

Genetically Predicted Endophenotypes:

**Getting to the Next Level in Understanding
How Genome Variation Drives Disease**



vanderbilt Genetics Institute

Nancy J. Cox, Ph.D.

Just a Few Remaining Questions on the Role of Genetic Variation in Disease...

- **What GENES contribute?**
- **What are the MECHANISMS by which they affect risk of disease?**
- **For any contributing gene, what is the DIRECTION OF EFFECT?**
- **For common diseases, WHERE ARE THOSE #@!% RARE VARIANTS WITH BIG EFFECTS?!?!**

Genomic Discoveries for Common Disease



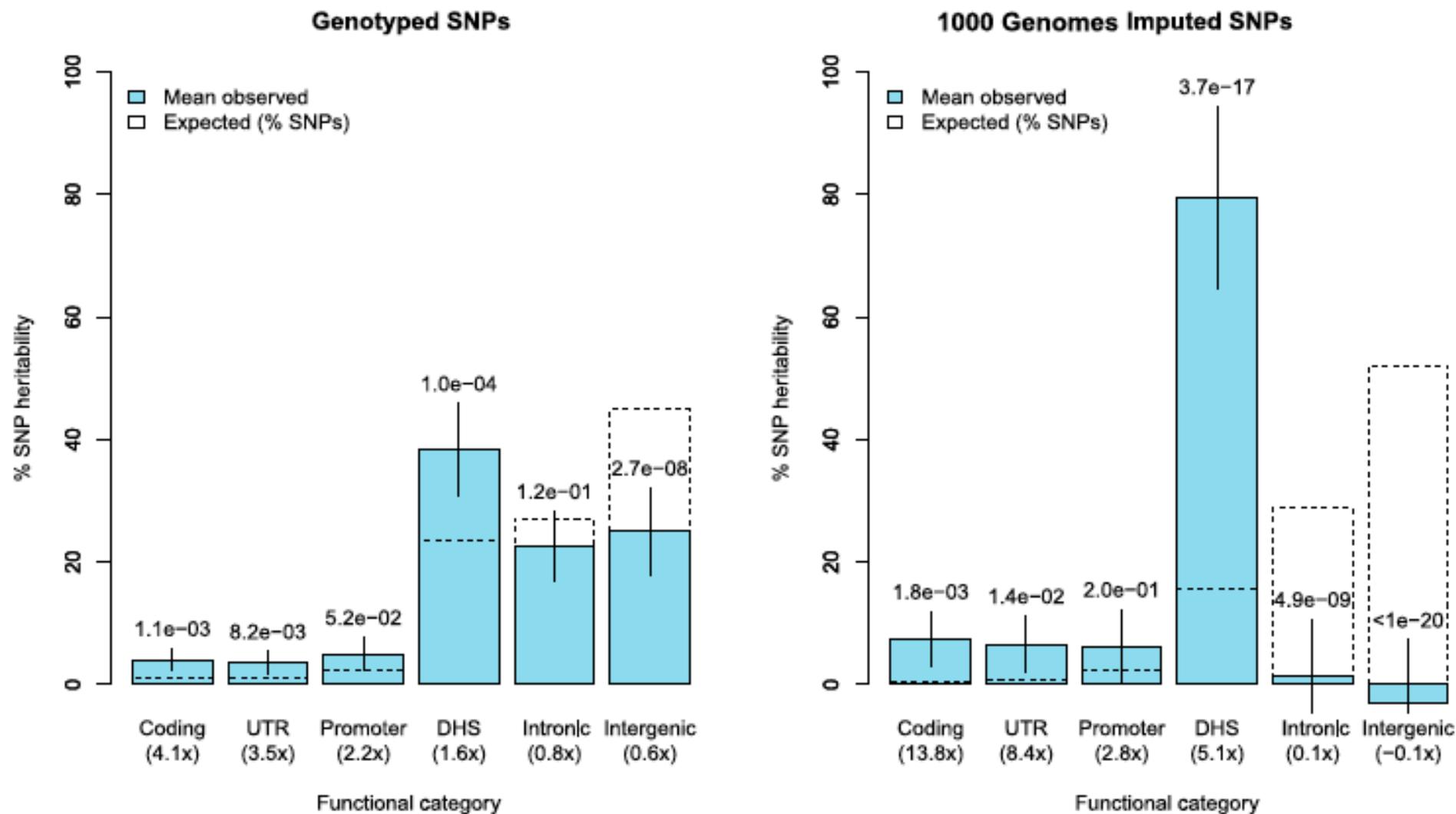
Genomic Discoveries for Common Disease



Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases

Alexander Gusev,^{1,*} S. Hong Lee,² Gosia Trynka,^{3,4,5,6,16} Hilary Finucane,⁷ Bjarni J. Vilhjálmsson,¹

Yan Xue,⁸ Chao Li,⁷ Zhen Zhang,⁸ Guohua Bi,^{9,10} Bradon B. Whitfield,^{9,10} Erik S. Lander,¹¹



