

Phenotypic Consequences of Gene Expression Regulation in 40 Human Tissues

Hae Kyung Im, PhD



THE UNIVERSITY OF
CHICAGO

April 29, 2016

Acknowledgement

- Alvaro Barbeira
- Jason Torres
- Jiamao Zheng
- Scott Dickinson
- Kaanan Shah
- Heather Wheeler
- Tzintzuni Garcia

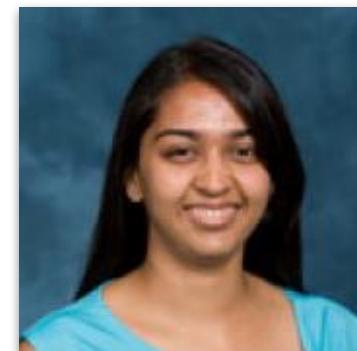


- Nancy Cox
- Dan Nicolae
- Graeme Bell



- Funding
 - R01MH107666
 - P30DK020595

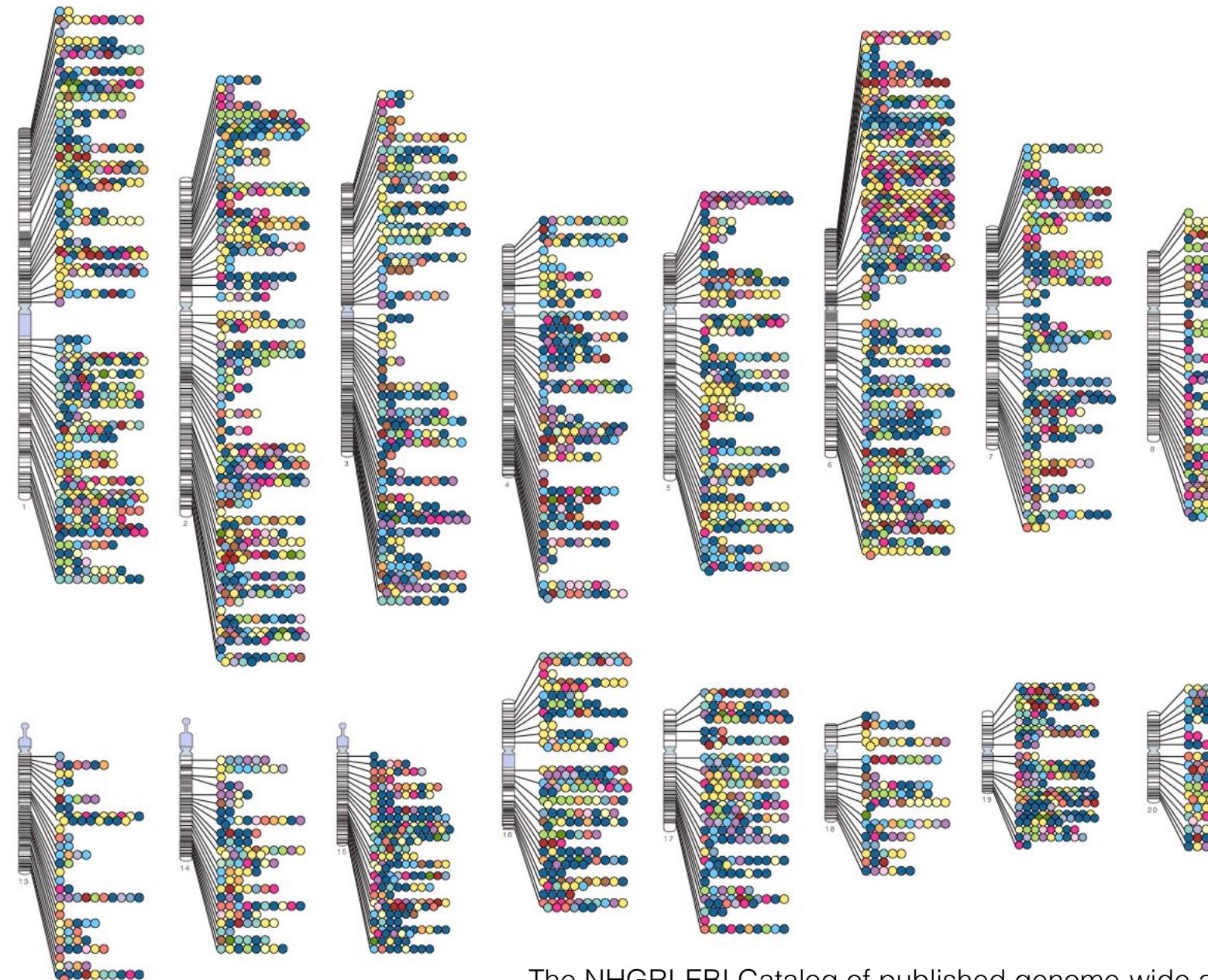
No conflicts of interests



Huge Progress in Genome Technology

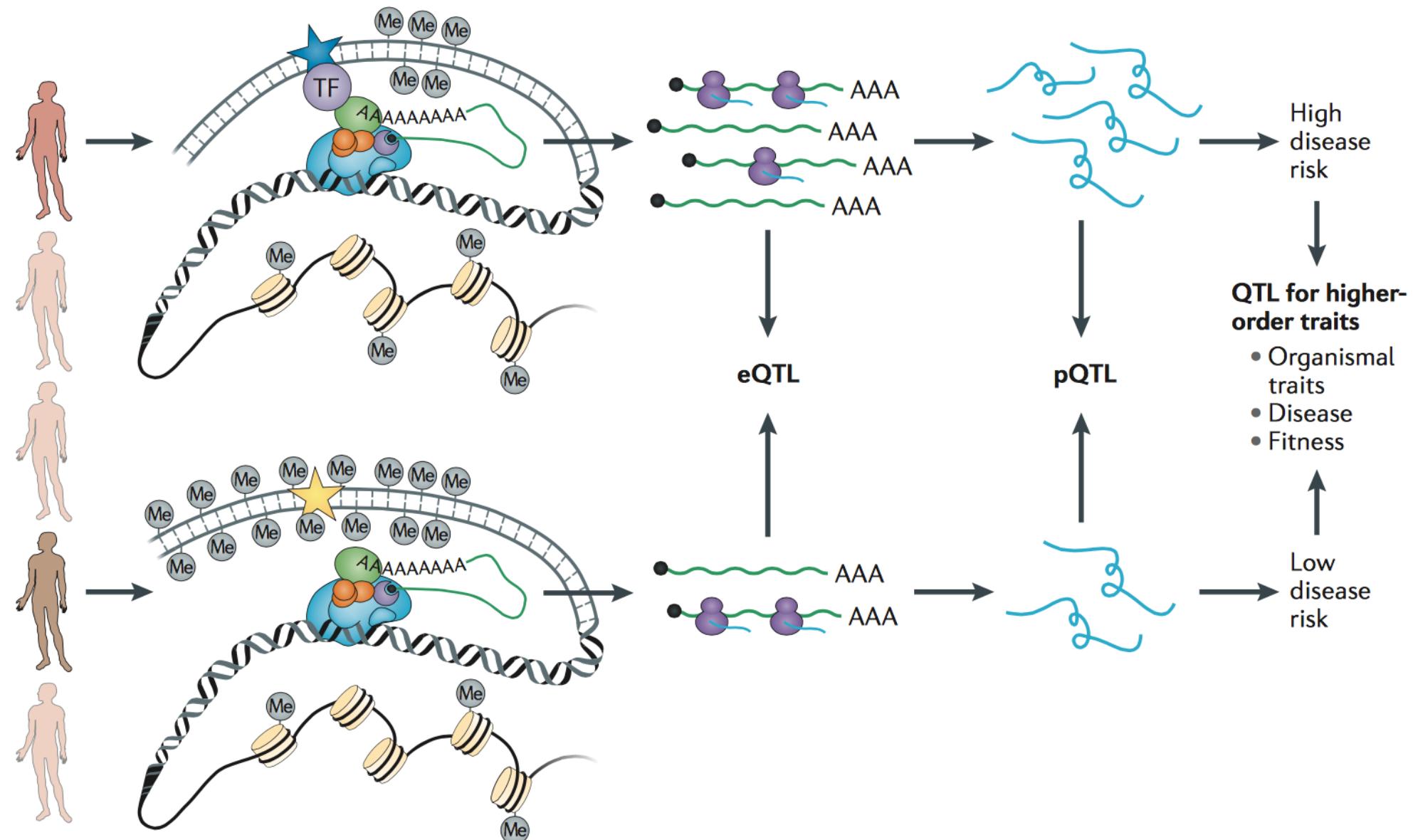
- Unprecedented advances in high throughput technology
- Translation to human health less successful
- Genome wide association studies successful in identifying variants
- Mechanism not well understood

Most GWAS Hits are in Noncoding Regions



The NHGRI-EBI Catalog of published genome-wide association studies

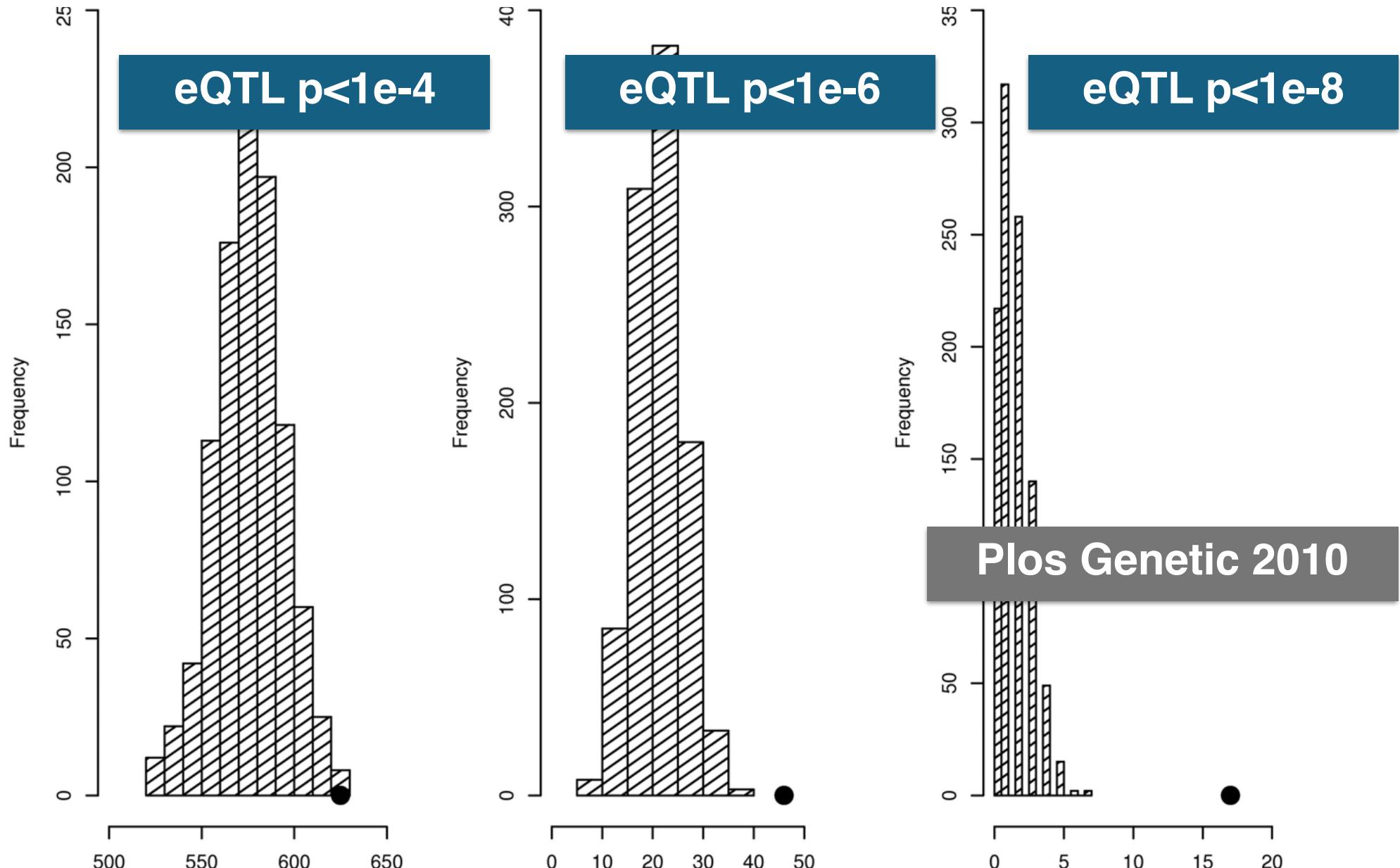
Altered Protein Levels Influences Disease Risk



