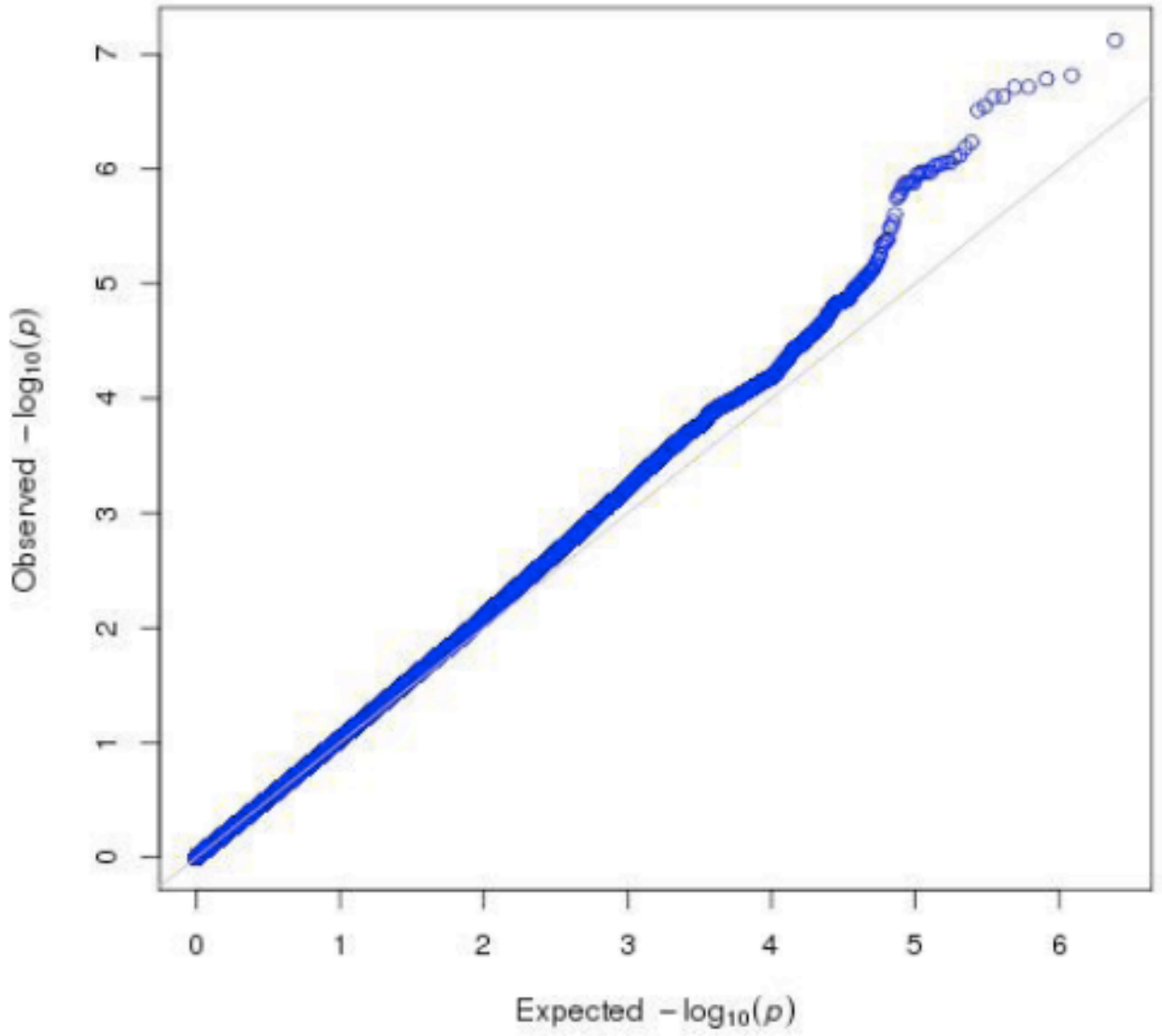
Getting to Real Data Integration Across -Omics

anderbilt Genetics Institute

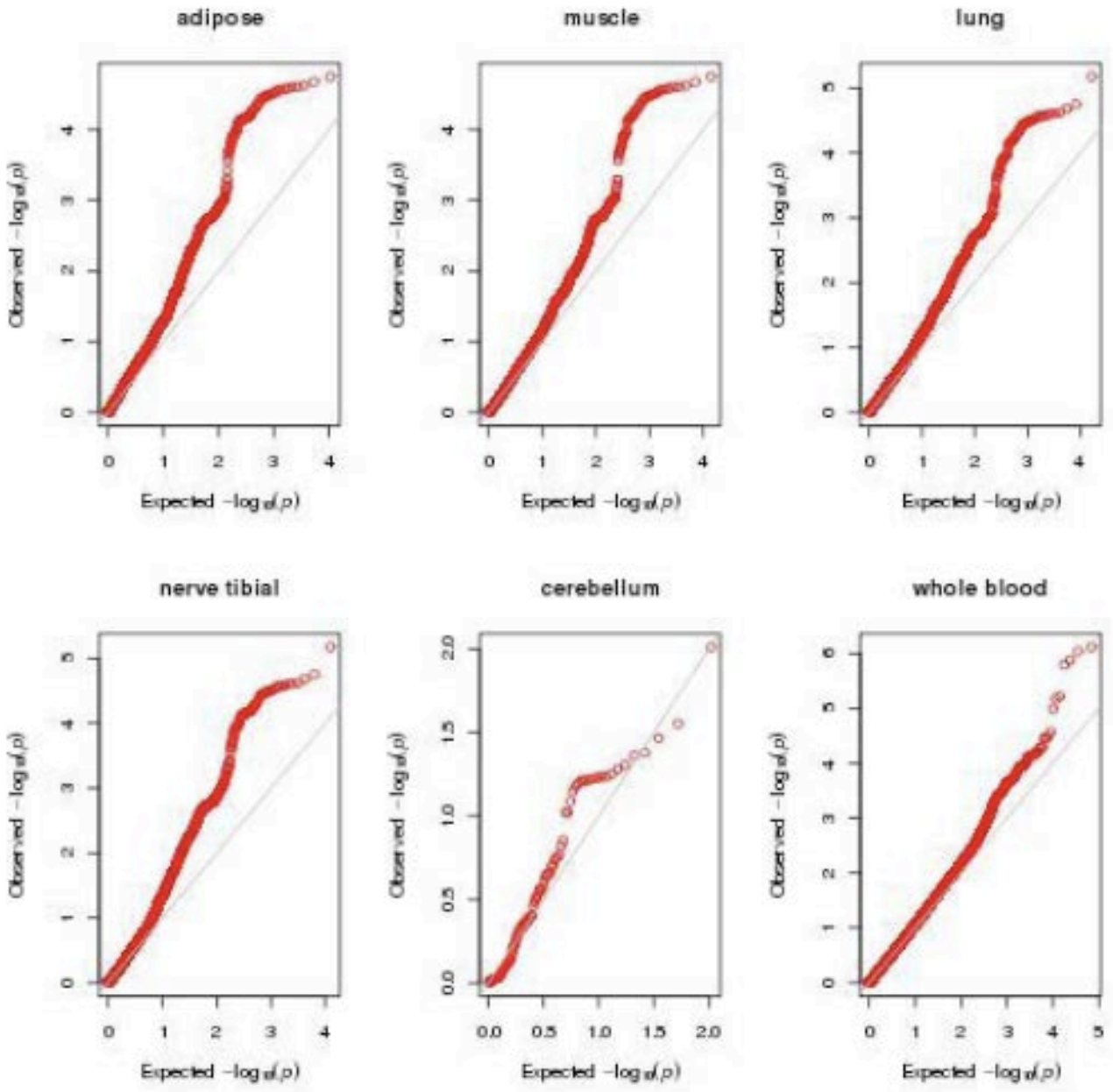
Nancy J. Cox, Ph.D.

What Does Real Data Integration Across -Omics Look Like?

