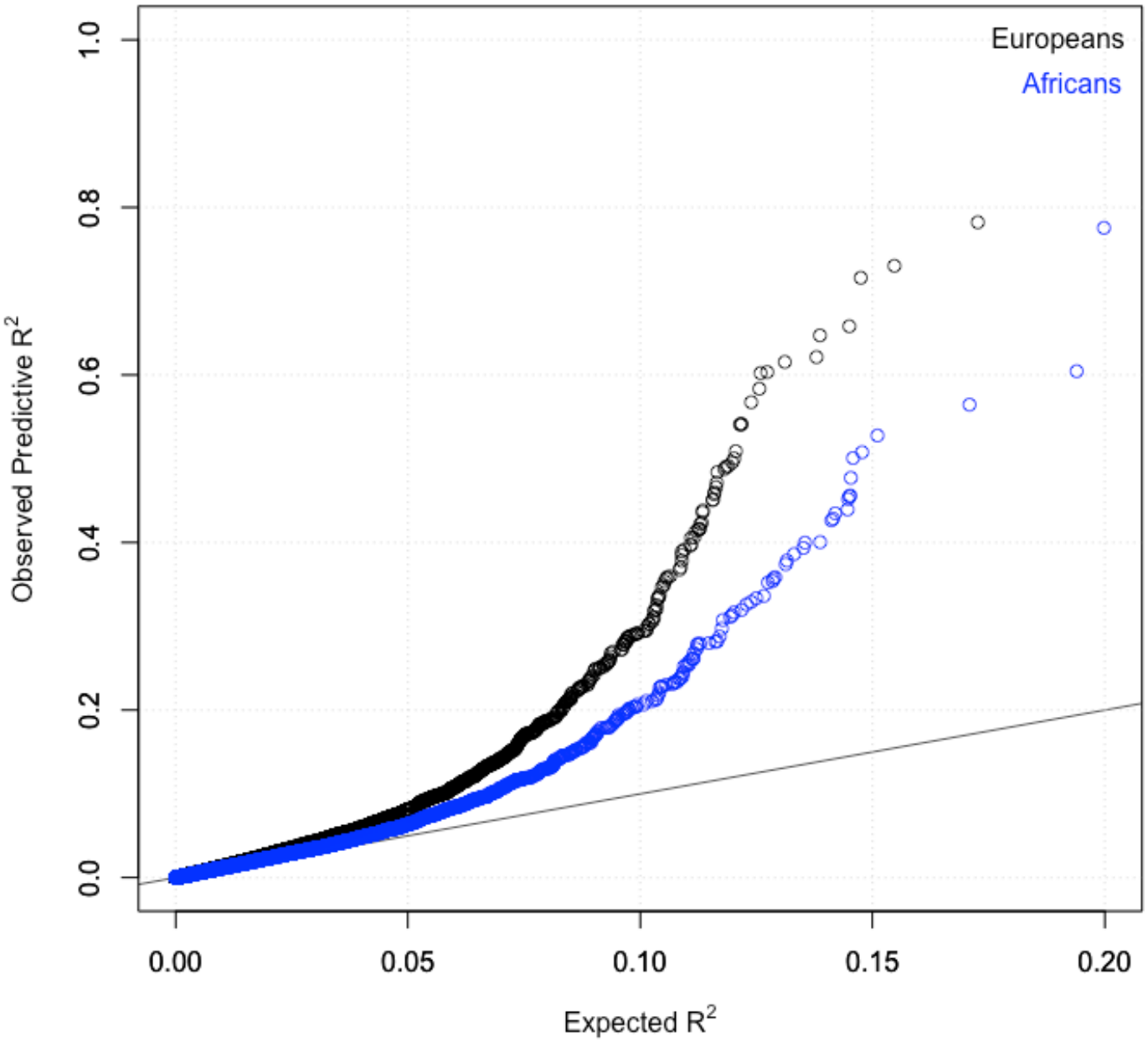
Funding

- 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

PrediXcan vs. MetaXcan - T1D WTCCC

$$Z_g = \sum_{l \in \text{Model}_g} w_{lg} \frac{\sigma_l}{\hat{\sigma}_g} \frac{\hat{\beta}_l}{\text{se}(\hat{\beta}_l)} \sqrt{\frac{1 - R_l^2}{1 - R_g^2}}$$