Albert & Kruglyak 2015 NGReviews

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

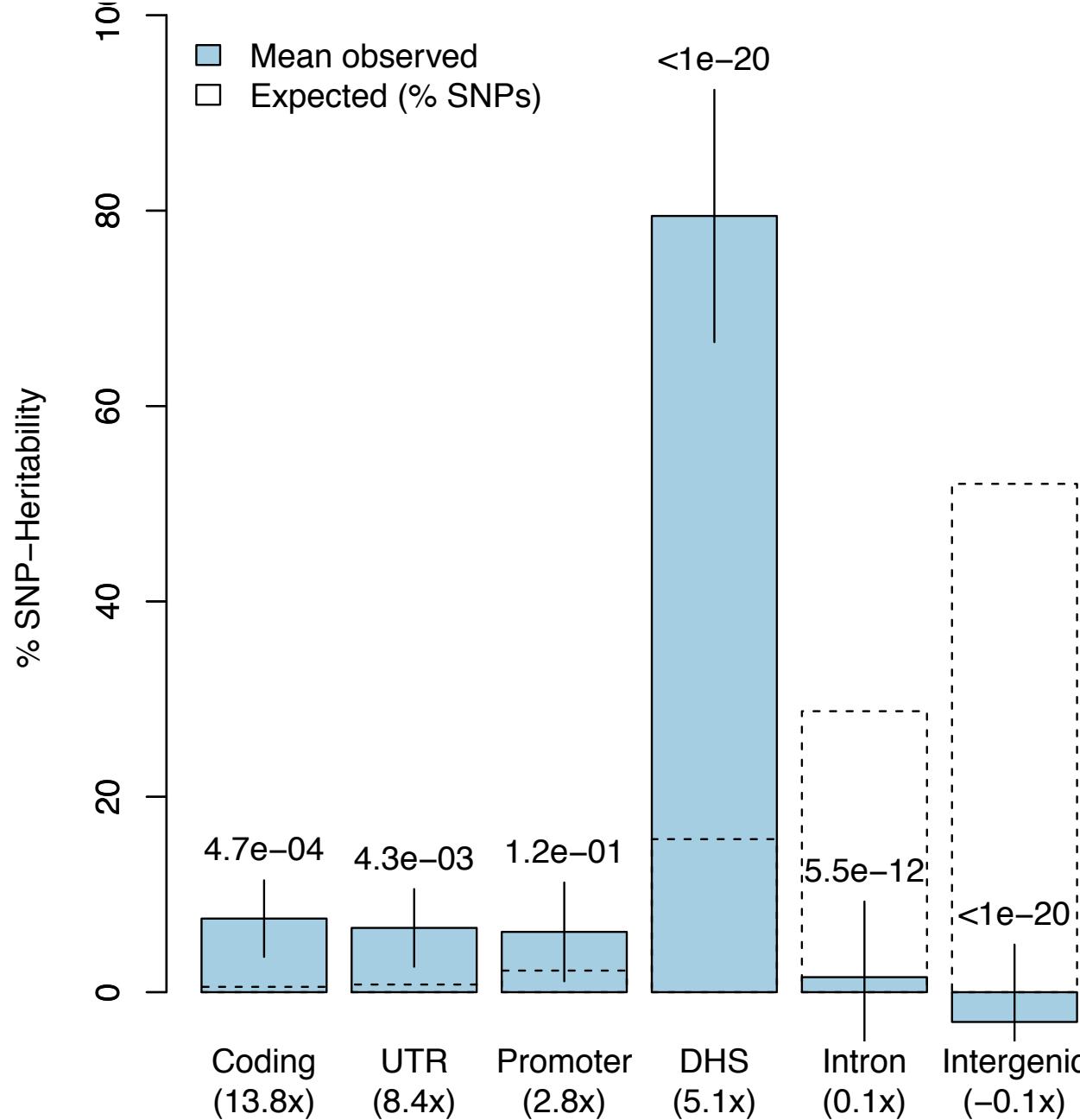
Dan L. Nicolae^{1,2,3}, Eric Gamazon¹, Wei Zhang¹, Shiwei Duan^{1*}, M. Eileen Dolan^{1,2}, Nancy J. Cox¹,



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

AJHG 2014

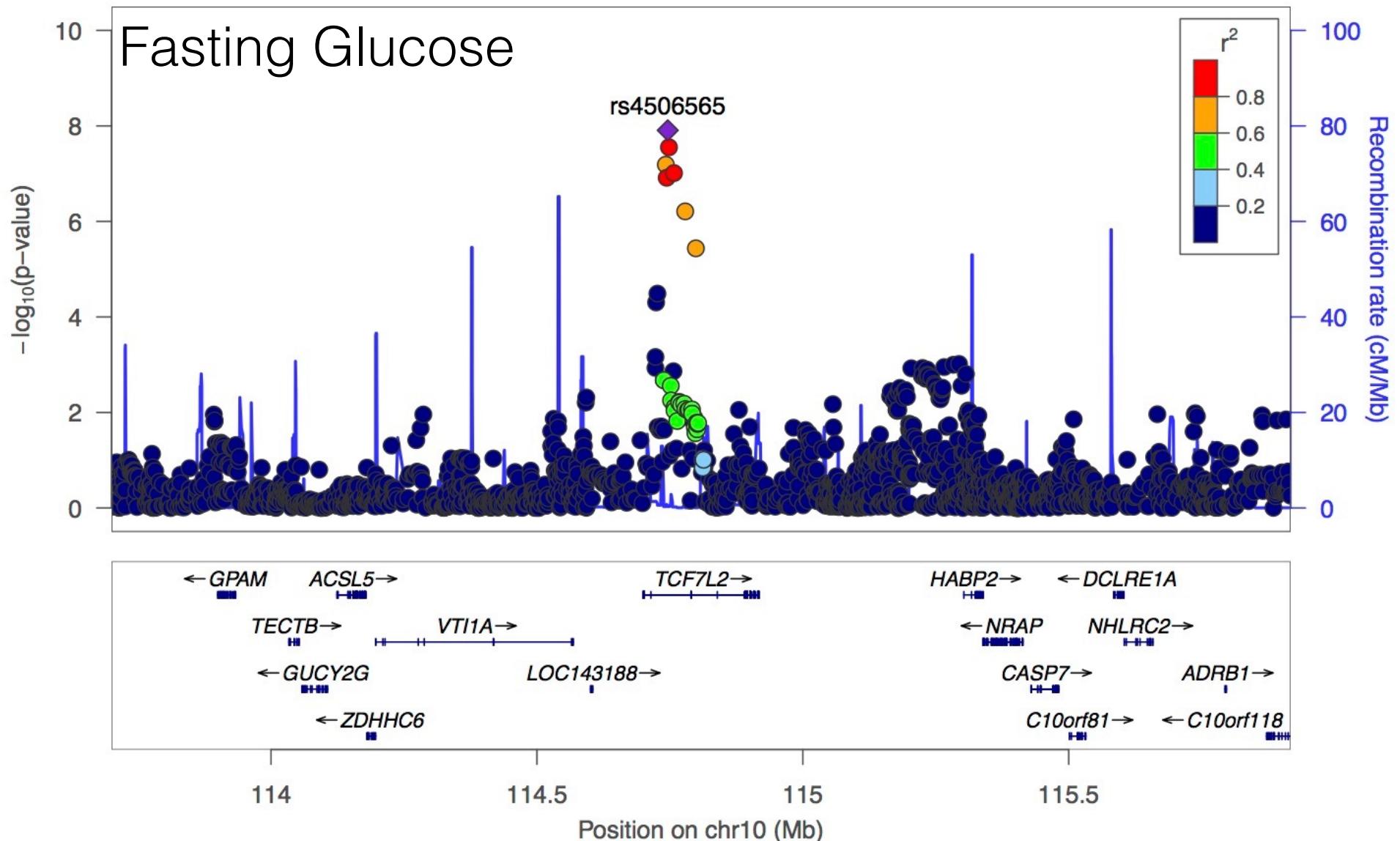
Alexander Gusev, S Hong Lee, Benjamin M Neale, et al.



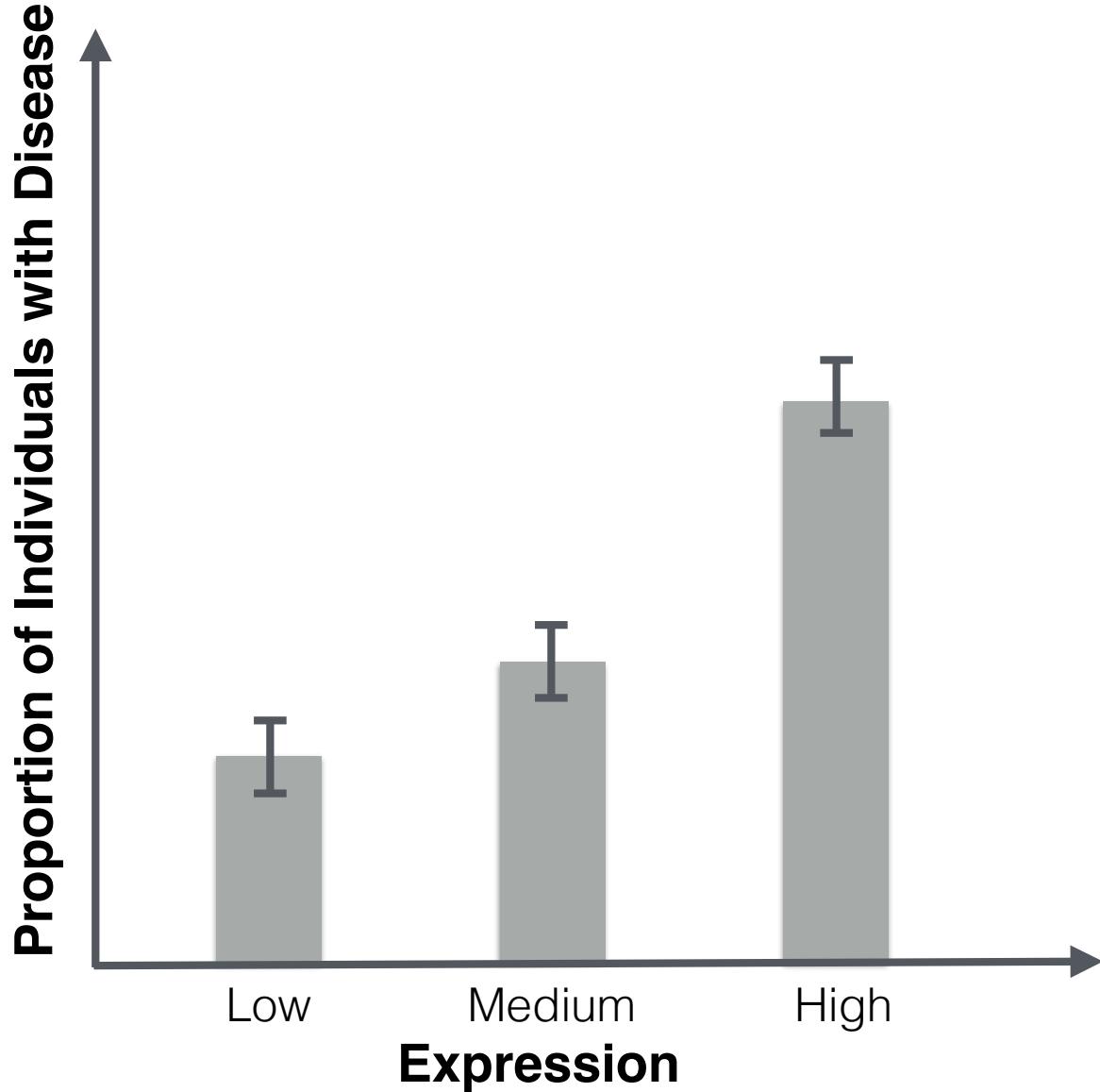
DHS: DNase hypersensitivity sites, control accessibility of the region thus levels of transcription

