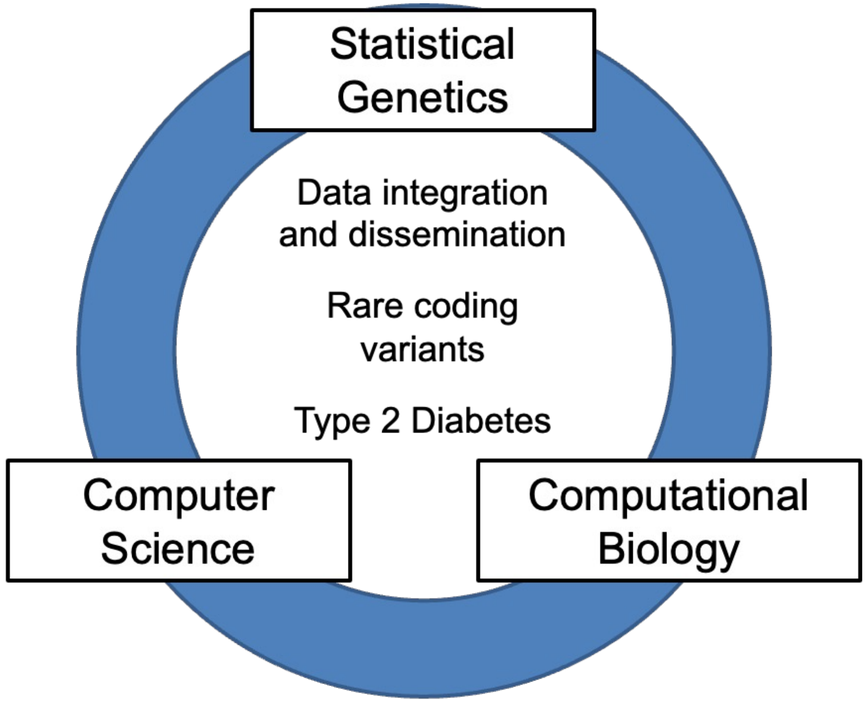
Making Human Genetic Data More Useful

Jason Flannick
flannick@broadinstitute.org
flannicklab.org

BROAD INSTITUTE

Boston Children's Hospital
Until every child is well™

HARVARD MEDICAL SCHOOL



BROAD INSTITUTE

Boston Children's Hospital
Until every child is well™

HARVARD MEDICAL SCHOOL

FlannickLab













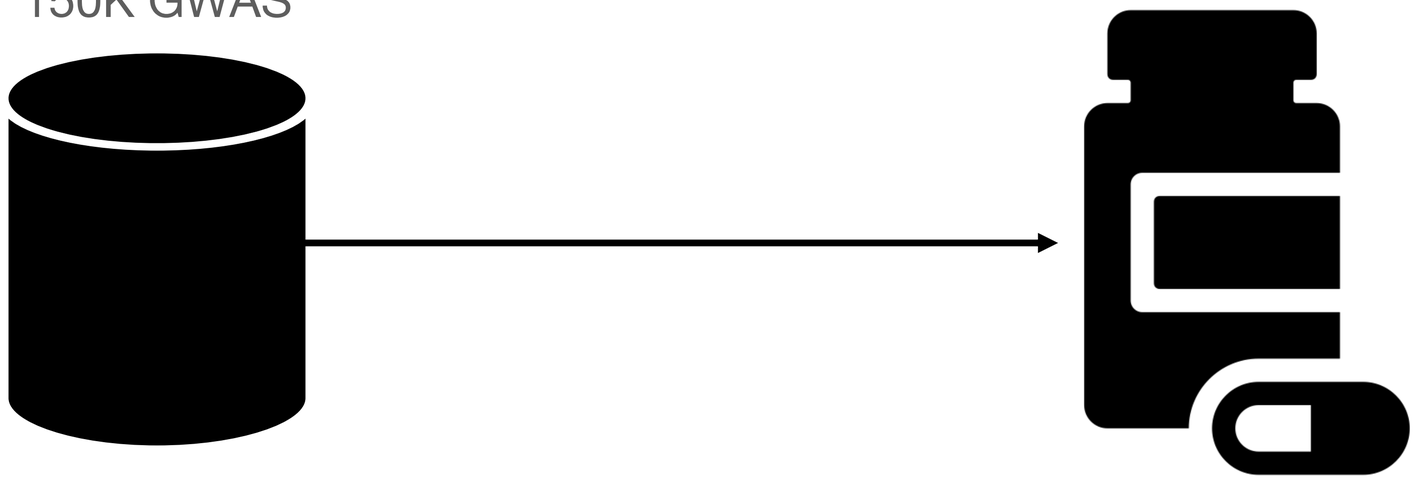

In 2015…

Type 2 Diabetes

26K exomes

80K exome chips

150K GWAS



52K target prioritization

Overview presentation

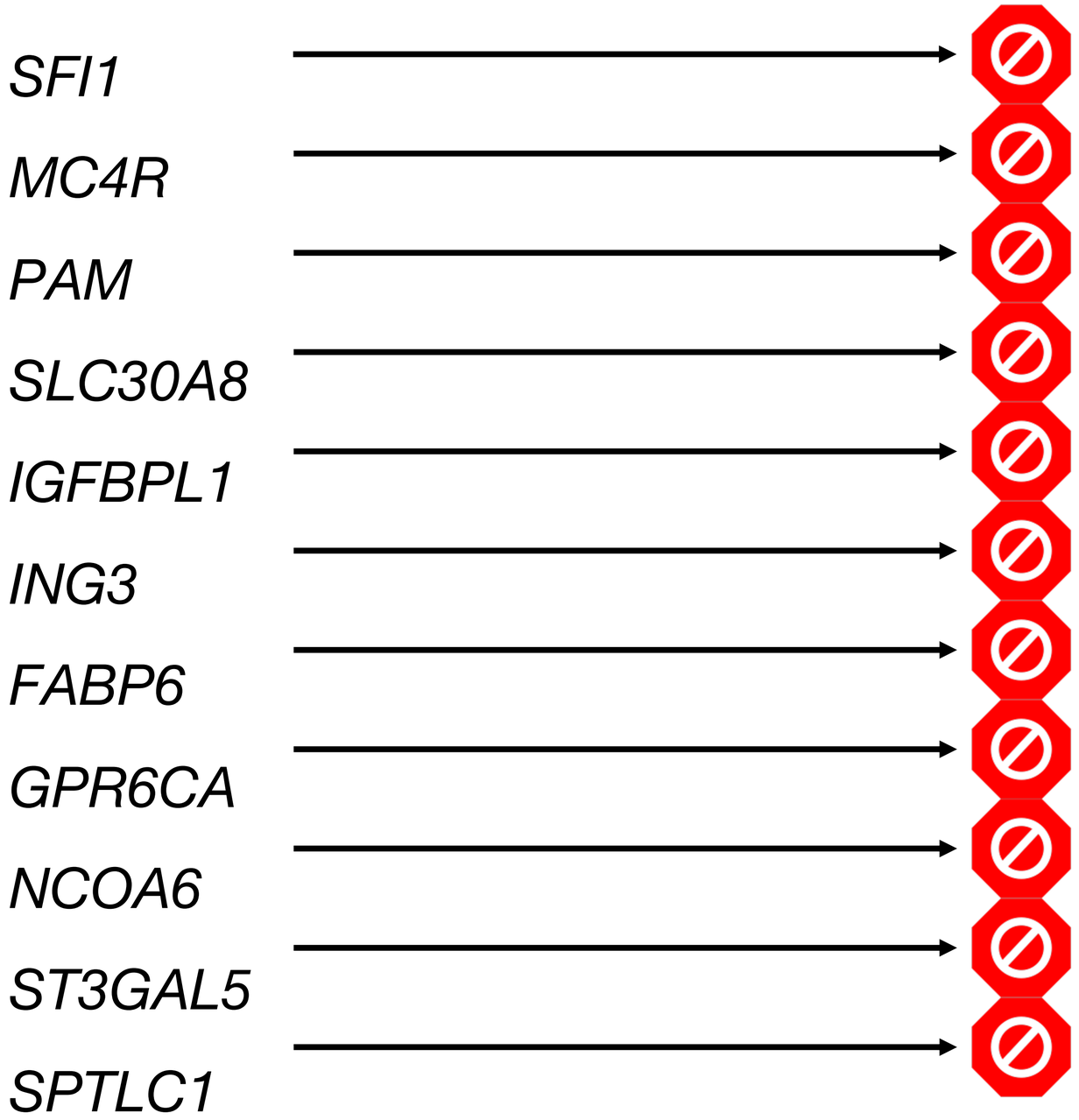
6/9/17

52K target prioritization

Appendix

6/9/17

Table 1.	Unbiased	Exome-W	
Gene	Entrez		
Symbol	Gene ID	Variant	K


SFI1
MC4R
PAM
SLC30A8
IGFBPL1
ING3
FABP6
GPR6CA
NCOA6
ST3GAL5
SPTLC1

Why were these data not useful?

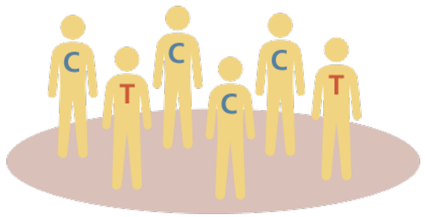
And what can we do about it?

Back in 2008…

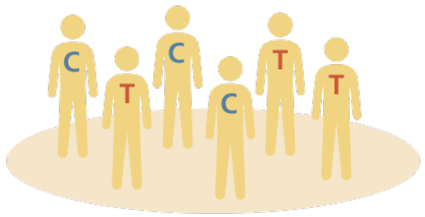
The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Genetic associations

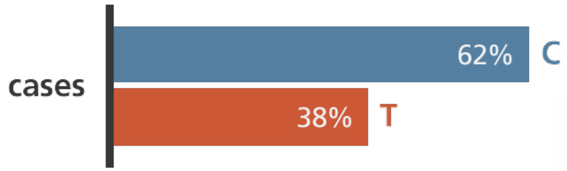


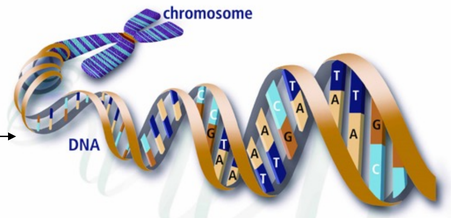
cases (n=1,000)
people with heart disease



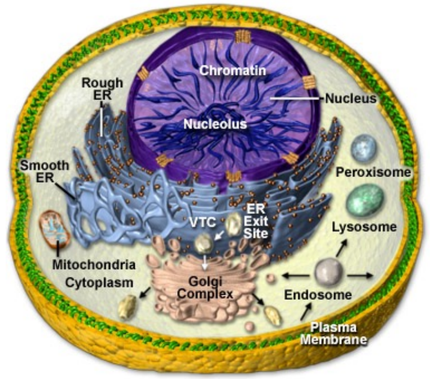
controls (n=1,000)
people without heart disease

Genome Research Limited





Disease Associated Region



GWAS

- Unbiased interrogation of all **common** variants in the genome
  - Responsible for ~90% of variants in any one individual
  - Analyze all variants by only genotyping <500k

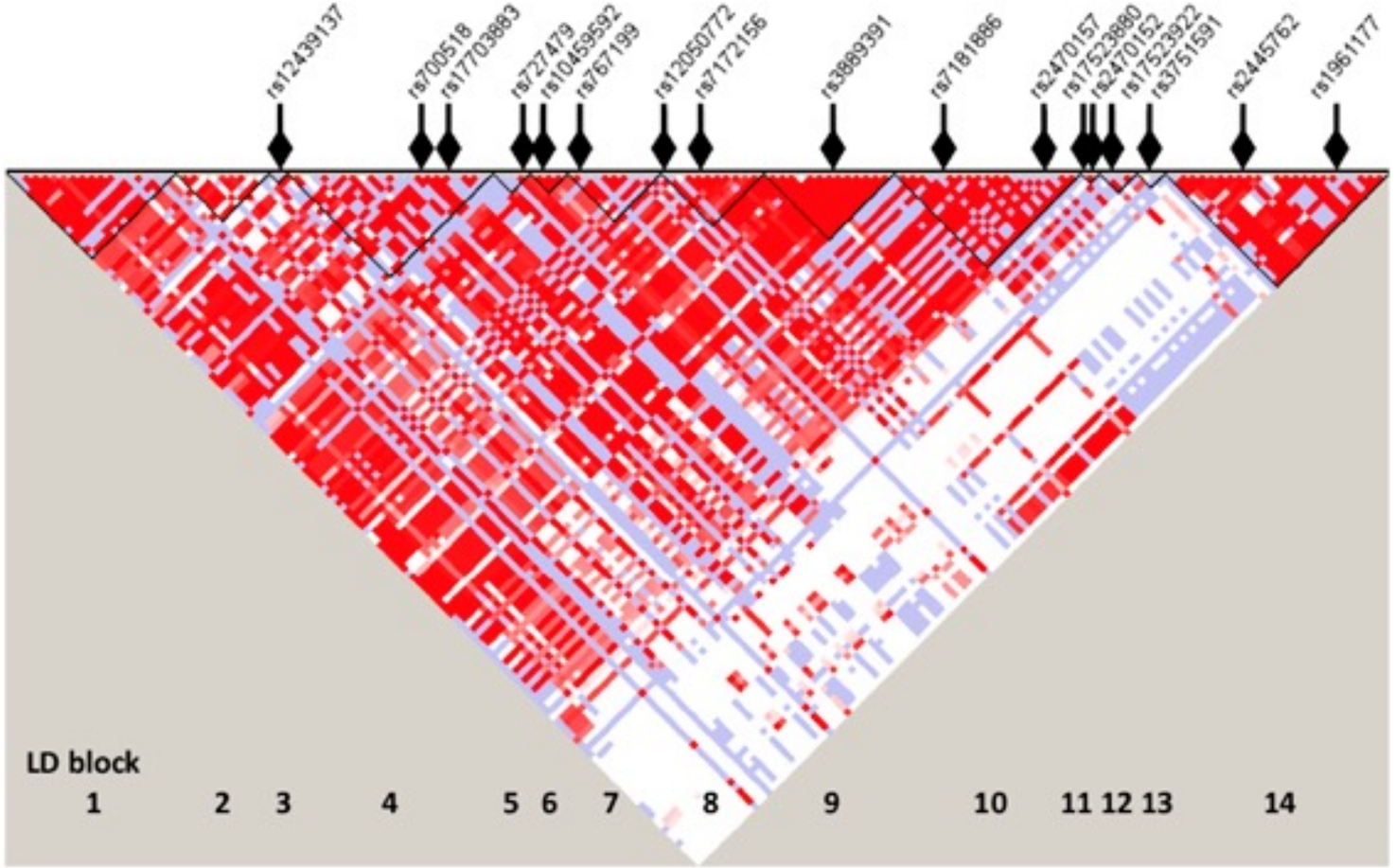


Table 1 | Estimates of heritability and number of loci for several complex traits

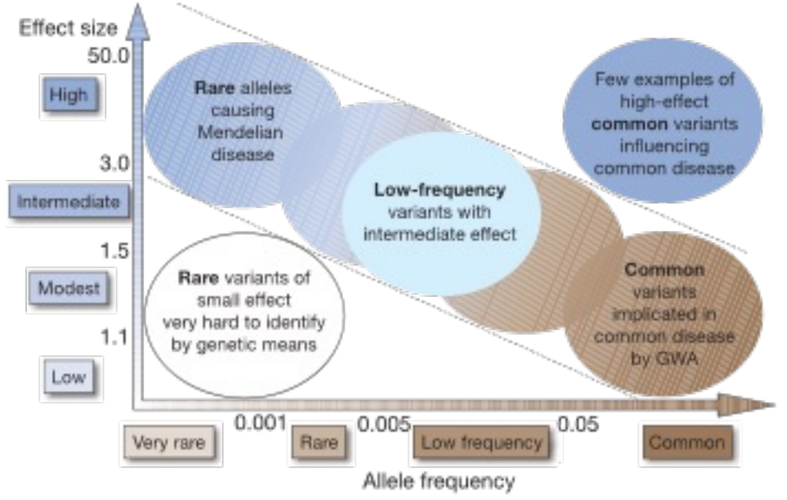
Disease	Number of loci	Proportion of heritability explained	Heritability measure
Age-related macular degeneration[72]	5	50%	Sibling recurrence risk
Crohn's disease[21]	32	20%	Genetic risk (liability)
Systemic lupus erythematosus[73]	6	15%	Sibling recurrence risk
Type 2 diabetes[74]	18	6%	Sibling recurrence risk
HDL cholesterol[75]	7	5.2%	Residual* phenotypic variance
Height[15]	40	5%	Phenotypic variance
Early onset myocardial infarction[76]	9	2.8%	Phenotypic variance
Fasting glucose[77]	4	1.5%	Phenotypic variance

* Residual is after adjustment for age, gender, diabetes.

Manolio et al, Nature 2009

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.



Are Rare Variants Responsible for Susceptibility to Complex Diseases?

Jonathan K. Pritchard

Department of Statistics, University of Oxford, Oxford

Common Genetic Variation and Human Traits

David B. Goldstein, Ph.D.

Genetic Heterogeneity in Human Disease

Jon McClellan[1,*] and Mary-Claire King[2,*]
[1]Department of Psychiatry
[2]Departments of Medicine and Genome Sciences
University of Washington, Seattle, WA 98195-7720, USA
*Correspondence: drjack@uw.edu (J.M.), mcking@uw.edu (M.-C.K.)
DOI 10.1016/j.cell.2010.03.032

Clan Genomics and the Complex Architecture of Human Disease

James R. Lupski,[1,2,3,*] John W. Belmont,[1,2] Eric Boerwinkle,[4,5] and Richard A. Gibbs[1,5,*]

OPEN ACCESS Freely available online

PLoS BIOLOGY

Rare Variants Create Synthetic Genome-Wide Associations

Samuel P. Dickson[1,2], Kai Wang[3], Ian Krantz[3,4,5], Hakon Hakonarson[3,4,5], David B. Goldstein[1]*

Common and rare variants in multifactorial susceptibility to common diseases

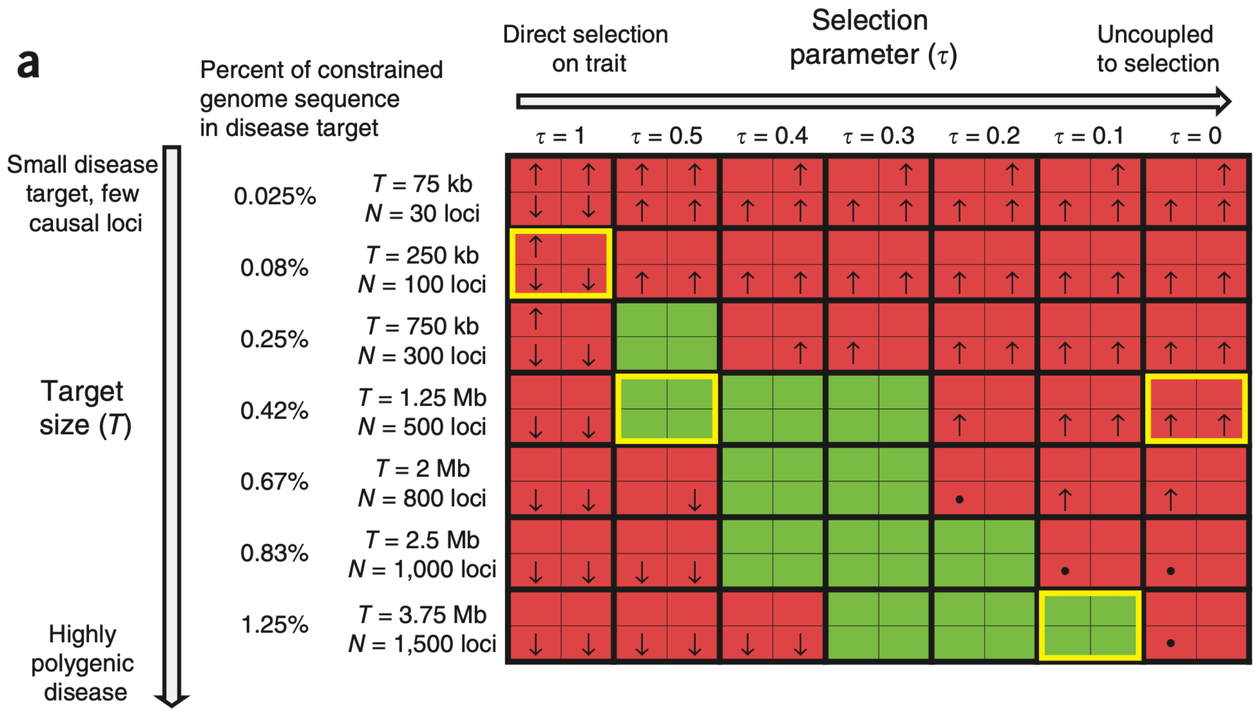
Walter Bodmer & Carolina Bonilla

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Uncovering the roles of rare variants in common disease through whole-genome sequencing

Elizabeth T. Cirulli and David B. Goldstein

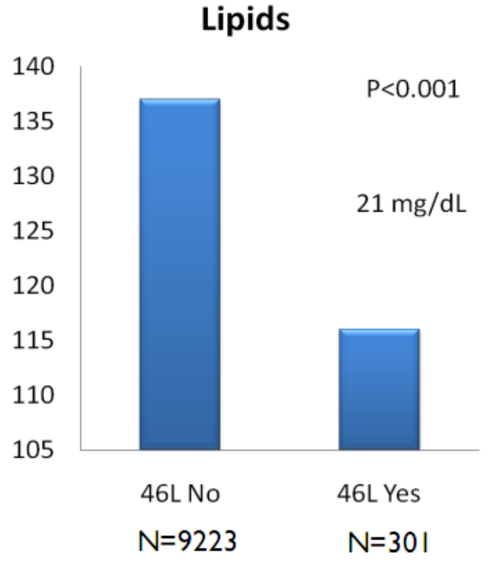
Are rare variants responsible for a significant fraction of disease heritability?



Could rare "high impact" variants suggest biological or clinical insights?

Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*

Jonathan Cohen[1–3], Alexander Pertsemlidis[2,3], Ingrid K Kotowski[4], Randall Graham[1], Christine Kim Garcia[1–3] & Helen H Hobbs[1–4]



Outcome

Hazard Ratio for incident CHD:

0.50 (0.32 – 0.79)

P=0.003

1108 events/9524

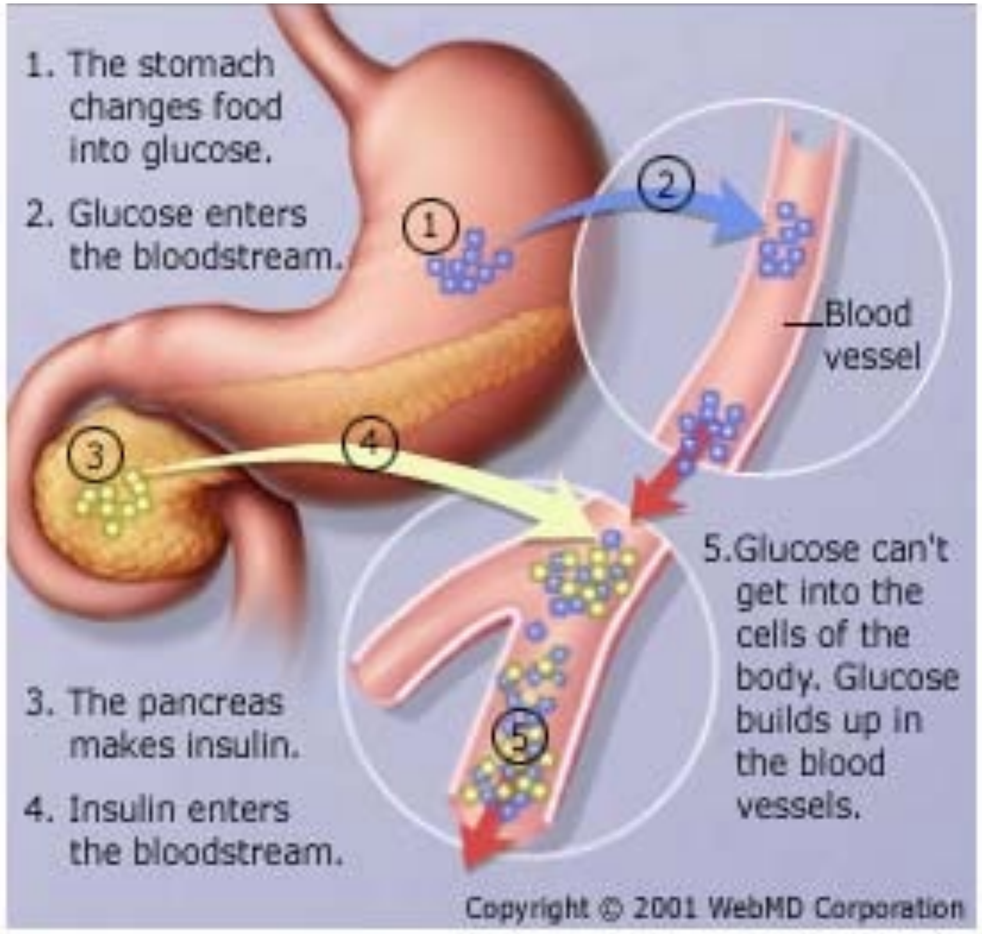
Cohen, *N Engl J Med* 2006
Kathiresan, *N Engl J Med* 2008

Type 2 diabetes is the most common metabolic disease

- Affects ≈415m people (10% of population) with prevalence increasing

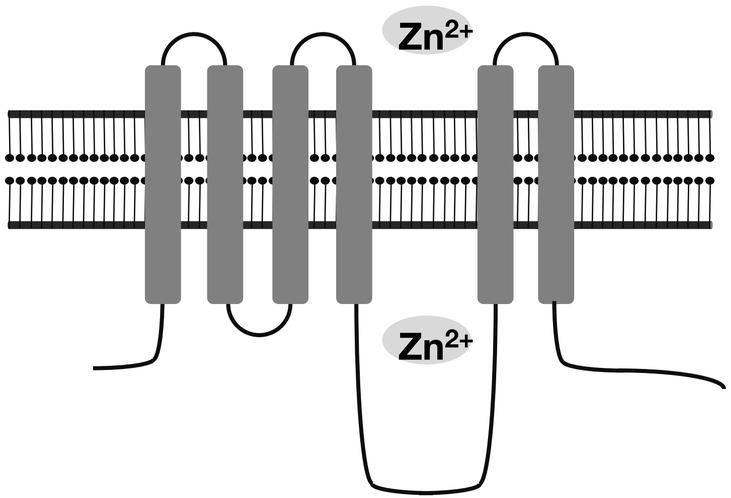
- T2D is characterized by high blood glucose due to insulin resistance and insulin deficiency

- Leading cause of cardiovascular disease, kidney failure, blindness, and premature death



Example: SLC30A8

- Encodes zinc transporter (ZnT8) expressed in pancreatic islets
  - Zinc enables proper processing, storage, and secretion of insulin
- GWAS association with unclear direction of effect



How would we target it?

Study	Alteration	Protein activity	Zinc content	Insulin secretion	Glucose levels
Chimienti 2006	Overexpression *in vitro*	↑	↑	↑	
Sladek 2007 ***Dupuis 2010*** ***Nicolson 2009***	**p.R325W**	↓	↓	—	↑
Pound 2009	Global mouse deletion	↓	↓	↓	—
Lemaire 2009	Global mouse deletion	↓	↓	—	↑ *
Nicolson 2009	Global mouse deletion	↓	↓	↑ **	↑ **
Wijesekara 2010	Beta-cell mouse deletion	↓	↓	↓	↑ ***
Pound 2012	**Global mouse deletion**	↓	↓	—	—

Note: Table is a crude simplification

* Glucose intolerant after high fat feeding
** Dependent on age, sex, and diet
*** Normal fasting glucose but glucose intolerant

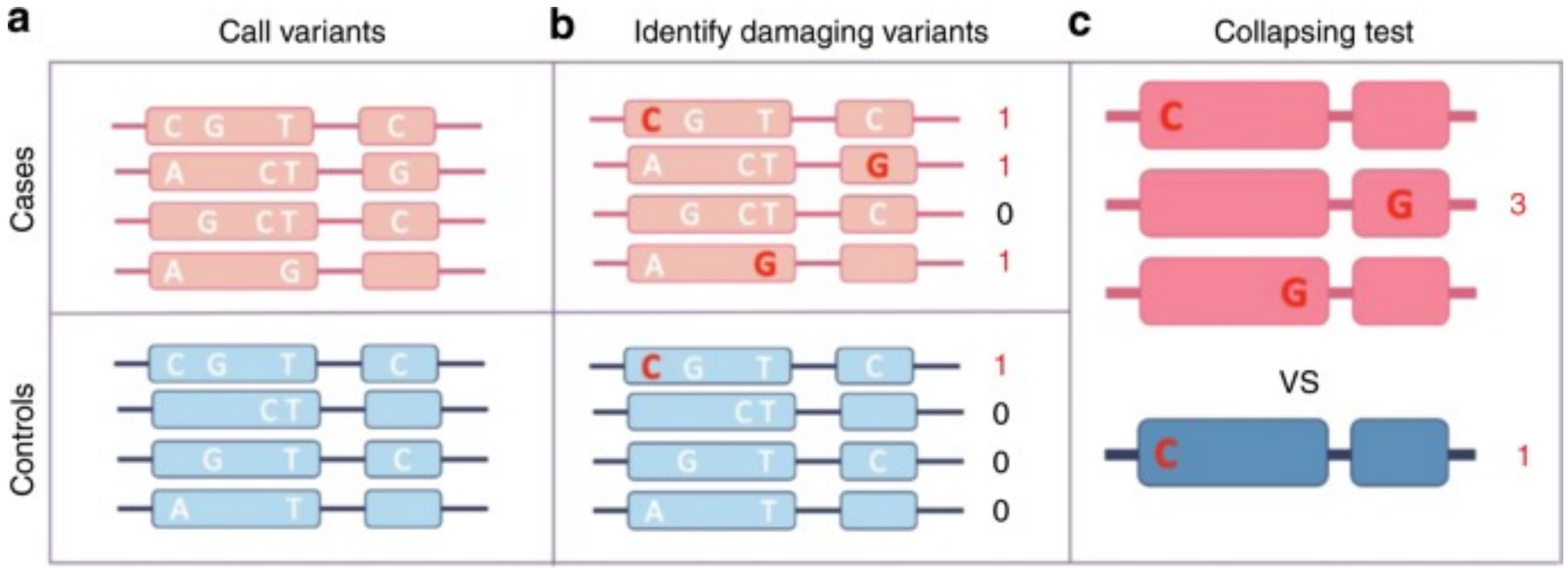
Identified from targeted sequencing

12 loss-of-function *SLC30A8* mutations in **149,134** individuals

Flannick et al. Nature Genetics, 2014

Rare variants in complex disease: gene-level associations

- Reflect aggregate effects of all variants in a gene

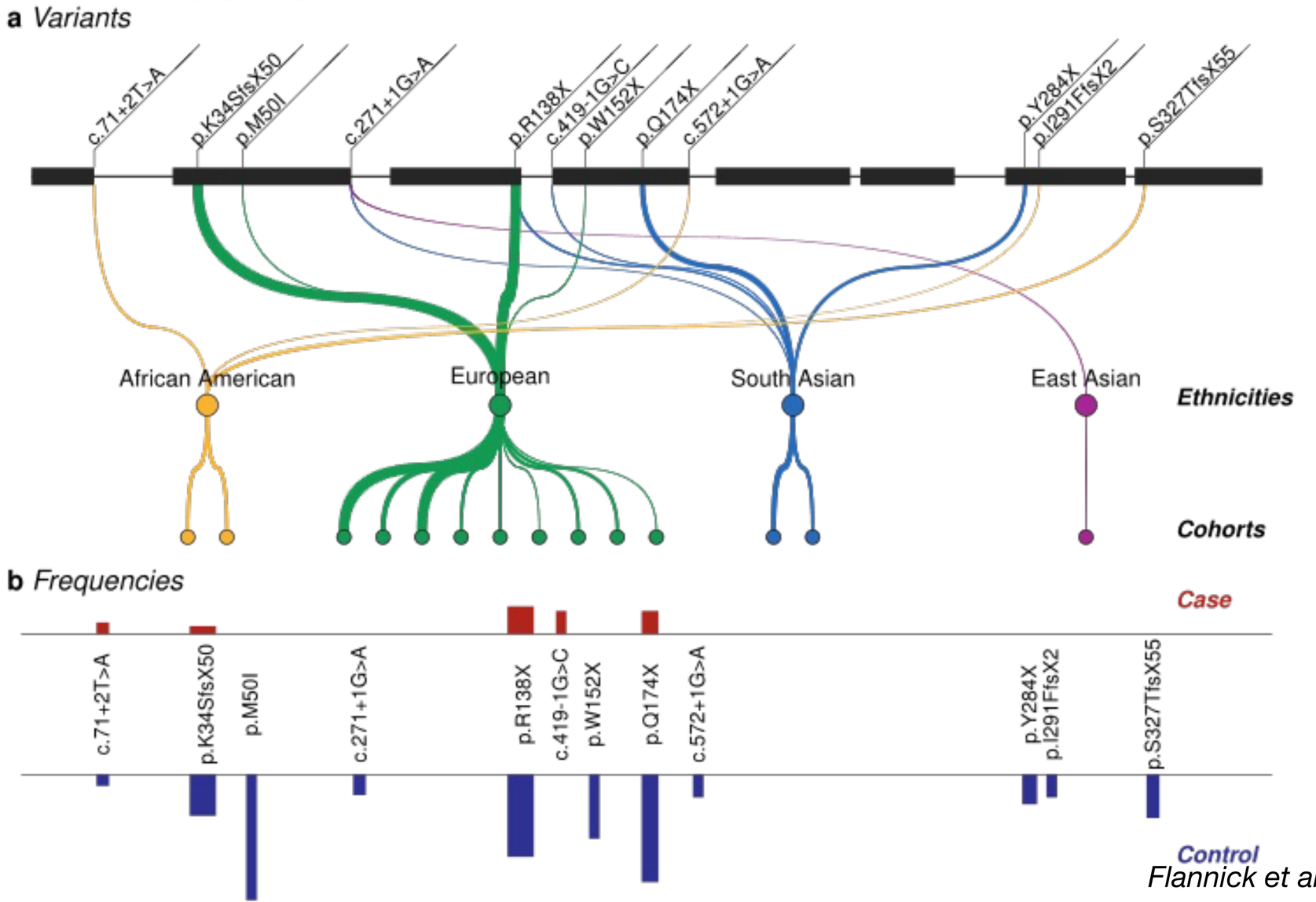


Cirulli et al, Nature Communications, 2020

An early success from targeted sequencing

12 loss-of-function *SLC30A8* mutations in **149,134** individuals

Aggregate odds ratio: **0.34** p=**1.7 x 10^{-6}**



Flannick et al. Nature Genetics, 2014

A more complete look

Fuchsberger, Flannick, Teslovich, Mahajan, Gaulton, Agarwala et al, Nature 2016

Coding variation

13,000 samples from diverse ancestries

Exome sequencing

Noncoding variation

2,700 samples of European ancestry

Whole genome sequencing

Follow-up

82,000 samples

Exome array genotyping

44,000 samples

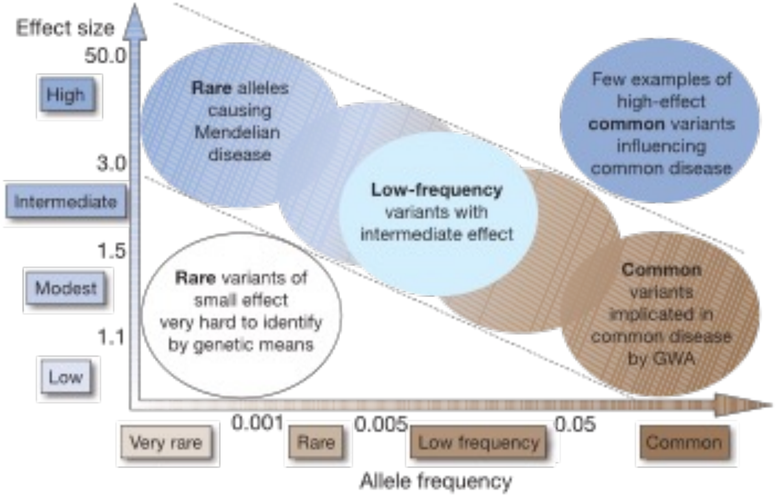
WGS imputation

Main finding:

a polygenic, common variant model for T2D

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.



Are Rare Variants Responsible for Susceptibility to Complex Diseases?

Jonathan K. Pritchard

Department of Statistics, University of Oxford, Oxford

Common Genetic Variation and Human Traits

David B. Goldstein, Ph.D.

Genetic Heterogeneity in Human Disease

Jon McClellan[1,*] and Mary-Claire King[2,*]
[1]Department of Psychiatry
[2]Departments of Medicine and Genome Sciences
University of Washington, Seattle, WA 98195-7720, USA
*Correspondence: drjack@uw.edu (J.M.), mcking@uw.edu (M.-C.K.)
DOI 10.1016/j.cell.2010.03.032

Clan Genomics and the Complex Architecture of Human Disease

James R. Lupski,[1,2,3,*] John W. Belmont,[1,2] Eric Boerwinkle,[4,5] and Richard A. Gibbs[1,5,*]

OPEN ACCESS Freely available online

PLoS BIOLOGY

Rare Variants Create Synthetic Genome-Wide Associations

Samuel P. Dickson[1,2], Kai Wang[3], Ian Krantz[3,4,5], Hakon Hakonarson[3,4,5], David B. Goldstein[1]*

APPLICATIONS OF NEXT-GENERATION SEQUENCING

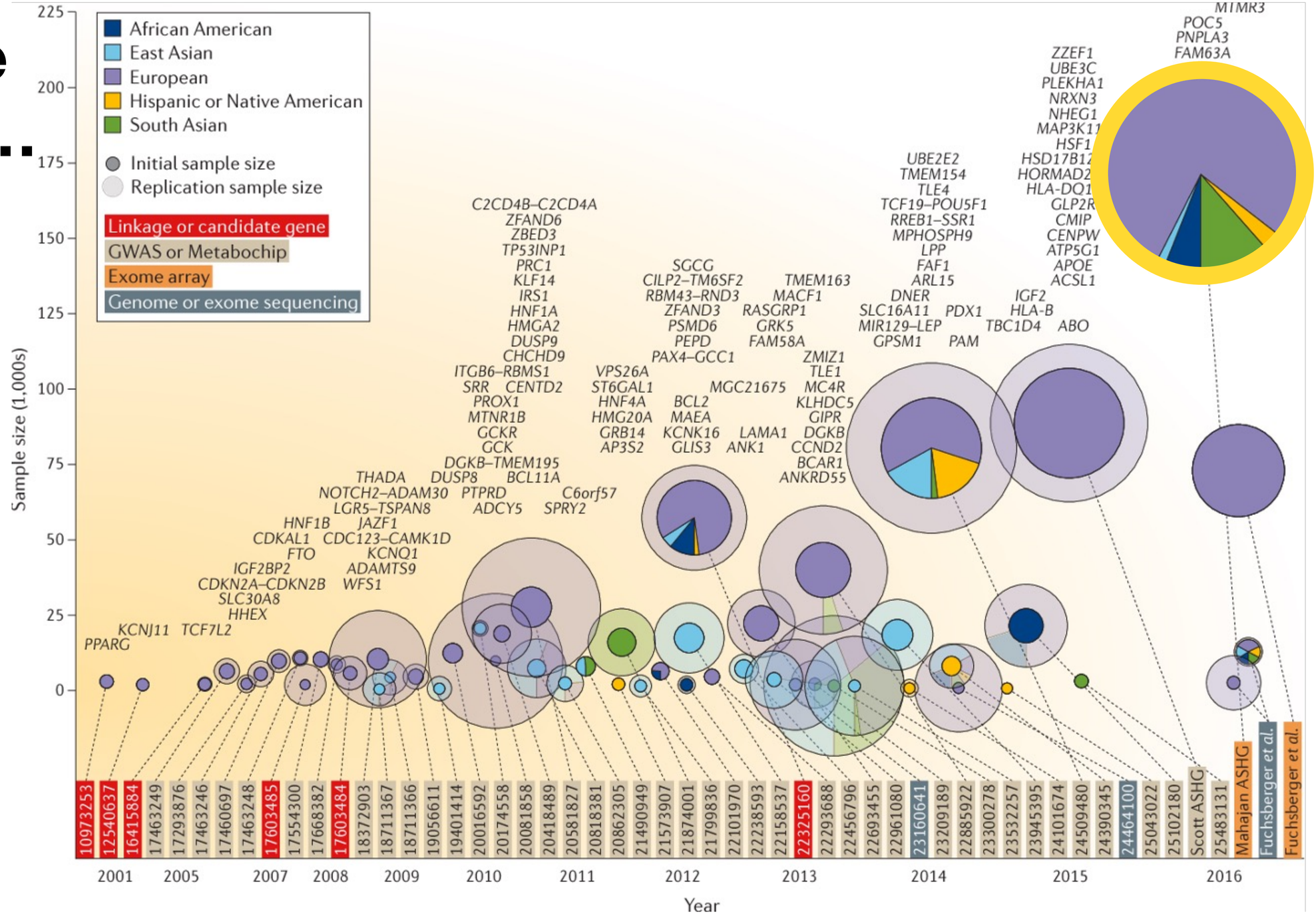
Common and rare variants in multifactorial susceptibility to common diseases

Walter Bodmer & Carolina Bonilla

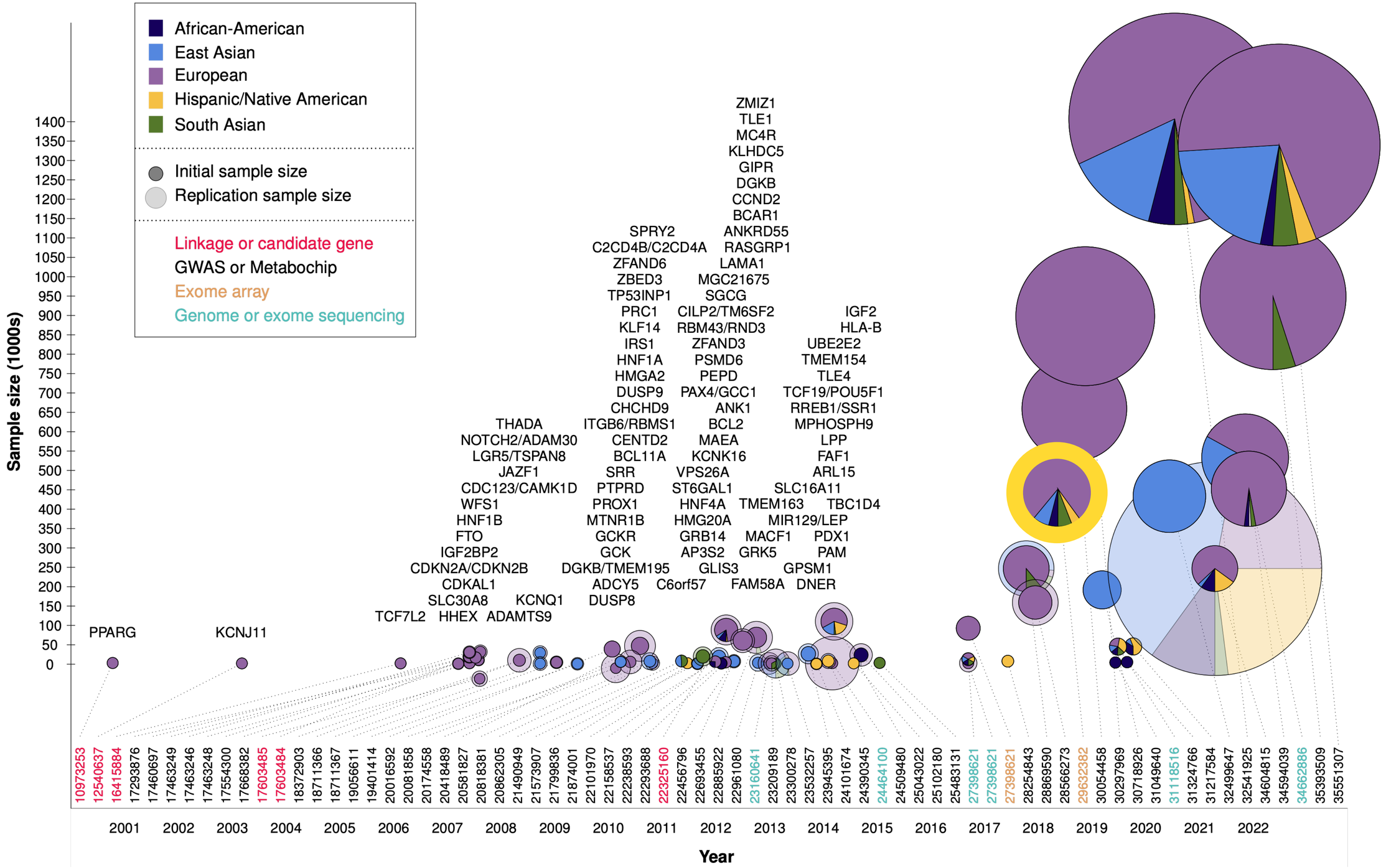
Uncovering the roles of rare variants in common disease through whole-genome sequencing

Elizabeth T. Cirulli and David B. Goldstein

Since then…

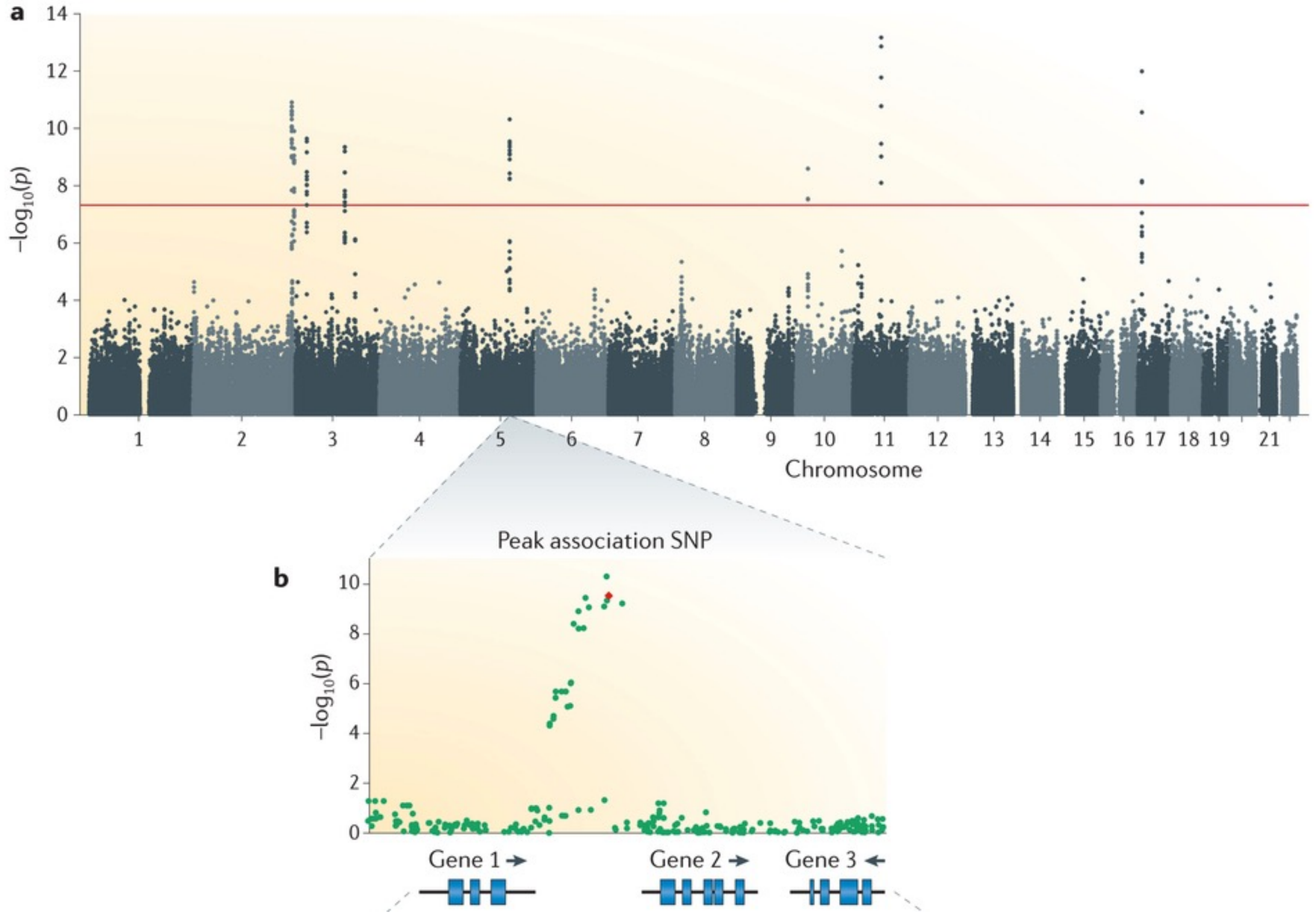


Flannick and Florez, Nature Reviews Genetics, 2016


African-American
East Asian
European
Hispanic/Native American
South Asian
Initial sample size
Replication sample size
Linkage or candidate gene
GWAS or Metabochip
Exome array
Genome or exome sequencing
Sample size (1000s)
Year
PPARG
KCNJ11
TCF7L2
HHEX
ADAMTS9
SLC30A8
KCNQ1
DUSP8
CDKAL1
ADCY5
C6orf57
FAM58A
DNER
CDKN2A/CDKN2B
DGKB/TMEM195
GLIS3
GPSM1
IGF2BP2
GCK
AP3S2
GRK5
PAM
FTO
GCKR
GRB14
MACF1
PDX1
HNF1B
MTNR1B
HMG20A
MIR129/LEP
WFS1
PROX1
HNF4A
TMEM163
TBC1D4
CDC123/CAMK1D
PTPRD
ST6GAL1
SLC16A11
JAZF1
SRR
VPS26A
ARL15
LGR5/TSPAN8
BCL11A
KCNK16
FAF1
NOTCH2/ADAM30
CENTD2
MAEA
LPP
THADA
ITGB6/RBMS1
BCL2
MPHOSPH9
CHCHD9
ANK1
RREB1/SSR1
DUSP9
PAX4/GCC1
TCF19/POU5F1
HMGA2
PEPD
TLE4
HNF1A
PSMD6
TMEM154
IRS1
ZFAND3
UBE2E2
KLF14
RBM43/RND3
HLA-B
PRC1
CILP2/TM6SF2
IGF2
TP53INP1
SGCG
ZBED3
MGC21675
ZFAND6
LAMA1
C2CD4B/C2CD4A
RASGRP1
SPRY2
ANKRD55
BCAR1
CCND2
DGKB
GIPR
KLHDC5
MC4R
TLE1
ZMIZ1

Problem: what are the genes?