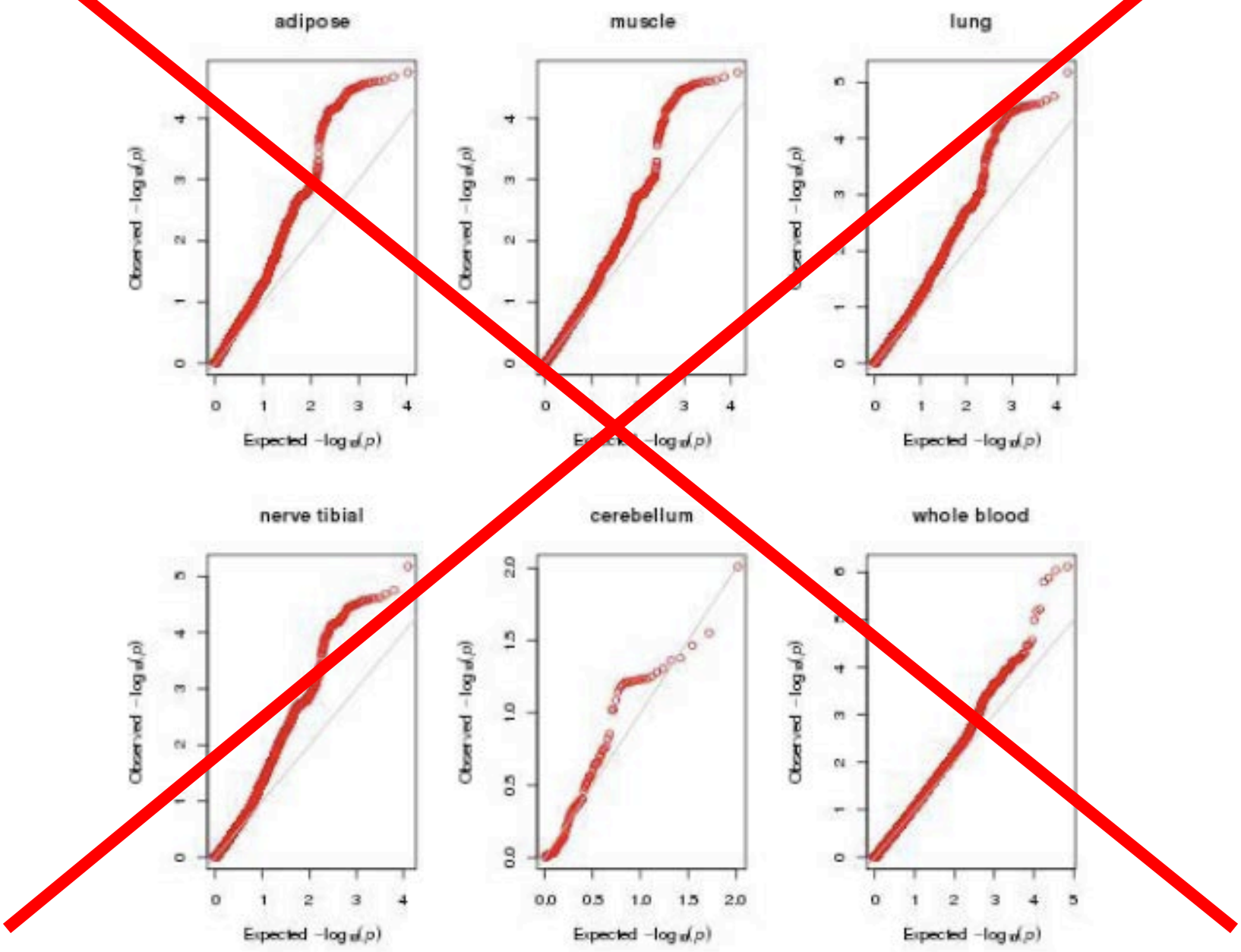
MAGIC: HOMA-IR (all SNPs)