What Are Phenotypic Consequences of GWAS Loci

Plotted SNPs

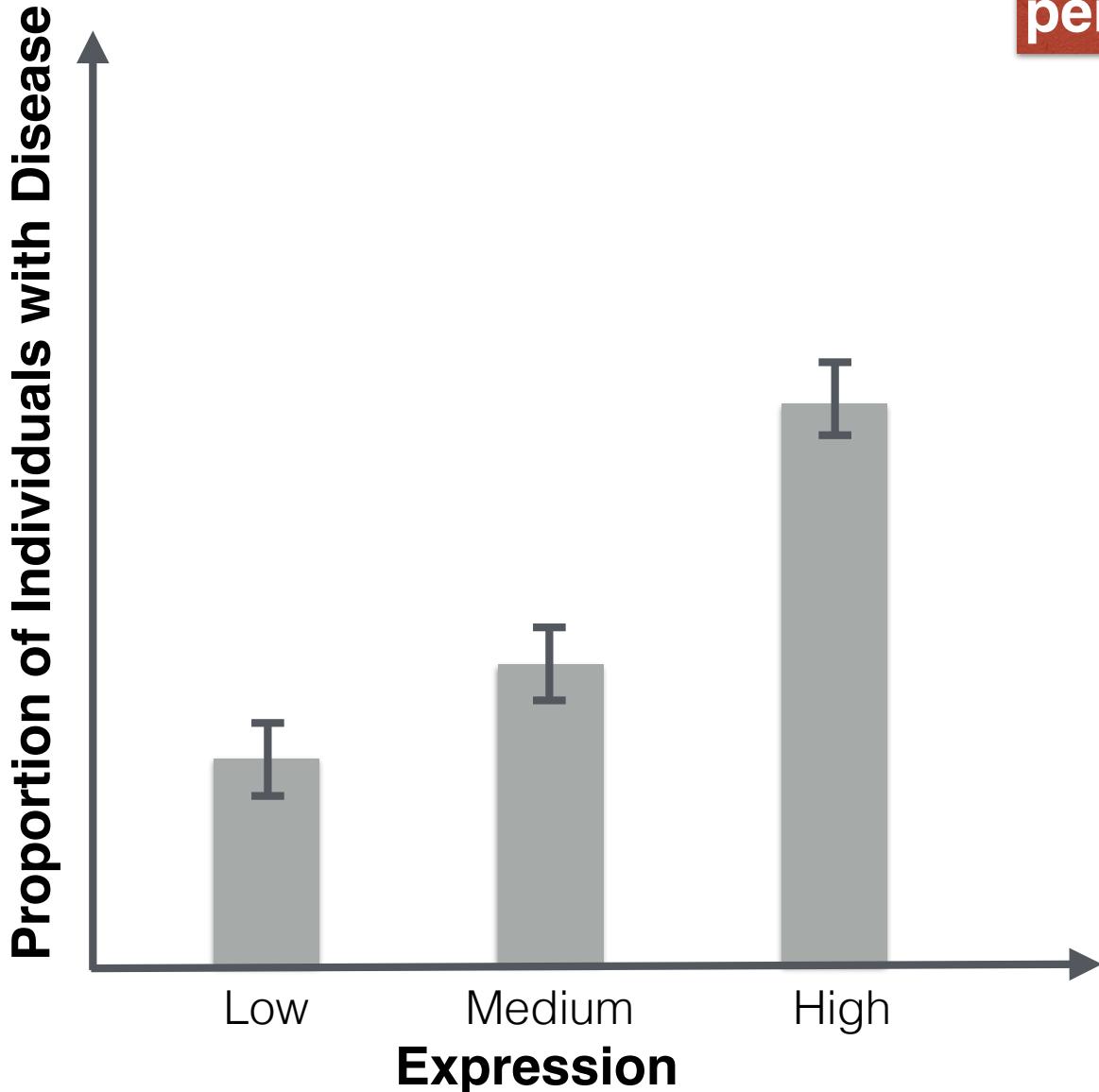


Measuring Phenotypic Consequence of Regulated Expression

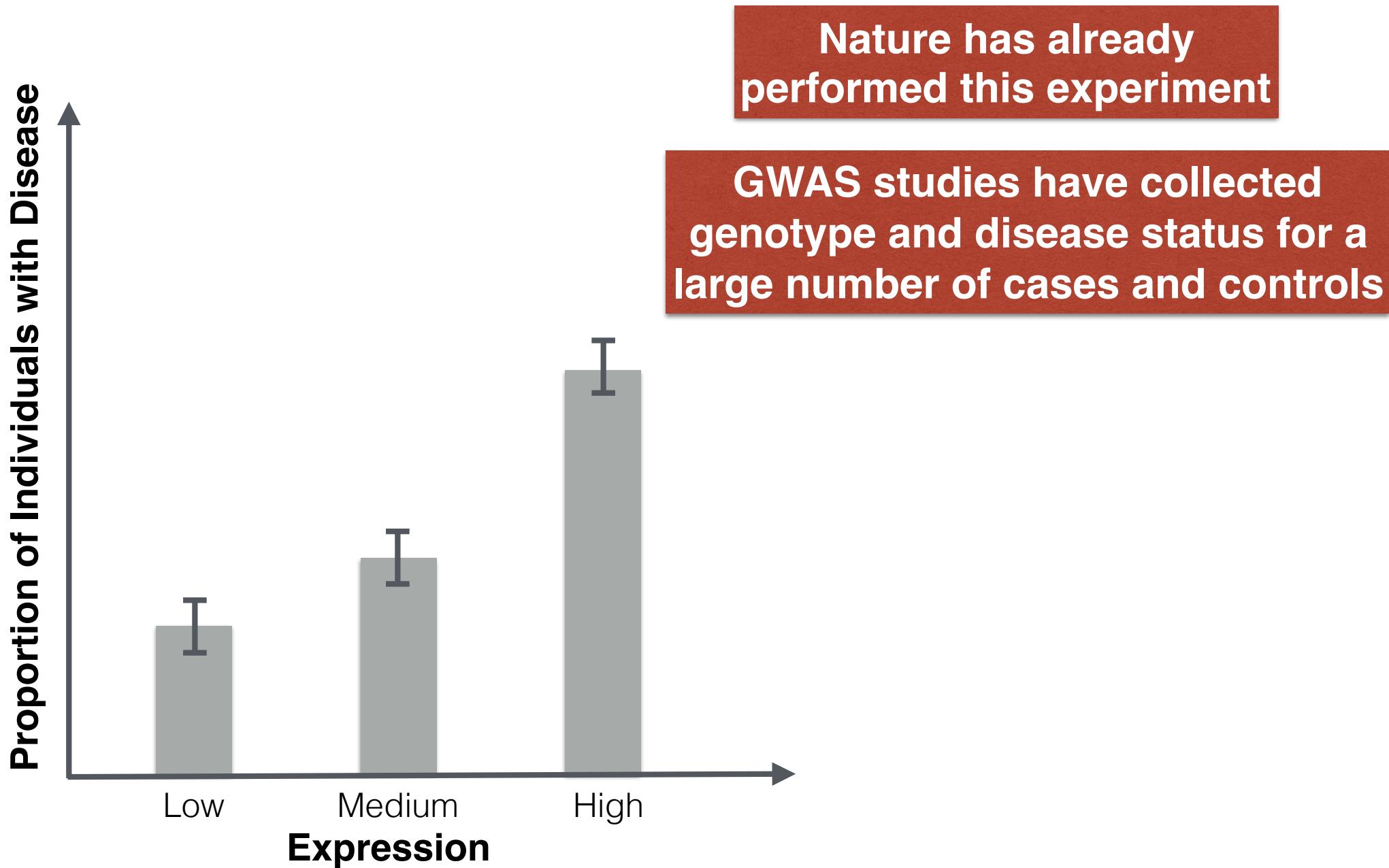


Measuring Phenotypic Consequence of Regulated Expression

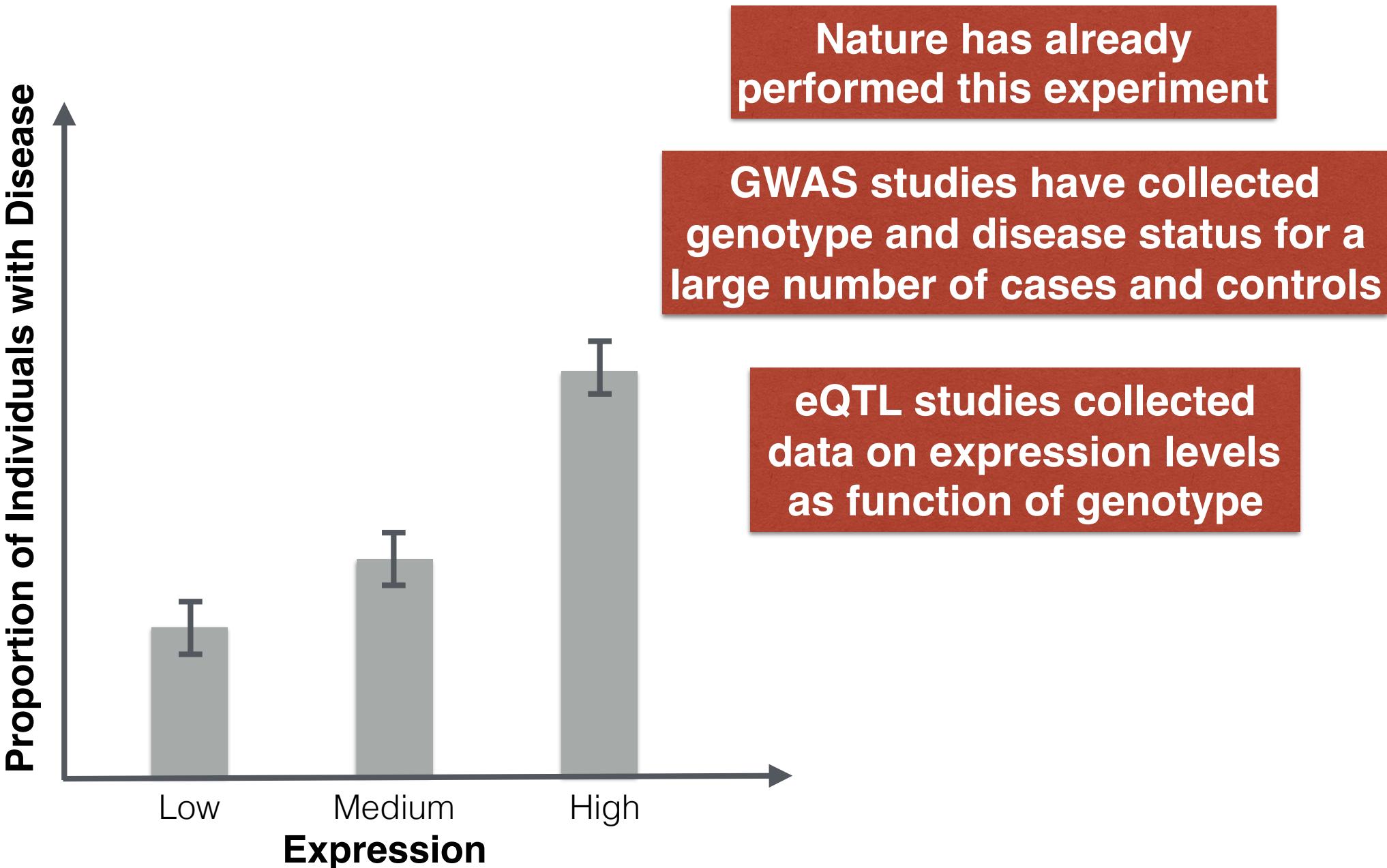
Nature has already performed this experiment



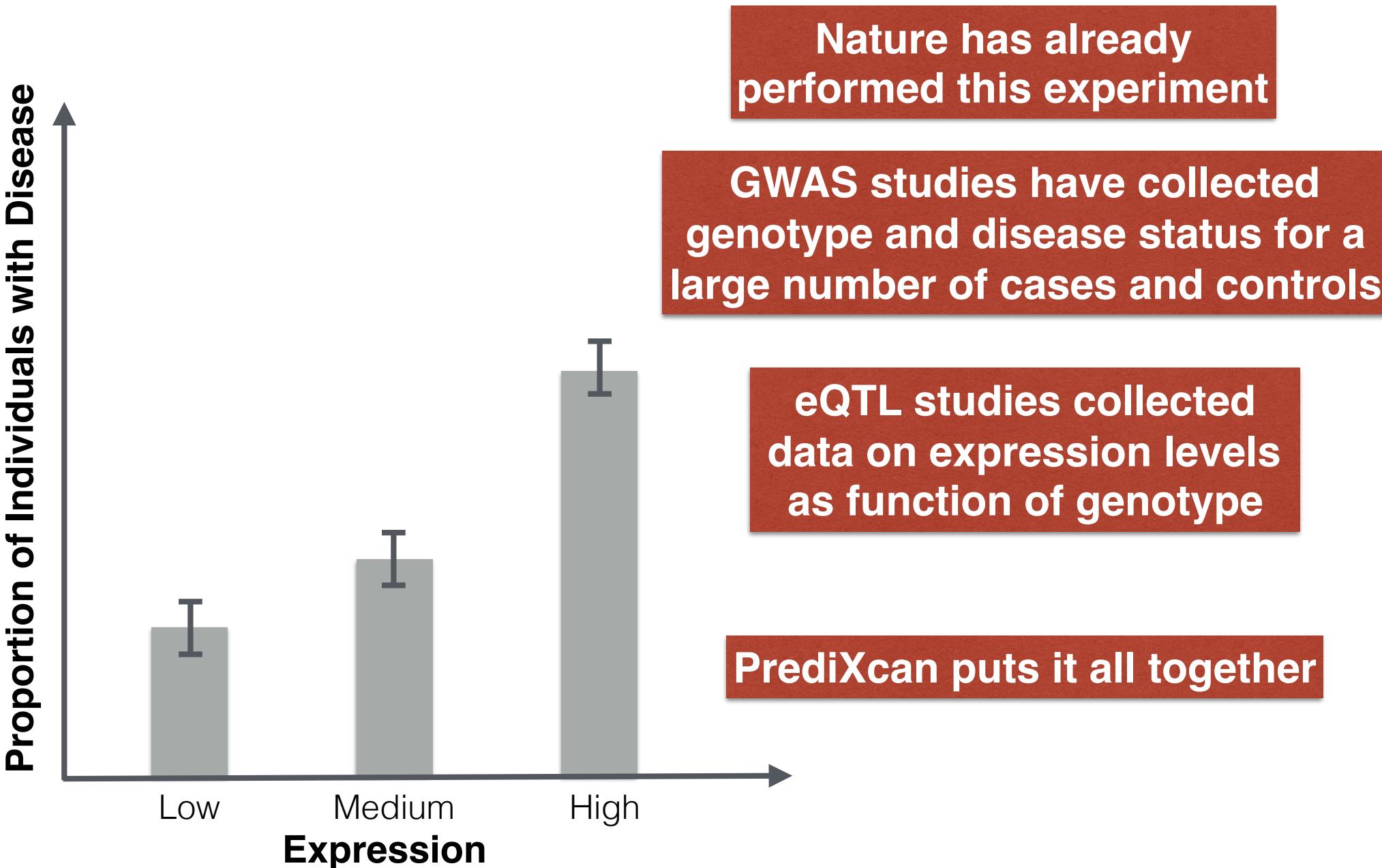
Measuring Phenotypic Consequence of Regulated Expression



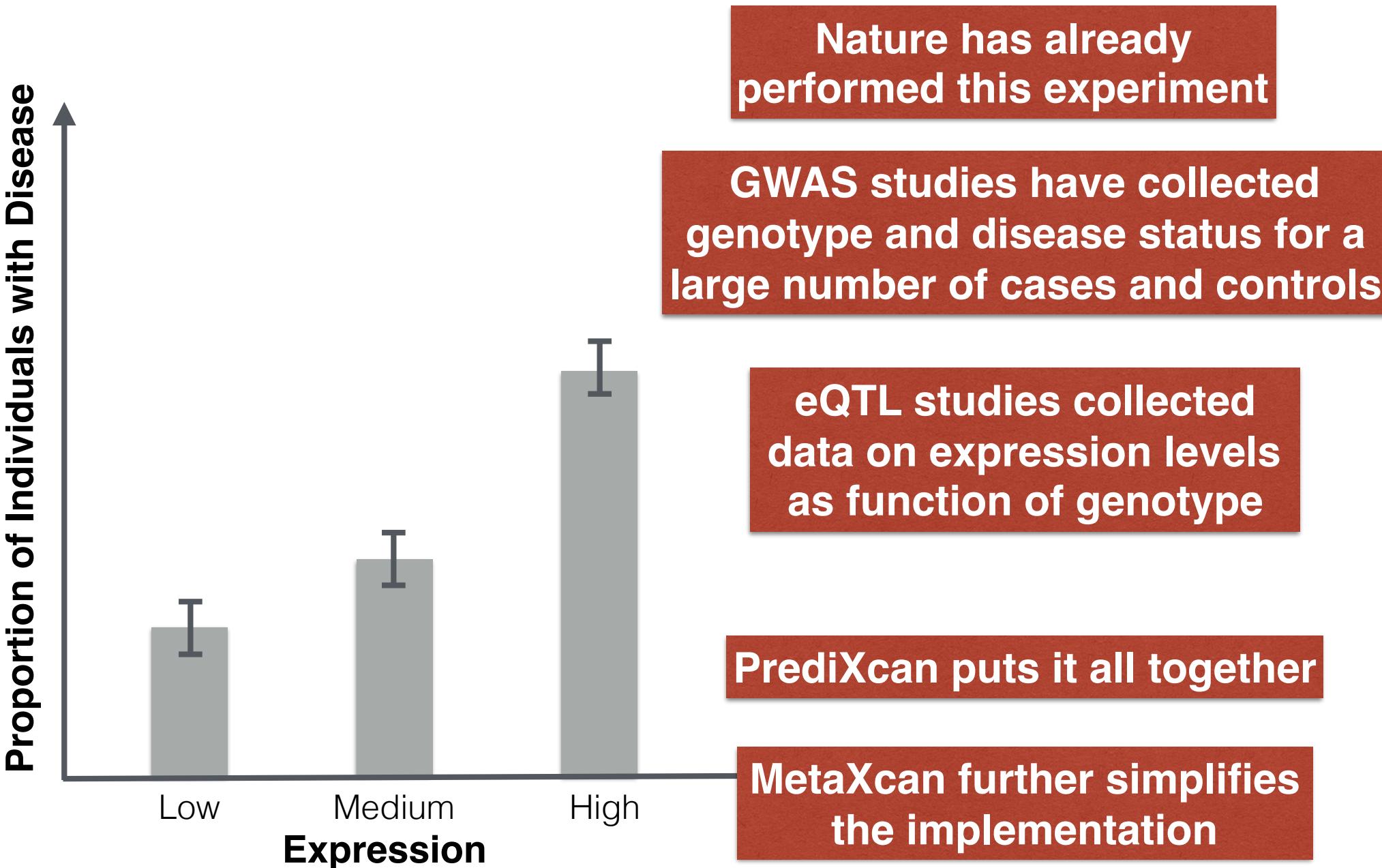
Measuring Phenotypic Consequence of Regulated Expression