- Usually only one (or a few) variants are causal

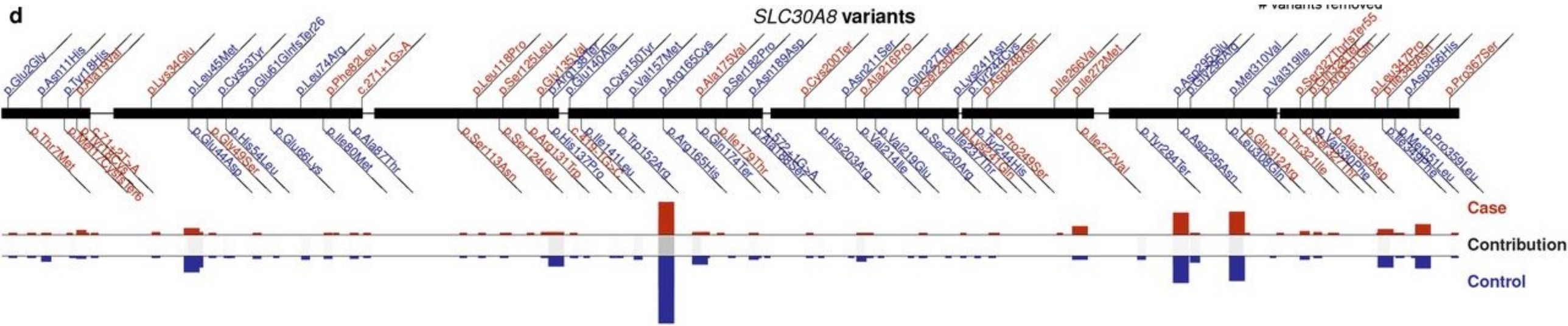


Civelek and Lusis, 2014

Data type 1: GWAS (common variants, high power, unclear gene)

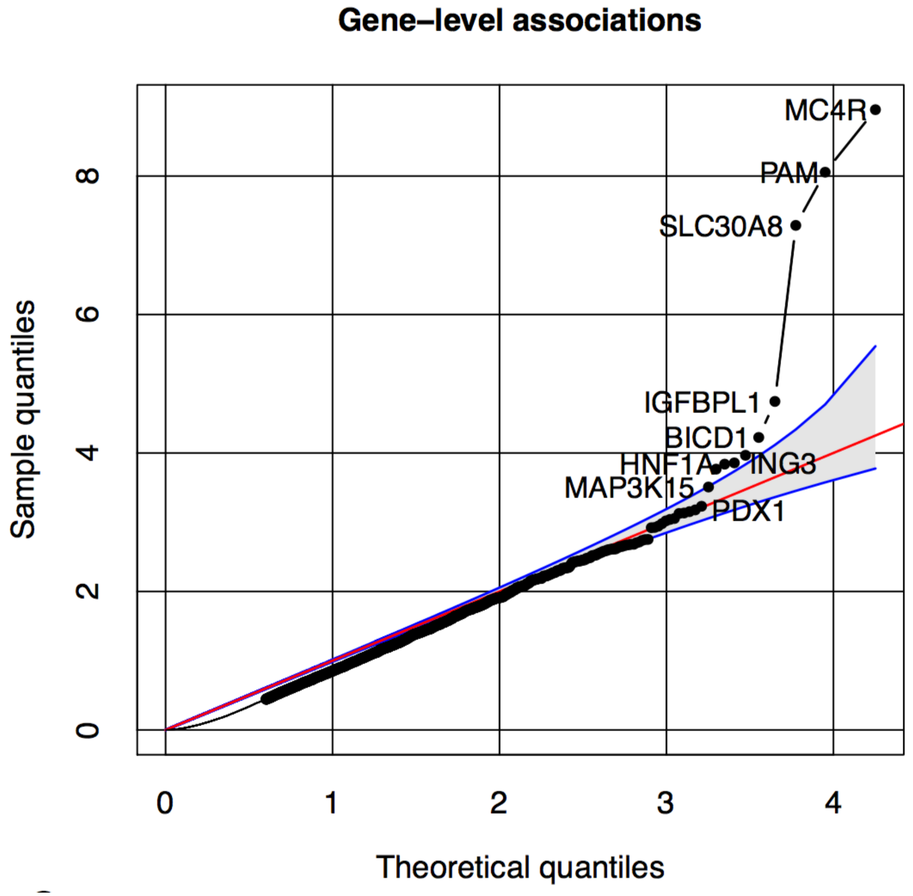
Meanwhile, rare variants…

- When significant: **direct links** from disease to genes
- When variants inactivate protein: **directional link** with disease risk



More recent (45,000 sample) exome sequencing

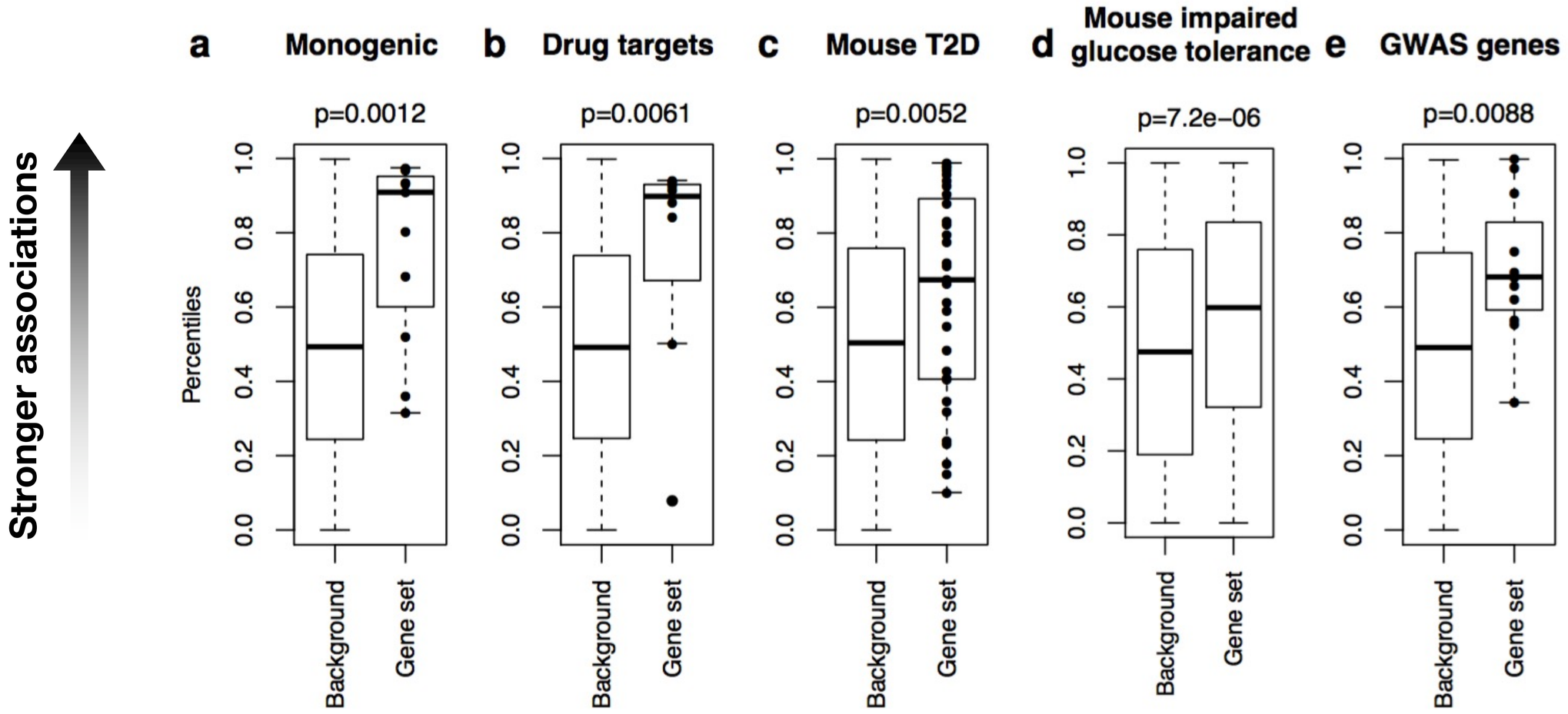
- Three exome-wide significant gene-level associations



Flannick et al, Nature 2019

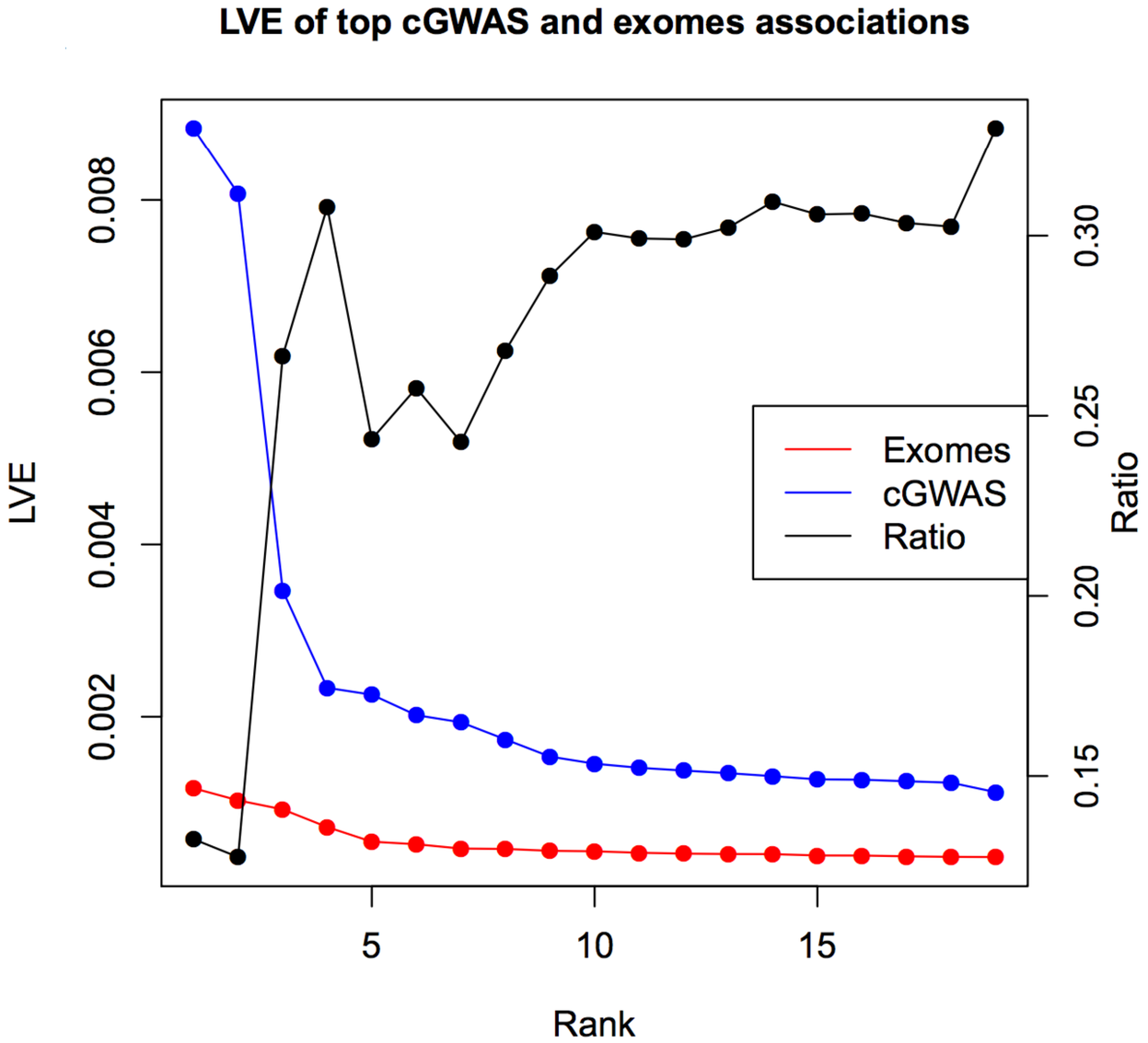
More recent (45,000 sample) exome sequencing

- Enrichment of associations across diabetes-relevant genes

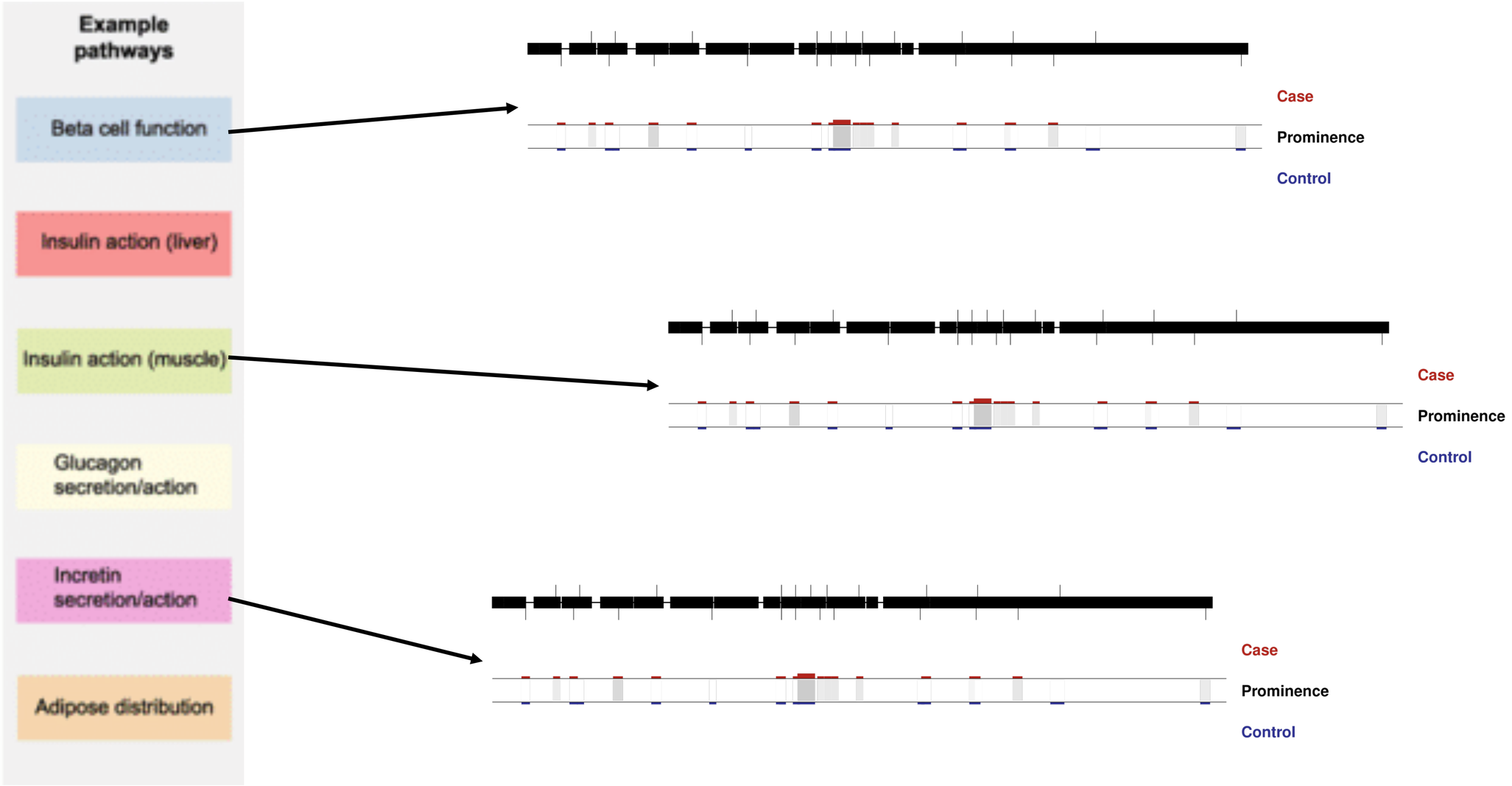


Flannick et al, Nature 2019

Even though the strongest rare variants (even in aggregate) contribute less to T2D than do the strongest common variants…



…rare variant signals are likely widespread in disease genes and pathways



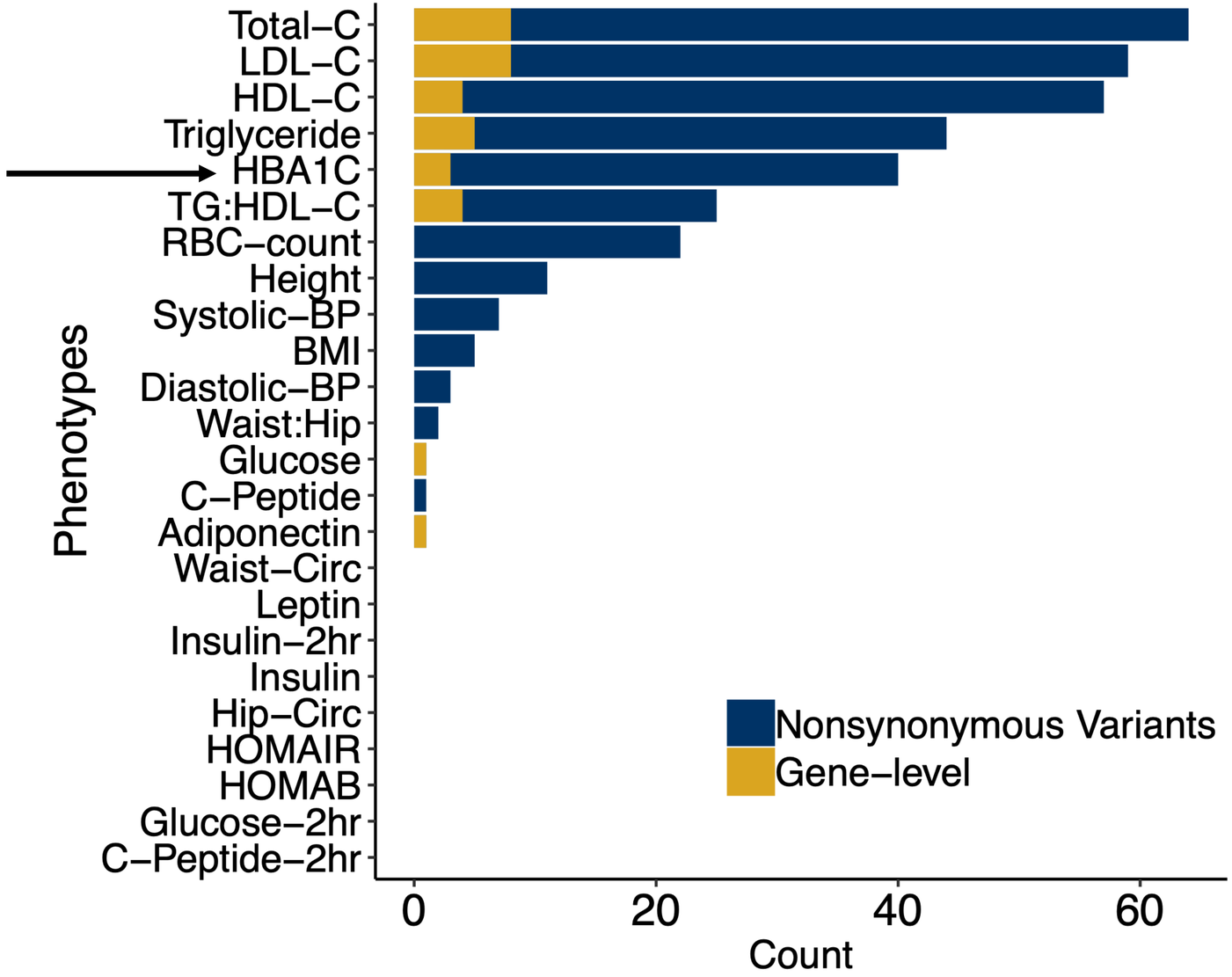
Flannick, Curr Diab Rep 2019

Further support for this model

Beyond type 2 diabetes

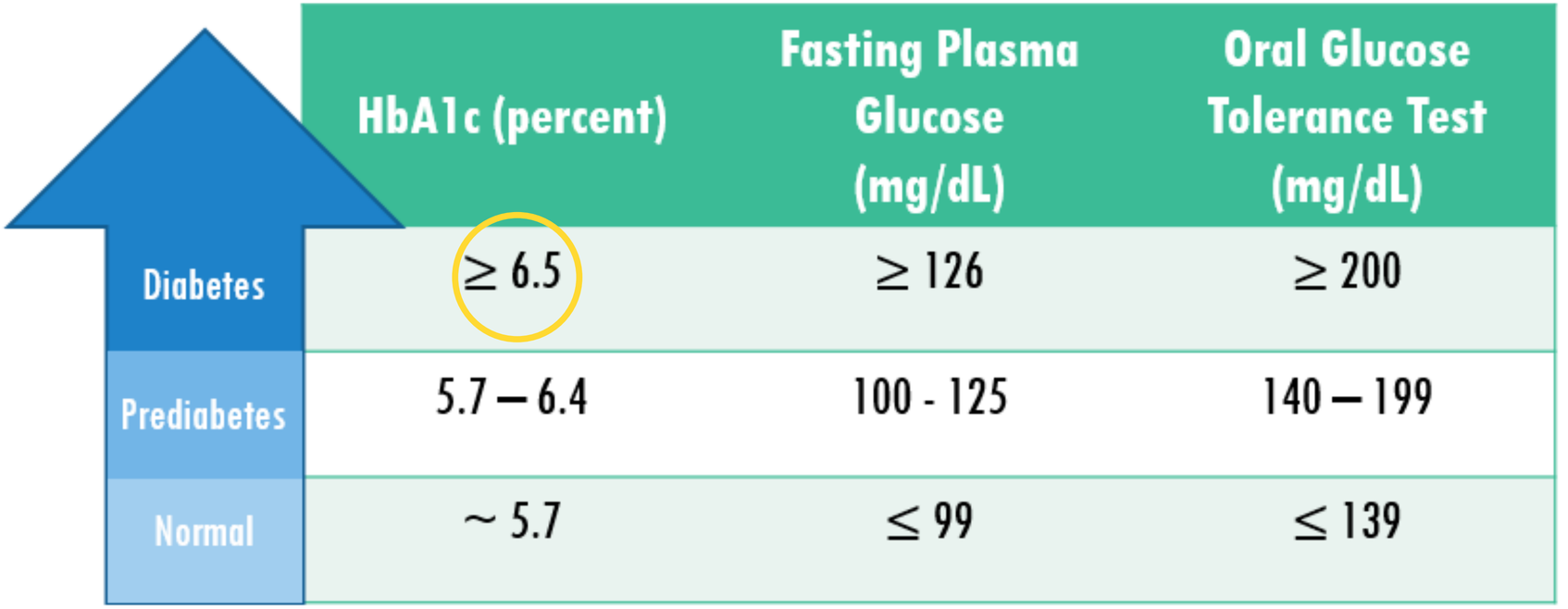
- **Study design:**
  - Exome sequencing of 85,474 non-diabetic individuals
    - UKBB exomes (n=40,151; 100% European)
    - AMP-T2D exomes (n=45,323; 15.8% African American; 25.6% East Asian; 18.7% European; 18.0% Hispanic; and 22.2% South Asian)
  - 24 quantitative traits
  - Single variant analysis (mostly for common variants)
  - Gene-level analysis (for rare variants)