MAGIC: HOMA-IR



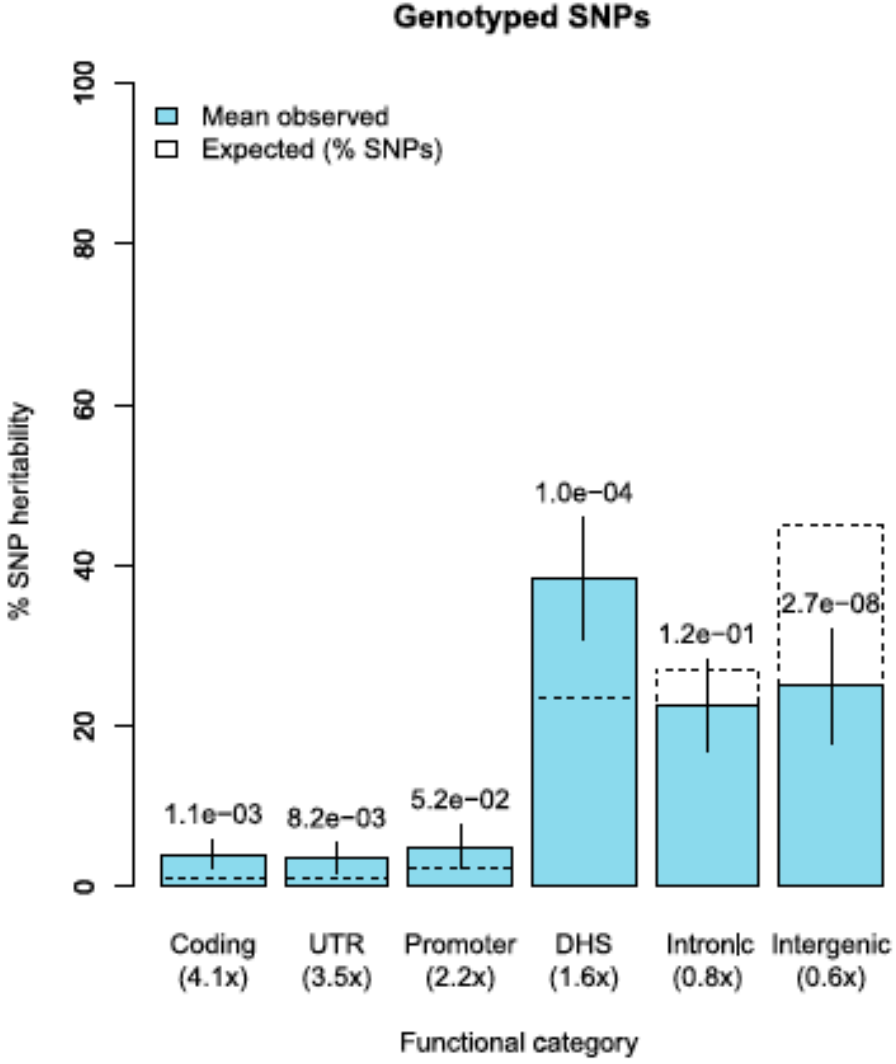
MAGIC: HOMA-IR

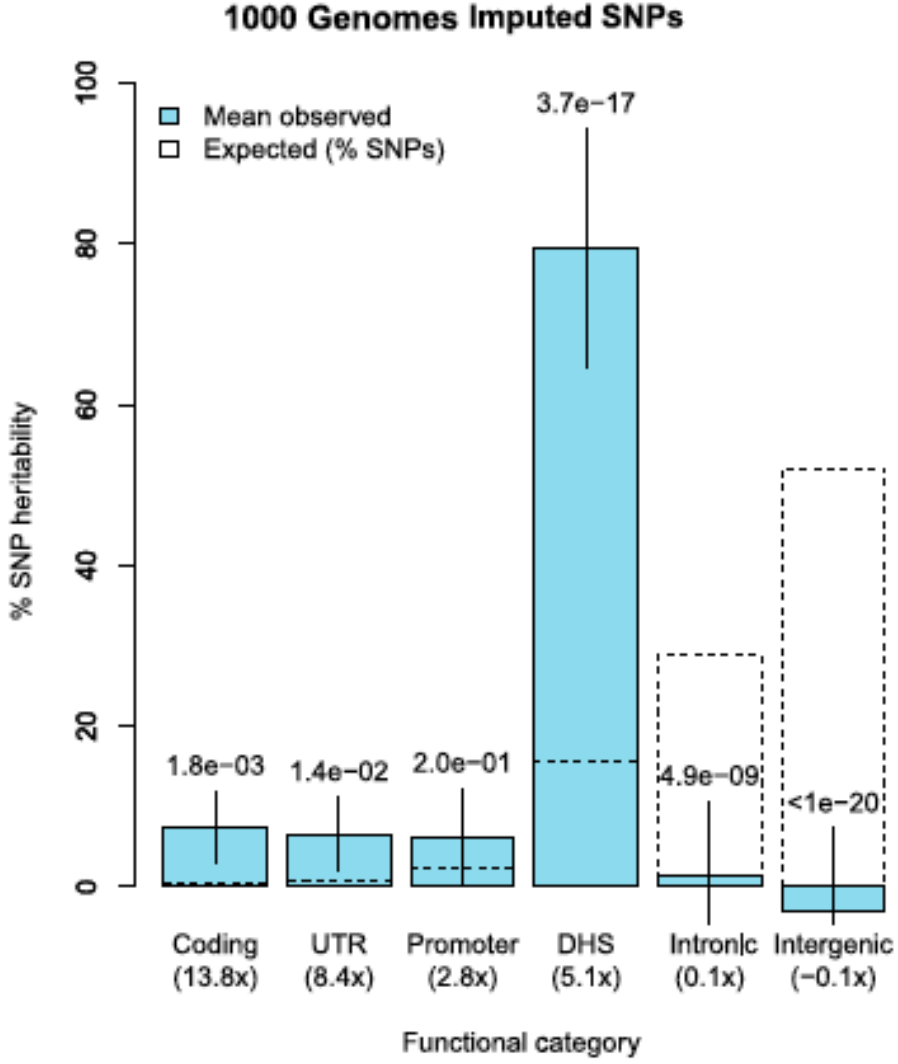


	Type 1 Diabetes			Crohns Disease	
	V(G)/V(P)	SE		V(G)/V(P)	SE
adipose	0.21	0.019		0.03	0.008
heart	0.199	0.02		0.017	0.006
lung	0.192	0.018		0.02	0.007
muscle	0.188	0.018		0.028	0.008
nerve	0.191	0.018		0.025	0.008
whole blood	0.187	0.023		0.17	0.024
Overall	**0.48**	**0.06**		**0.50**	**0.07**

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]

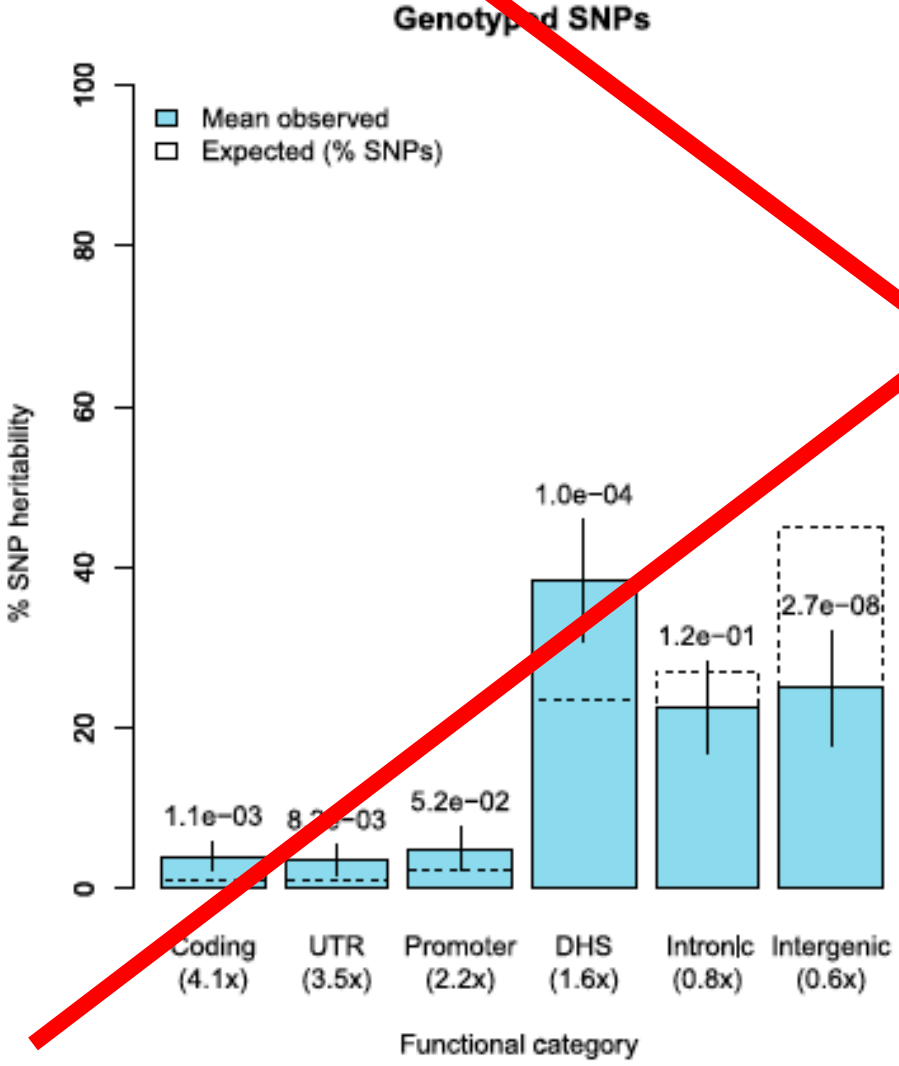


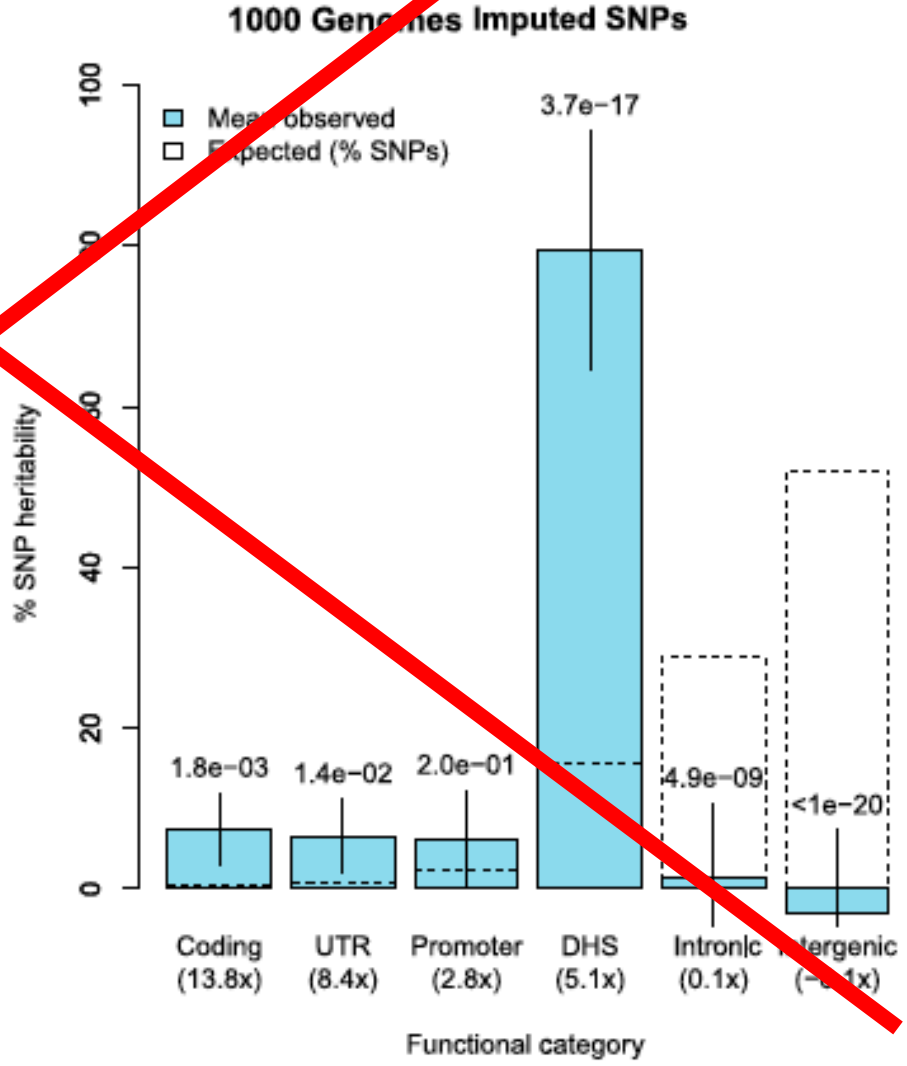


	Type 1 Diabetes		Crohns Disease	
	V(G)/V(P)	SE	V(G)/V(P)	SE
adipose	0.21	0.019	0.03	0.008
heart	0.199	0.02	0.017	0.006
lung	0.192	0.018	0.02	0.007
muscle	0.188	0.018	0.028	0.008
nerve	0.191	0.018	0.025	0.008
whole blood	0.187	0.023	0.17	0.024
Overall	**0.48**	**0.06**	**0.50**	**0.07**

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]