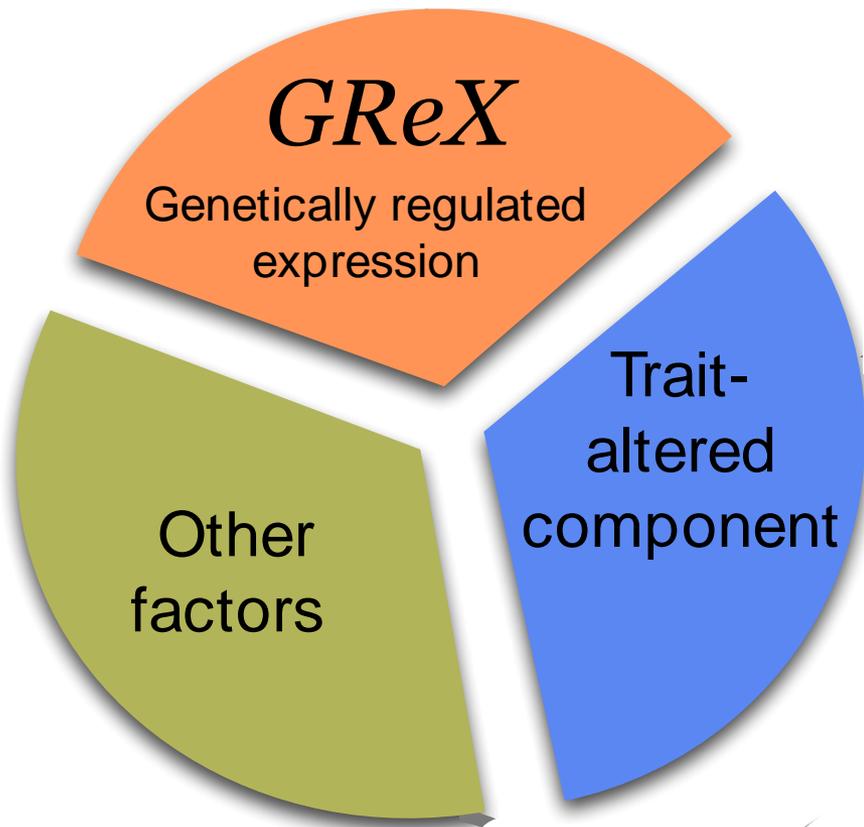
Type 1 Diabetes

Crohns Disease

	V(G)/V(P)	SE			V(G)/V(P)	SE
adipose	0.21	0.019			0.03	0.008
heart	0.199	0.02			0.017	0.006
lung	0.192	0.018			0.02	0.007
muscle	0.188	0.018			0.028	0.008
nerve	0.191	0.018			0.025	0.008
whole blood	0.187	0.023			0.17	0.024
Overall	0.48	0.06			0.50	0.07

A Missing Data Problem?

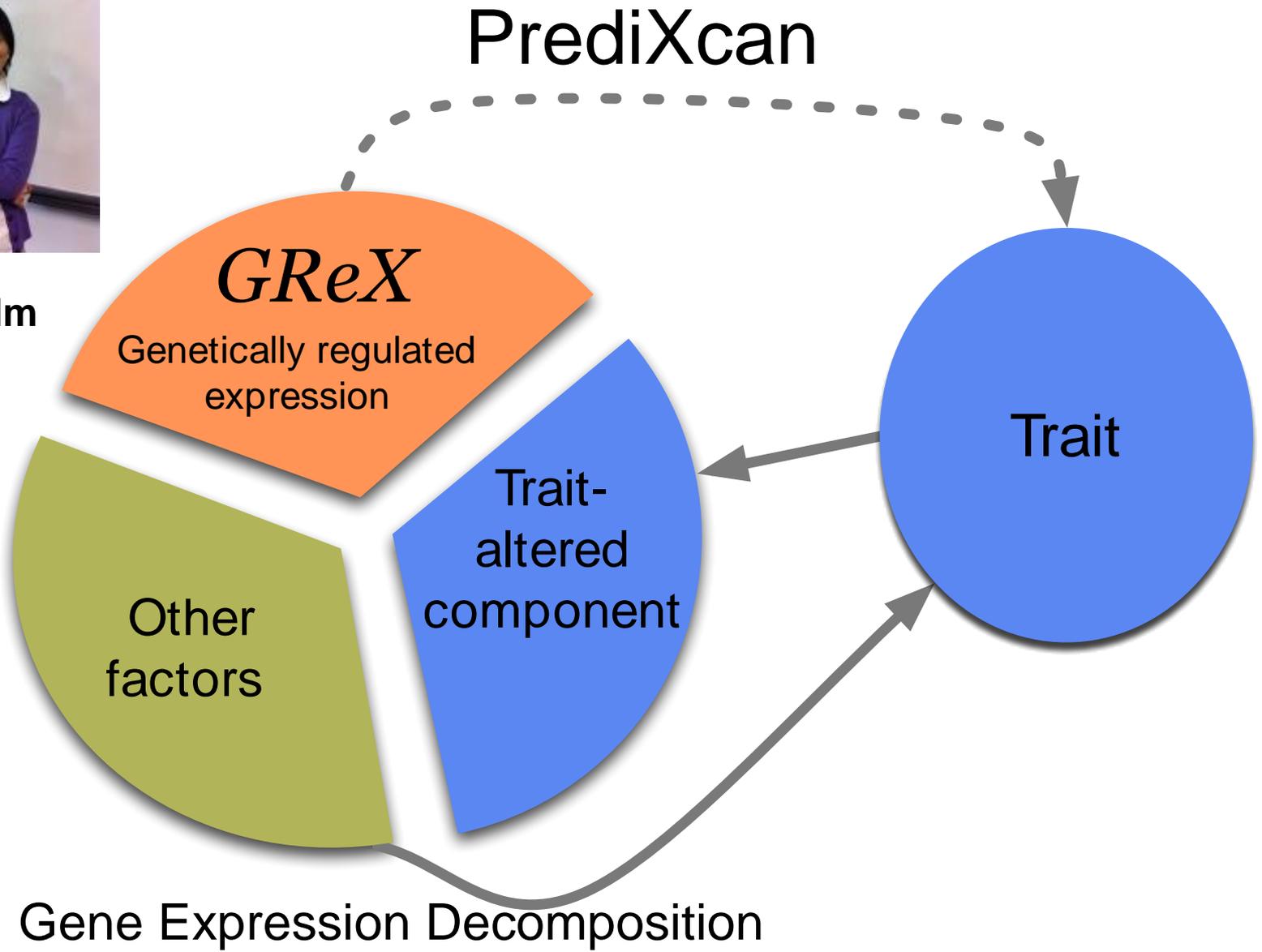
- If a substantial fraction of the genome variation affecting risk of common disease is regulatory, why not alter focus of analyses to endophenotypes that are more direct measures of what we really want – the *genetically determined part* of protein levels (transcript levels)
- Instead of testing individual variants, aggregate variants into SNP-based predictors of transcript levels (and ultimately protein levels) and test those directly for association with disease