Dornbos et al, Nature Genetics 2022



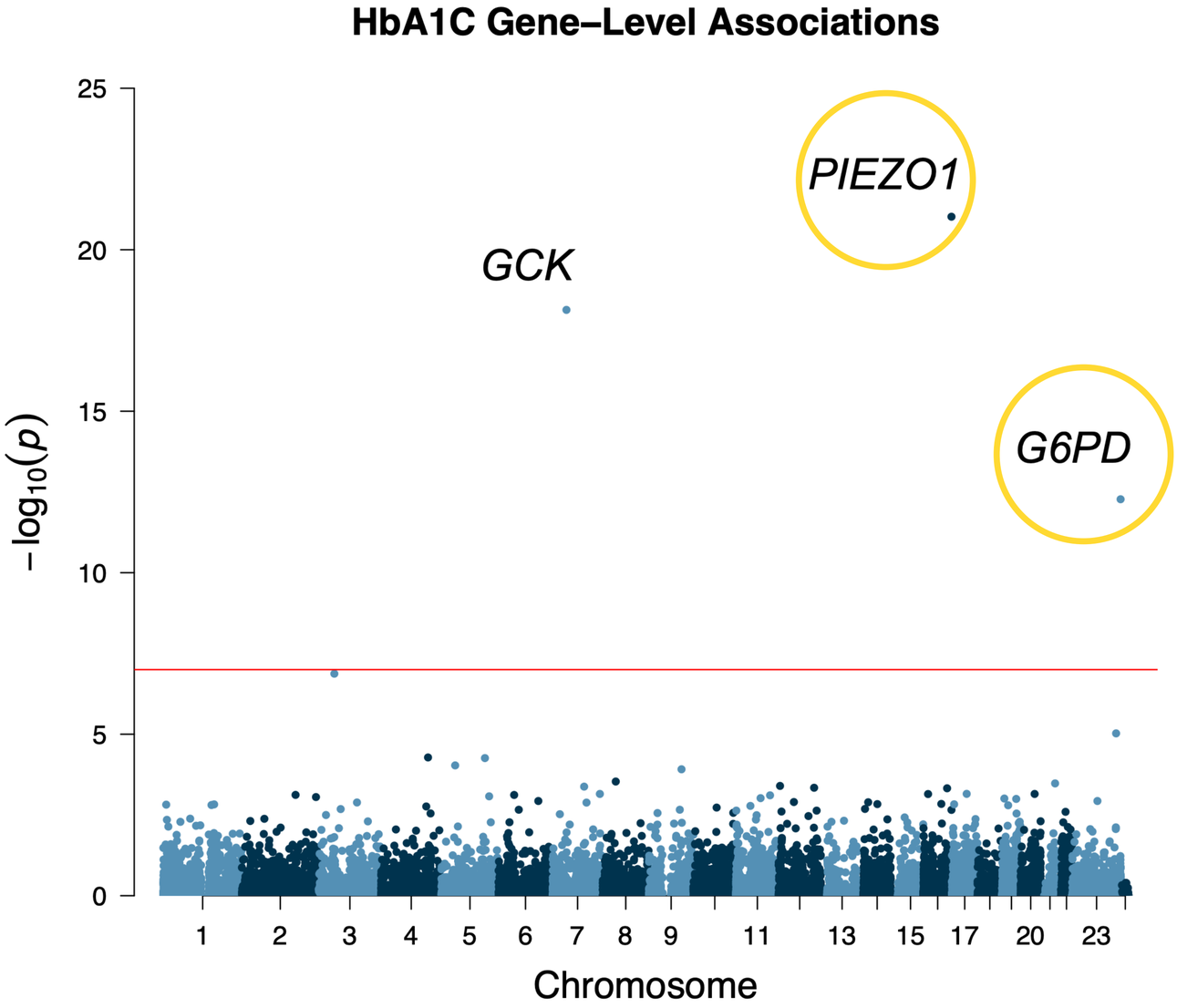
Dornbos et al, Nature Genetics 2022

HbA1C is commonly used in T2D diagnosis



	HbA1c (percent)	Fasting Plasma Glucose (mg/dL)	Oral Glucose Tolerance Test (mg/dL)
Diabetes	≥ 6.5	≥ 126	≥ 200
Prediabetes	5.7 – 6.4	100 - 125	140 – 199
Normal	~ 5.7	≤ 99	≤ 139

Dornbos et al, Nature Genetics 2022

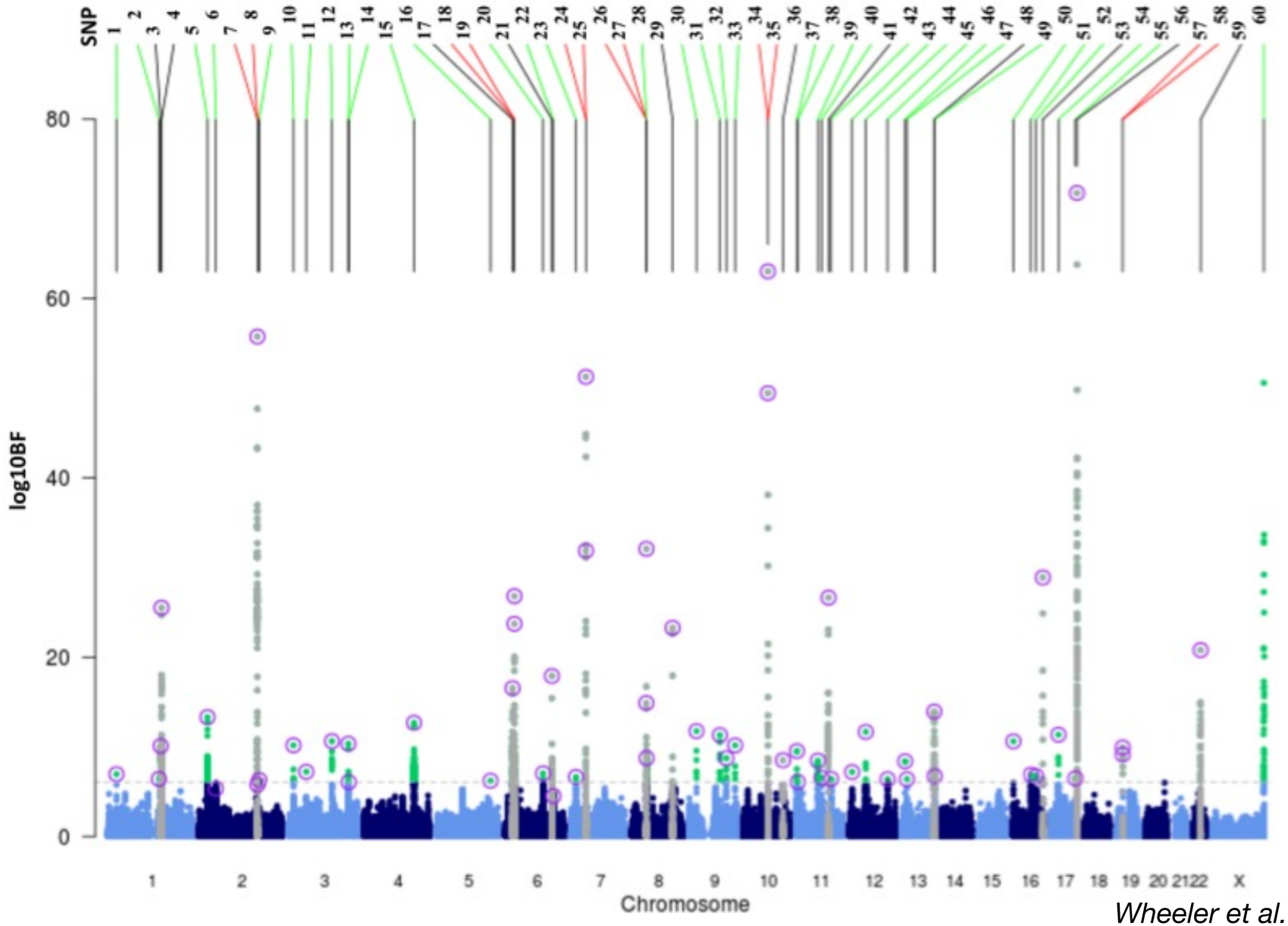


Dornbos et al, Nature Genetics 2022

Common variants and HbA1C

~150,000 individuals

60 common variant associations



G6PD variant 11% of African-Americans

0.81% **reduction** in HbA1c

Wheeler et al. Nature Genetics, 2017

Common variants and HbA1c

- Associations can be grouped into two classes

Associated with red blood cell traits

Associated with other glycemic traits

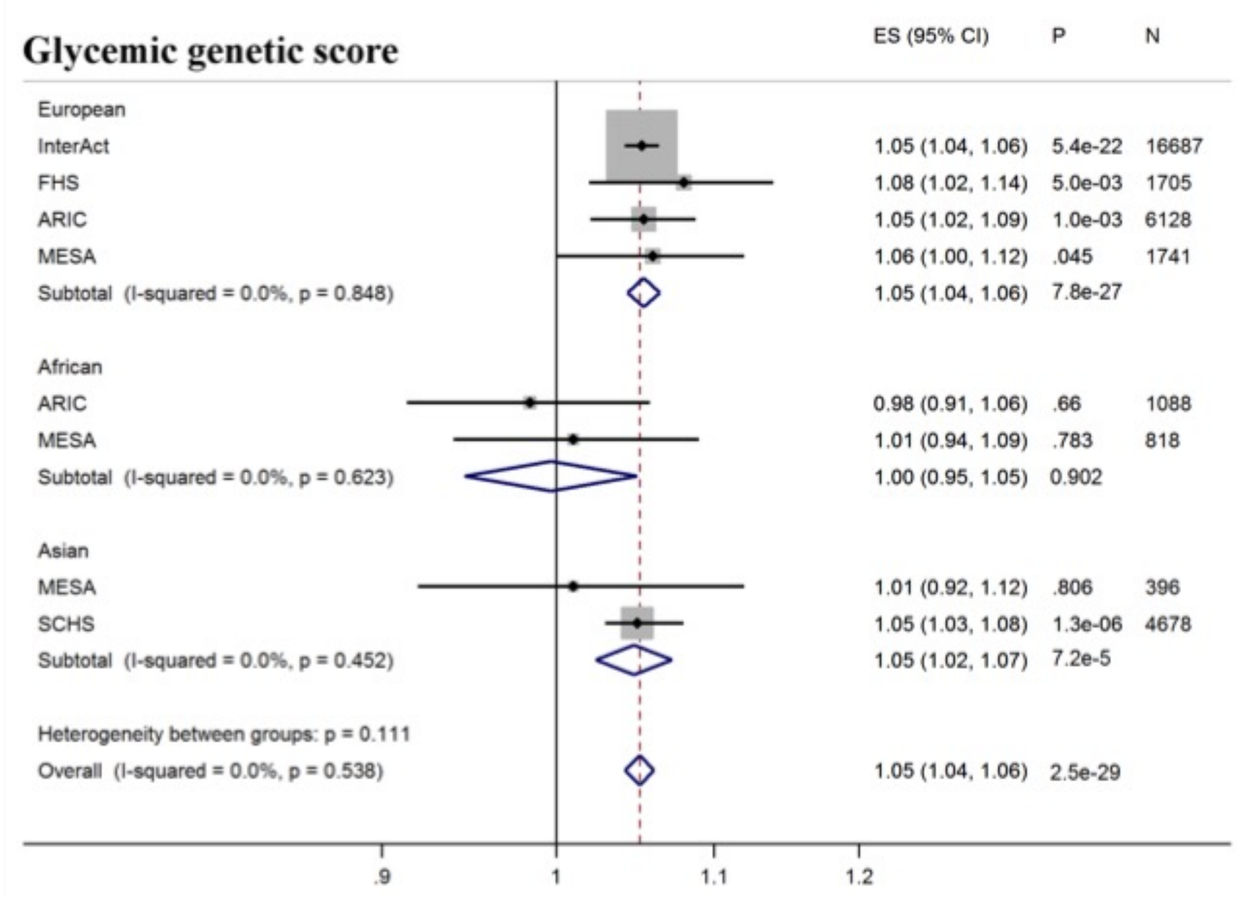
Table 1. (*Continued*)

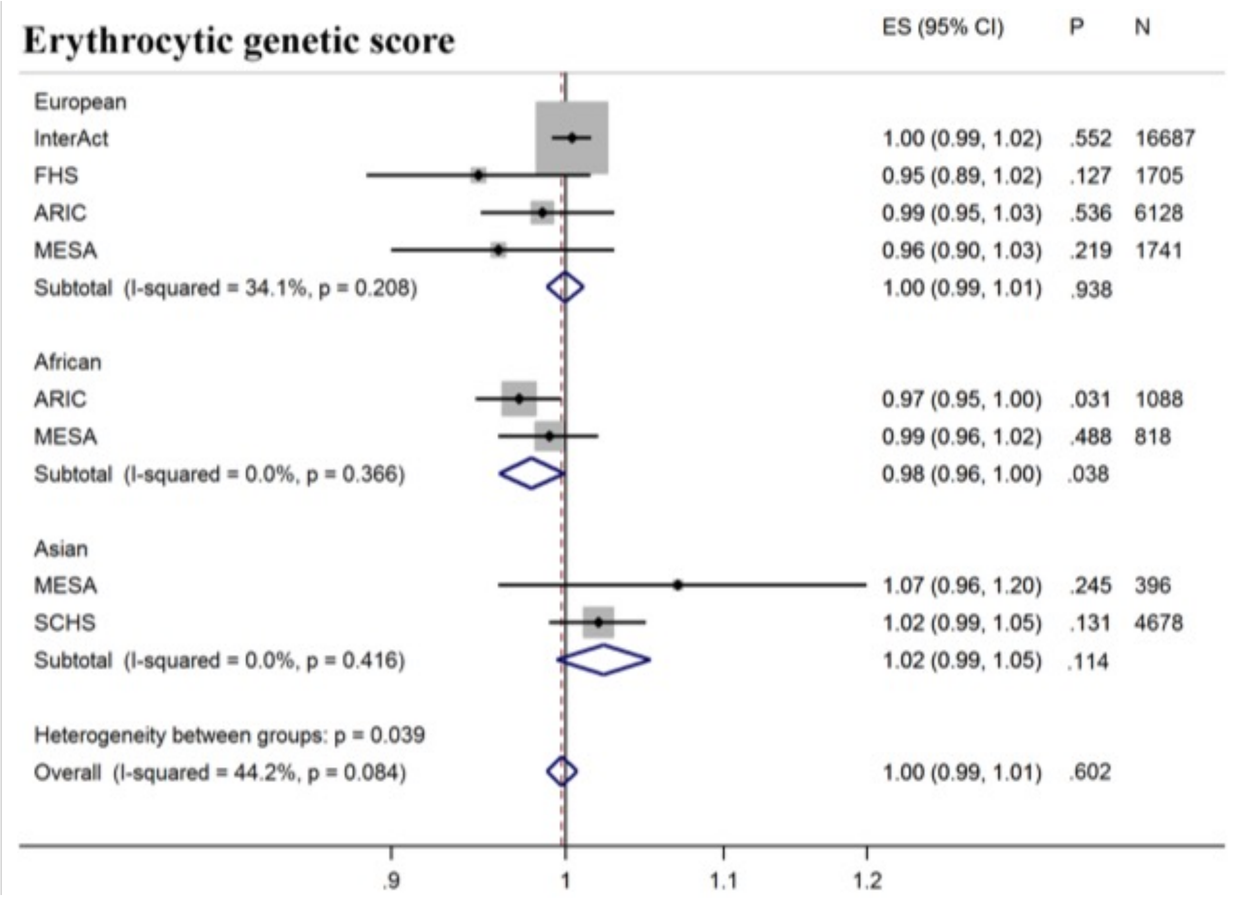
SNP	Markername	Chr.	Position (bp)	Effect Allele	Other Allele	Gene	Status	Signals	Classification	European ancestry METAL *p*-value	Trans-ethnic MANTRA log10BF
45	**rs10774625**	12	110394602	G	A	ATXN2	Novel	Single	**Erythrocytic**	1.46×10^{-8}	6.38
46	**rs11619319**	13	27385599	G	A	PDX1	Novel	Single	**Glycemic**	4.58×10^{-7}	8.38
47	**rs576674**	13	32452302	G	A	KL	Novel	Single	**Glycemic**	1.39×10^{-5}	6.38
48	**rs282587**	13	112399663	G	A	ATP11A	Known	Single	Unclassified	1.70×10^{-12}	13.92
49	**rs9604573**	13	113571085	T	C	GAS6	Novel	Single	Unclassified	9.60×10^{-9}	6.72
50	**rs11248914**	16	233563	T	C	ITFG3	Novel	Single	**Erythrocytic**	2.56×10^{-14}	10.60
51	**rs1558902**	16	52361075	A	T	FTO	Novel	Single	Unclassified	3.27×10^{-8}	6.88
52	**rs4783565**	16	67307691	A	G	CDH3	Novel	Single	**Erythrocytic**	1.73×10^{-7}	6.73
53	**rs837763**	16	87381230	T	C	CDT1	Known	Single	**Erythrocytic**	1.68×10^{-28}	28.89
54	**rs9914988**	17	24207230	A	G	ERAL1	Novel	Single	**Erythrocytic**	2.77×10^{-11}	11.34
55	**rs2073285**	17	73628956	C	T	TMC6	Novel	Single	Unclassified	1.27×10^{-4}	6.47
56	**rs1046896**	17	78278822	T	C	FN3KRP	Known	Single	Unclassified	4.46×10^{-64}	71.79
57	**rs11086054**	19	17107737	A	T	MYO9B	Novel	Multiple	Unclassified	8.16×10^{-6}	9.12
58	**rs17533903**	19	17117523	A	G	MYO9B	Known	Multiple	**Erythrocytic**	5.27×10^{-12}	9.912
59	**rs4820268**	22	35799537	G	A	TMPRSS6	Known	Single	**Erythrocytic**	1.40×10^{-22}	20.79
60	**rs1050828**	X	153417411	T	C	G6PD	Novel	Single	**Erythrocytic**	NA*	NA

Wheeler et al. Nature Genetics, 2017

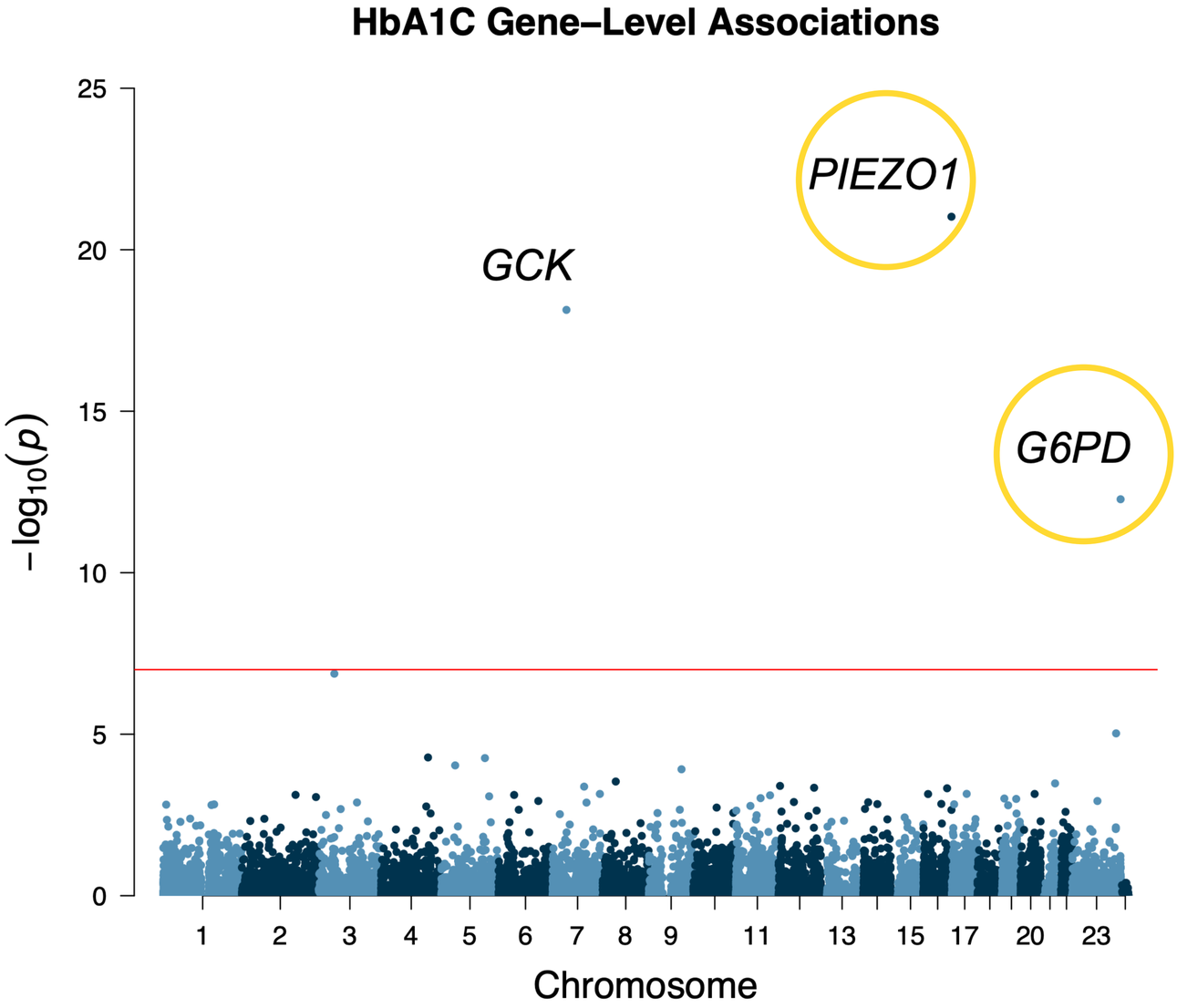
Different biological effects

- Glycemic associations, but not erythrocytic associations, predict future development of T2D
  - ~2% of African-Americans could be misclassified due to *G6PD* variant



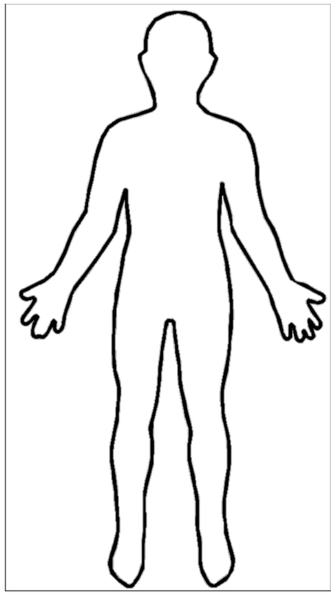


Wheeler et al. Nature Genetics, 2017



Dornbos et al, Nature Genetics 2022

How do these variants affect HbA1c diagnosis?