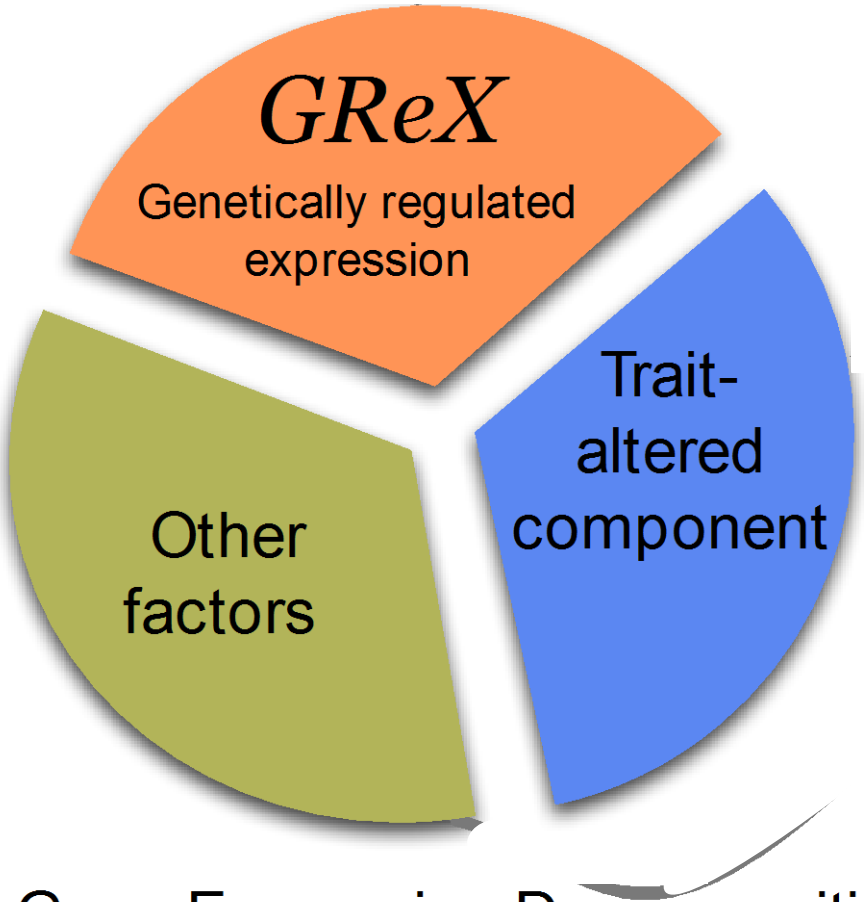


A Missing Data Problem?

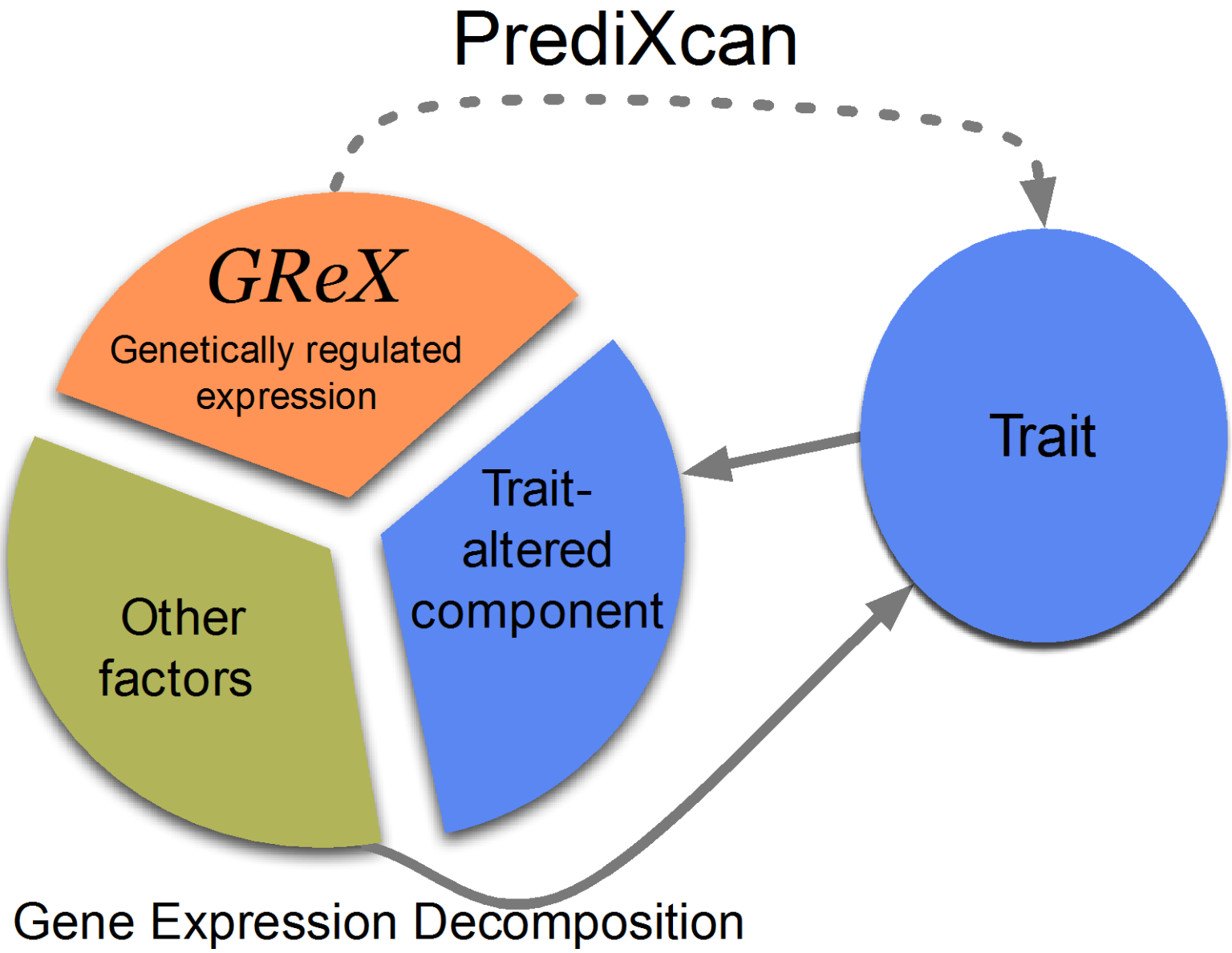
- **What we want to know:**
  - **What are the genes, the mechanisms, and the direction of effects?**
- **What we have:**
  - **Knowledge that transcriptome regulation drives much of common variant heritability, and we have measurement of genome and transcriptome variation in many tissues (GTEx) and large samples**

A Missing Data Problem?

- **If we believe that genome variation affects risk of common disease largely through transcript regulation, why not use what information we have to look directly at that endophenotype?**
- **That is – test the association of *genetically regulated* transcript levels with disease**