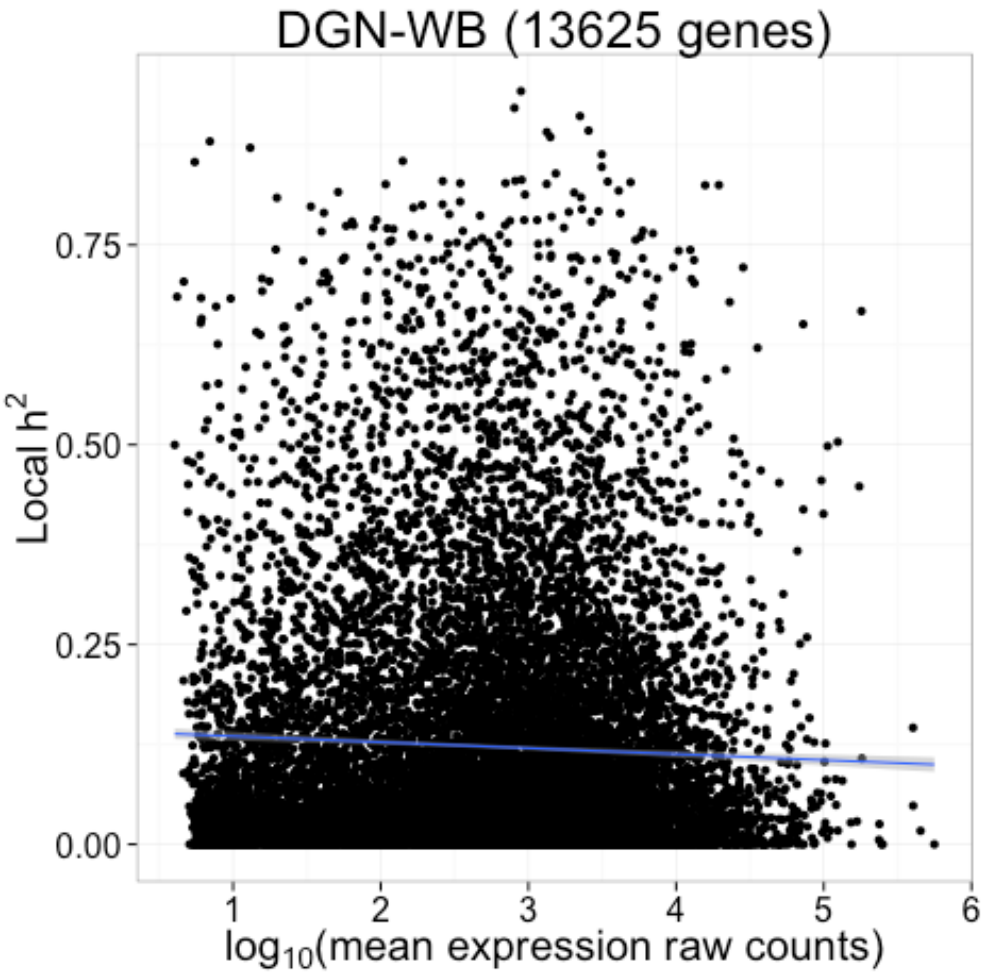
Cross Population Prediction Performance

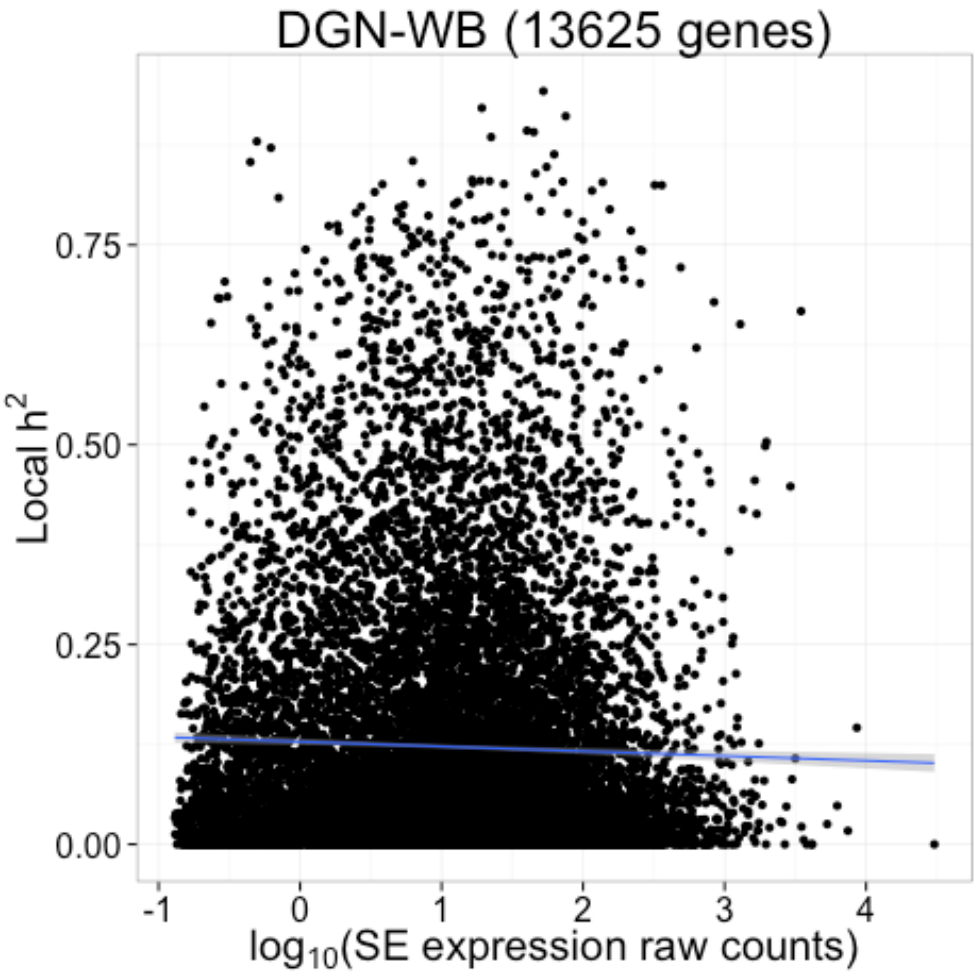