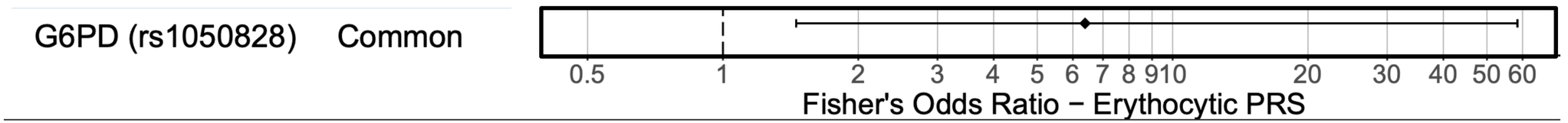


HbA1C 6.4%

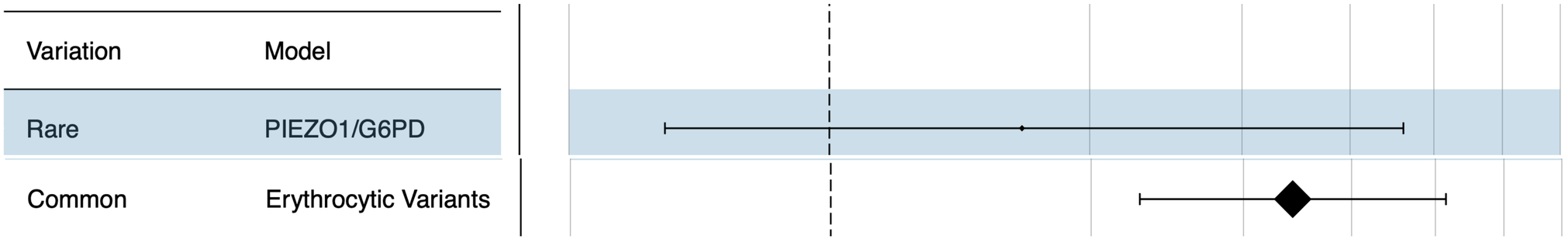
Carries variants with HbA1C effect -0.2%

Evaluate via independent diagnosis

	HbA1c (percent)	Fasting Plasma Glucose (mg/dL)	Oral Glucose Tolerance Test (mg/dL)
Diabetes	≥ 6.5	≥ 126	≥ 200
Prediabetes	5.7 – 6.4	100 - 125	140 – 199
Normal	~ 5.7	≤ 99	≤ 139



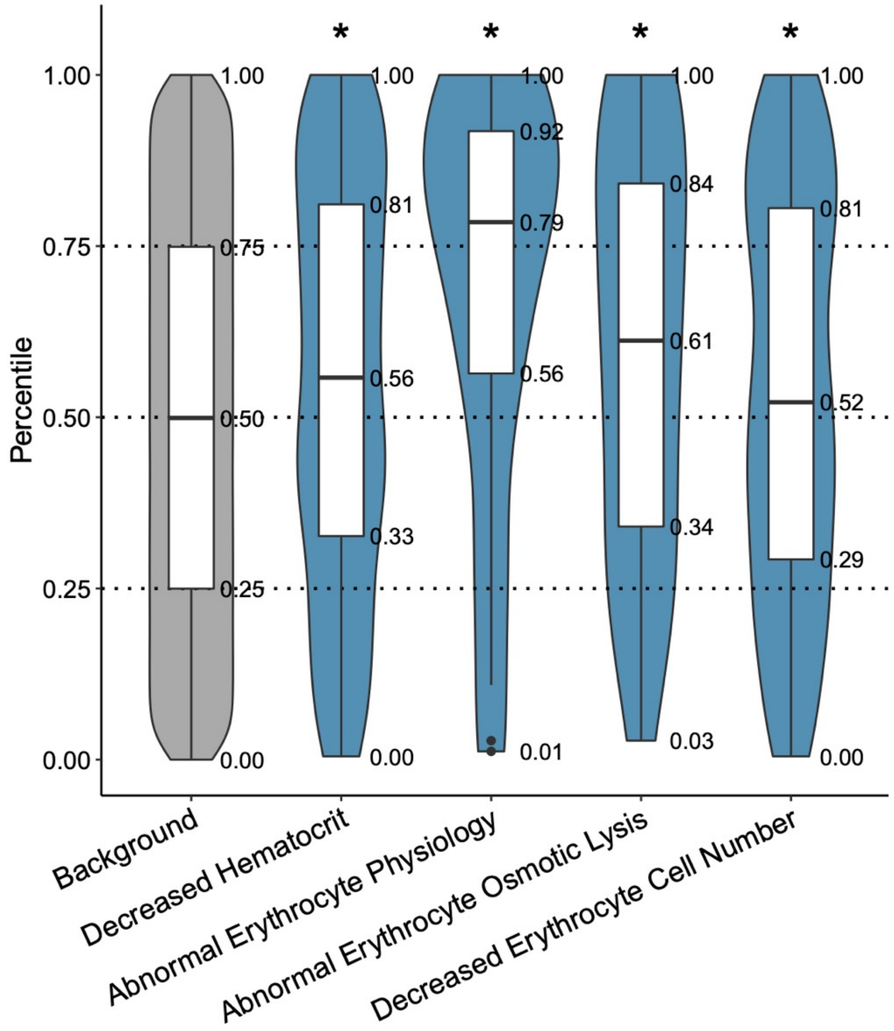
How do these variants affect HbA1c diagnosis?



Dornbos et al, Nature Genetics 2022

Expanding the model

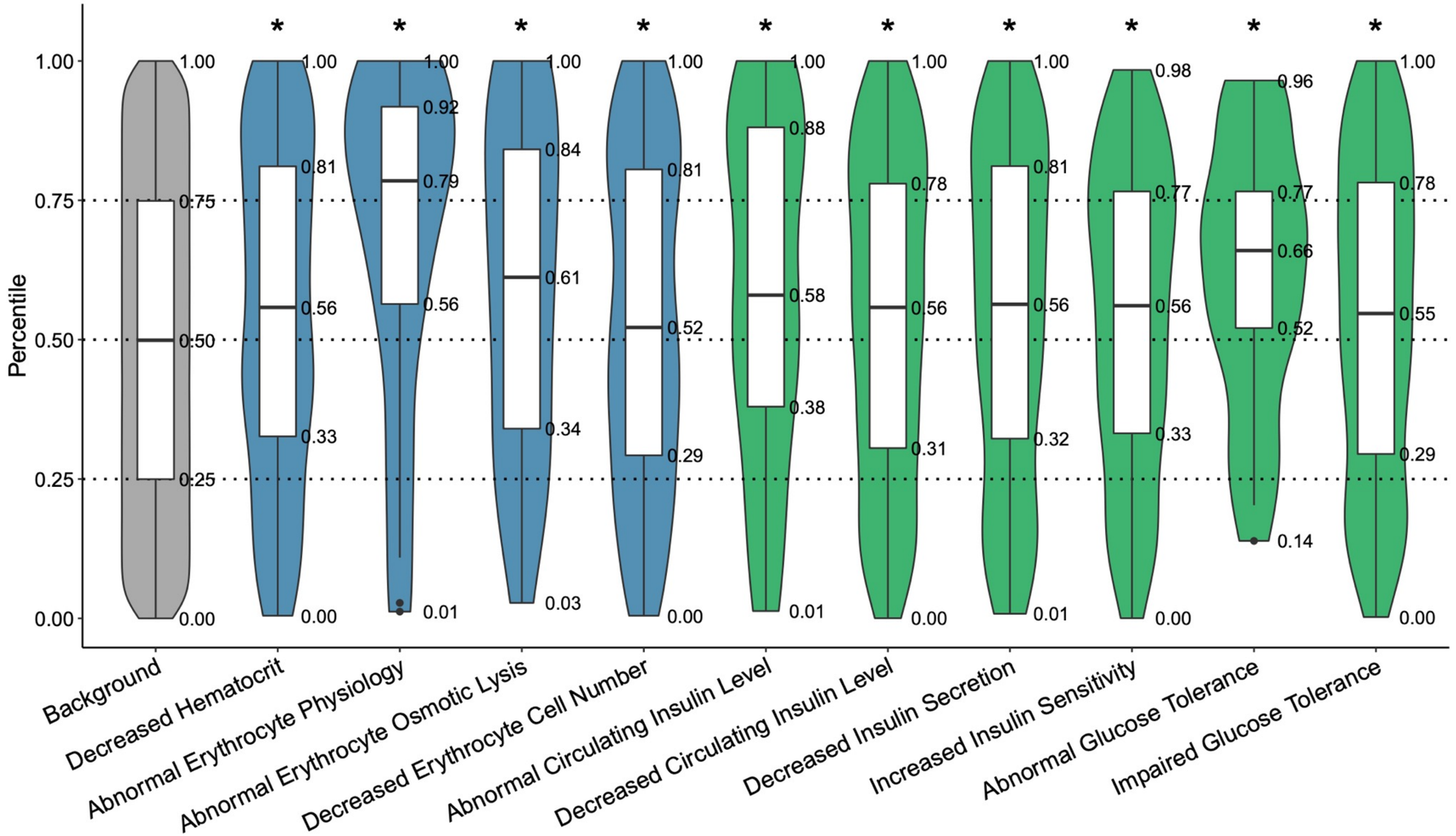
- Significant associations in many sets of genes with known function on erythrocytic lifespan in mice

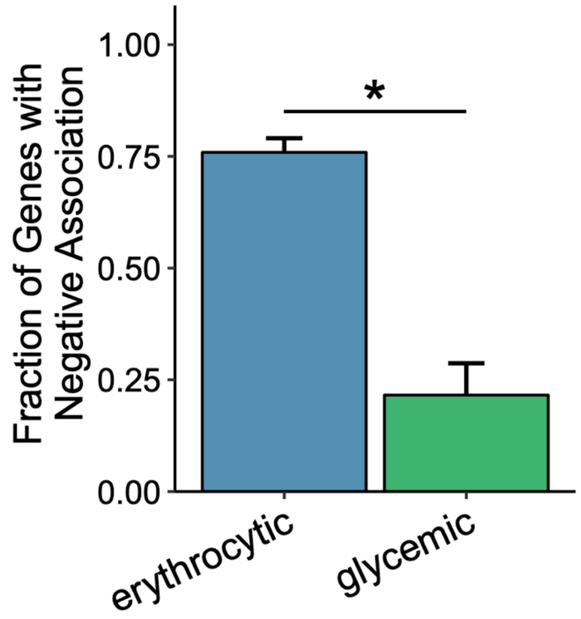


Dornbos et al, Nature Genetics 2022

Evidence for associations across many genes

- Compared to rare variants in genes involved in glycemia in mice, rare variants in erythrocytic genes are more likely to decrease HbA1c





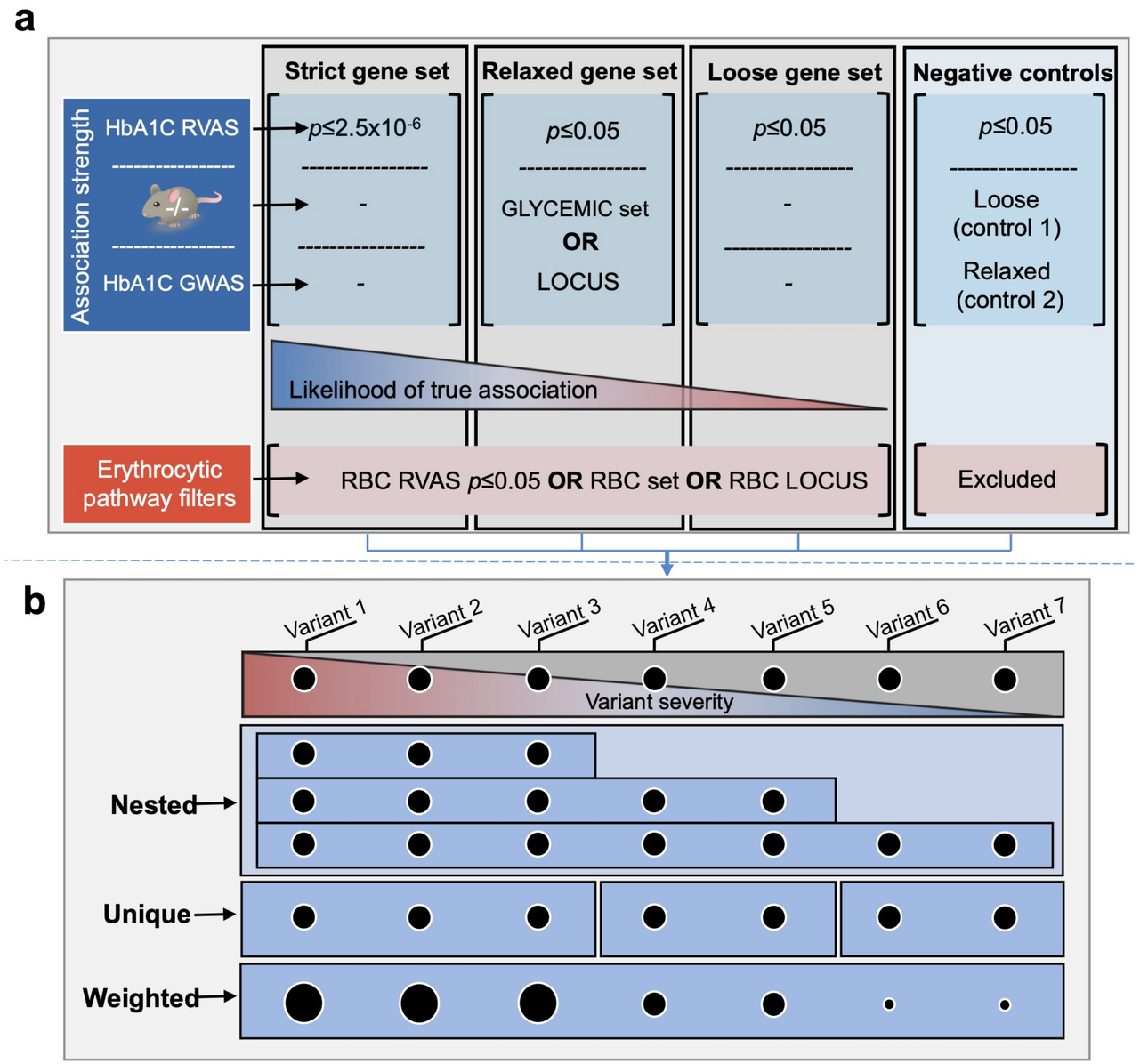
Dornbos et al, Nature Genetics 2022

Putting this together in a polygenic score

Filtering to true associations

Filtering to erythrocytic variants

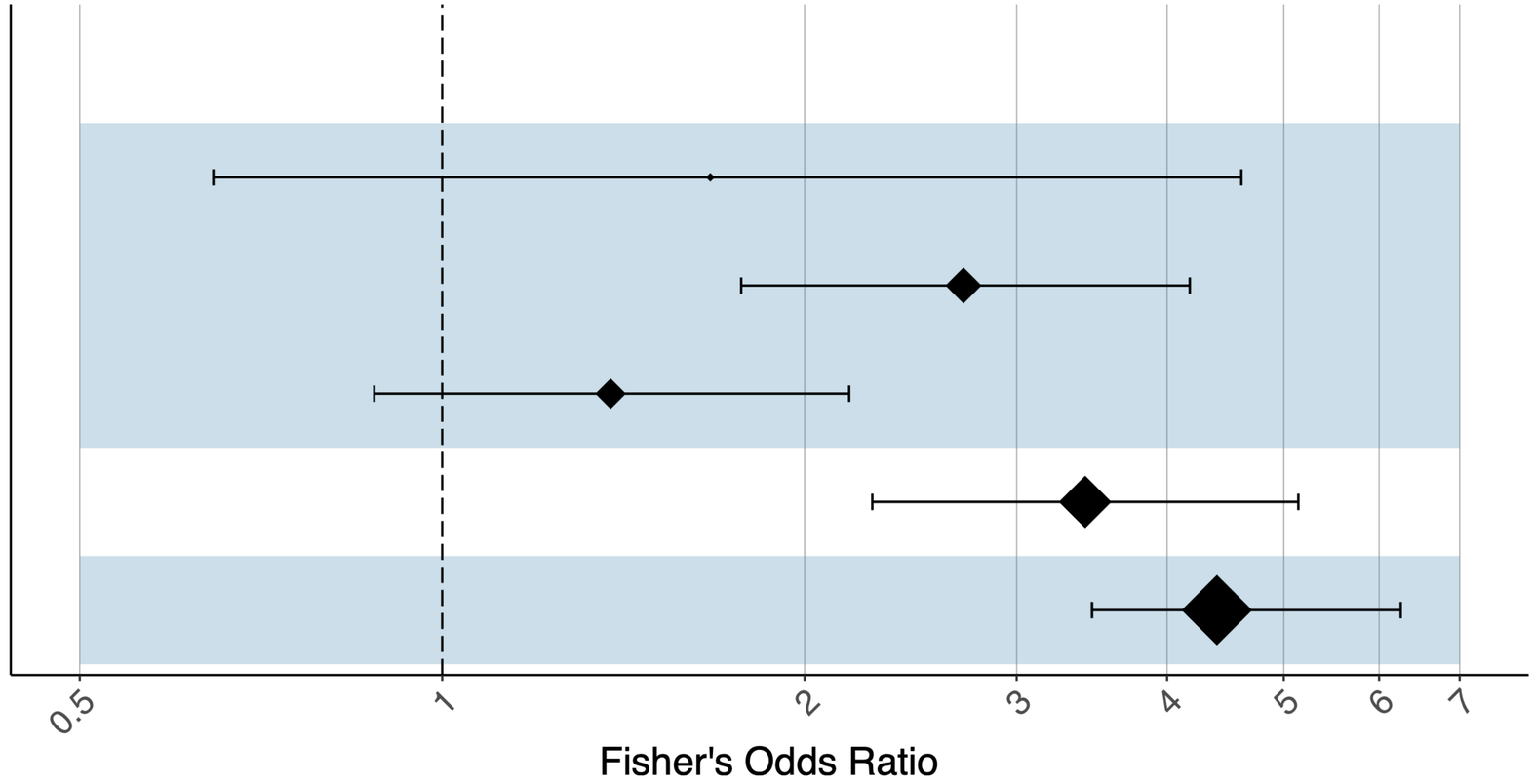
Estimate variant effect size



Dornbos et al, Nature Genetics 2022

Final model: 21,293 variants

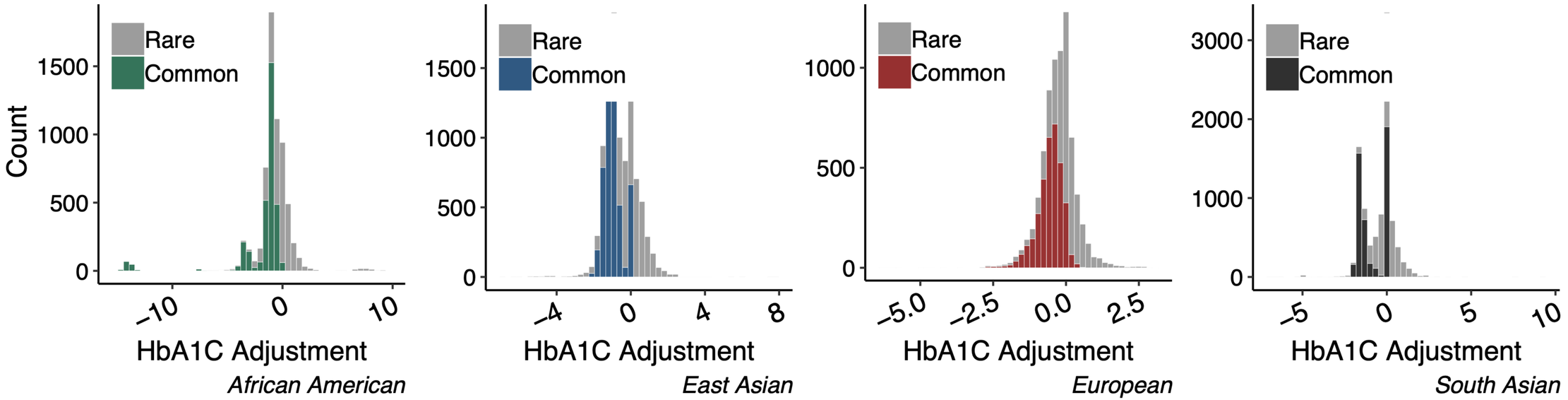
Variation	Model
Rare	PIEZO1/G6PD
	Erythrocytic Genes
	Glycemic Genes
Common	Erythrocytic Variants
Combined	Erythrocytic PGS



**Diamonds scaled to number of reclassified individuals in the test sample

Dornbos et al, Nature Genetics 2022

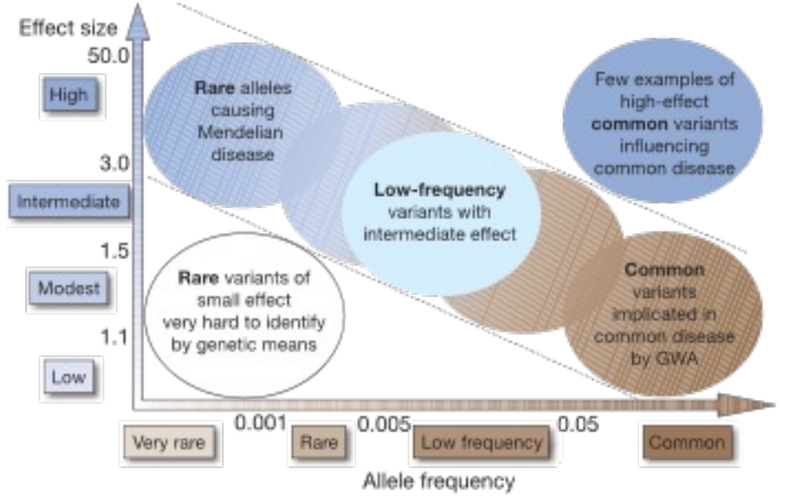
Model is highly polygenic



Dornbos et al, Nature Genetics 2022

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.



Are Rare Variants Responsible for Susceptibility to Complex Diseases?

Jonathan K. Pritchard

Department of Statistics, University of Oxford, Oxford

Common Genetic Variation and Human Traits

David B. Goldstein, Ph.D.