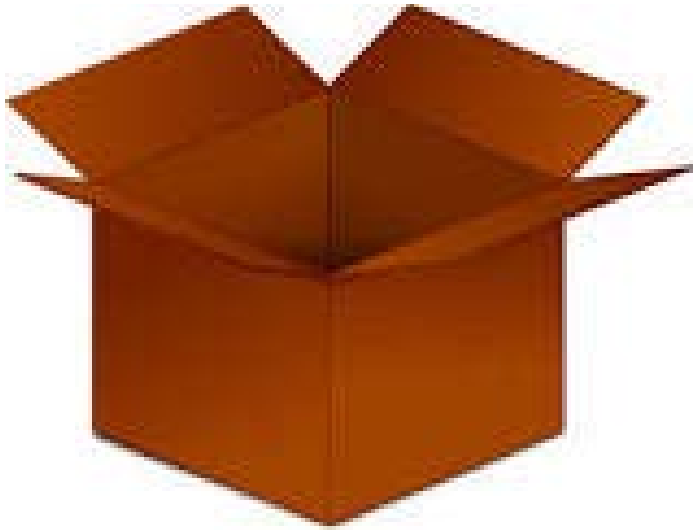


Gene Expression Decomposition

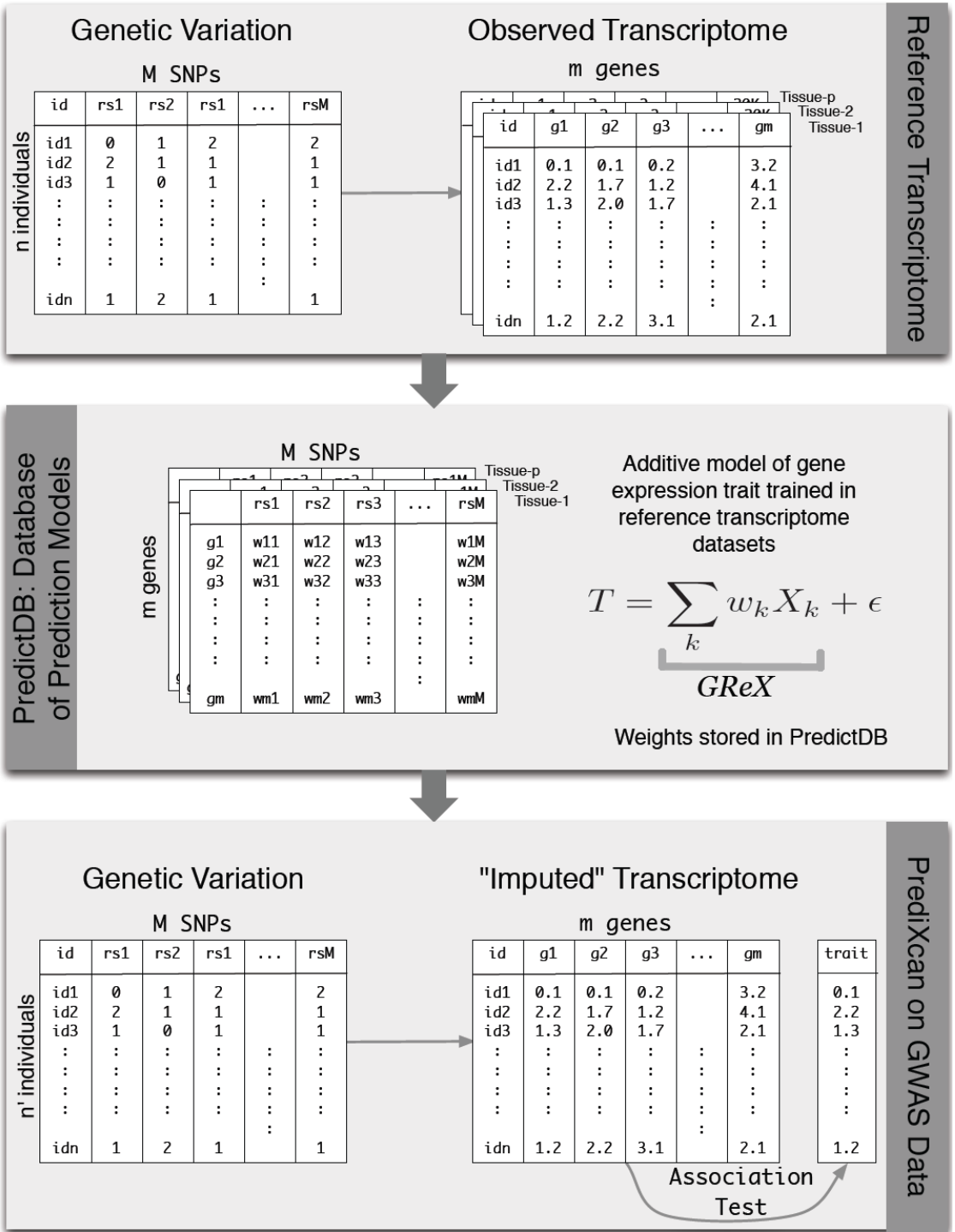

PrediXcan
GReX
Genetically regulated expression
Other factors
Trait-altered component
Trait
Gene Expression Decomposition