Gene Expression Decomposition



Haky Im



Submitted Nat Genet as GTEx companion paper

Genetic Variation

M SNPs

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

Observed Transcriptome

m genes

id	g1	g2	g3	...	gm
id1	0.1	0.1	0.2		3.2
id2	2.2	1.7	1.2		4.1
id3	1.3	2.0	1.7		2.1
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
idn	1.2	2.2	3.1		2.1

Reference Transcriptome

Analogous to Imputation

Learn relationship of genome variation to transcriptome in reference sample (GTEx)

PredictDB: Database of Prediction Models

M SNPs

g1	rs1	rs2	rs3	...	rsM
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

$$T = \underbrace{\sum_k w_k X_k}_{GReX} + \epsilon$$

Weights stored in PredictDB

Store weights from prediction equations

Genetic Variation

M SNPs

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

"Imputed" Transcriptome

m genes

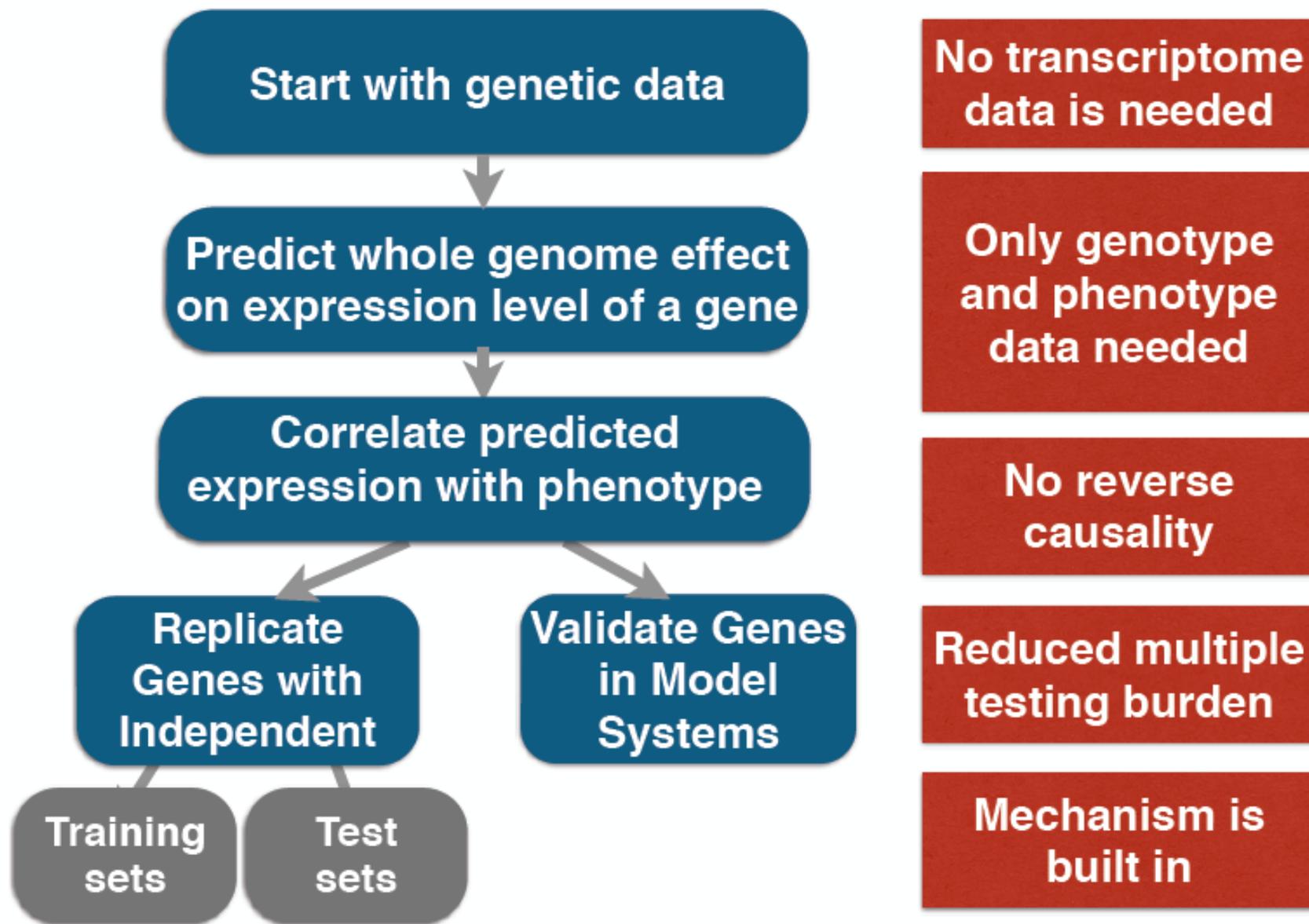
id	g1	g2	g3	...	gm	trait
id1	0.1	0.1	0.2		3.2	0.1
id2	2.2	1.7	1.2		4.1	2.2
id3	1.3	2.0	1.7		2.1	1.3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
idn	1.2	2.2	3.1		2.1	1.2