Measuring Phenotypic Consequence of Regulated Expression



Measuring Phenotypic Consequence of Regulated Expression

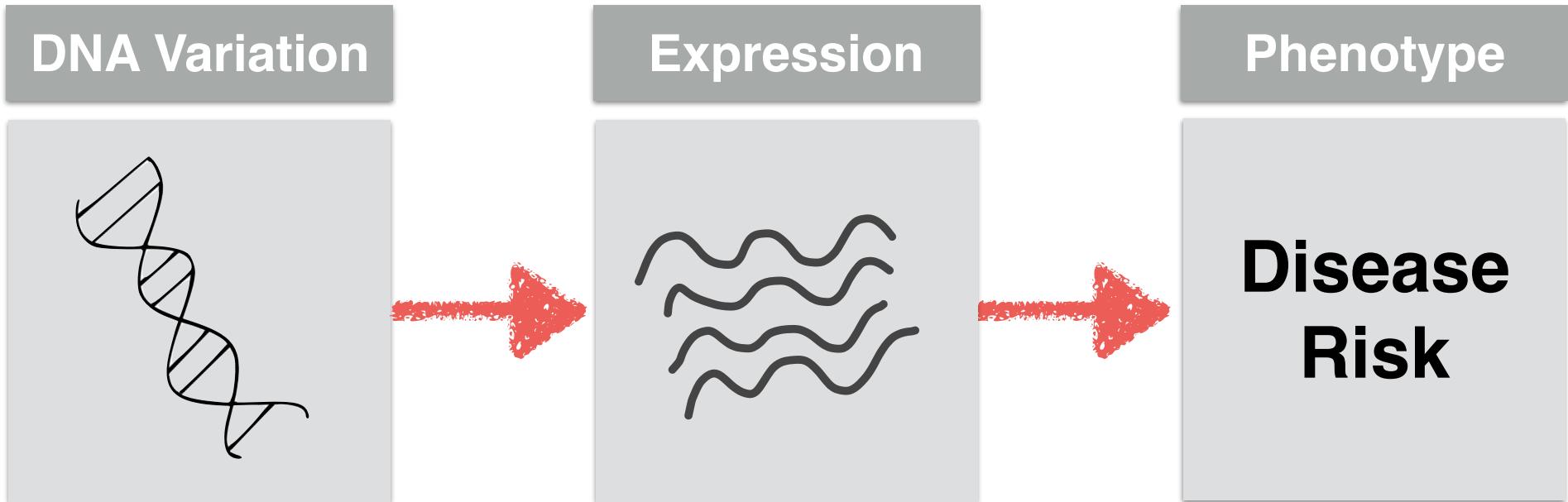


PrediXcan

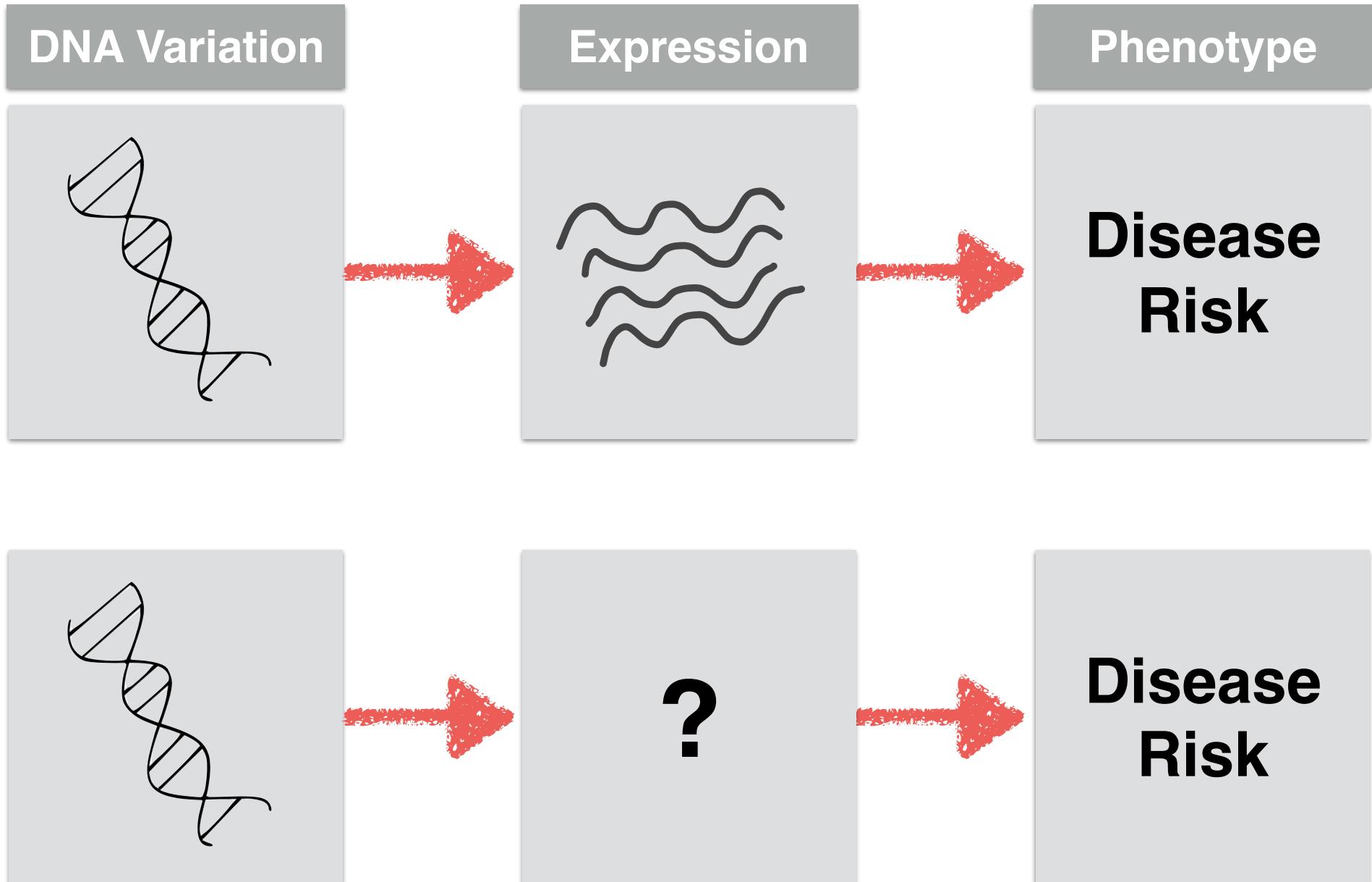


Nature
Genetics
2015

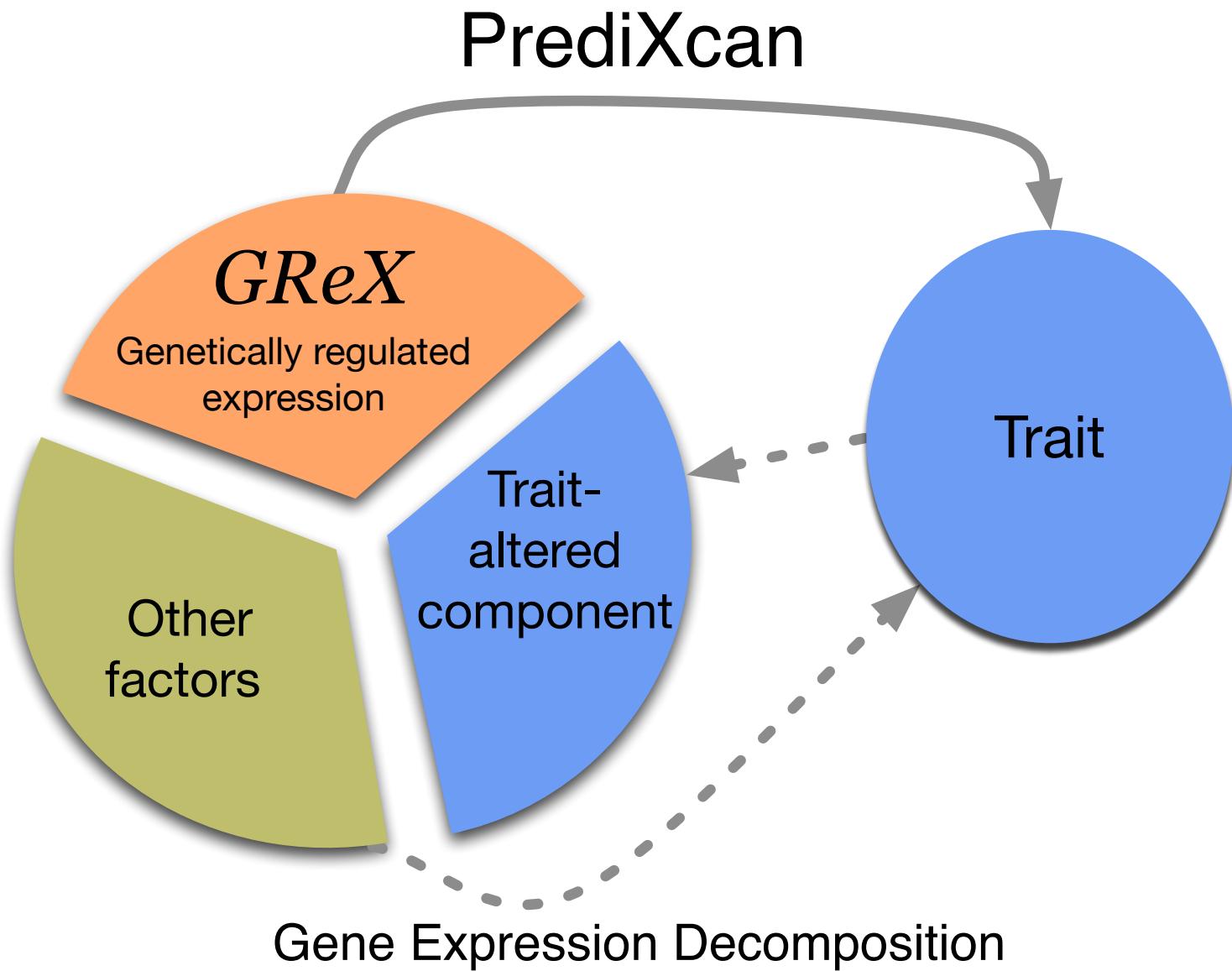
PrediXcan: Gene-Level Association Test



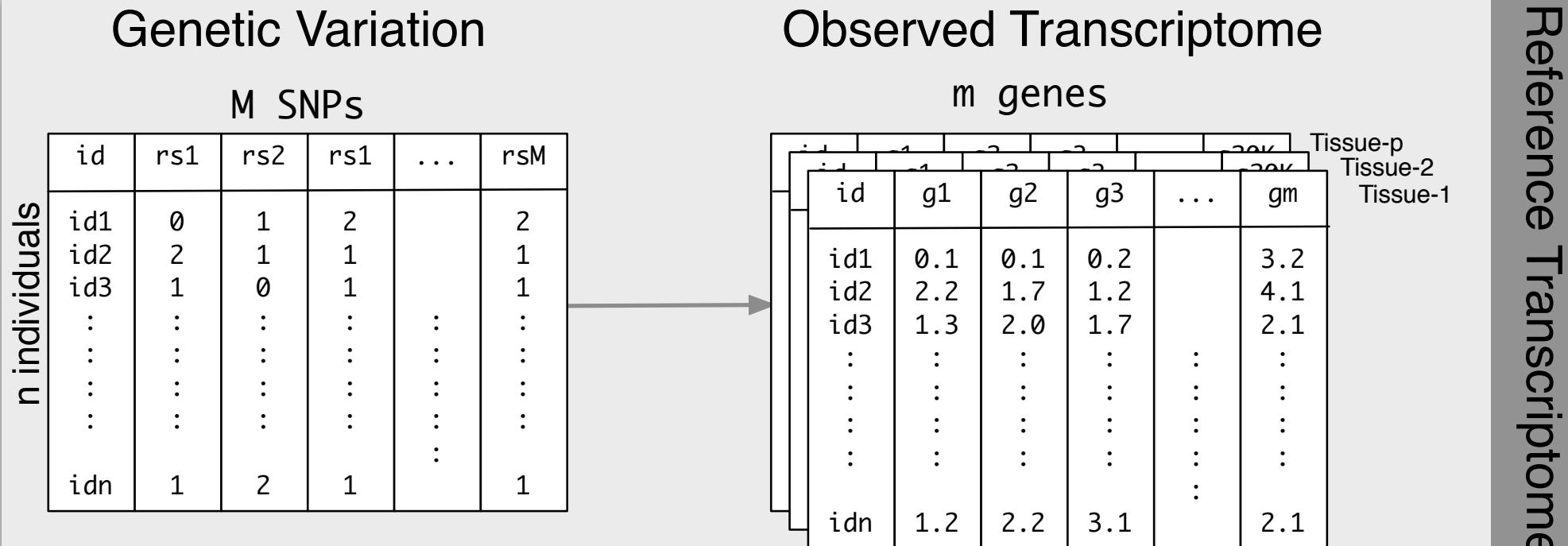
PrediXcan: Gene-Level Association Test



Mechanisms Tested by PrediXcan



PrediXcan uses Reference Transcriptome to Train Prediction Models



PredictDB Database of Weights for Prediction

PredictDB: Database
of Prediction Models



		M SNPs							
		rs1	rs2	rs3	...	rsM			
m genes	g1	w11	w12	w13		w1M	Tissue-p	Tissue-2	Tissue-1
	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