Genetic Heterogeneity in Human Disease

Jon McClellan[1,*] and Mary-Claire King[2,*]
[1]Department of Psychiatry
[2]Departments of Medicine and Genome Sciences
University of Washington, Seattle, WA 98195-7720, USA
*Correspondence: drjack@uw.edu (J.M.), mcking@uw.edu (M.-C.K.)
DOI 10.1016/j.cell.2010.03.032

Clan Genomics and the Complex Architecture of Human Disease

James R. Lupski,[1,2,3,*] John W. Belmont,[1,2] Eric Boerwinkle,[4,5] and Richard A. Gibbs[1,5,*]

OPEN ACCESS Freely available online

PLoS BIOLOGY

Rare Variants Create Synthetic Genome-Wide Associations

Samuel P. Dickson[1,2], Kai Wang[3], Ian Krantz[3,4,5], Hakon Hakonarson[3,4,5], David B. Goldstein[1]*

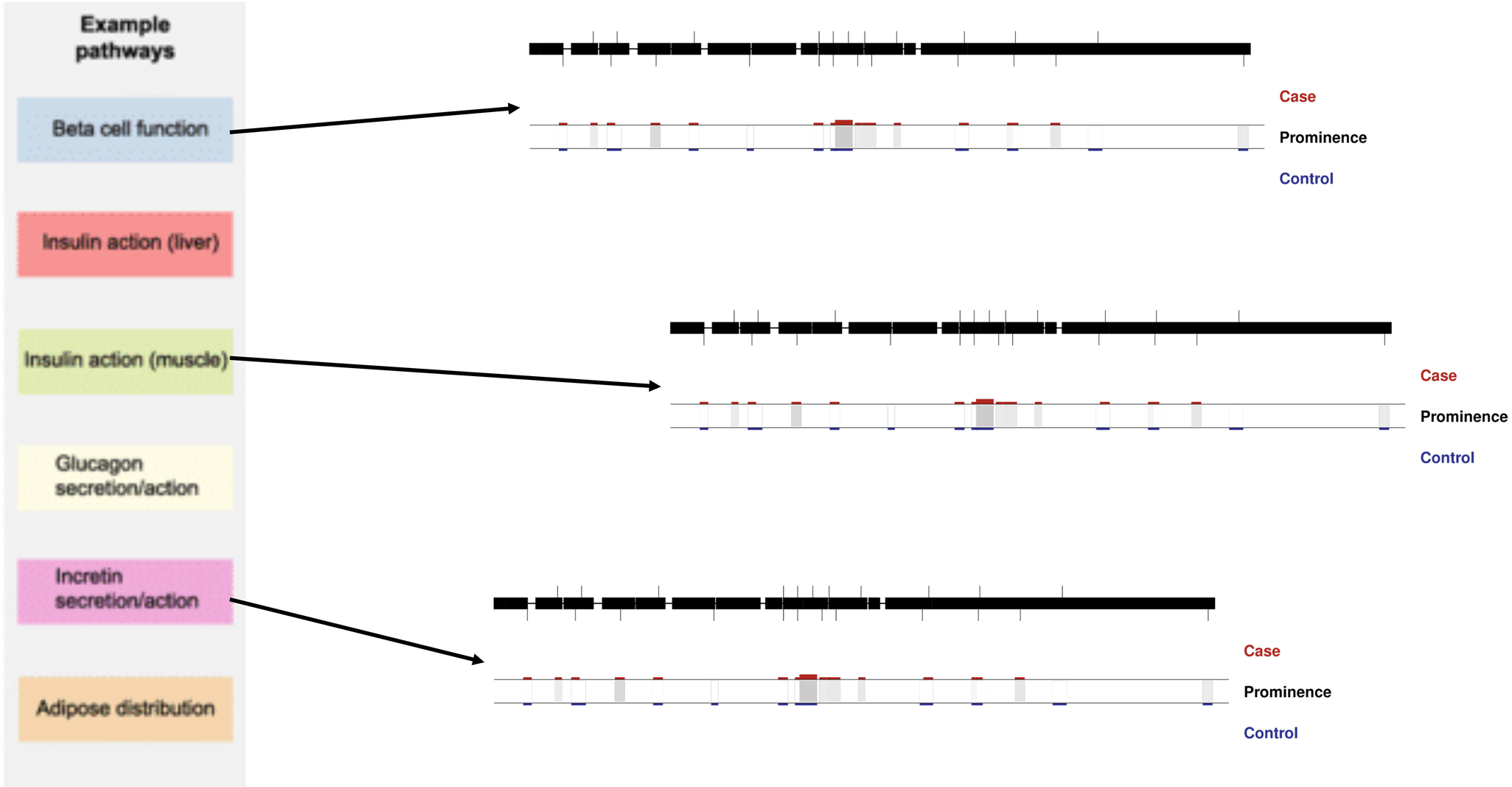
Common and rare variants in multifactorial susceptibility to common diseases

Walter Bodmer & Carolina Bonilla

APPLICATIONS OF NEXT-GENERATION SEQUENCING

Uncovering the roles of rare variants in common disease through whole-genome sequencing

Elizabeth T. Cirulli and David B. Goldstein



Flannick, Curr Diab Rep 2019

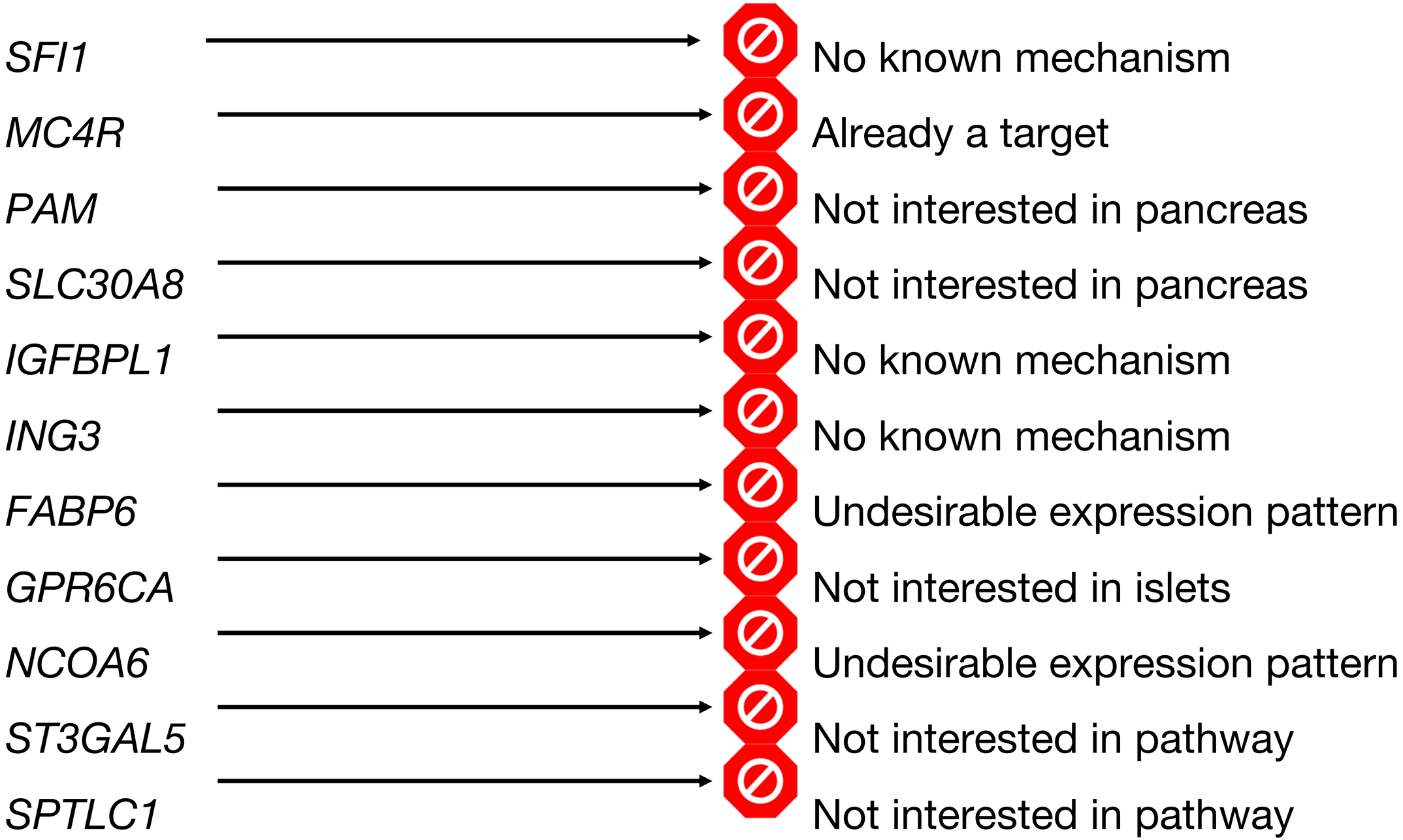
Data type 1: GWAS (common variants, high power, unclear gene)

Data type 2: Exomes (rare variants, low power, clear gene)

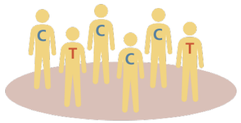
Signals across many genes

Why were these not useful?

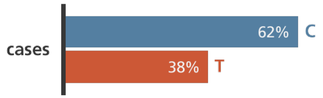

SFI1
MC4R
PAM
SLC30A8
IGFBPL1
ING3
FABP6
GPR6CA
NCOA6
ST3GAL5
SPTLC1
?


SFI1
No known mechanism
MC4R
Already a target
PAM
Not interested in pancreas
SLC30A8
Not interested in pancreas
IGFBPL1
No known mechanism
ING3
No known mechanism
FABP6
Undesirable expression pattern
GPR6CA
Not interested in islets
NCOA6
Undesirable expression pattern
ST3GAL5
Not interested in pathway
SPTLC1
Not interested in pathway

Forward vs. reverse genetics

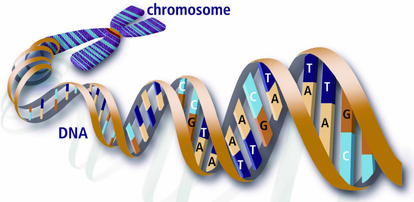


cases (n=1,000)
people with heart disease

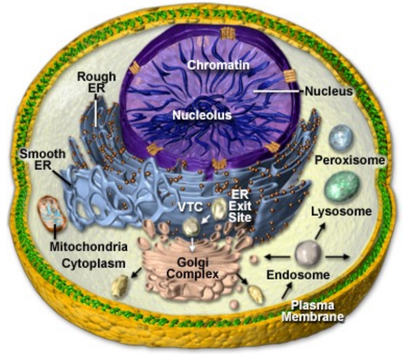


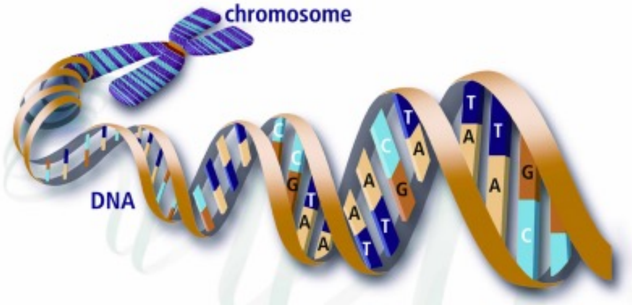
controls (n=1,000)
people without heart disease

Forward genetics



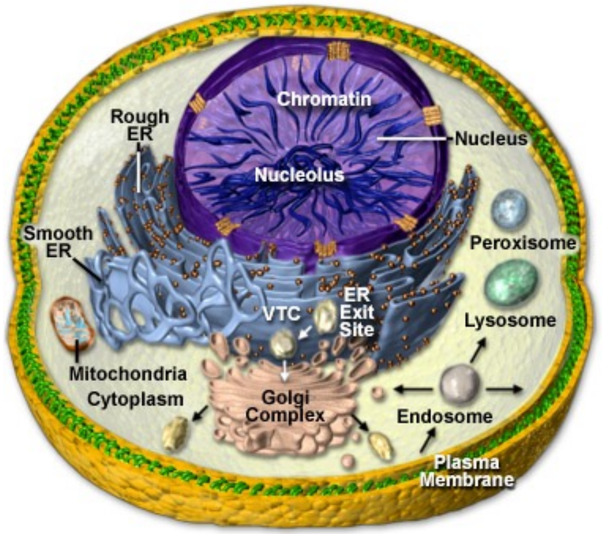
Associated locus

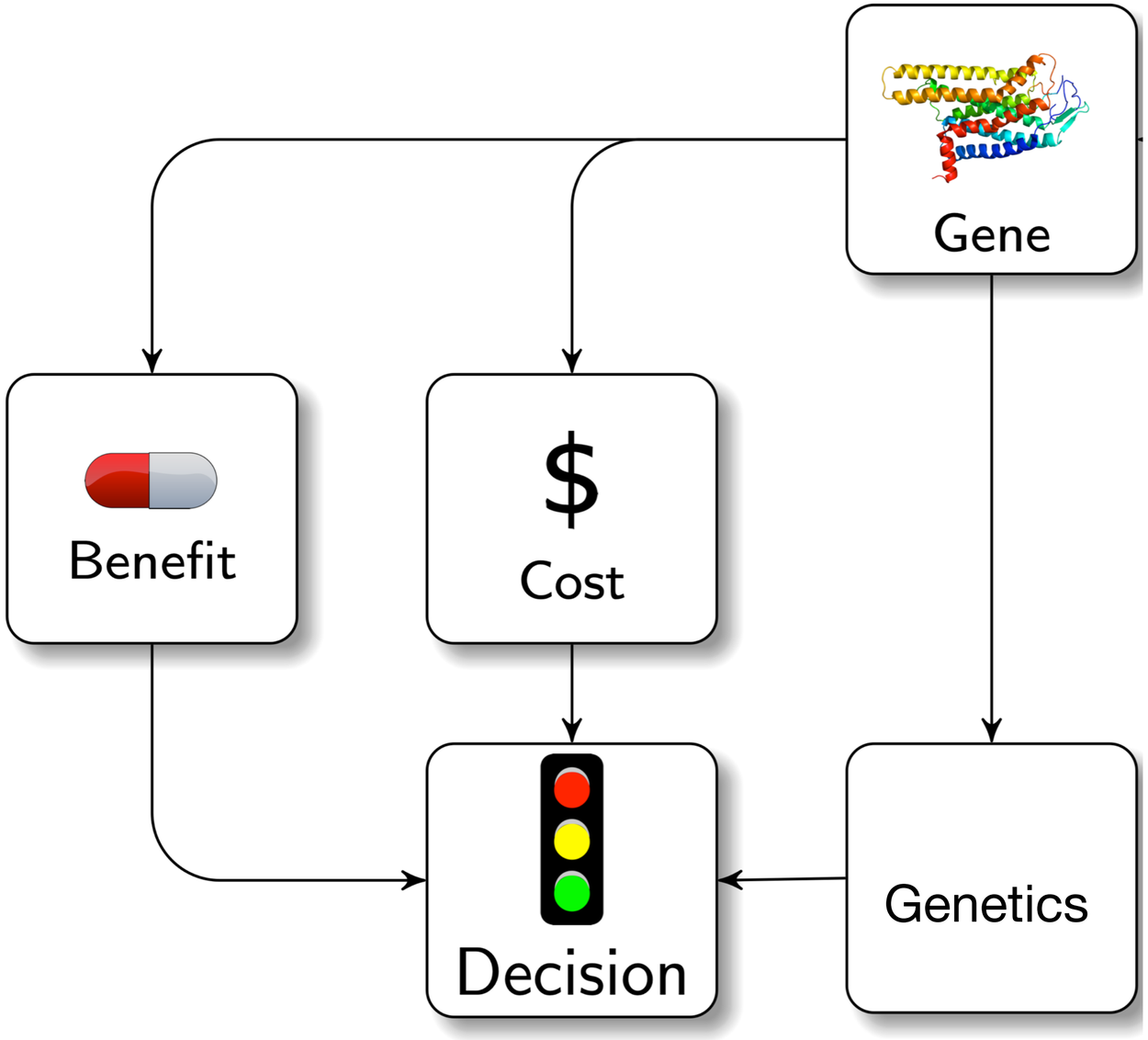




Disease associated genes

Reverse genetics

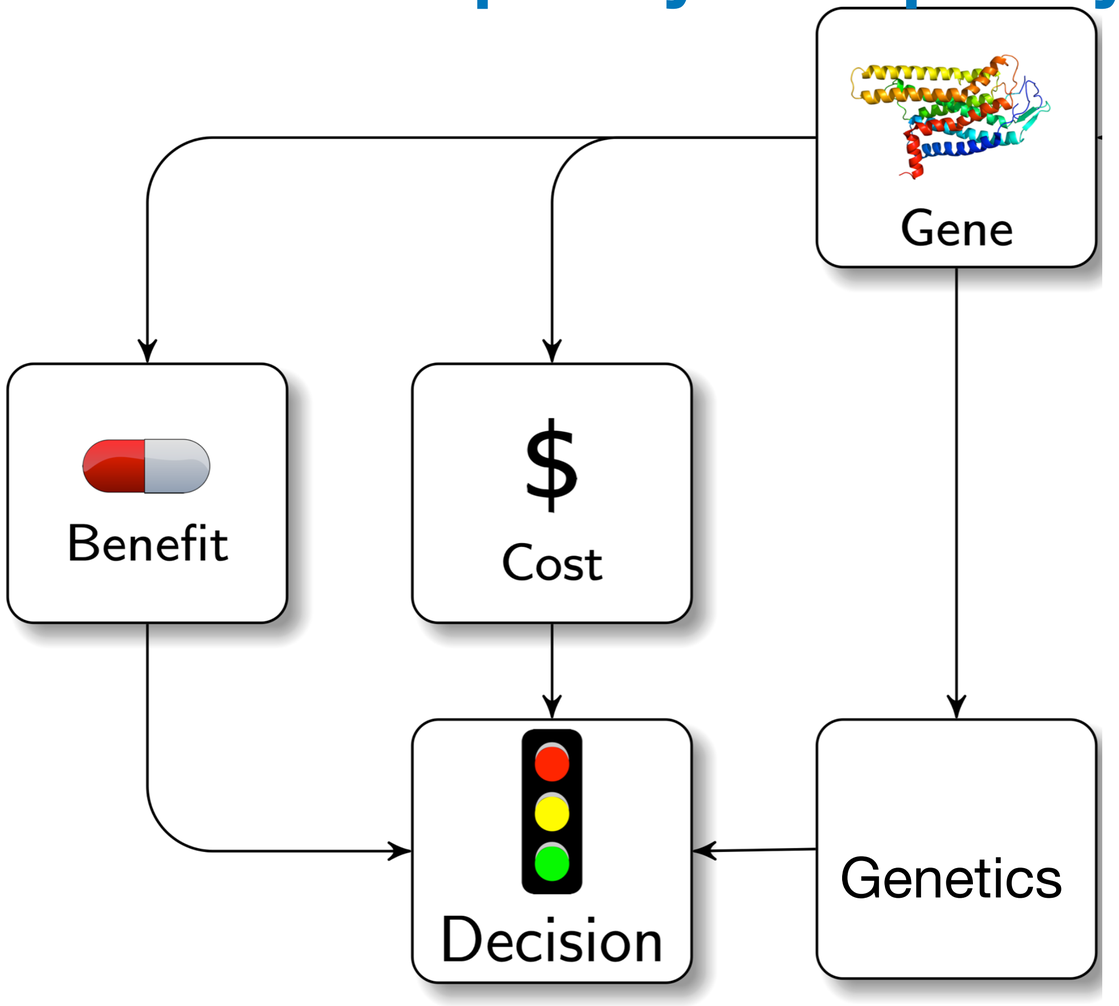



Gene
Benefit
$
Cost
Decision
Genetics

Our organizing question

What does human genetic data tell us about a gene?

explicitly or implicitly



Do people use human genetics in this way?

- Reviewed Nature/Science/Cell papers from last 3 years
- Catalogued every paper that reported a diabetes/glucose/insulin gene
- Examined how many referenced human genetic support
- **Result**: 4 out of 35 articles referenced human genetics

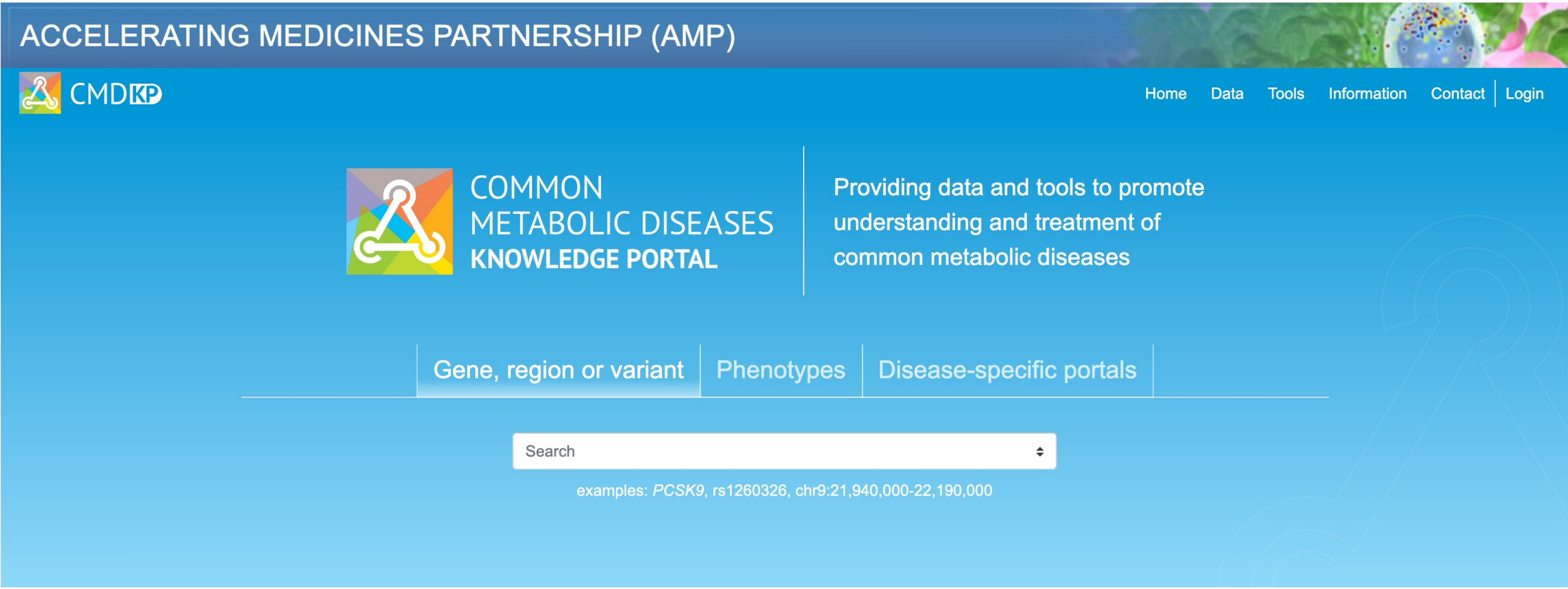
Gene	PMID/Citation	Journal	Type of Evidence	Reference Human Genetics
PPARGC1A	28340340 (Sharabi et al., 2017)	Cell	mouse, cell culture	-
STUB1	28431247 (Tawo et al., 2017)	Cell	C elegans, Drosophila melanogaster, cell culture	Yes
TBK1	29425491 (Zhao et al., 2018)	Cell	mouse, cell culture	-
ZMPSTE24	29526462 (Kayatekin et al., 2018)	Cell	yeast, cell culture	Yes
LEPR	29670283 (Xu et al., 2018)	Nature	mouse, cell culture	-
VDR BRD9 BRD7	29754817 (Wei et al., 2018)	Cell	mouse, cell culture	Yes
PIK3CA PIK3CB PIK3CG PIK3CD PTEN	30051890 (Hopkins et al., 2018)	Nature	mouse, cell culture	-
PIK3CA PIK3CB HRAS NRAS KRAS	30982732 (Molinaro et al., 2019)	Cell Metabolism	mouse, cell culture	-
ALOX12	31353262 (Leiria et al., 2019)	Cell Metabolism	human study, mouse, cell culture	-
PAX6	31607563 (Singer et al., 2019)	Cell Metabolism	human study, mouse, cell culture	-
EIF2AK3	31543404 (Chen et al., 2019)	Cell Metabolism	mouse, cell culture	-
CPT1A SLC25A20	31378464 (Nicholas et al., 2019)	Cell Metabolism	human study, Cell Culture	-
GSK3A GSK3B	30879985 (Sacco et al., 2019)	Cell Metabolism	human study, mouse, cell culture	-
VDAC1	30293774 (Zhang et al., 2019)	Cell Metabolism	human study, cell culture	-
C3 ATG16L1	30293775 (King et al., 2019a)	Cell Metabolism	human study, mouse, rat, cell culture	-
PRKCE	30318338 (Brandon et al., 2019)	Cell Metabolism	mouse	-
OR4M1	31230984 (Li et al., 2019)	Cell Metabolism	mouse, cell culture	-
TREM2	31257031 (Jaitin et al., 2019)	Cell	mouse	-
CERS6 MFF	31150623 (Hammerschmidt et al., 2019)	Cell	mouse, cell culture	-
FOXK1 FOXK2	30700909 (Sukonina et al., 2019)	Nature	mouse, cell culture	-
SLC25A5	31528845 (Seo et al., 2019)	Nature Metabolism	human study, mouse, cell culture	-