Association Test

PrediXcan on GWAS Data

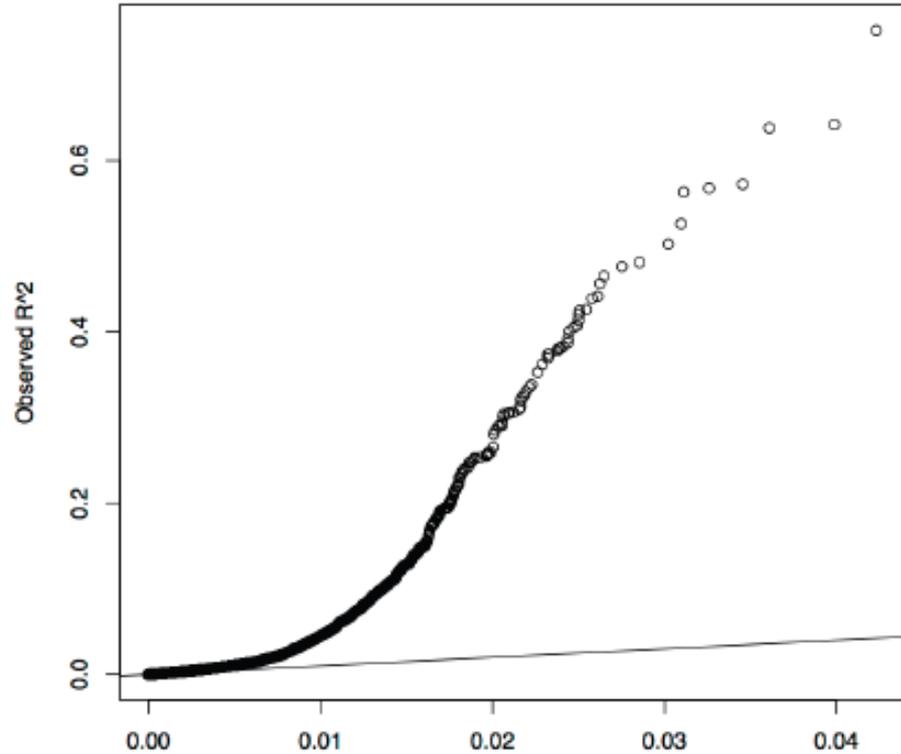
Apply to any dataset with genome interrogation

PrediXcan Flow



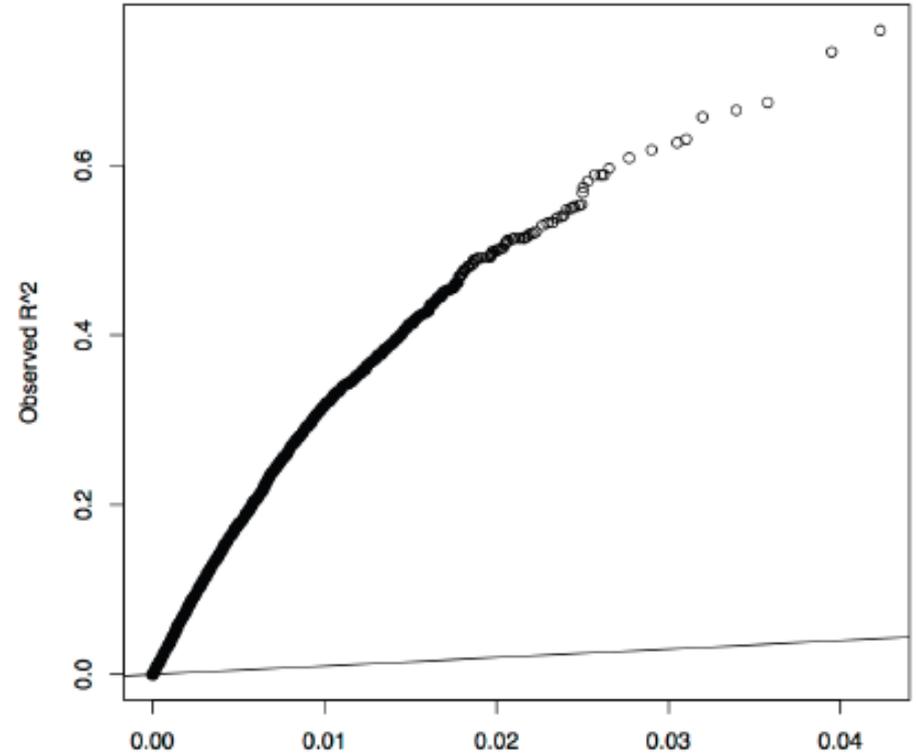
Good Prediction Performance

Prediction R^2



Training with GTEx
Testing in 1K Genomes

Replication R^2



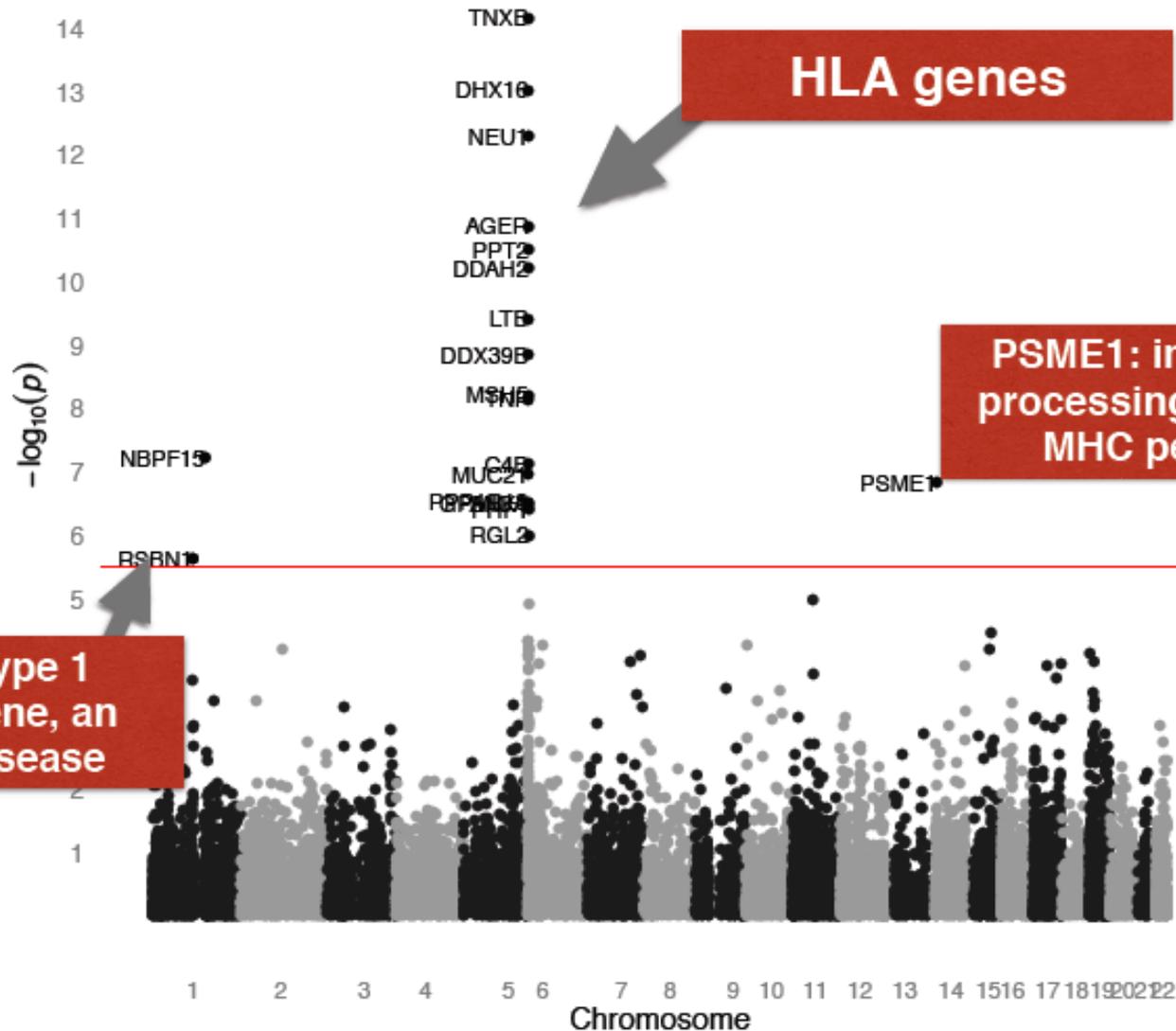
Replicate RNAseq
Pickrell et al 2010 vs.
1K Genomes 2013