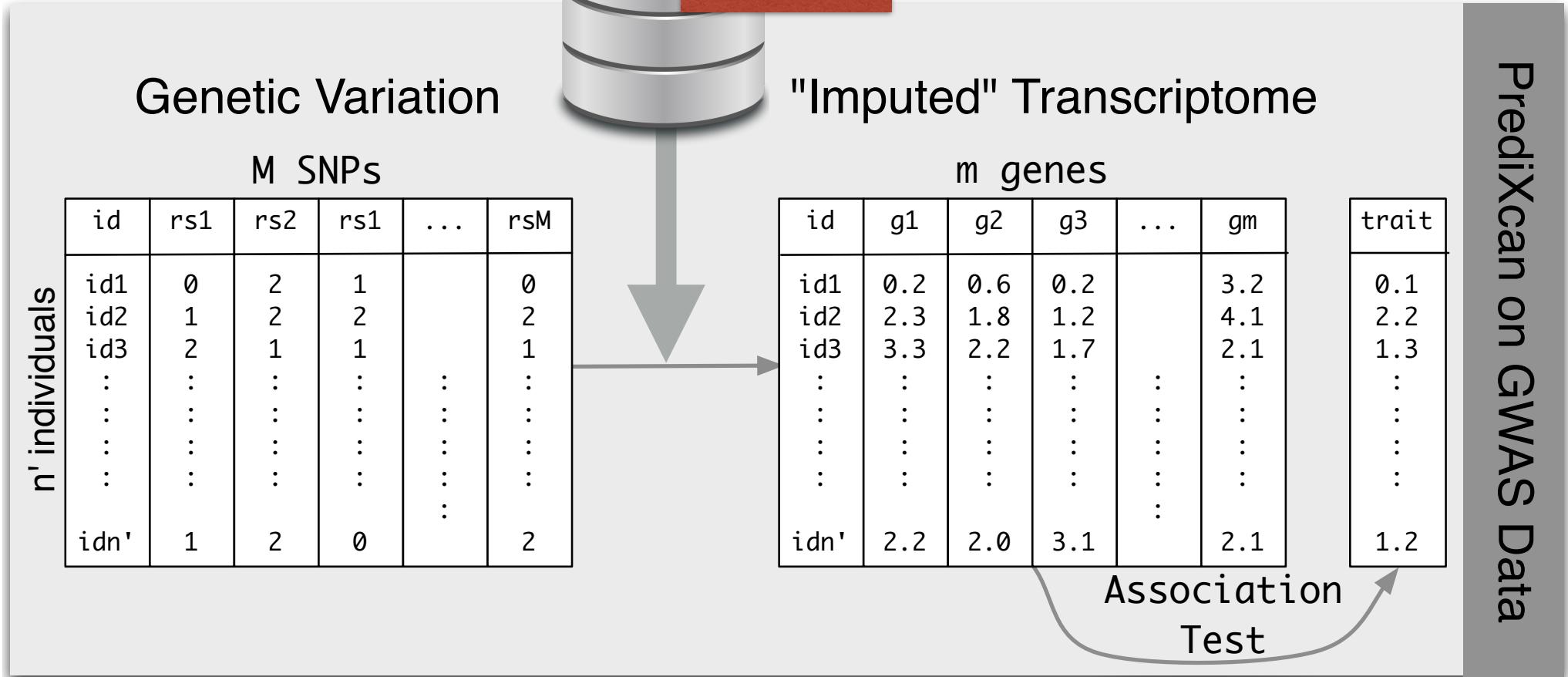
$$T = \sum_k w_k X_k + \epsilon$$

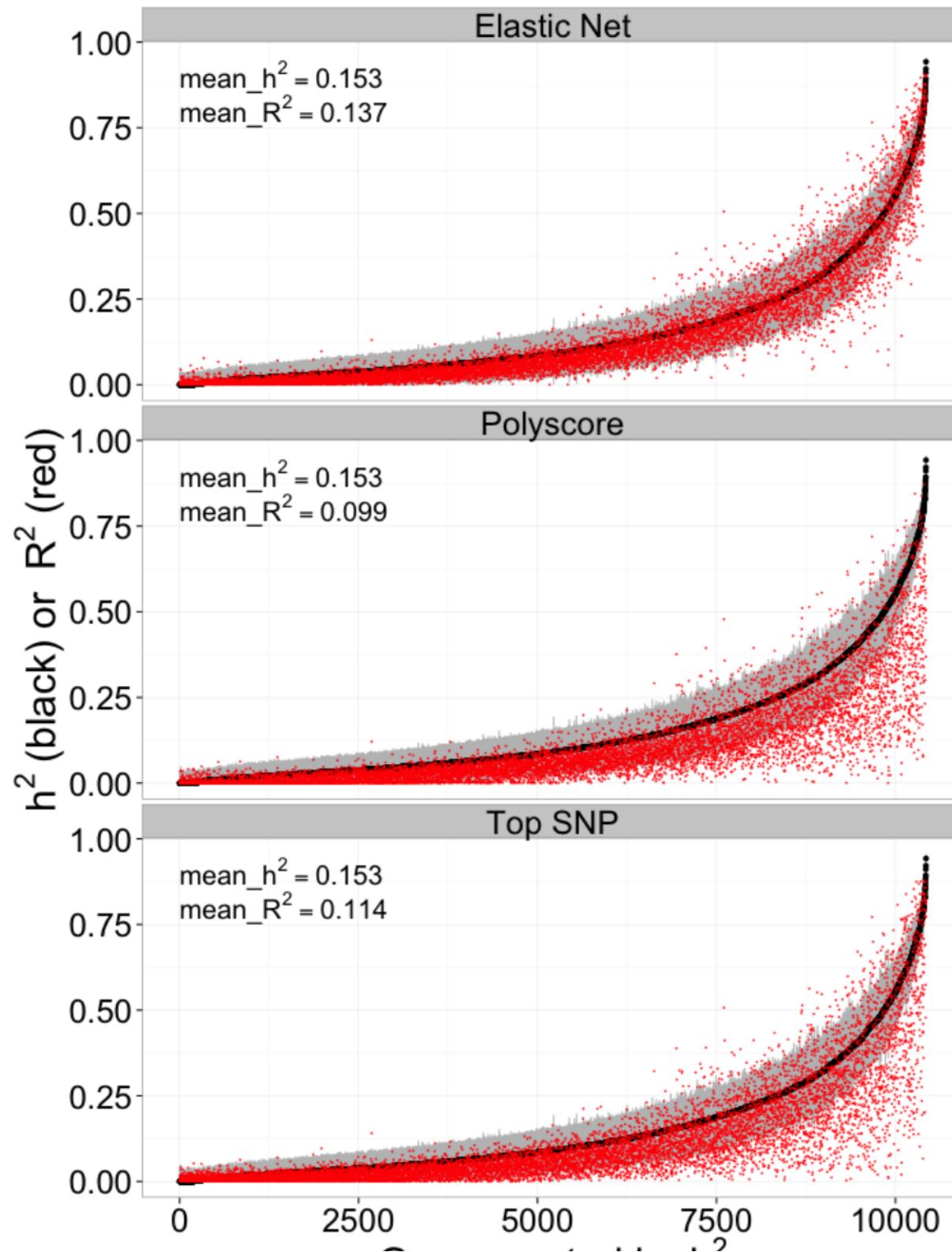
GReX

Weights stored in PredictDB

Weights go into
PredictDB

PrediXcan Imputes Transcriptome & Tests Assoc.

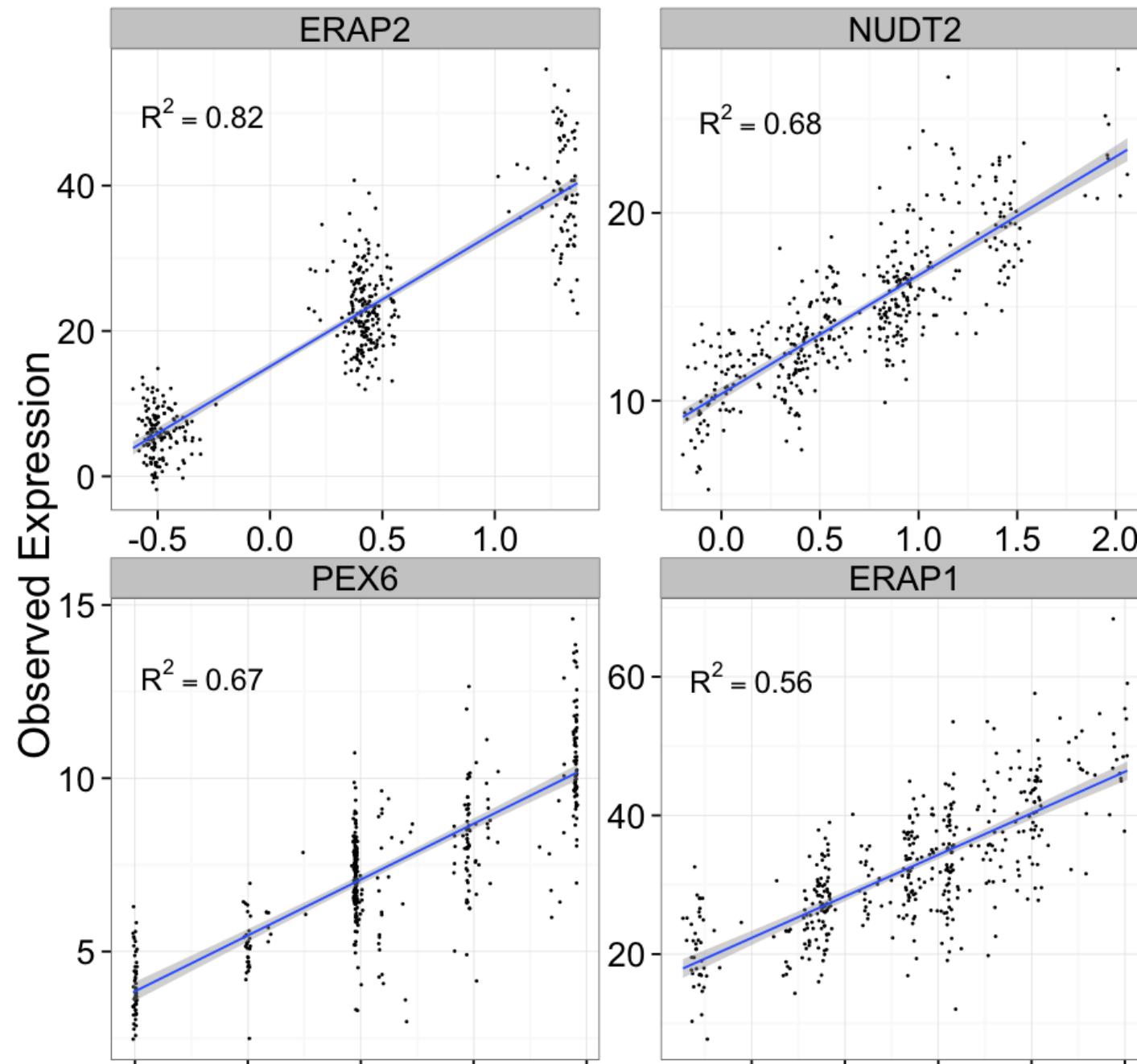




**Elastic Net
Outperform
Polygenic Models**

**Using top eQTL
suboptimal**

Examples of well predicted genes



This repository Search Pull requests Issues Gist

Unwatch 1 Star 0 Fork 2

branch: master PrediXcan / README.md

hakyim 2 days ago modified README files

2 contributors

44 lines (24 sloc) 2.076 kb

Raw Blame History

PrediXcan

PrediXcan is a gene-based association test that prioritizes genes that are likely to be causal for the phenotype.

PredictDB

PredictDB hosts genetic prediction models of transcriptome levels to be used with PrediXcan.

The following models are available for download.