PrediXcan

Haky Im



Revision at Nat Genet as GTEx companion paper



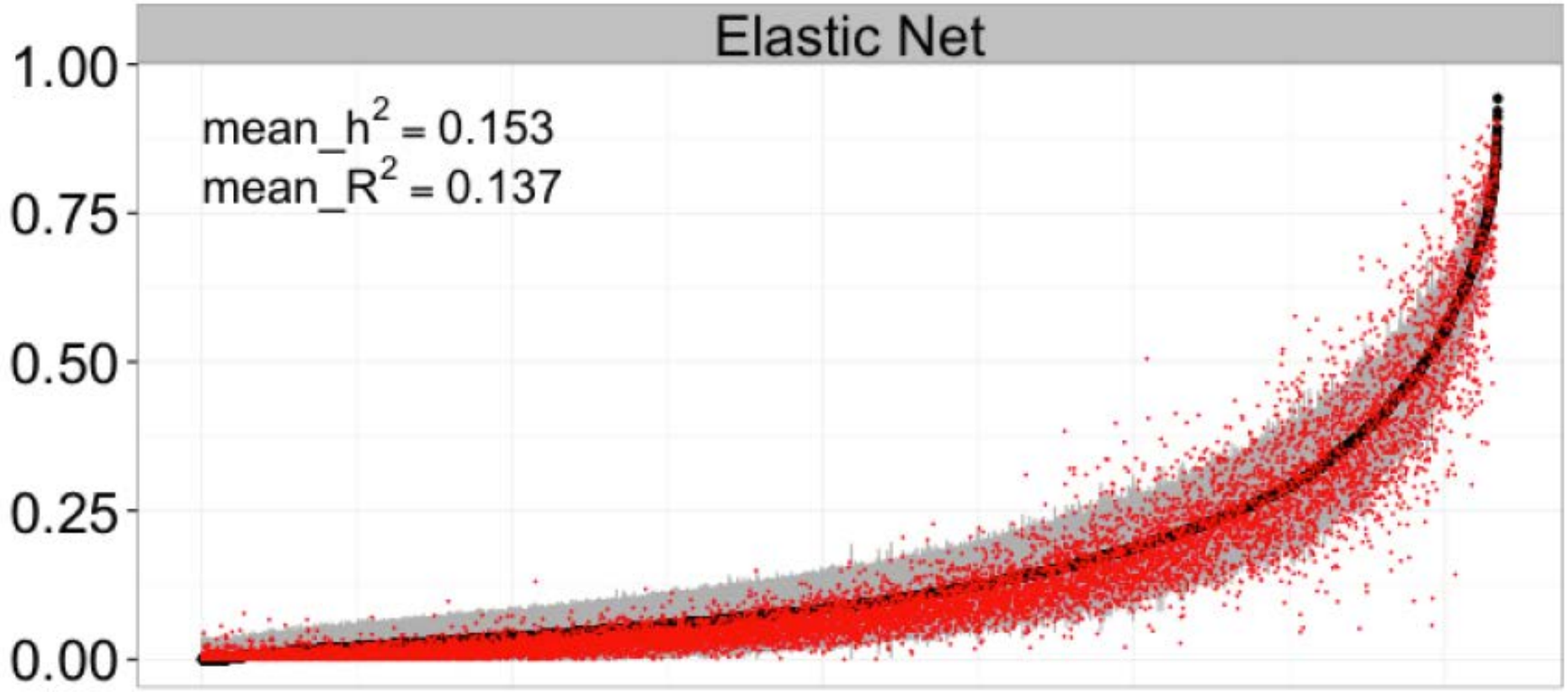
Analogous to Imputation

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

Store weights from prediction equations

Apply to any dataset with genome interrogation

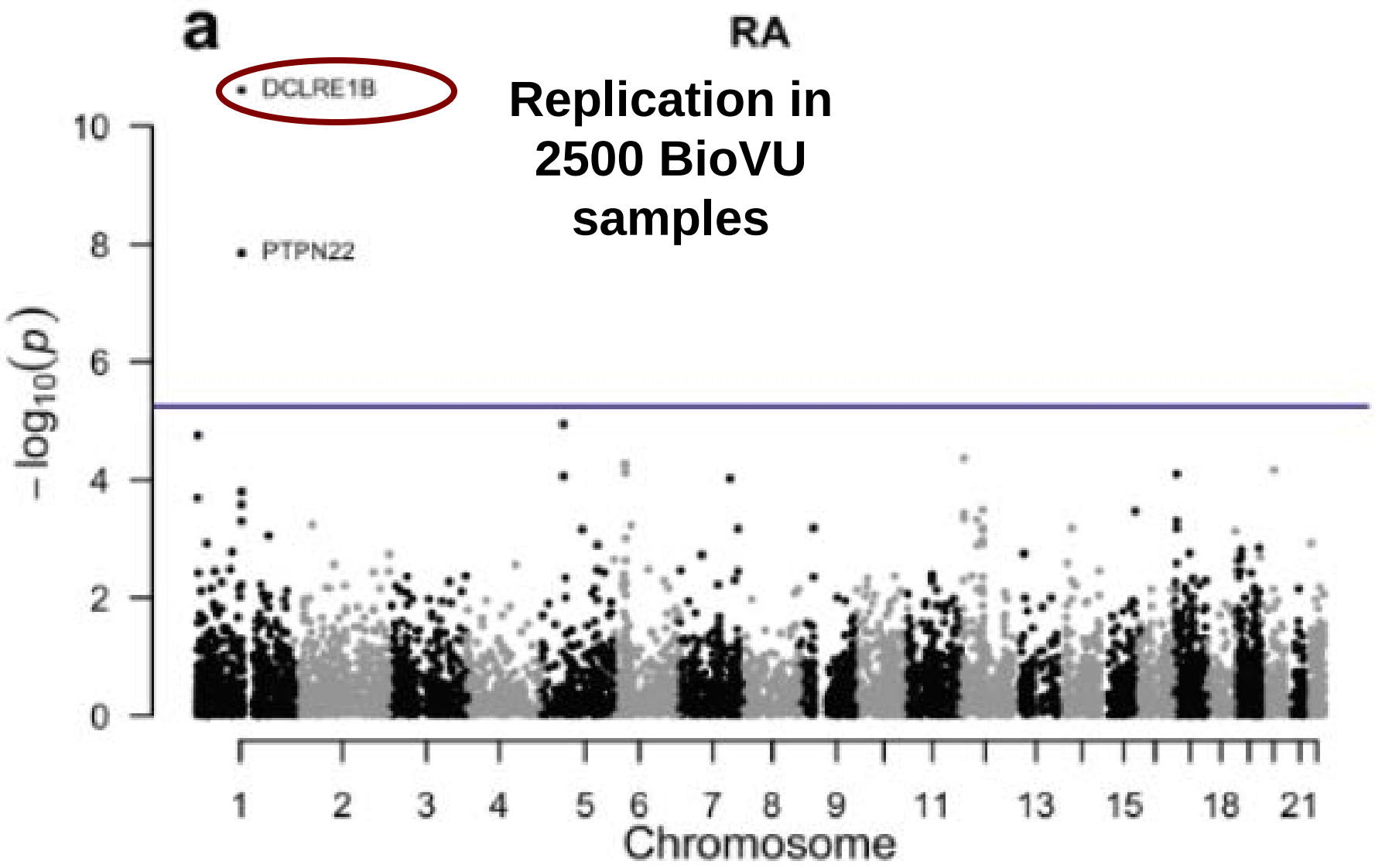
Prediction Performance R^2 by Heritability



Prediction in an Independent Sample

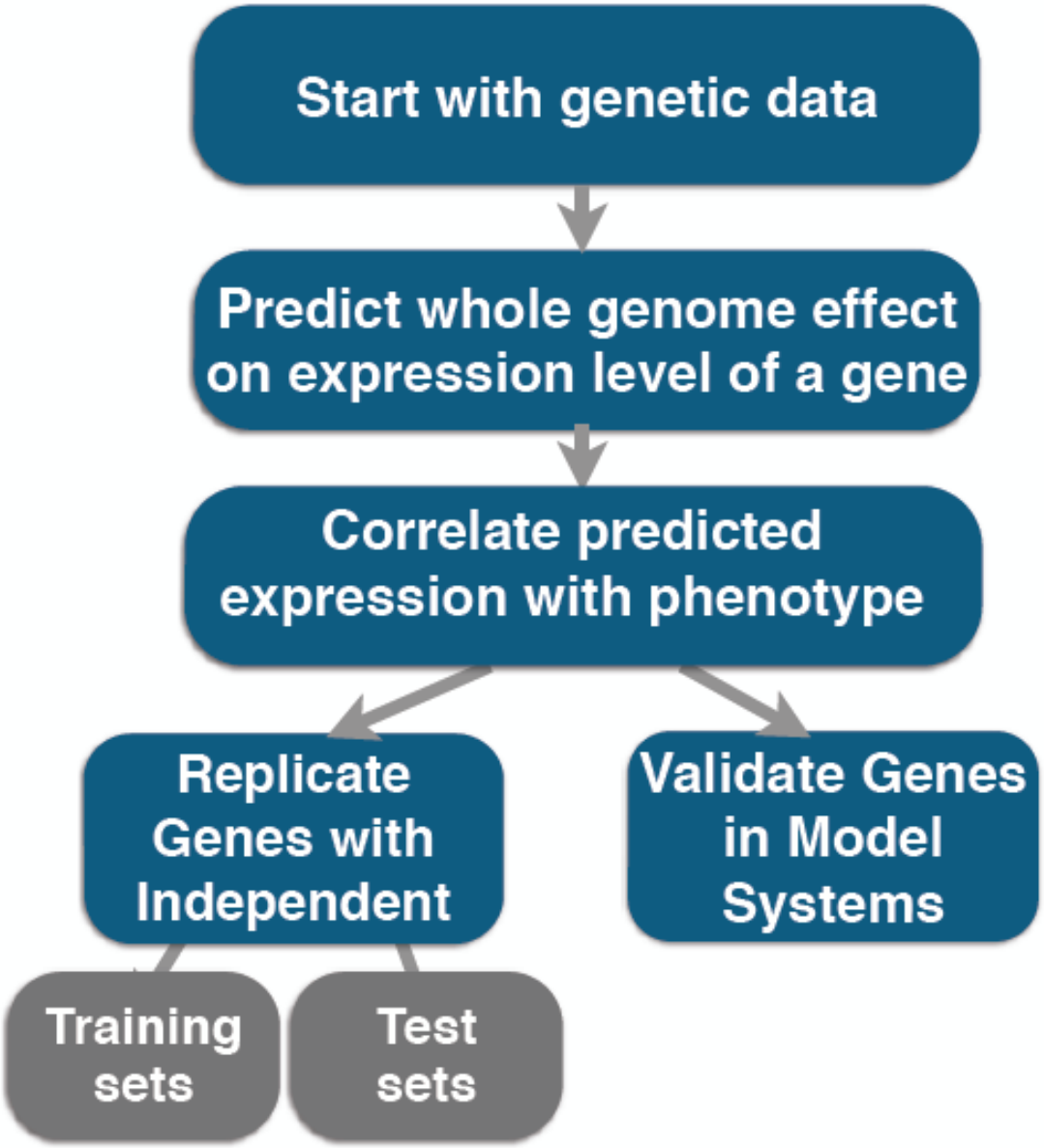
- **Significance of correlation between predicted and directly measured expression levels: q-value < 0.05 for 40-50% of genes, < 0.1 for 60-70%**
- **80% of genes have correlation between predicted and measured expression > 0.1, 50% > 0.2**
- **Polygenic prediction < {lasso, elastic net} – genetic architecture**