Prediction in an Independent Sample

- **Significance of correlation between predicted and directly measured expression levels: q-value < 0.05 for 40-50% of genes, < 0.1 for 60-70%**

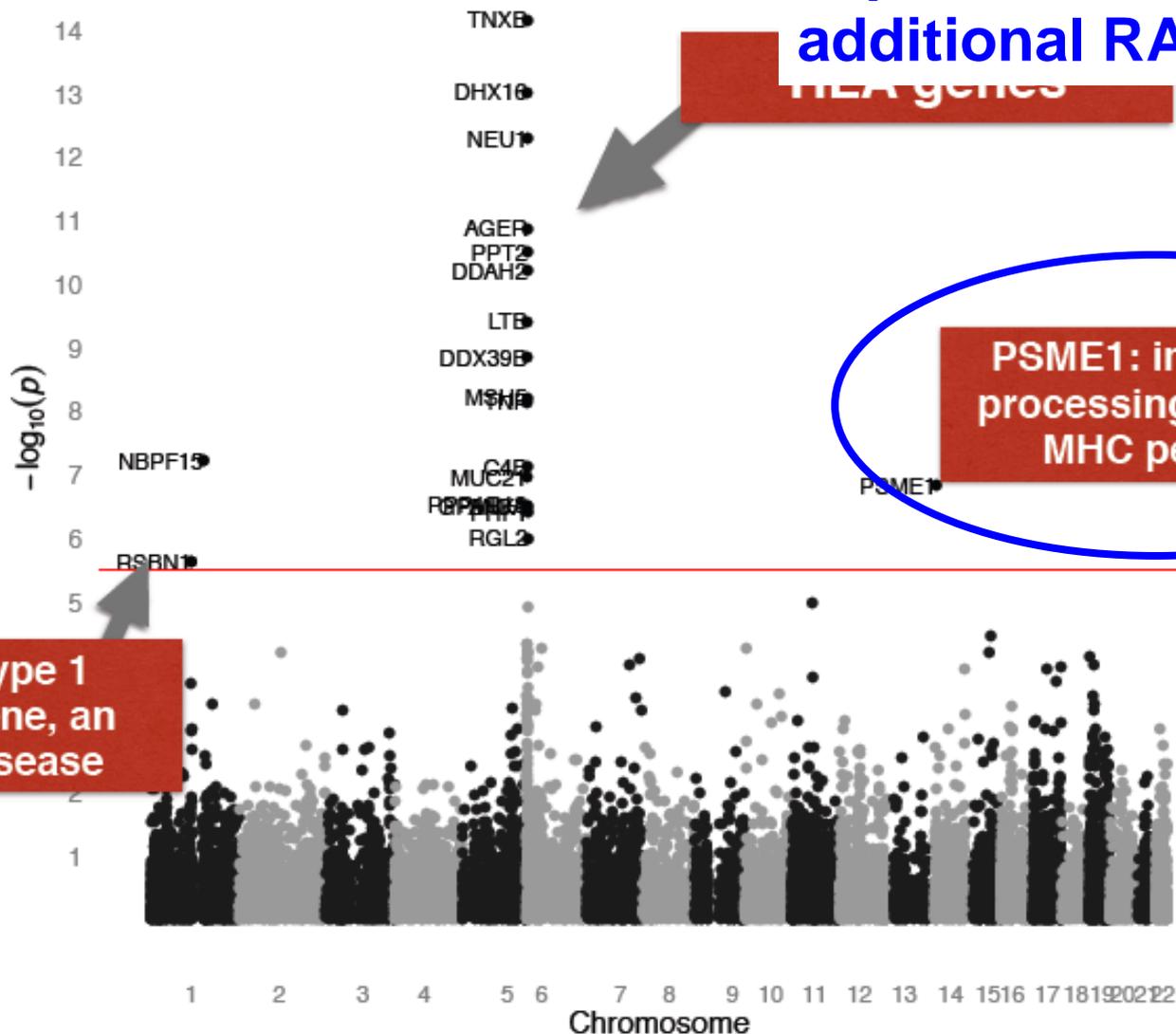
We predict only the genetically determined part of gene expression