Dornbos et al, Cell Metabolism 2022

Gap 1: human genetic data is not accessible

- The T2D Knowledge portal allows users to interrogate a gene across hundreds of traits through a simple web interface

hugeamp.org



Costanzo et al, Accepted in principle at Cell Metabolism

View complementary data at the Common Metabolic Diseases Genome Atlas

457 datasets, 473 traits

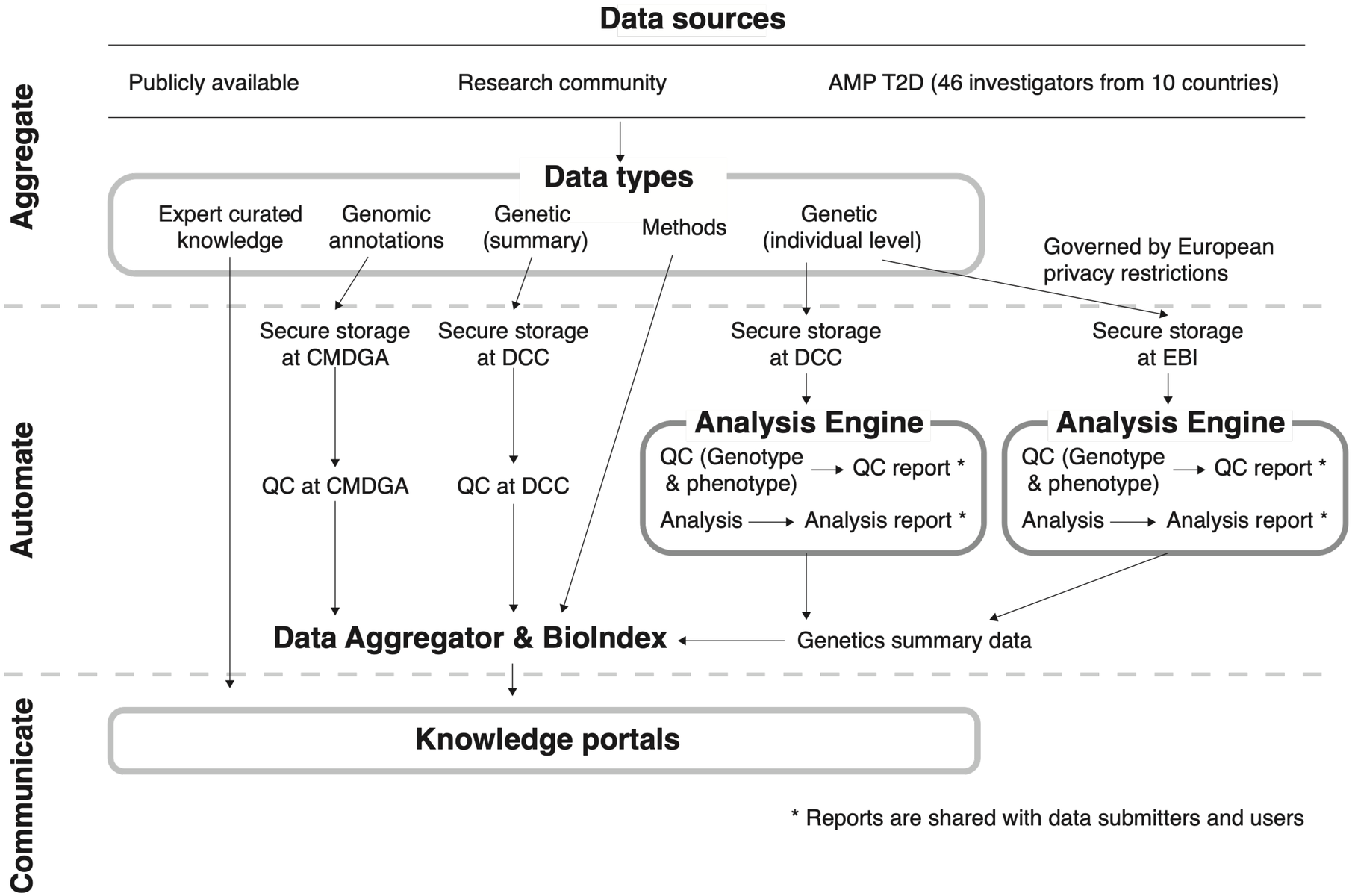
Filter Datasets

Search by phenotypes | Filter by phenotype group | Filter by data type

New Datasets (Click datasets for description)

Dataset	Access	Samples	Ancestry	Technology	Contributing community
GIANT 2022 height GWAS: trans-ancestry	Open access	5314291	Mixed ancestry	GWAS	
Heart failure 2022 GWAS: trans-ancestry	Open access	1665481	Mixed ancestry	GWAS	
GIGASTROKE 2022 GWAS: trans-ancestry	Open access	1614080	Mixed ancestry	GWAS	
Global Biobank Meta-analysis Initiative 2022 POAG GWAS: trans-ancestry		1487447	Mixed ancestry	GWAS	
Global Biobank Meta-analysis Initiative 2022 stroke GWAS: trans-ancestry		1370901	Mixed ancestry	GWAS	
Global Biobank Meta-analysis Initiative 2022 heart failure GWAS: trans-ancestry		1354739	Mixed ancestry	GWAS	
GIGASTROKE 2022 GWAS: European ancestry	Open access	1308460	European	GWAS	
Global Biobank Meta-analysis Initiative 2022 cardiomyopathy GWAS: trans-ancestry		1193060	Mixed ancestry	GWAS	
Global Biobank Meta-analysis Initiative 2022 POAG GWAS: European ancestry		1172905	European	GWAS	
Global Biobank Meta-analysis Initiative 2022 stroke GWAS: European ancestry		1039382	European	GWAS	
Global Biobank Meta-analysis Initiative 2022 heart failure GWAS: European ancestry		1020441	European	GWAS	
Global Biobank Meta-analysis Initiative 2022 cardiomyopathy GWAS: European ancestry		922988	European	GWAS	
Global Biobank Meta-analysis Initiative 2022 heart failure GWAS: females, trans-ancestry		633306	Mixed ancestry	GWAS	
Global Biobank Meta-analysis Initiative 2022 stroke GWAS: females. trans-ancestry		601704	Mixed ancestry	GWAS	
Global Biobank Meta-analysis Initiative 2022 POAG GWAS: females, trans-ancestry		582981	Mixed ancestry	GWAS	

Processing pipeline



Costanzo et al, Accepted in principle at Cell Metabolism

Example: combining multiple associations for a variant?

rs1234567
T2D

N=10,000

p=1.7e-10

N=30,000

p=2.4e-2

N=50,000

p=0.27

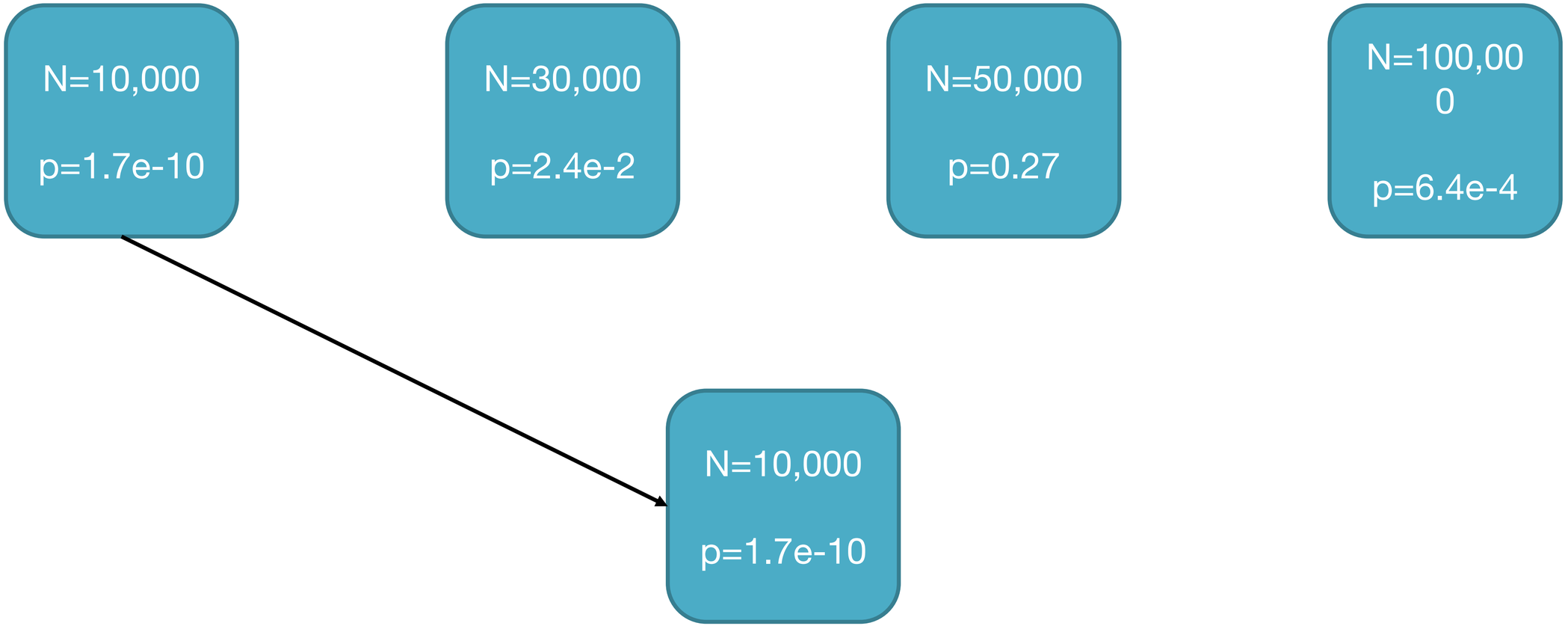
N=100,000

p=6.4e-4

Example: combining multiple associations for a variant?

1. Minimum p-value

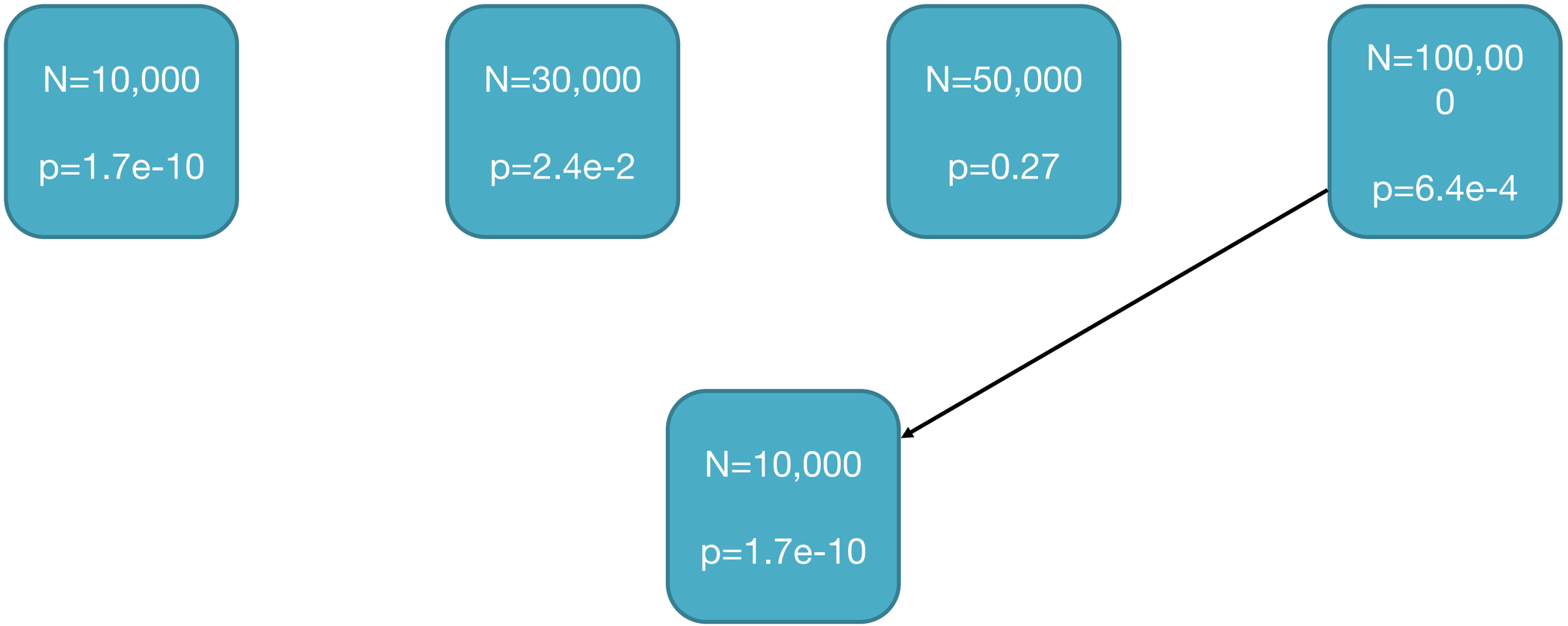
rs1234567
T2D



Example: combining multiple associations for a variant?

1. Minimum p-value
2. Largest dataset

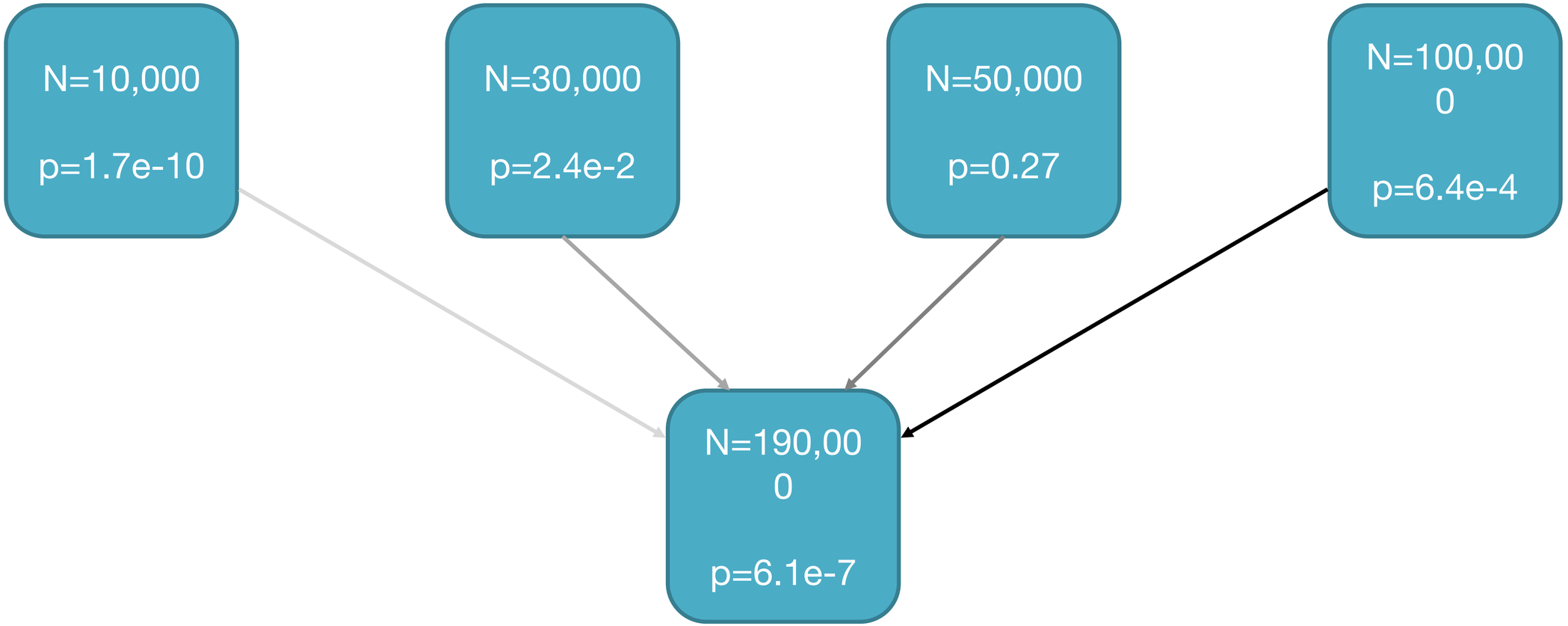
rs1234567
T2D



Example: combining multiple associations for a variant?

1. Minimum p-value
2. Largest dataset
3. Meta-analysis

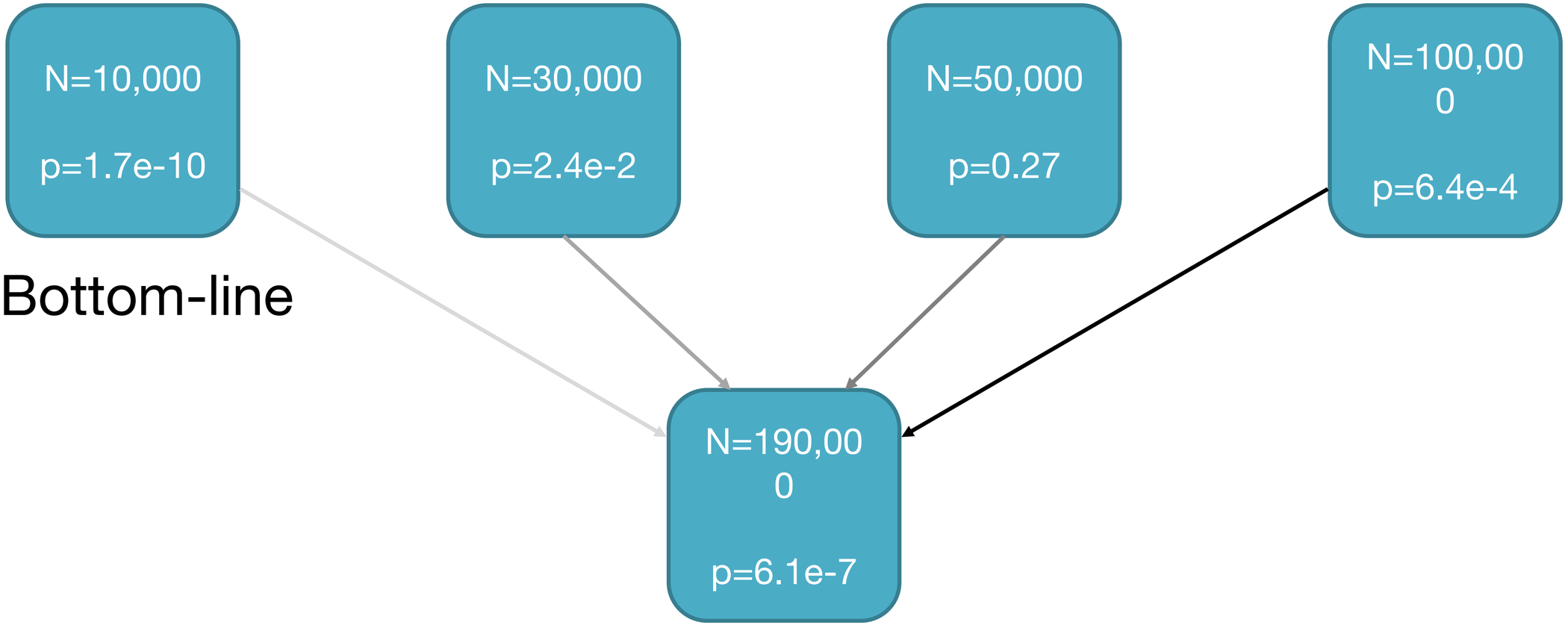
rs1234567
T2D



Example: combining multiple associations for a variant?

1. Minimum p-value
2. Largest dataset
3. Meta analysis
4. Bottom-line

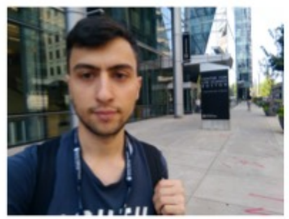
rs1234567
T2D

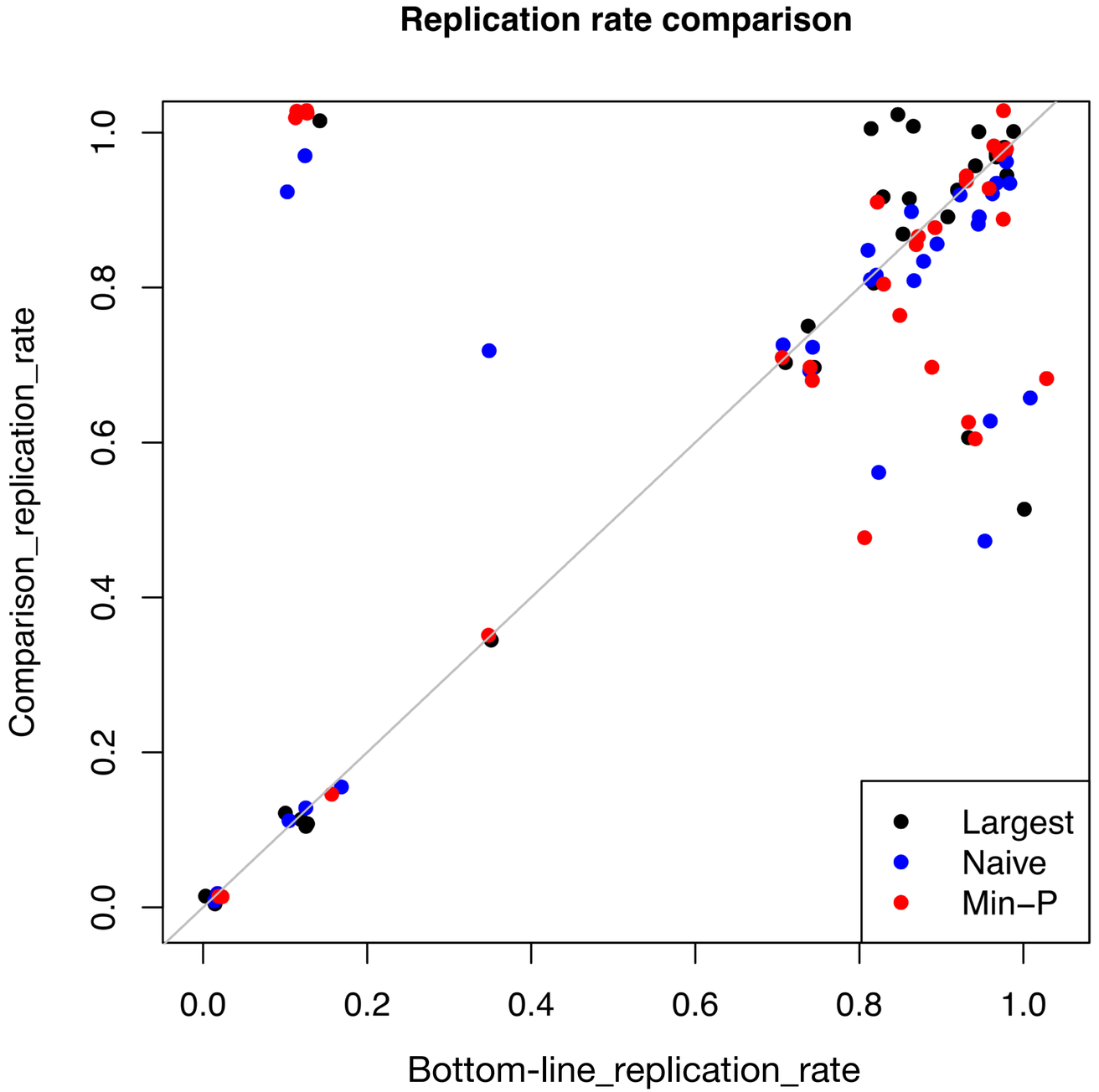


Bottom-line method is more accurate

Method	Replicated	Total	Rate
Bottom-line	**1628**	**2001**	**81.4%**
Largest	1561	2200	71.0%
Min-P	1681	2579	65.2%
Naïve	1903	3058	62.2%