Rheumatoid Arthritis (RA) in WTCC

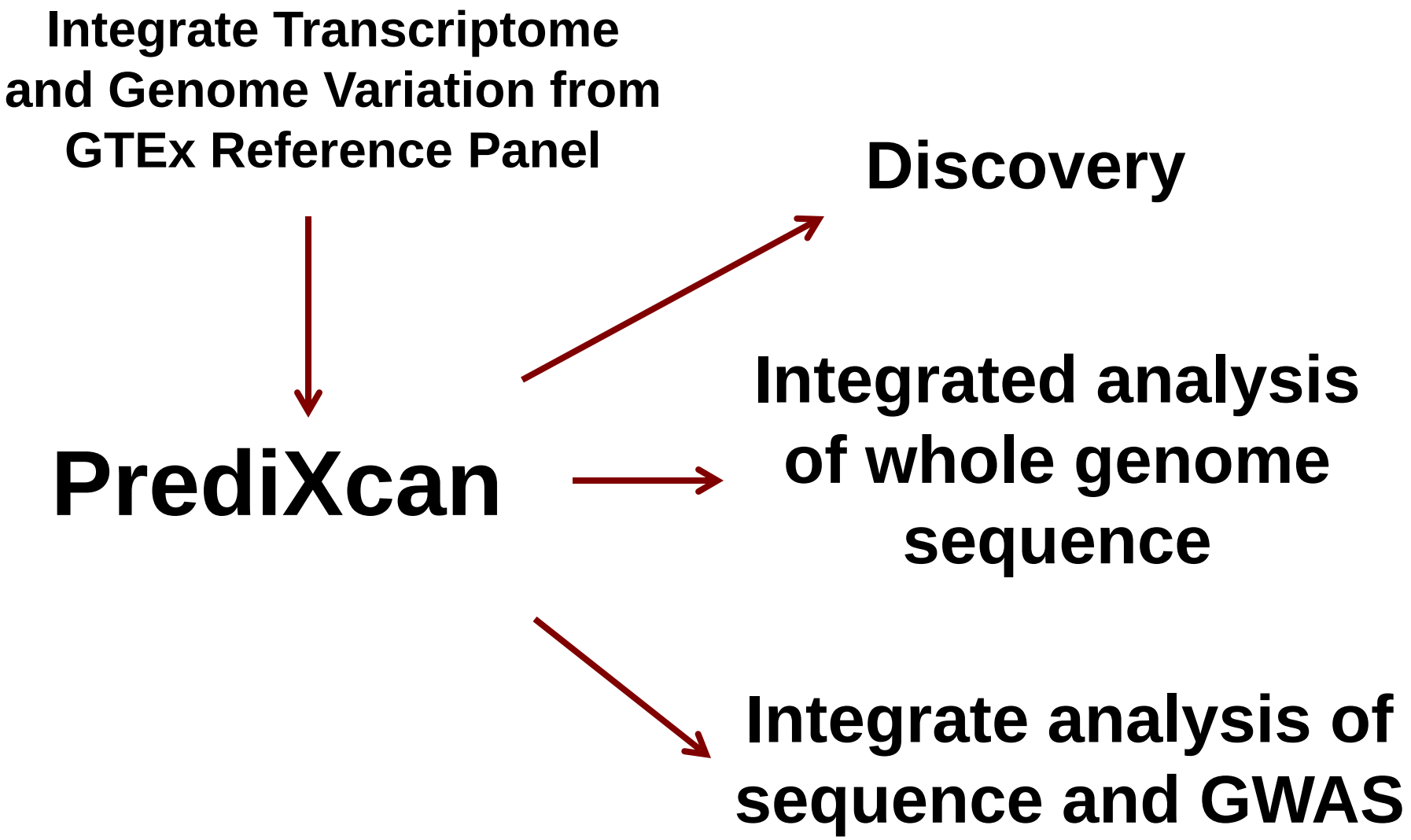


Replication in 2500 BioVU samples

PrediXcan Flow



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


Integrate Transcriptome and Genome Variation from GTEx Reference Panel
PrediXcan
Discovery
Integrated analysis of whole genome sequence
Integrate analysis of sequence and GWAS

Integrate Transcriptome and Genome Variation from GTEx Reference Panel

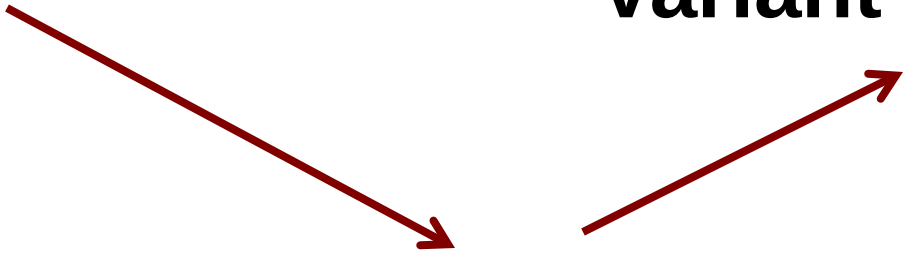
PrediXcan  **Validate and Prioritize Rare Variant Discoveries**

Requires Further Integration with a “Phenome” Reference Panel

Integrate Transcriptome and Genome Variation from GTEx Reference Panel

PrediXcan

Validate and Prioritize Rare Variant Discoveries



Integration with a “Phenome” Reference Panel

Resources for EMR-based research at Vanderbilt

The Synthetic Derivative

A de-identified and continuously-updated image of the EMR: 2,358,760 subjects

BioVU

Subjects with DNA: >200,000

- **Dense (GWAS-level) genotypes: ~20,000**
- **Exome chip data: 42,000**

Resources for EMR-based research at Vanderbilt … end 2015

The Synthetic Derivative

A de-identified and continuously-updated image of the EMR: 2,500,000 subjects

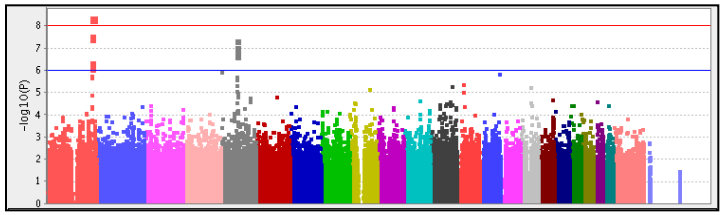
BioVU

Subjects with DNA: ~225,000

- **Dense genotypes: >100,000**
- **Exome chip data: >100,000**

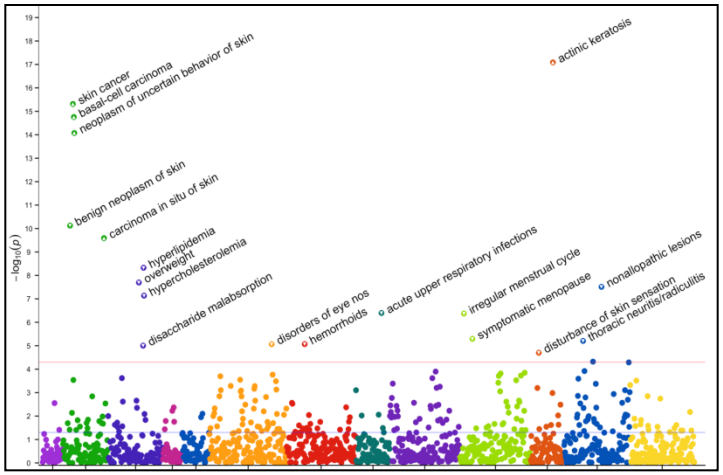
The genome-wide association study

Target phenotype → **association P value**



The phenome-wide association study

Target genotype → **association P value**



PheWAS requirement: A large cohort of patients with genotype data and many diagnoses

An Engine Enabling Basic Discovery Science

- **Test with PrediXcan the association of predicted gene expression with EMR phenotypes (gene-based PheWAS)**

An Engine Enabling Basic Discovery Science

- **Test with PrediXcan the association of predicted gene expression with EMR phenotypes (gene-based PheWAS)**

A “Phenome” Reference Panel!

Cell
Volume 155
Number 1
September 26, 2013
www.cell.com

A Nondegenerate Code of Deleterious Variants in Mendelian Loci Contributes to Complex Disease Risk

Blair DR, Lyttle CS, Mortensen JM, Bearden CF, Jensen AB, Khiabanian H, Melamed R, Rabadan R, Bernstam EV, Brunak S, Jensen LJ, Nicolae D, Shah NH, Grossman RL, Cox NJ, White KP, Rzhetsky A

Continuum from Mendelian to Complex

Continuum from LOF to deleterious to ↓expression

What Phenotypes Are Associated with Reduced GReX of Mendelian Genes?

PrediXcan X BioVU

- **Preliminary studies on those genes in which all SNPs included in the prediction equation are included in the Illumina 1M SNP set (~5000 BioVU subjects)**
- **Studies completed for whole blood (built in >900 from DGN) and cardiac tissue (built from > 300 in GTEx)**
- **125 genes**

PrediXcan in BioVU for *PEX19*

- **Mutations in *PEX19* lead to a peroxisomal biogenesis disorder, Zellweger Syndrome spectrum of Mendelian phenotypes**
  - **Hypotonic, seizures, bony stippling at the patella and other long bones, kidney and liver cysts, coagulopathies, stone formation and renal failure**
  - **What BioVU phenotypes are associated with reduced GReX of *PEX19*?**

Reduced GReX of *PEX19*

758	Chromosomal anomalies, genetic disorders	27	0.000179
401.2	Hypertensive heart and/or renal disease	1137	0.00473
401.22	Hypertensive chronic kidney disease	823	0.0064
110.13	Dermatophytosis of the body	28	0.0065
458.1	Orthostatic hypotension	141	0.0109
345.1	Epilepsy	148	0.0111
587	Kidney replaced by transplant	696	0.0122
443.1	Raynaud's syndrome	26	0.0136
701.4	Keloid scar	21	0.0161
275.5	Disorders of calcium/phosphorus metabolism	428	0.0174
252.1	Hyperparathyroidism	89	0.0177
585.32	End stage renal disease	516	0.0263
585.3	Chronic renal failure [CKD]	1235	0.0273
252	Disorders of parathyroid gland	103	0.0277
720	Spinal stenosis	299	0.0309
353.2	Nerve root lesions	26	0.0314
800.4	Fracture of patella	49	0.0315
345.12	Partial epilepsy	81	0.0386
420	Carditis	261	0.0401
531.2	Gastric ulcer	81	0.0441
345	Epilepsy, recurrent seizures, convulsions	424	0.0483
411.41	Aneurysm and dissection of heart	25	0.0487