Genes Associated with Rheumatoid Arthritis



Genes Associated with Rheumatoid Arthritis

Replicate predicted expression of PSME1 in additional RA datasets



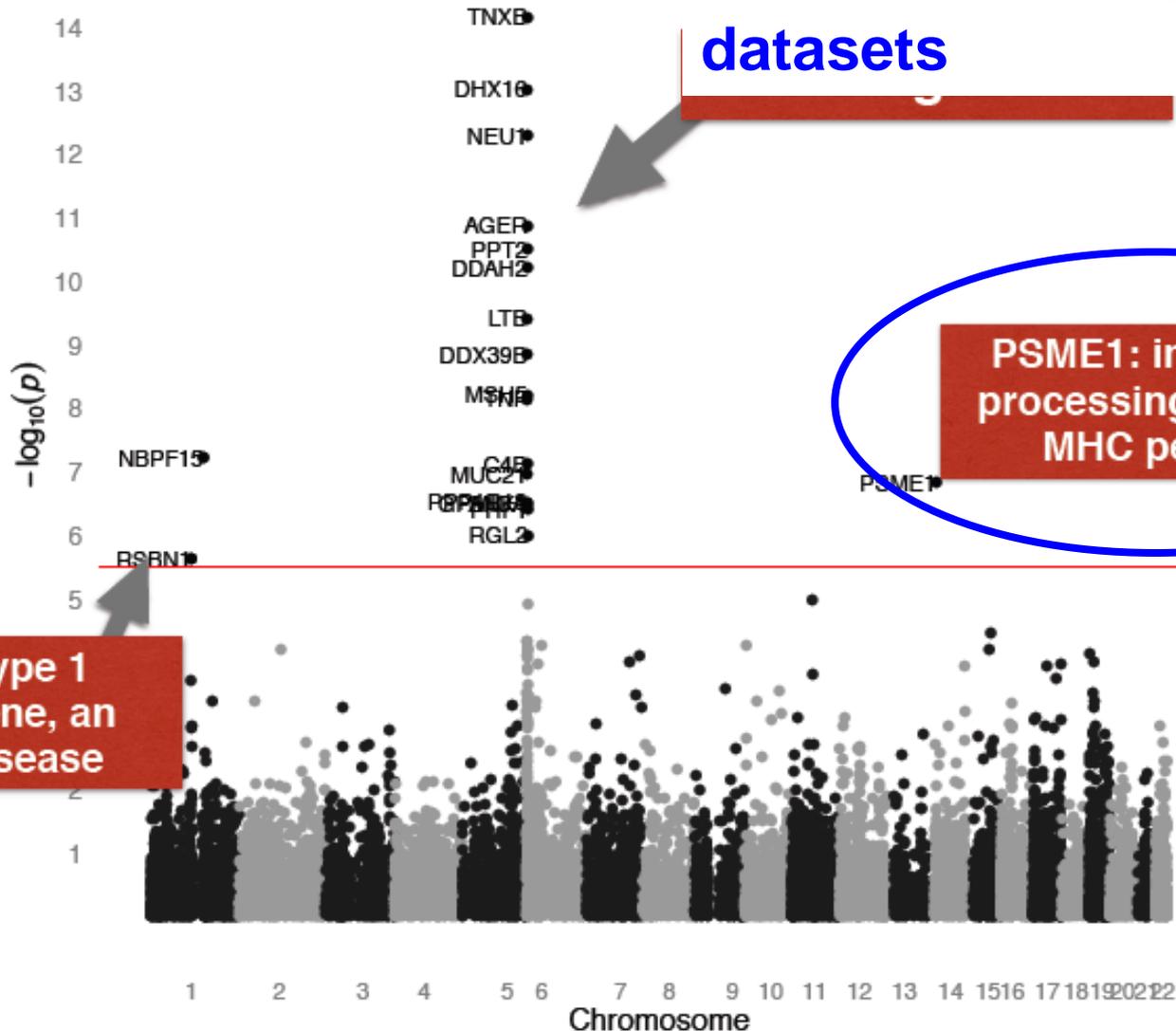
RSN1: Type 1 diabetes gene, an immune disease

PSME1: involved in processing of class I MHC peptides

HLA genes

Genes Associated with

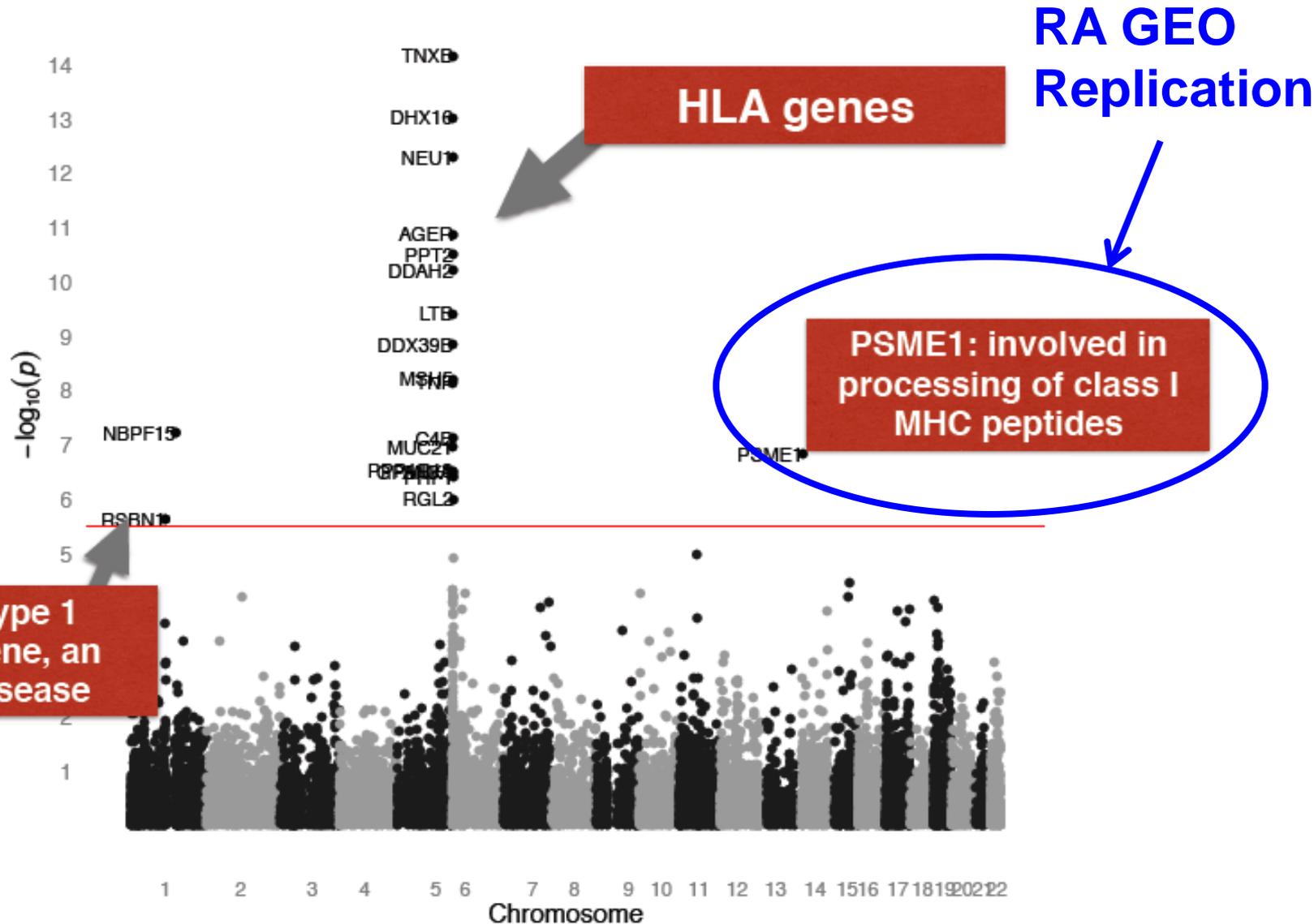
Replicate SNP predictors as cis- and trans-eQTLs in additional transcriptome datasets