- DGN Whole Blood Elastic Net SQLite db [link](#) (A ~28MB file will be downloaded)
- DGN Whole Blood Elastic Net text [link](#) (A ~18MB file will be downloaded)
- DGN Whole Blood LASSO SQLite db [link](#) (A ~18MB file will be downloaded)

Reference

Eric R. Gamazon†, Heather E. Wheeler†, Sahar Mozaffari†, Kaanan P. Shah, Keston Aquino-Michaels, Robert J. Carroll, Anne E. Eyler, Joshua C. Denny, GTEx Consortium, Dan L. Nicolae, Nancy J. Cox, and Hae Kyung Im* **PrediXcan: Trait Mapping Using Human Transcriptome Regulation** (2015 Under revision)

Software <https://github.com/hakyimlab/PrediXcan>

Benefits of PrediXcan

- Directly tests the molecular mechanism through which genetic variants affect phenotype
- Genes more attractive than genetic variants
 - A lot is known about their function
 - Animal model results can be integrated seamlessly
 - 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
- Reverse causality largely avoided
- Can be systematically applied to existing GWAS studies
- Tissue-specificity can be inferred

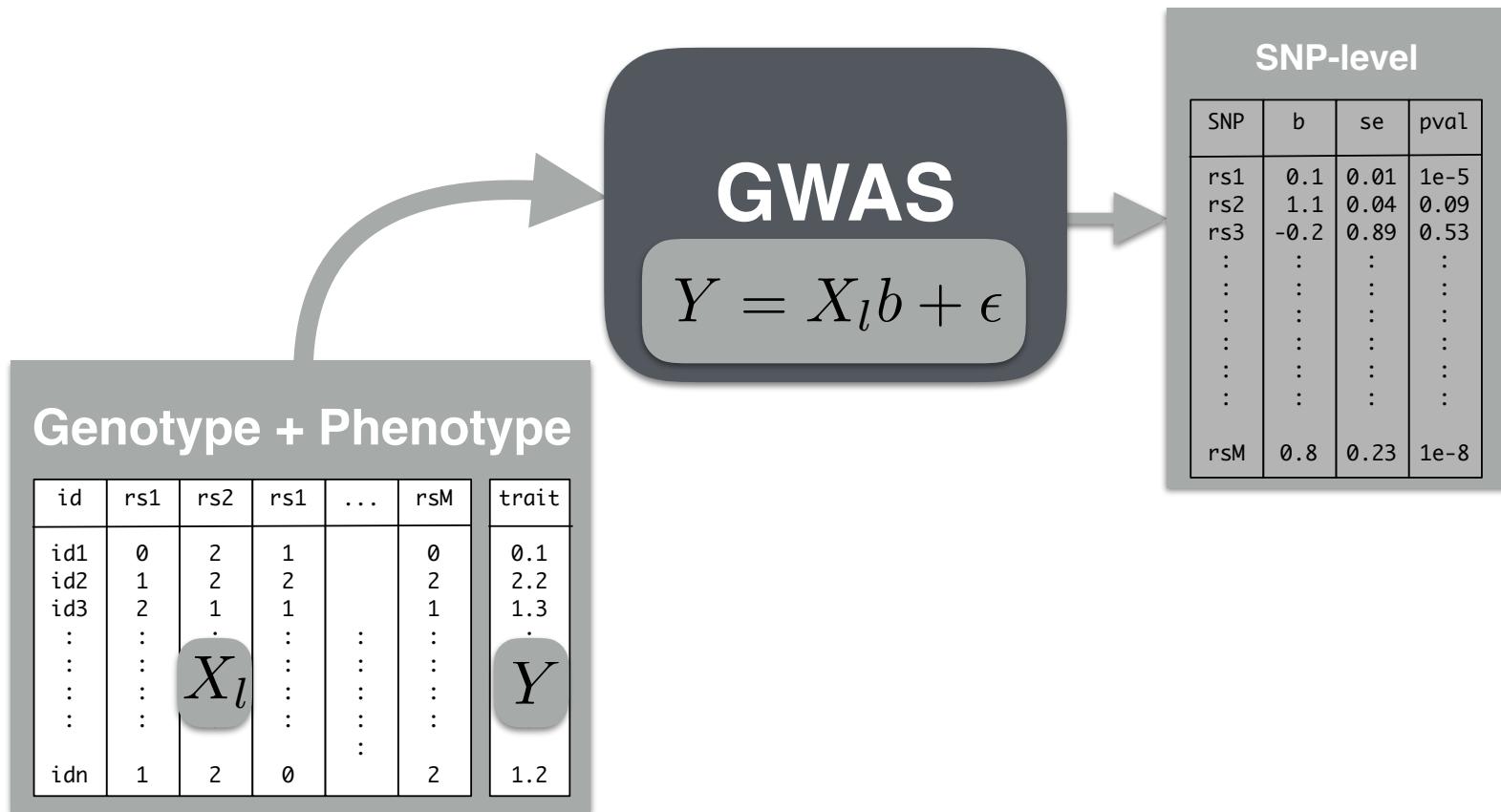
MetaXcan

a summary statistics based gene-level association test

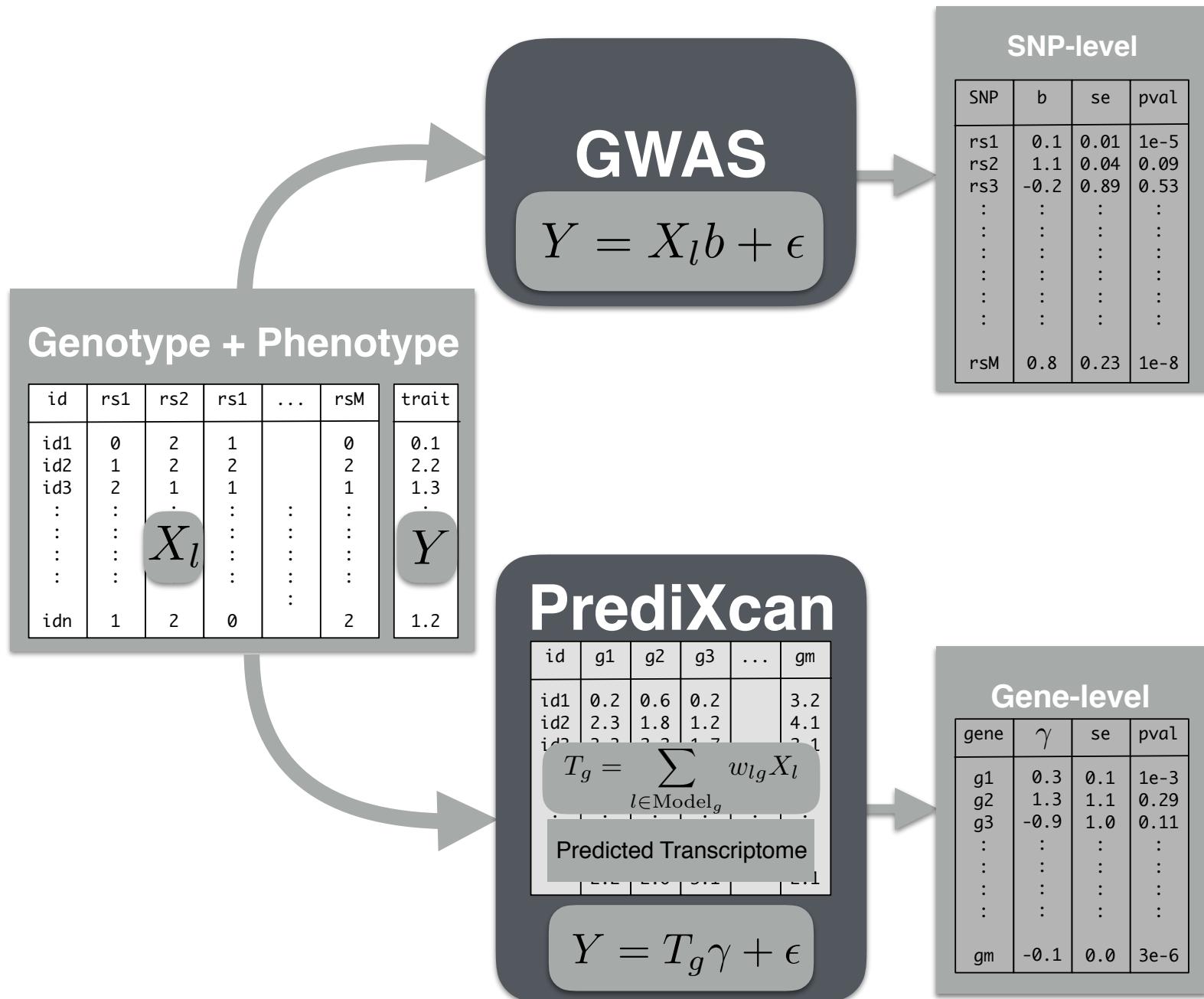


Alvaro Barbeira

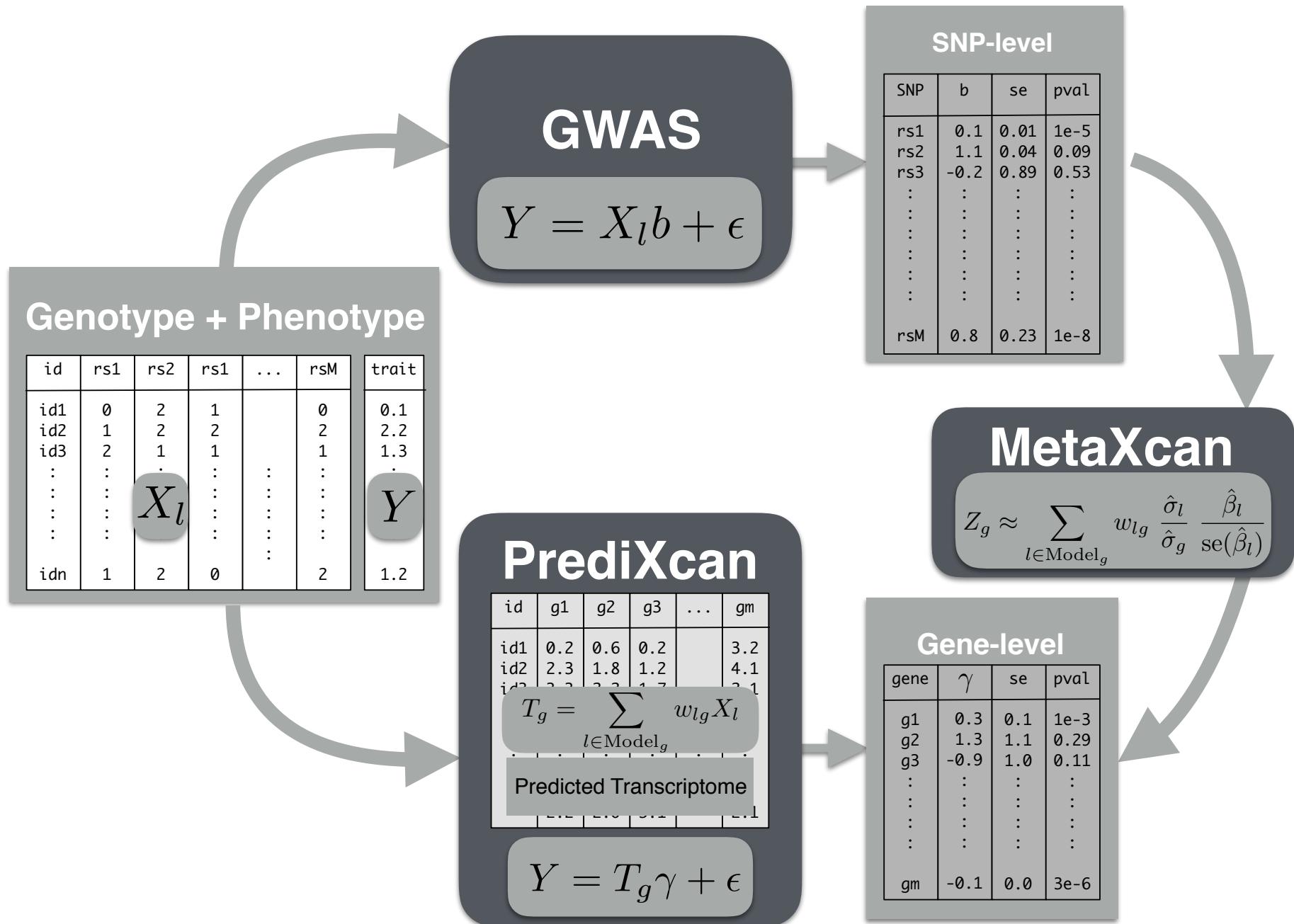
GWAS - PrediXcan - MetaXcan



GWAS - PrediXcan - MetaXcan



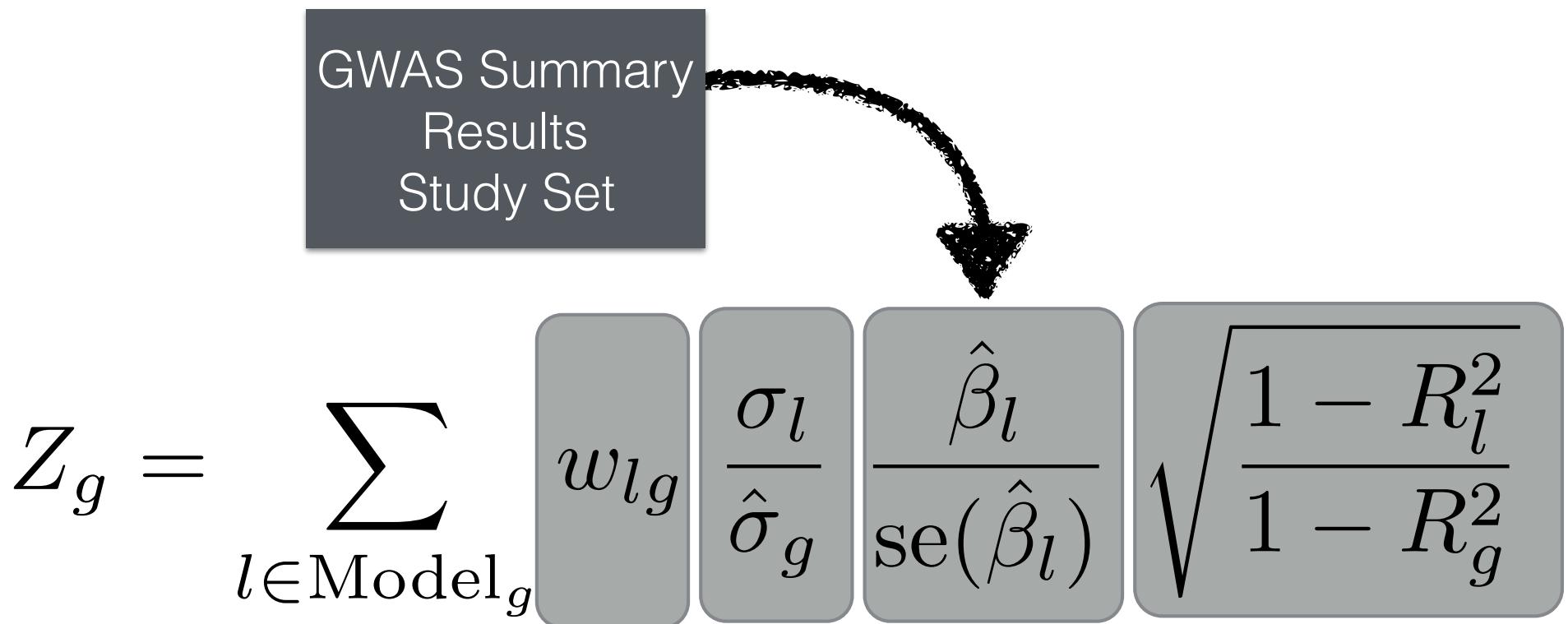
GWAS - PrediXcan - MetaXcan



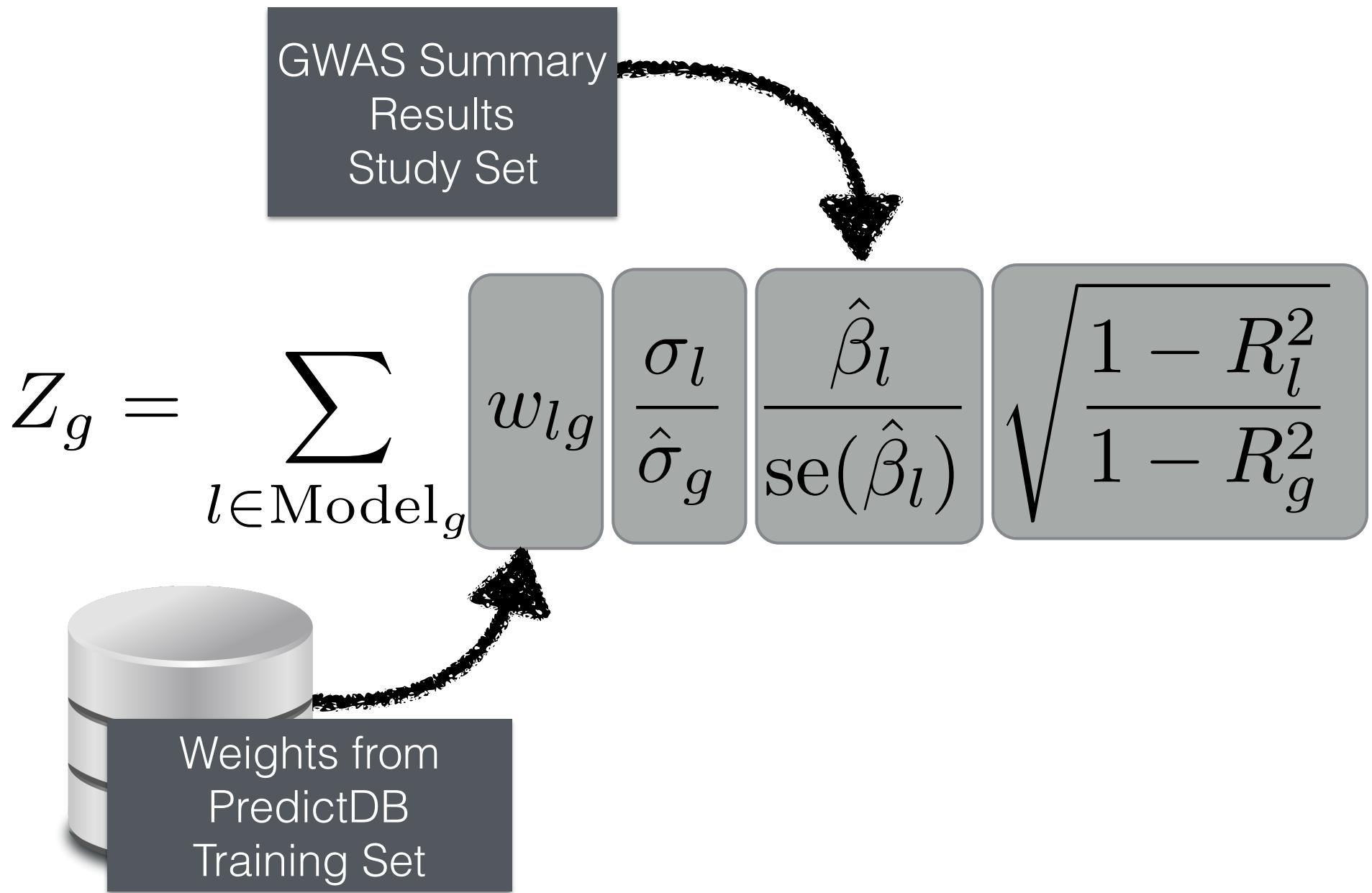
MetaXcan Formula

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

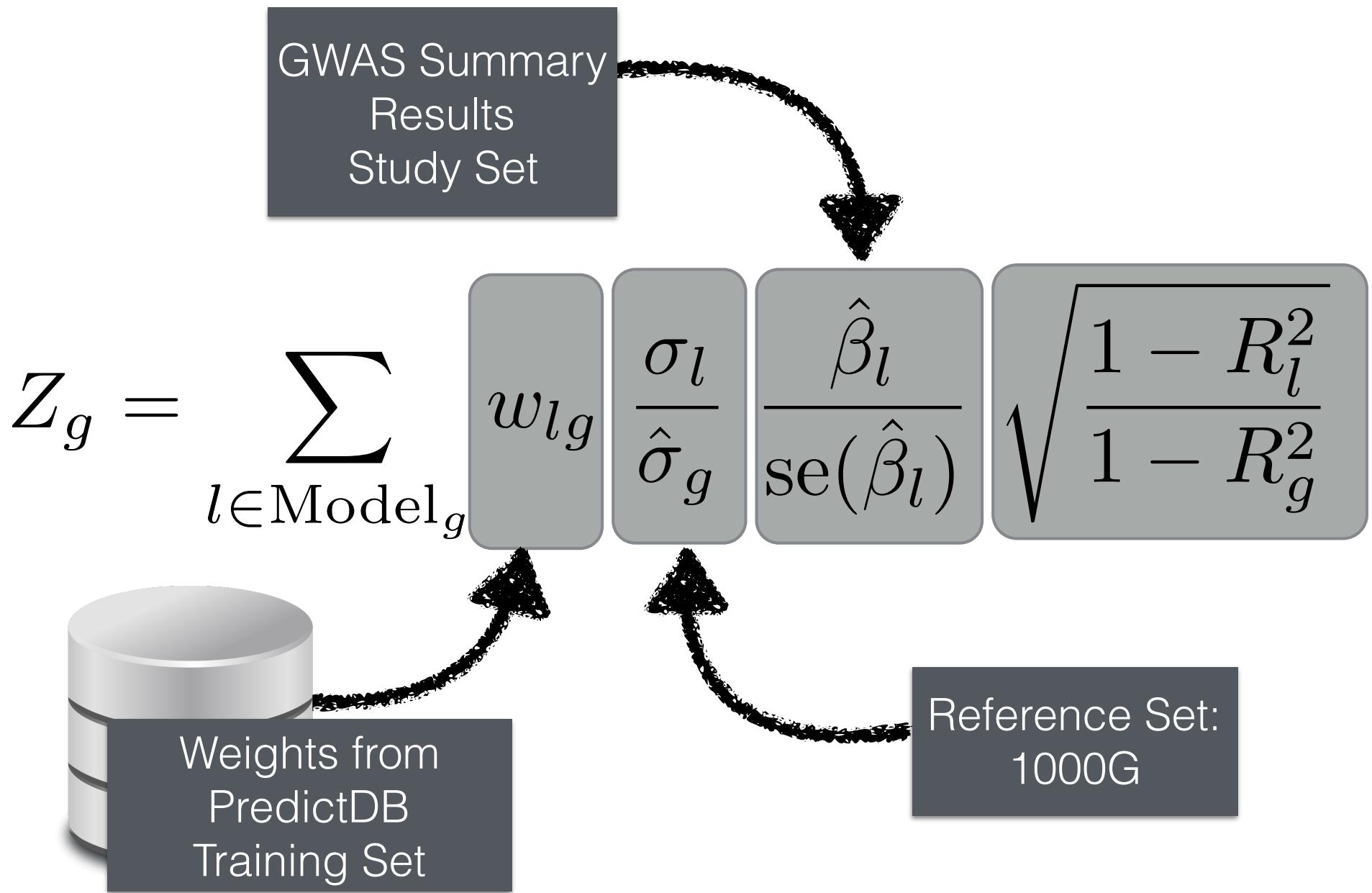
MetaXcan Formula



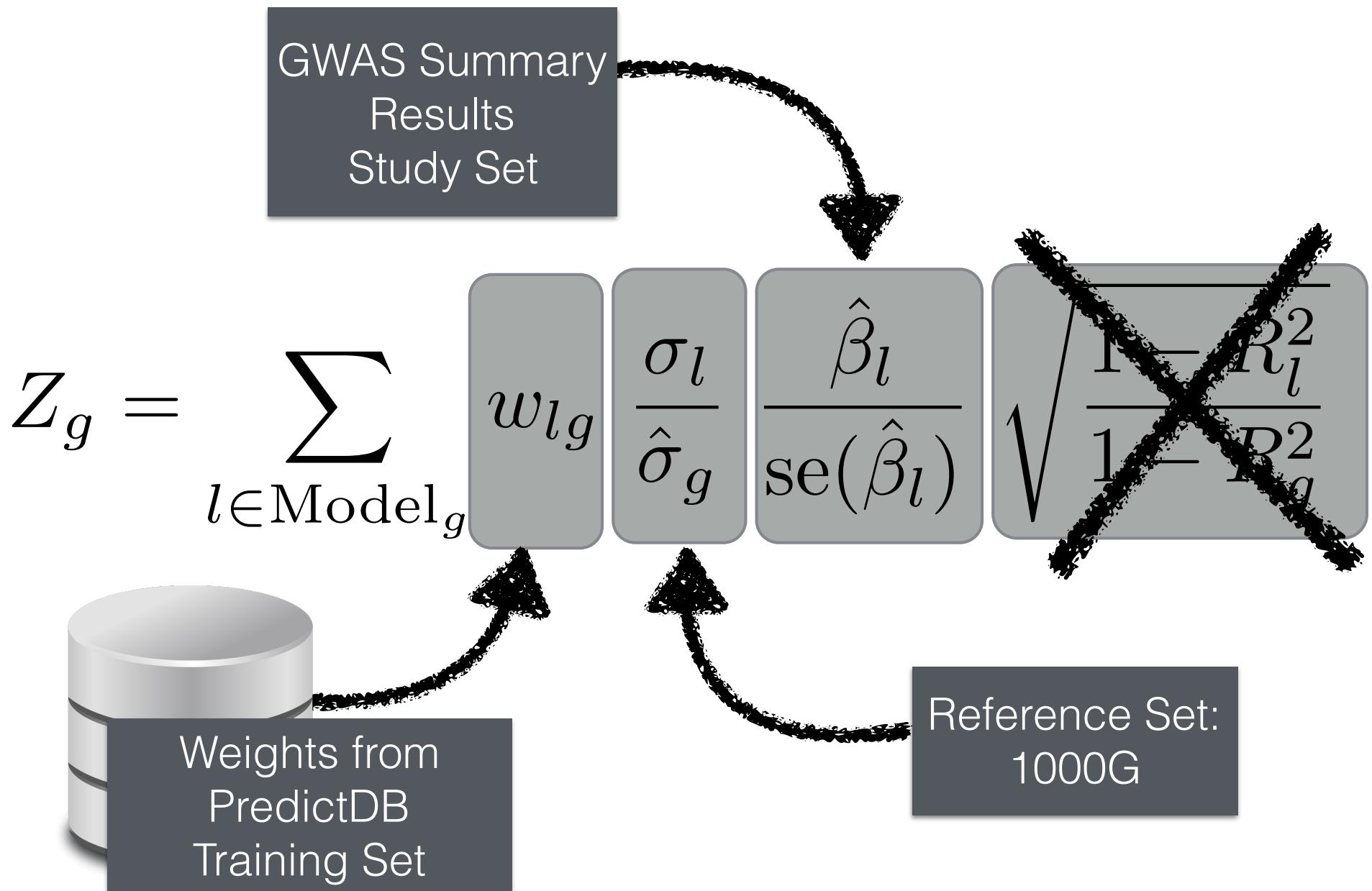
MetaXcan Formula



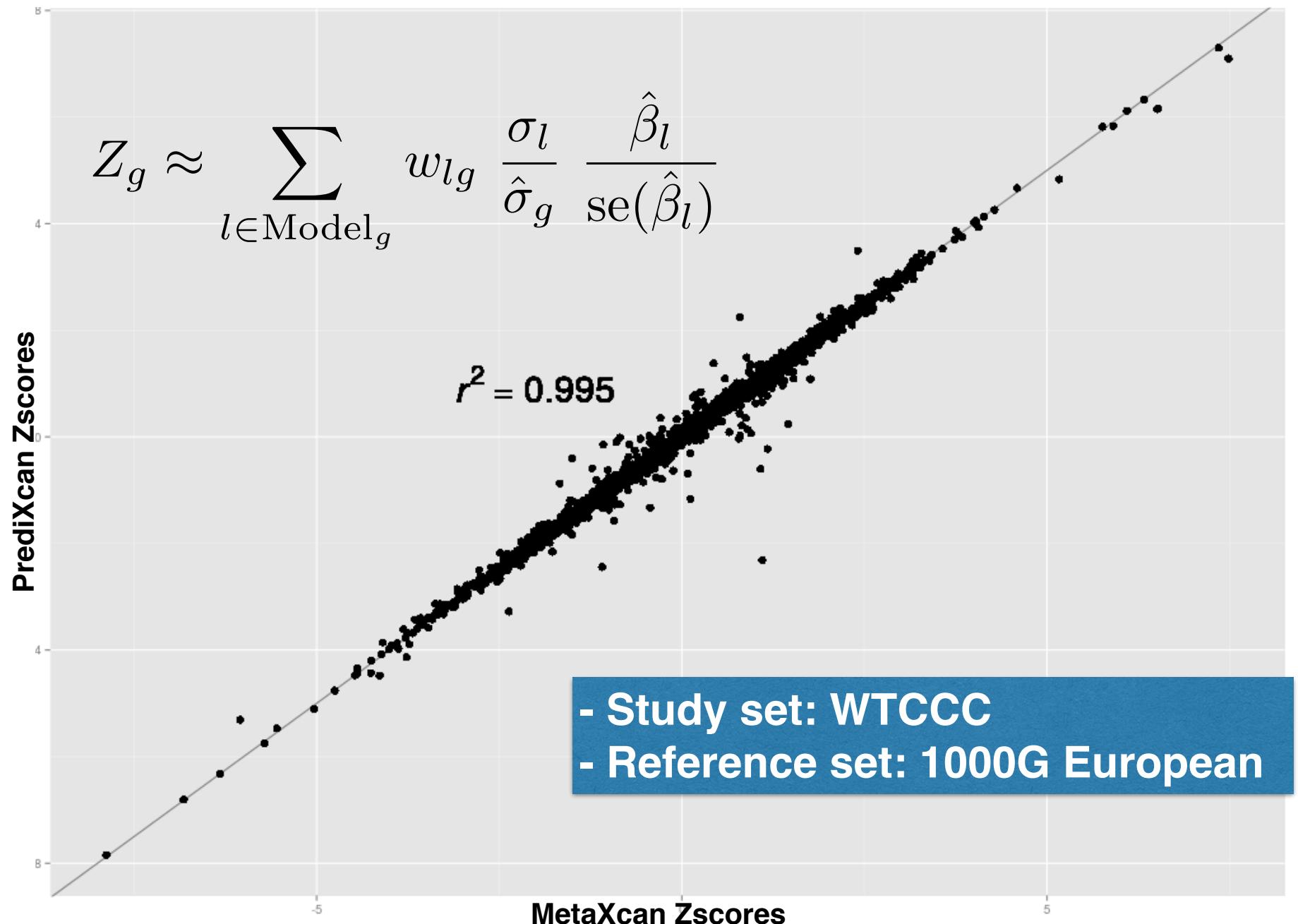
MetaXcan Formula



MetaXcan Formula



PrediXcan vs MetaXcan - WTCCC Type 1 Diabetes



MetaXcan Web App - Beta Version

The screenshot shows the MetaXcan Web App interface. At the top, there's a blue header bar with the text "MetaXcan Client" on the left and "test-hki-2 Logout" on the right. Below the header, a green callout box contains the text: "You can check [here](#) for examples on how to set the following parameters." The main content area is titled "Metaxcan Job:gtx-jamboree-test". It contains several input fields and dropdown menus:

- SNP Column Name: SNP
- Odd Ratio Column Name: (empty)
- P-Value Column Name: (empty)