MetaXcan Formula

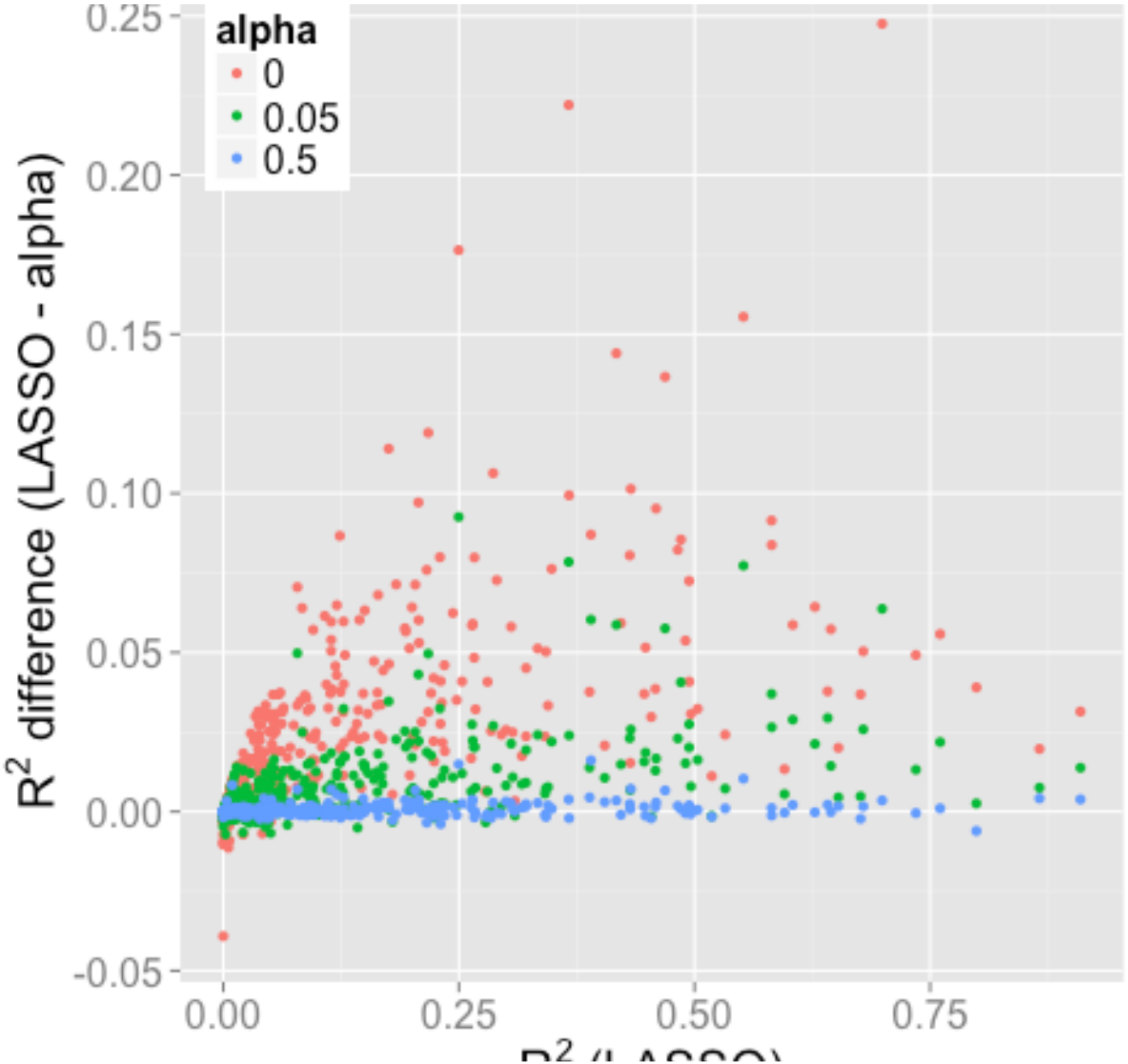
$$Z_g = \sum_{l \in \text{Model}_g} w_{lg} \frac{\sigma_l}{\sqrt{\mathbf{W}_g' \Gamma_g \mathbf{W}_g}} \frac{\hat{\beta}_l}{\text{se}(\hat{\beta}_l)} \sqrt{\frac{1 - R_l^2}{1 - R_g^2}}$$