RSN1: Type 1 diabetes gene, an immune disease

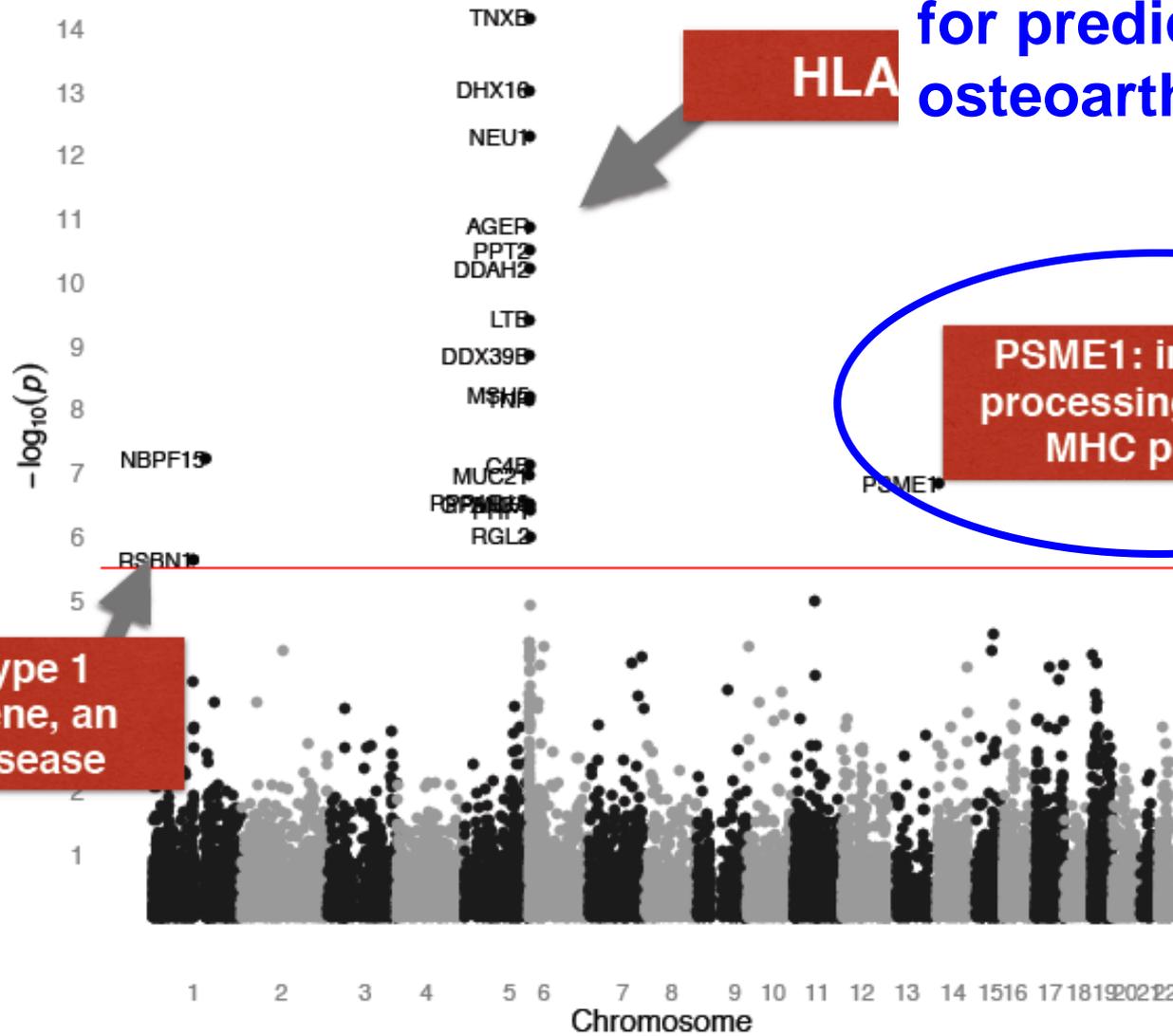
PSME1: involved in processing of class I MHC peptides