Increased GReX *PEX19*

170.2	Cancer of connective tissue	120	0.00128
170	Cancer of bone and connective tissue	170	0.00219
371.3	Inflammation of eyelids	118	0.00397
781	Symptoms involving nervous and musculoskeletal systems	55	0.00759
198.1	Secondary malignancy of lymph nodes	268	0.00833
627.22	Need for Hormone replacement therapy (postmenopausal)	25	0.00868
287.32	Secondary thrombocytopenia	144	0.0107
371	Inflammation of the eye	237	0.0117
290.13	Senile dementia	24	0.015
174.1	Breast cancer [female]	189	0.0168
90	Sexually transmitted infections (not HIV or hepatitis)	20	0.018
174.11	Malignant neoplasm of female breast	187	0.0199
759	Other and unspecified congenital anomalies	54	0.021
277.7	Dysmetabolic syndrome X	40	0.0214
174	Breast cancer	208	0.024
170.1	Bone cancer	103	0.0257
772.1	Muscular wasting and disuse atrophy	20	0.0269
352.2	Facial nerve disorders [CN7]	36	0.0314
228.1	Hemangioma of skin and subcutaneous tissue	23	0.0382
611.3	Lump or mass in breast	156	0.0403
426.2	Atrioventricular [AV] block	269	0.0406
627.2	Symptomatic menopause	235	0.0409
442.11	Abdominal aortic aneurysm	105	0.0421
352	Disorders of other cranial nerves	79	0.0474

PrediXcan in BioVU for *TK2*

- ***TK2*-related mitochondrial DNA depletion syndrome, myopathic form (TK2-MDS) is an inherited condition that causes progressive muscle weakness (myopathy).**
- **The signs and symptoms of TK2-MDS typically begin in early childhood. Development is usually normal early in life, but as muscle weakness progresses, people with TK2-MDS lose motor skills such as standing, walking, eating, and talking. Some affected individuals have increasing weakness in the muscles that control eye movement, leading to droopy eyelids (progressive external ophthalmoplegia).**
- **Study in Finland noted increase in fractures. Due to increase in falls caused by myopathy? Bones weak?**

Reduced Predicted Expression *TK2*

728.1	Muscular calcification and ossification	21	0.00734
990	Effects radiation NOS	72	0.00861
496.1	Emphysema	163	0.0097
872	Traumatic amputation	38	0.0127
800.3	Fracture of tibia and fibula	145	0.0182
807	Fracture of ribs	118	0.0202
809	Fracture of unspecified bones	287	0.0207
733	Other disorders of bone and cartilage	218	0.0235
596.5	Functional disorders of bladder	164	0.0252
374.3	Ptosis of eyelid	50	0.0314
800	Fracture of lower limb	375	0.0318
627.21	Symptomatic artificial menopause	29	0.0325
962	Poisoning by hormones and synthetic substitutes	87	0.0372
819	Skull and face fracture and other intercranial injury	130	0.0378
149.5	Hx of malignant neoplasm of oral cavity and pharynx	58	0.04
716.9	Arthropathy NOS	402	0.0407
800.2	Fracture of unspecified part of femur	111	0.0422
196	Radiotherapy	123	0.0436
716	Other arthropathies	418	0.0488

Increased Predicted Expression *TK2*

270	Disorders of protein plasma/amino-acid transport and metabolism	180	1.50E-05
270.3	Disorders of plasma protein metabolism	152	1.00E-04
270.32	Paraproteinemia	98	0.00147
270.38	Other specified disorders of plasma protein metabolism	40	0.00261
550.6	Incisional hernia	149	0.00314
286.8	Hypercoagulable state	79	0.00316
204.4	Multiple myeloma	125	0.00527
510.2	Lung transplant	59	0.00574
480	Pneumonia	1311	0.00646

Does Altered Expression of Mendelian Disease Genes Contribute Disproportionately to Common Disease?

If Reduced GReX of Mendelian Genes Is Associated with Components of Mendelian Disease, Can't We …

- **Test genes implicated in sequencing studies on rare disorders for phenotypes associated with GReX**
  - **"Undiagnoses Disease" sequencing often yields 6-12 genes at which LOF or deleterious mutations could be causing disease; prioritize and validate**
- **Test genes implicated in sequencing studies on common diseases that fail to meet genome-wide criteria**
- **"Predict" new Mendelian diseases**

Increased Predicted Expression *CCKBR* cholecystokinin B receptor

575	Other biliary tract disease	176	1.66E-05
411.8	Other chronic ischemic heart disease, unspecified	567	0.000915
575.8	Other disorders of biliary tract	98	0.00115
513	Respiratory abnormalities	424	0.00228
362.26	Macular puckering of retina	62	0.00232
593	Hematuria	476	0.00375
297	Suicidal ideation or attempt	50	0.00391
70.3	Viral hepatitis C	135	0.00415
70	Viral hepatitis	225	0.00459
204.12	Lymphoid leukemia, chronic	67	0.00608
362.2	Degeneration of macula and posterior pole of retina	193	0.00663
297.1	Suicidal ideation	25	0.00671
594	Urinary calculus	320	0.007
573	Other disorders of liver	802	0.00701
300.9	Posttraumatic stress disorder	78	0.00954

***CCKBR* is a receptor for regulatory peptides of the brain and gastrointestinal tract**

Increased Predicted Expression *GRIK5*

361	Retinal detachments and defects	54	0.000629
366	Cataract	629	0.000642
365	Glaucoma	219	0.00105
379	Other disorders of eye	233	0.00131
250.6	Polyneuropathy in diabetes	276	0.0014
365.11	Primary open angle glaucoma	72	0.00153
365.1	Open-angle glaucoma	150	0.00226
79	Viral infection	246	0.00379
627	Menopausal and postmenopausal disorders	365	0.00401
250.3	Insulin pump user	449	0.00422
530.1	Esophagitis, GERD and related diseases	1408	0.00455
366.2	Senile cataract	530	0.00507
627.2	Symptomatic menopause	235	0.0052
476	Allergic rhinitis	527	0.00525
379.2	Disorders of vitreous body	188	0.00627
530	Diseases of esophagus	1551	0.00636
763	Thoracic or lumbosacral neuritis or radiculitis, unspecified	134	0.00649
362	Other retinal disorders	321	0.00739
613	Other nonmalignant breast conditions	99	0.00752
577.3	Cyst and pseudocyst of pancreas	40	0.00756
530.11	GERD	1268	0.00812
514.2	Solitary pulmonary nodule	20	0.00831

An Eye Super Gene?

Reduced Predicted Expression *ST6GALNAC4*

295	Schizophrenia and other psychotic disorders	145	0.000174
295.3	Psychosis	119	0.000257
343	Infantile cerebral palsy	32	0.000573
292.5	Transient alteration of awareness	25	0.0016
242	Thyrotoxicosis with or without goiter	121	0.00165
381.1	Otitis media	141	0.00312
264	Lack of normal physiological development	254	0.00353
381.11	Suppurative and unspecified otitis media	112	0.00379
291.8	Alteration of consciousness	489	0.00495
291	Other specified nonpsychotic and/or transient mental disorders	502	0.00578
264.2	Failure to thrive	210	0.00676
175	Acquired absence of breast	50	0.00897
381	Otitis media and Eustachian tube disorders	243	0.0109
303	Psychogenic and somatoform disorders	44	0.0187
320	Meningitis	66	0.0192
601.1	Prostatitis	50	0.0221
345.1	Epilepsy	148	0.0273
174.11	Malignant neoplasm of female breast	187	0.0274

Results on 125 genes in 5000 individuals

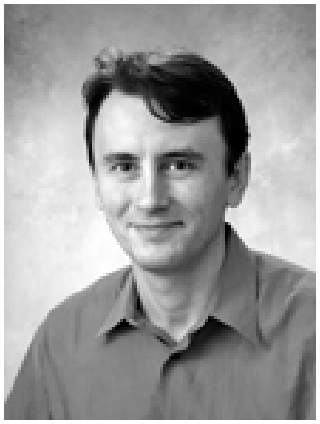
Results on all genes in 20,000

Results in 100,000+

An Engine Enabling Basic Discovery Science

- Conduct systematic evaluation of animal model knock-out phenotypes with human
- Test predicted expression of genes targeted by drugs for phenotypes related to ADRs

Our GTEx Team



Dan Nicolae

Lin Chen

Lea Davis (&Bridget)

Richard Jones

Eric Gamazon

Hae Kyung "Haky" Im

Anuar Konkashbaev

Barbara Stranger

Younghee Lee


W615
NASHVILLE

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, Timothèe 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 Regulaton, 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 Shabalin, *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*