- Other Allele Column Name: A1
- Beta Column Name: (empty)
- Effect Allele Column Name: A2
- Beta Sign Column Name: (empty)
- Transcriptome Model: Tissue-Wide Whole Blood
- Covariance: Tissue-Wide Whole Blood, European Reference
- Column Separator: Any Whitespace (including spaces or tabs)
- Files Are GZipped Compressed

At the bottom of the page, a large red banner displays the URL: <http://hakyimlab.org/metaxcan>.

MetaXcan Preprint on bioRxiv



Cold
Spring
Harbor
Laboratory



[HOME](#) | [ABOUT](#)

Search

New Results

MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results

Alvaro Barbeira, Kaanan P Shah, Jason M Torres, Heather E Wheeler, Eric S Torstenson, Todd Edwards, Tzintzuni Garcia, Graeme I Bell, Dan Nicolae, Nancy J Cox, Hae Kyung Im

doi: <http://dx.doi.org/10.1101/045260>

This article is a preprint and has not been peer-reviewed [what does this mean?].

Abstract

[Info/History](#)

[Metrics](#)

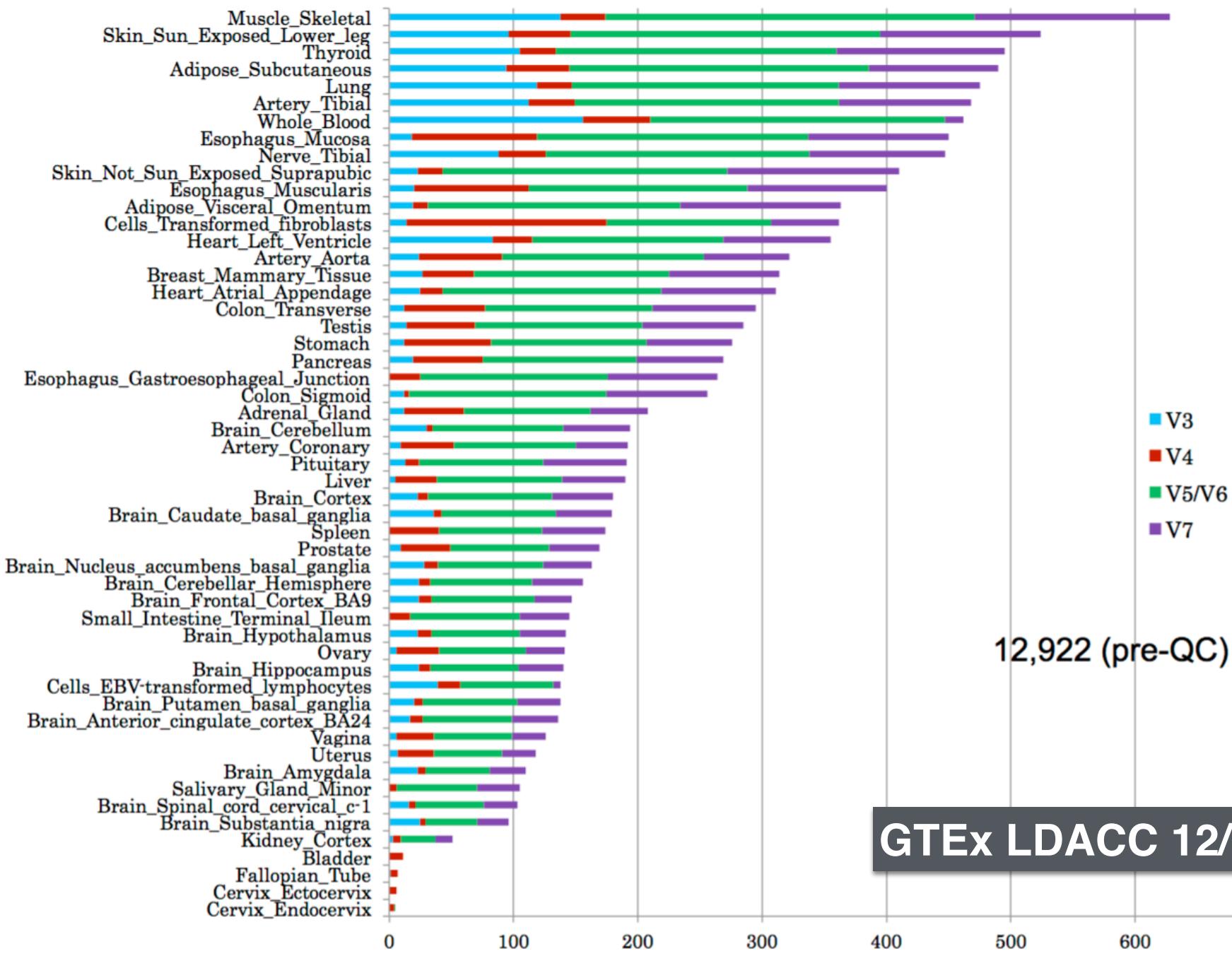
[Preview PDF](#)

Abstract

PredictDB

Gene Expression Prediction Models for 40 Human Tissues

Trancriptome Data: GTEx Tissues



GTEx LDACC 12/2015

Prediction Models for 40 Human Tissues

Screenshot of the PredictDB Data Files page on predictdb.org.

The page displays a table of prediction models for 40 human tissues. The columns are: Download, Tissue, Model, Training Set, Version Date, and Notes. All models listed are based on the GTEx training set and were created on November 12, 2015. The notes column indicates "run c" for all entries.

Download	Tissue ▲	Model	Training Set	Version Date	Notes
	Adipose, Subcutaneous	LASSO	GTEx	2015-11-12	run c
	Adipose, Subcutaneous	elastic net 0.5	GTEx	2015-11-12	run c
	Adrenal Gland	elastic net 0.5	GTEx	2015-11-12	run c
	Adrenal Gland	LASSO	GTEx	2015-11-12	run c
	Artery, Aorta	elastic net 0.5	GTEx	2015-11-12	run c
	Artery, Aorta	LASSO	GTEx	2015-11-12	run c
	Artery, Coronary	elastic net 0.5	GTEx	2015-11-12	run c
	Artery, Coronary	LASSO	GTEx	2015-11-12	run c
	Artery, Tibial	LASSO	GTEx	2015-11-12	run c
	Artery, Tibial	elastic net 0.5	GTEx	2015-11-12	run c

Page navigation: 1 / 9 | Next

Text at the bottom:

These data are provided for use with PrediXcan and MetaXcan from [the Im lab](#).