Genes Associated with Rheumatoid Arthritis



Genes Associated with Rheumatoid Arthritis

Part of a 20-gene expression signature for predicting osteoarthritis



RSN1: Type 1 diabetes gene, an immune disease

HLA

PSME1: involved in processing of class I MHC peptides

Chromosome

Advantages of Framework

- **We iteratively use more and more of what we do know to figure out what we most want to learn**
 - True no matter whether what you most want to learn about is transcript (protein) levels or genetics of common disease
 - Identifies key read-outs for discovering environmental risk factors
- **This sets up a (the?) natural framework for the analysis of whole genome sequence data**

Our GTEx Team



Dan Nicolae



Lin Chen



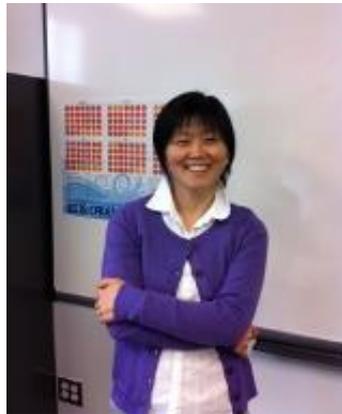
Lea Davis (&Bridget)



Richard Jones



Eric Gamazon



**Hae Kyung
"Haky" Im**



**Anuar
Konkashbaev**



Barbara Stranger



Younghee Lee

Acknowledgements

The GTEx Consortium Investigators (GTEx Pilot phase)

cancer Human Biobank (caHUB)

Biospecimen Source Sites (BSS)

John Lonsdale, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, *National Disease Research Interchange, Philadelphia, PA*

Richard Hasz, *Gift of Life Donor Program, Philadelphia, PA*

Gary Walters, *LifeNet Health, Virginia Beach, VA*

Nancy Young, *Albert Einstein Medical Center, Philadelphia, PA*

Laura Siminoff (ELSI Study), Heather Traino, Maghboeba Mosavel, Laura Barker, *Virginia Commonwealth University, Richmond, VA*

Barbara Foster, Mike Moser, Ellen Karasik, Bryan Gillard, Kimberley Ramsey, *Roswell Park Cancer Institute, Buffalo, NY*

Susan Sullivan, Jason Bridge, *Upstate New York Transplant Service, Buffalo, NY*

Comprehensive Biospecimen Resource (CBR)

Scott Jewell, Dan Rohr, Dan Maxim, Dana Filkins, Philip Harbach, Eddie Cortadillo, Bree Berghuis, Lisa Turner, Melissa Hanson, Anthony Watkins, Brian Smith, *Van Andel Institute, Grand Rapids, MI*

Pathology Resource Center (PRC)

Leslie Sobin, James Robb, *SAIC-Frederick, Inc., Frederick, MD*

Phillip Branton, *National Cancer Institute, Bethesda, MD*

John Madden, *Duke University, Durham, NC*

Jim Robb, Mary Kennedy, *College of American Pathologists, Northfield, IL*

Comprehensive Data Resource (CDR)

Greg Korzeniewski, Charles Shive, Liqun Qi, David Tabor, Sreenath Nampally, *SAIC-Frederick, Inc., Frederick, MD*

caHUB Operations Management

Steve Buia, Angela Britton, Anna Smith, Karna Robinson, Robin Burges, Karna Robinson,

Kim Valentino, Deborah Bradbury, *SAIC-Frederick, Inc., Frederick, MD*

Kenyon Erickson, *Sapient Government Services, Arlington, VA*

Brain Bank

Deborah Mash, PI; Yvonne Marcus, Margaret Basile *University of Miami School of Medicine, Miami, FL*

Laboratory, Data Analysis, and Coordinating Center (LDACC)

Kristin Ardlie, Gad Getz, co-PIs; David DeLuca, Taylor Young, Ellen Gelfand, Tim Sullivan, Yan Meng, Ayellet Segre, Jules Maller, Pouya Kheradpour, Luke Ward, Daniel MacArthur, Manolis Kellis, *The Broad Institute of Harvard and MIT, Inc., Cambridge, MA*

Statistical Methods Development (R01)

Jun Liu, co-PI, *Harvard University, Boston, MA, USA*

Jun Zhu, co-PI; Zhidong Tu, Bin Zhang, *Mt Sinai School of Medicine, New York, NY*

Nancy Cox, Dan Nicolae, co-PIs; Eric Gamazon, Haky Im, Anuar Konkashbaev, *University of Chicago, Chicago, IL*

Jonathan Pritchard, PI; Matthew Stevens, Timothée Flutre, Xiaoquan Wen, *University of Chicago, Chicago, IL*

Emmanouil T. Dermitzakis, co-PI; Tuuli Lappalainen, Pedro Ferreira, *University of Geneva, Geneva, Switzerland*

Roderic Guigo, co-PI; Jean Monlong, Michael Sammeth, *Center for Genomic Regulation, Barcelona, Spain*

Daphne Koller, co-PI; Alexis Battle, Sara Mostafavi, *Stanford University, Palo Alto, CA*

Mark McCarthy, co-PI; Manuel Rivas, Andrew Morris, *Oxford University, Oxford, United Kingdom*

Ivan Rusyn, Andrew Nobel, Fred Wright, Co-PIs; Andrey Shabalina, *University of North Carolina - Chapel Hill, Chapel Hill, NC*

US National Institutes of Health

NCBI dbGaP

Mike Feolo, Steve Sherry, Jim Ostell, Nataliya Sharopova, Anne Sturcke, *National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD*

Program Management

Leslie Derr, *Office of Strategic Coordination (Common Fund), Office of the Director, National Institutes of Health, Bethesda, MD*

Eric Green, Jeffery P. Struewing, Simona Volpi, Joy Boyer, Deborah Colantuoni, *National Human Genome Research Institute, Bethesda, MD*

Thomas Insel, Susan Koester, A. Roger Little, Patrick Bender, Thomas Lehner, *National Institute of Mental Health, Bethesda, MD*

Jim Vaught, Sherry Sawyer, Nicole Lockhart, Chana Rabiner, Joanne Demchok, *National Cancer Institute, Bethesda, MD*