Heritability vs. Expression Level

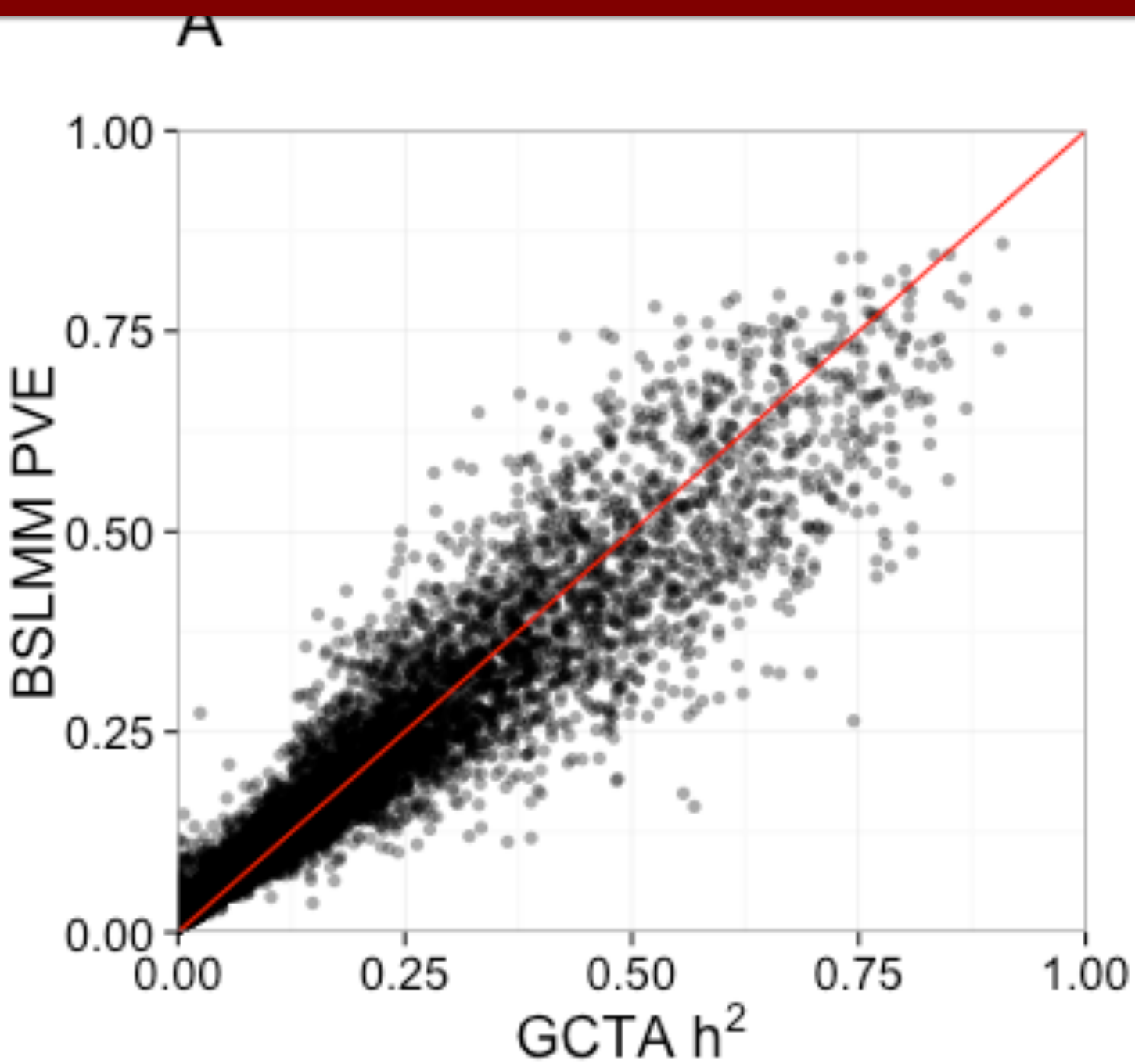