This work is supported by R01MH107666 (H.K.I.), K12 CA139160 (H.K.I.), R01 MH101820 (GTEx), P30 DK20595 and P60 DK20595 (Diabetes Research and Training Center), P50 DA037844 (Rat Genomics), P50 MH094267 (Conte).

Applying MetaXcan Phenome-wide

Phenome-wide Scan of Gene Level Associations

gene2pheno.org

Metaxcan Association results

'Tissue' stands for different transcriptome models used when generating the association.

Tissues built with GTEx data, covariances with 1000 Genomes.

Gene Name: Ordered Phenotype: Tissue: R2 threshold: Pvalue threshold: Record limit:

Show 25 entries

gene_name	zscore	pval	phenotype	tissue			
HLA-DQA2	37.8461	0.000000e+00	RA_OKADA_EUR	TW_Stomach_ElasticNet			
HLA-DQA2	38.1889	0.000000e+00	RA_OKADA_TRANS_ETHNIC	TW_Heart-LeftVentricle_ElasticNet			
HLA-DQA2	39.4297	0.000000e+00	RA_OKADA_TRANS_ETHNIC	TW_Colon-Sigmoid_ElasticNet			
HLA-DQA2	37.6550	2.708011e-310	RA_OKADA_TRANS_ETHNIC	TW_Artery-Aorta_ElasticNet			
HLA-DQA2	37.6003	2.125674e-309	RA_OKADA_EUR	TW_Colon-Sigmoid_ElasticNet			
HLA-DQA2	37.5029	8.267139e-308	RA_OKADA_EUR	TW_Heart-LeftVentricle_ElasticNet			
C6orf25	37.2617	6.850077e-304	RA_OKADA_TRANS_ETHNIC	TW_Heart-AtrialAppendage_ElasticNet			
C6orf25	36.7737	4.858262e-296	RA_OKADA_EUR	TW_Heart-AtrialAppendage_ElasticNet			
HLA-DQA2	36.7348	2.031999e-295	RA_OKADA_TRANS_ETHNIC	TW_SmallIntestine-TerminalIleum_ElasticNet	0.4674430	7	12

- 16 Large Consortia
- 6 million phenotype measurements
- 107 phenotypes
- 40 tissues
- gene2pheno.org

Conclusions

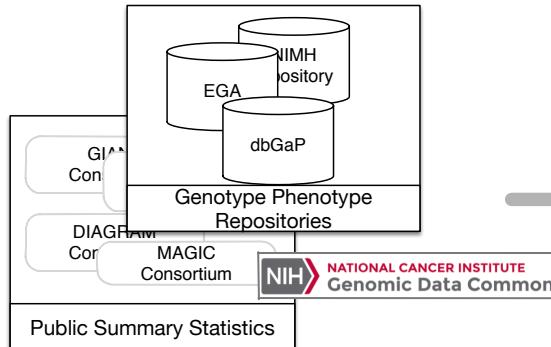
- We develop methods to understand disease biology
- PrediXcan
 - tests phenotypic consequences of genetic variation through gene regulation
 - discoveries at gene level -> immediate translational potential
- MetaXcan
 - allows us to scale up to massive sample sizes
 - phenome-wide scan of phenotypic consequences of gene expression alteration
- Public resources
 - prediction model (predictdb.org)
 - gene to phenotype database (gene2pheno.org)

Next Steps

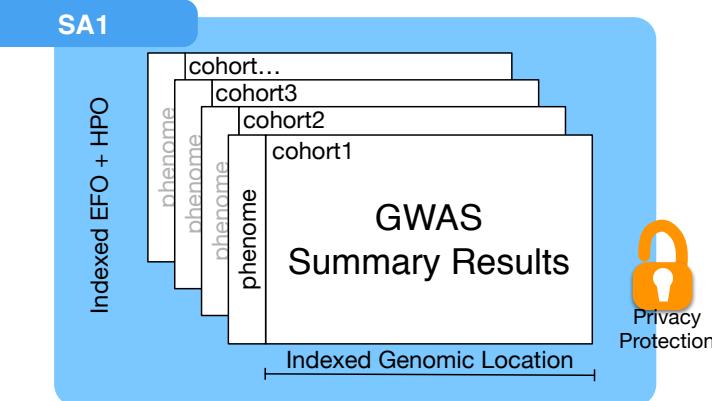
- Improve methods to integrate new genome/epigenome knowledge
- Generate genome-wide gene association results
 - UK Biobank, Framingham, MESA
- Build pipeline for prediction model training
 - Apply to GDC
 - Build other -omics models

Genomic Summary Data Commons: Proposal

Genotype+Phenotype Data



Genomic Summary Data Commons



Reference Data