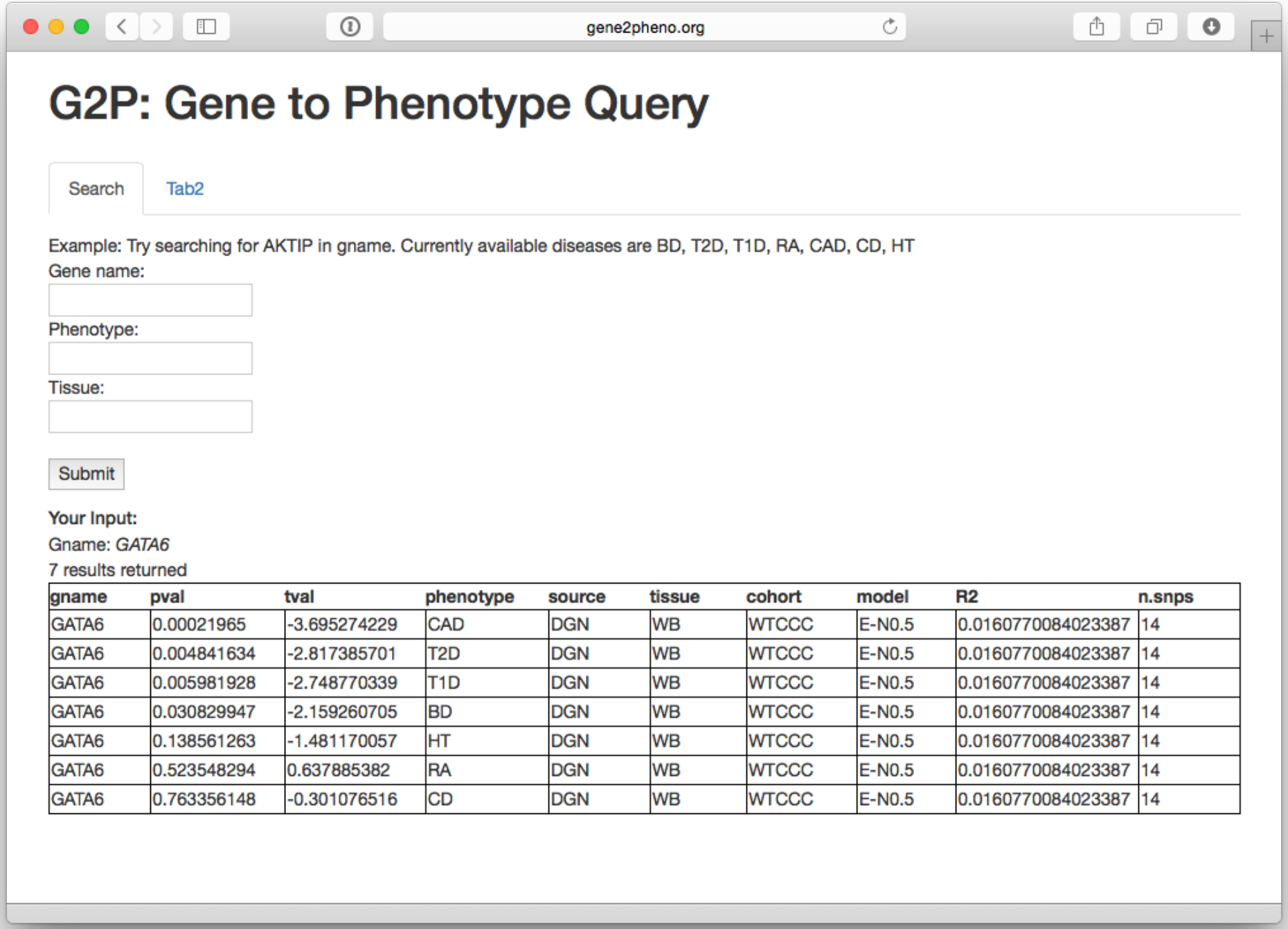
Sparse Models Outperform Polygenic Models



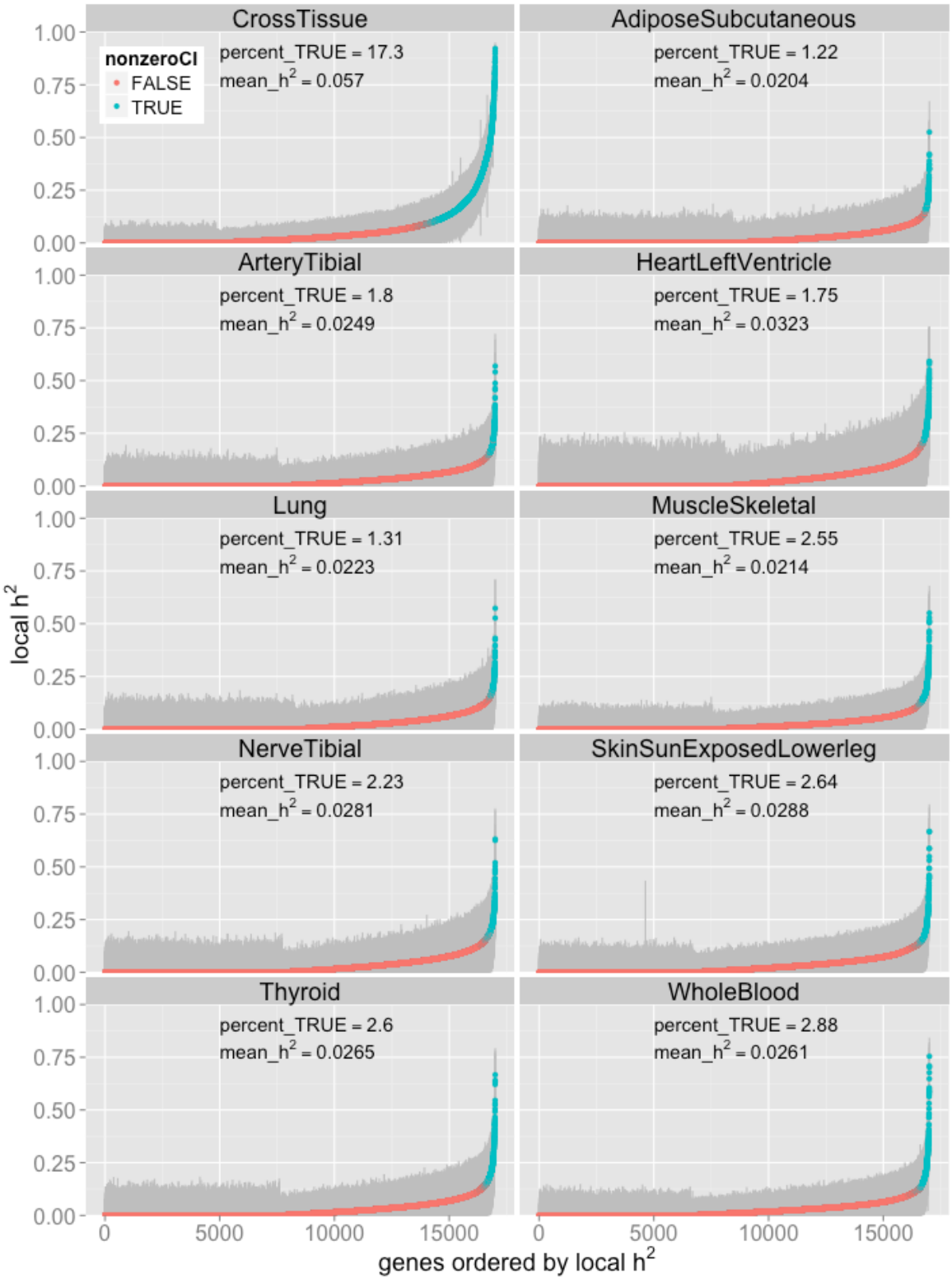
Similar Estimates of h2 with BSLMM and GCTA



PrediXcan Results WTCCC - Gene2Pheno.org



gname	pval	tval	phenotype	source	tissue	cohort	model	R2	n.snps
GATA6	0.00021965	-3.695274229	CAD	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14
GATA6	0.004841634	-2.817385701	T2D	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14
GATA6	0.005981928	-2.748770339	T1D	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14
GATA6	0.030829947	-2.159260705	BD	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14
GATA6	0.138561263	-1.481170057	HT	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14
GATA6	0.523548294	0.637885382	RA	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14
GATA6	0.763356148	-0.301076516	CD	DGN	WB	WTCCC	E-N0.5	0.0160770084023387	14


CrossTissue
percent_TRUE = 17.3
mean_h2 = 0.057
nonzeroCI
FALSE
TRUE
AdiposeSubcutaneous
percent_TRUE = 1.22
mean_h2 = 0.0204
ArteryTibial
percent_TRUE = 1.8
mean_h2 = 0.0249
HeartLeftVentricle
percent_TRUE = 1.75
mean_h2 = 0.0323
Lung
percent_TRUE = 1.31
mean_h2 = 0.0223
MuscleSkeletal
percent_TRUE = 2.55
mean_h2 = 0.0214
NerveTibial
percent_TRUE = 2.23
mean_h2 = 0.0281
SkinSunExposedLowerleg
percent_TRUE = 2.64
mean_h2 = 0.0288
Thyroid
percent_TRUE = 2.6
mean_h2 = 0.0265
WholeBlood
percent_TRUE = 2.88
mean_h2 = 0.0261
local h2
1.00
0.75
0.50
0.25
0.00
0
5000
10000
15000
genes ordered by local h2