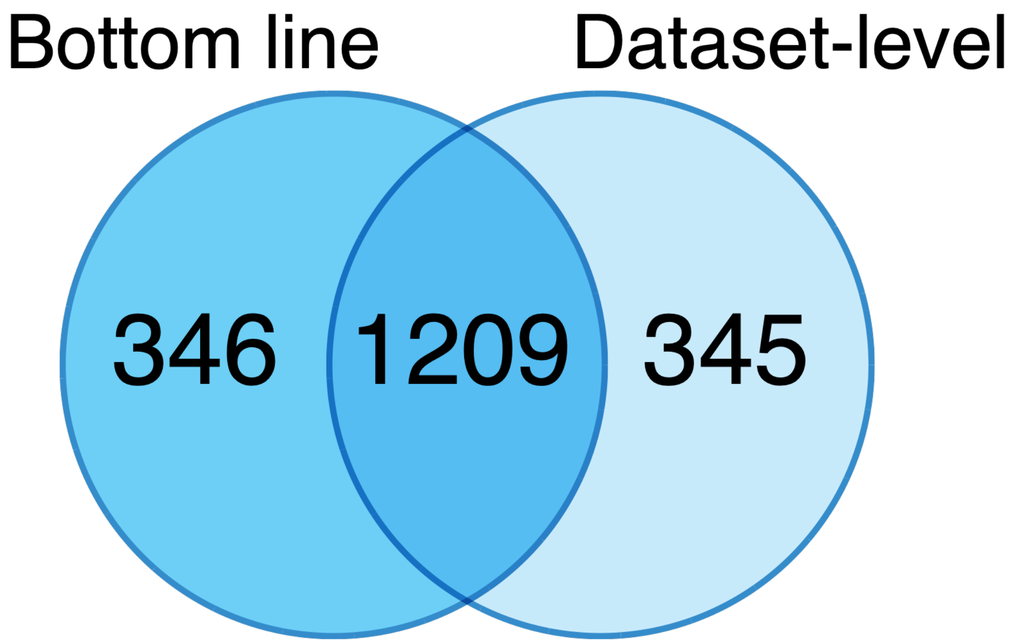
*In this case: Largest means “largest after validation dataset is removed”

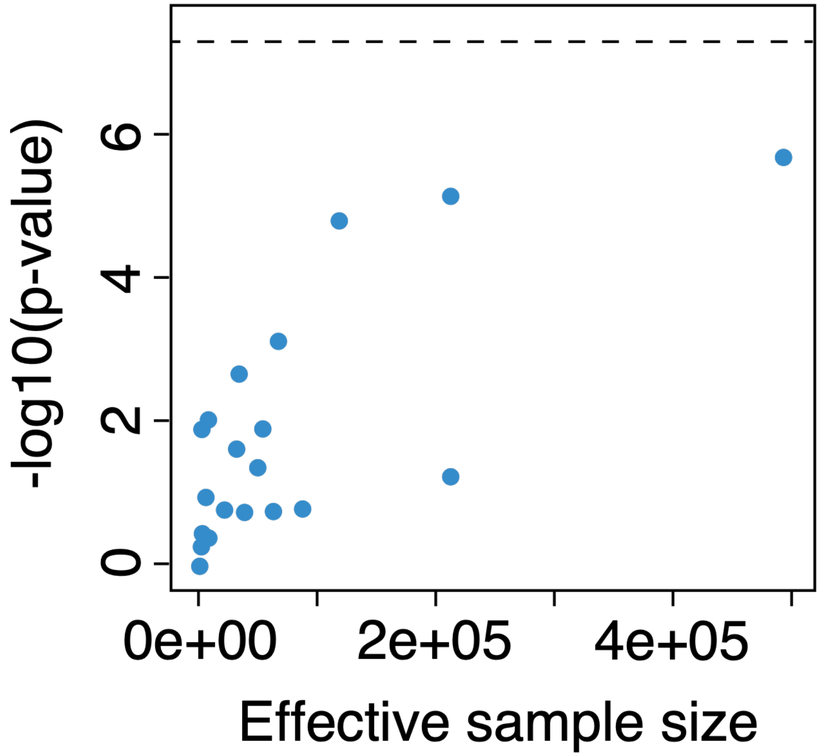




Associations in the T2DKP

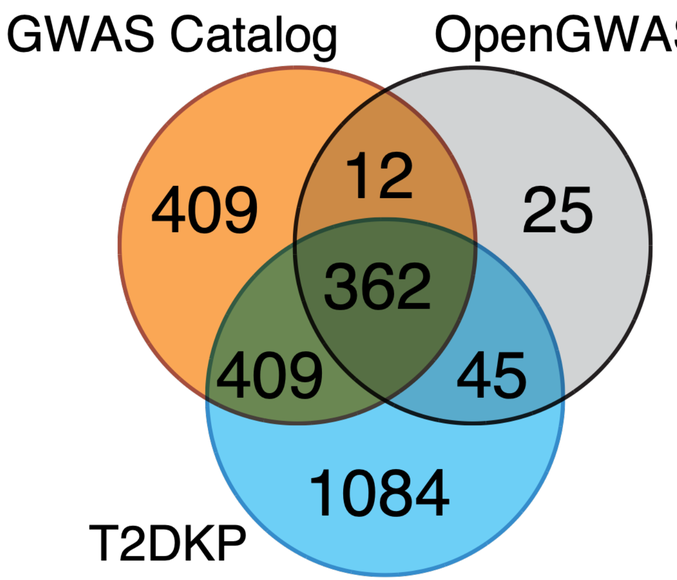
T2D, Fasting glucose, Fasting insulin, HbA1C



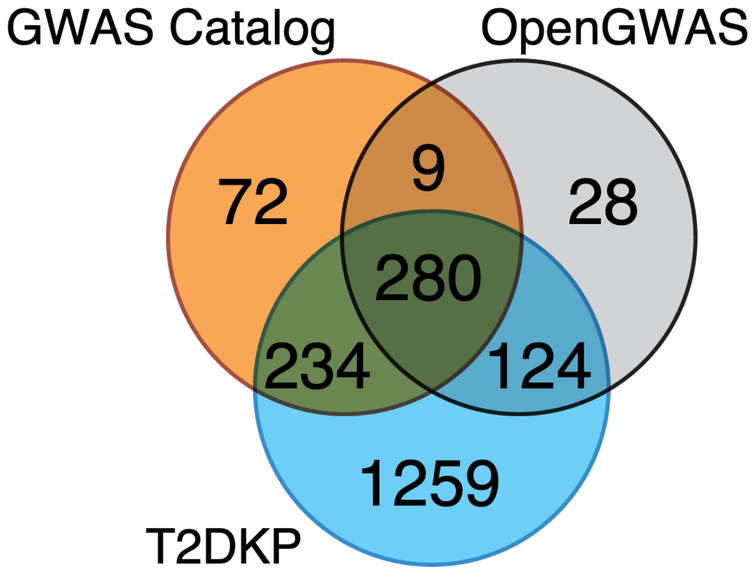


Associations in the T2DKP

T2D, Fasting glucose, Fasting insulin, HbA1C

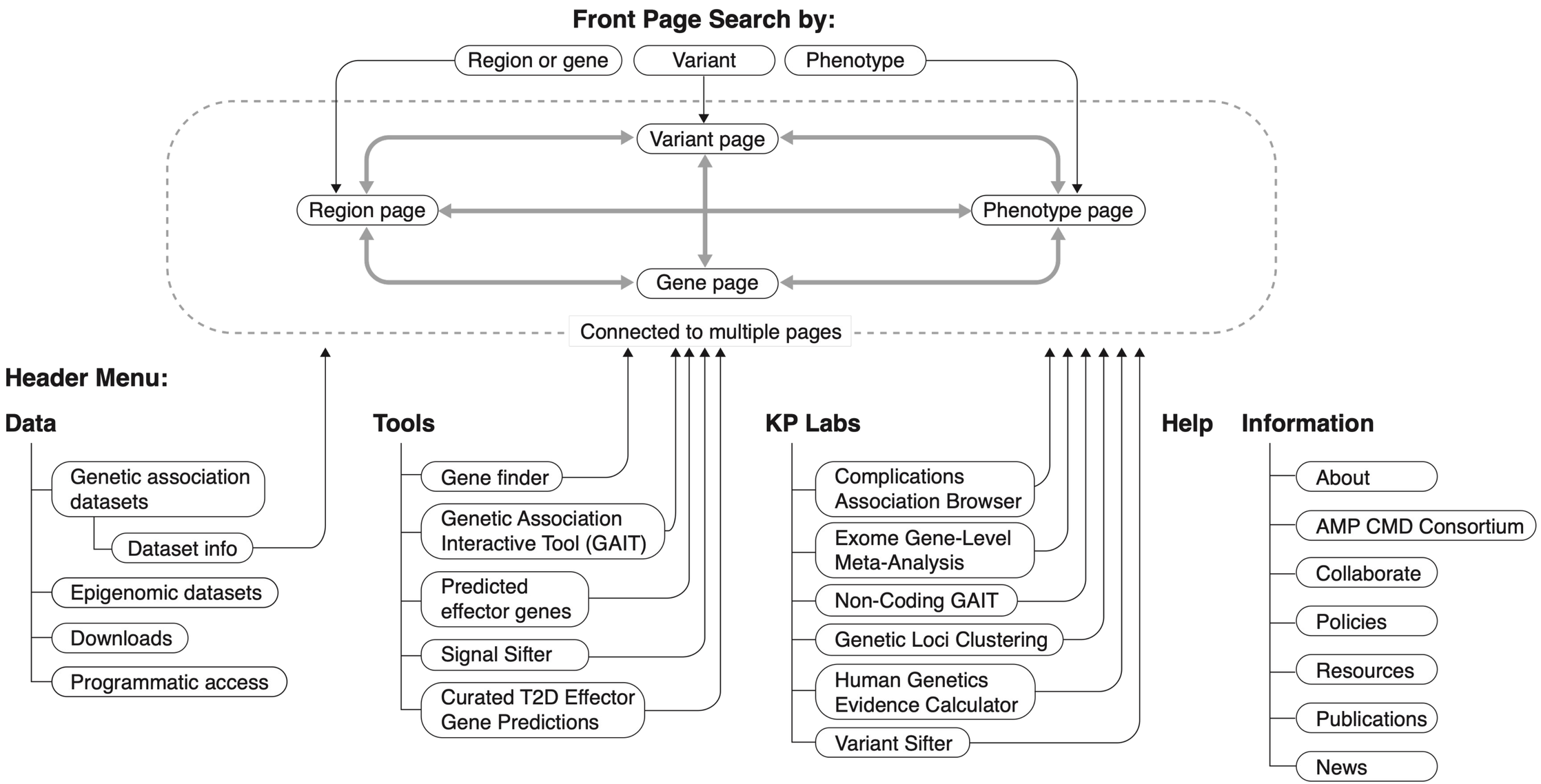


c Associations compared to other resources



e Associations from public summary statistic datasets

Access mechanisms

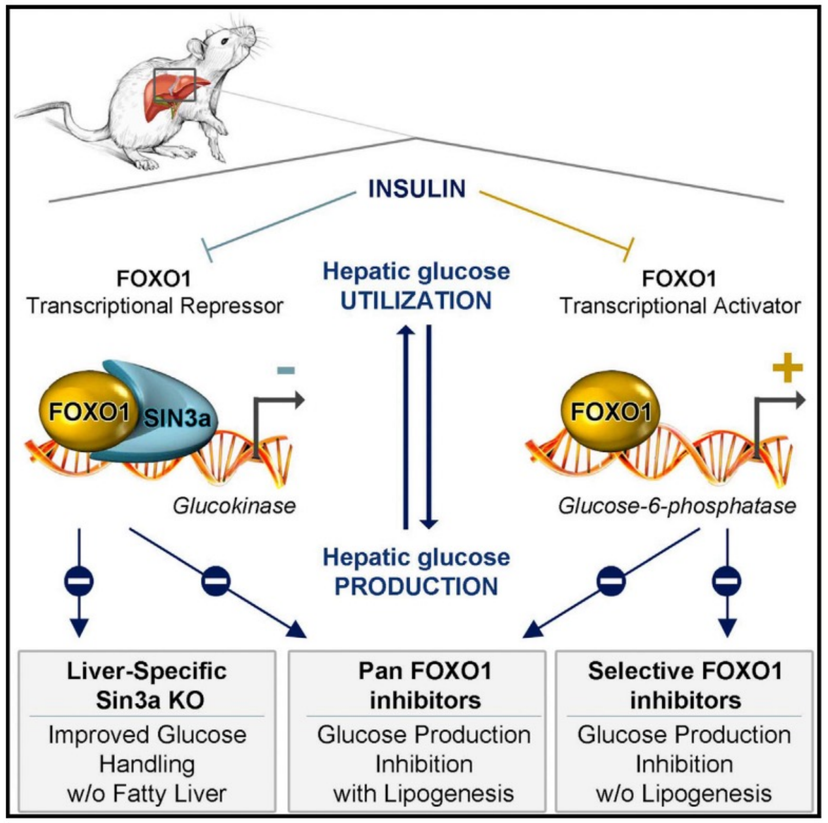


Costanzo et al, Accepted in principle at Cell Metabolism

Cell

Selective Inhibition of FOXO1 Activator/Repressor Balance Modulates Hepatic Glucose Handling

Graphical Abstract



Authors

Fanny Langlet, Rebecca A. Haeusler, Daniel Lindén, ..., Ling Wang, Christoph Buettner, Domenico Accili

Correspondence

da230@columbia.edu

In Brief

The transcriptional output of FOXO1 can be selectively modulated in a way that might reduce adverse effects of insulin sensitizers.

Set page level parameters

Chromosome: Start position - End position

15:75,611,720-75,798,183 Expand region by:

The Region page summarizes genetic associations, credible sets, and epigenomic annotations across a selected region. All coordinates are in the hg19 (GRCh37) genome build. Effect sizes and directions refer to the alternate (non-reference) allele. Genetic associations are generated by "bottom-line" integrative analysis. Click here for a complete guide to the Region page.

1–2–3 Modules on the Region page contribute to the Knowledge Portal Genetic support and Multi-trait analysis workflows.

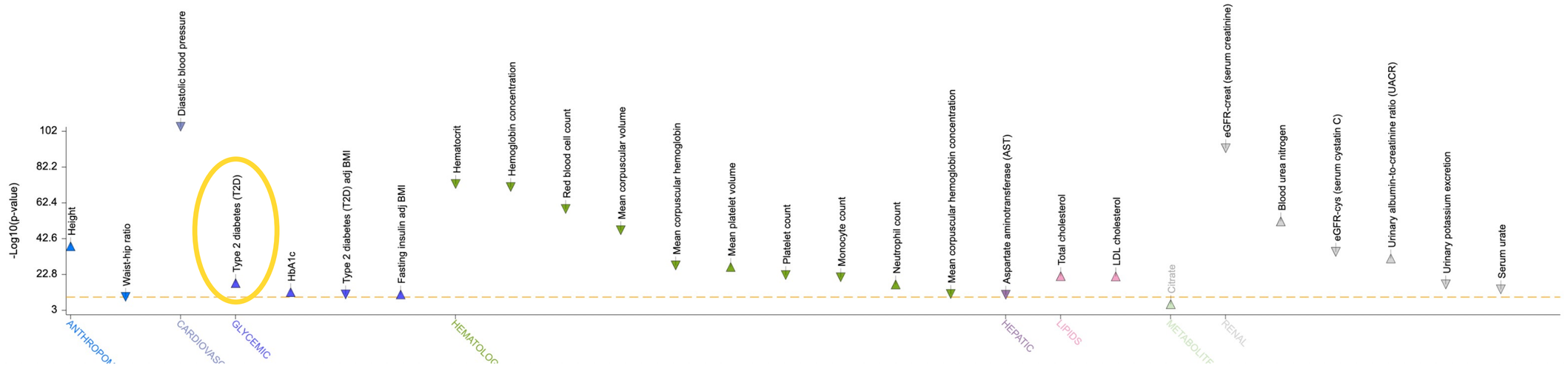
Genes overlapping region

COMMD4 NEIL1 MAN2C1 SIN3A PTPN9

Most significant variant associations in the region: 15:75,611,720-75,798,183 (Ancestry: All)

Associations are clumped by linkage disequilibrium. Click "View associations by phenotype group" for an alternative visualization.

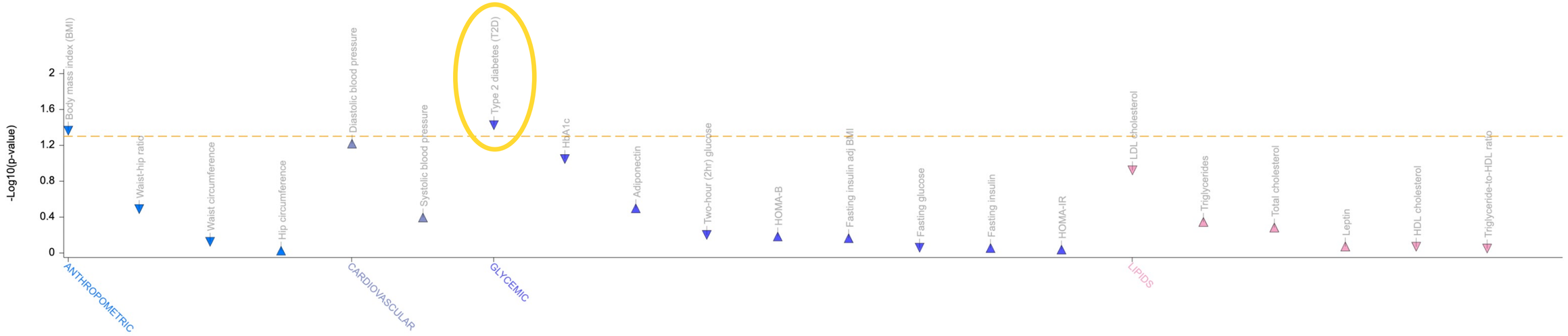
View associations by phenotype group



Rare variant gene-level associations for FOXO1

Transcript

FOXO1



Download CSV

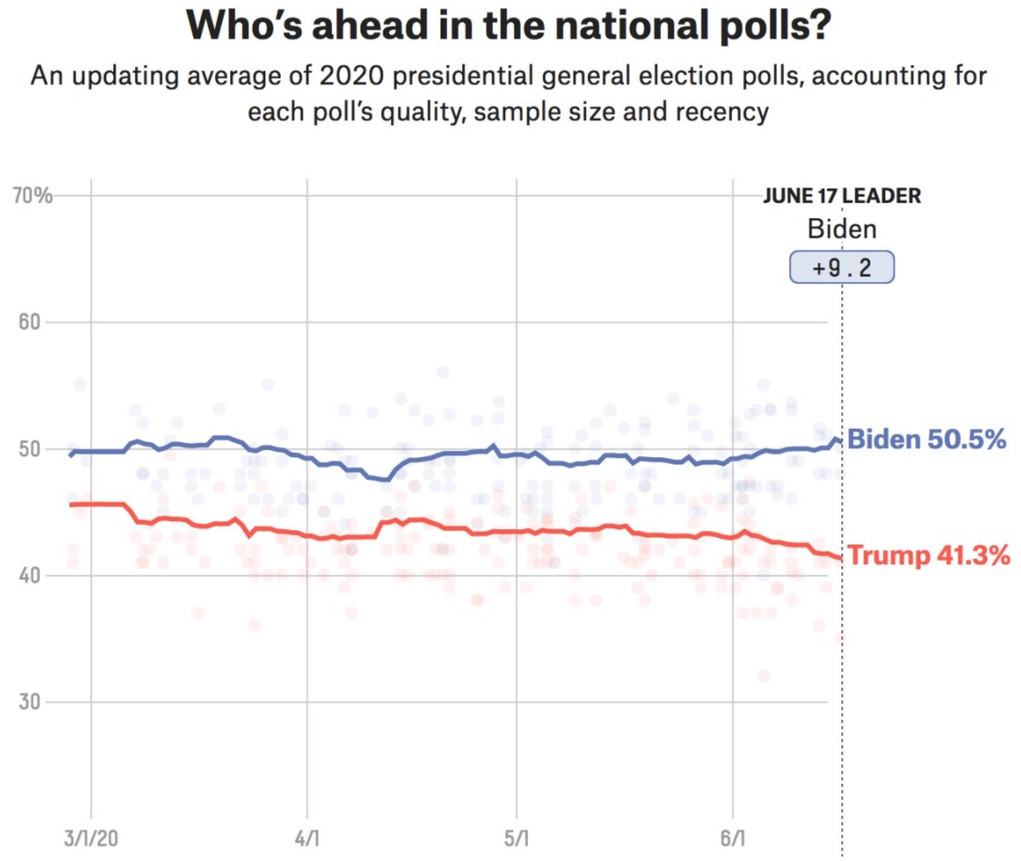
Phenotype	pValue	Beta	Odds Ratio	View
Type 2 diabetes (T2D)	0.0376		▼0.7672	Masks + Plot

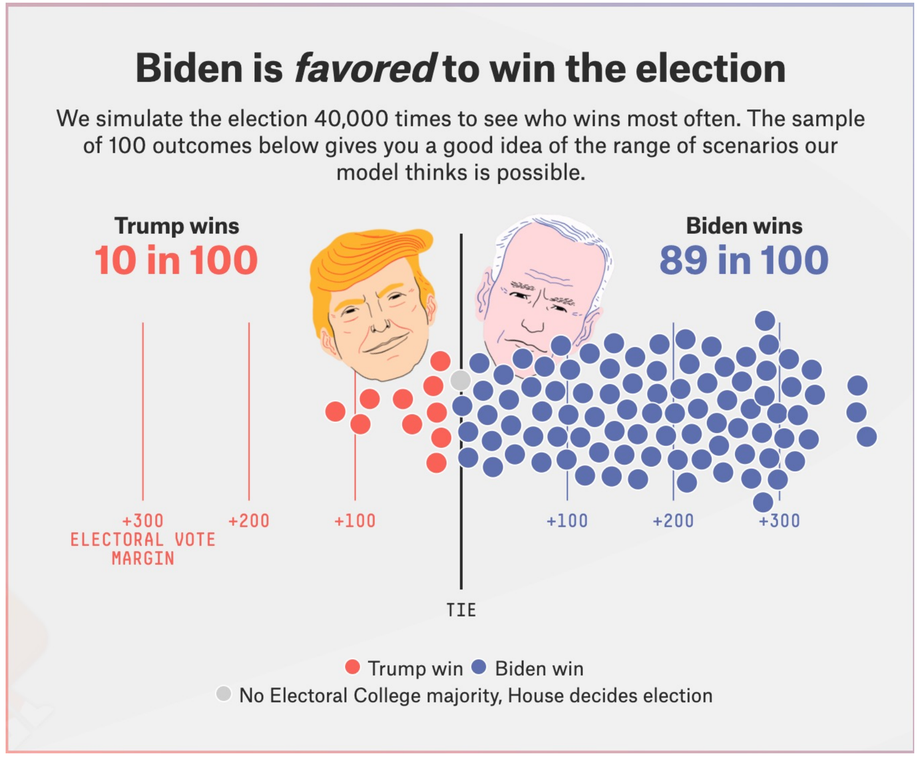
Gap 2: Interpreting genetic associations

Gene	GWAS	Exome
SIN3A	Minimum p=9.2e-16	p=0.59
FOXO1	Minimum p=1.91e-5	p=0.036

Converting frequency differences to probabilities

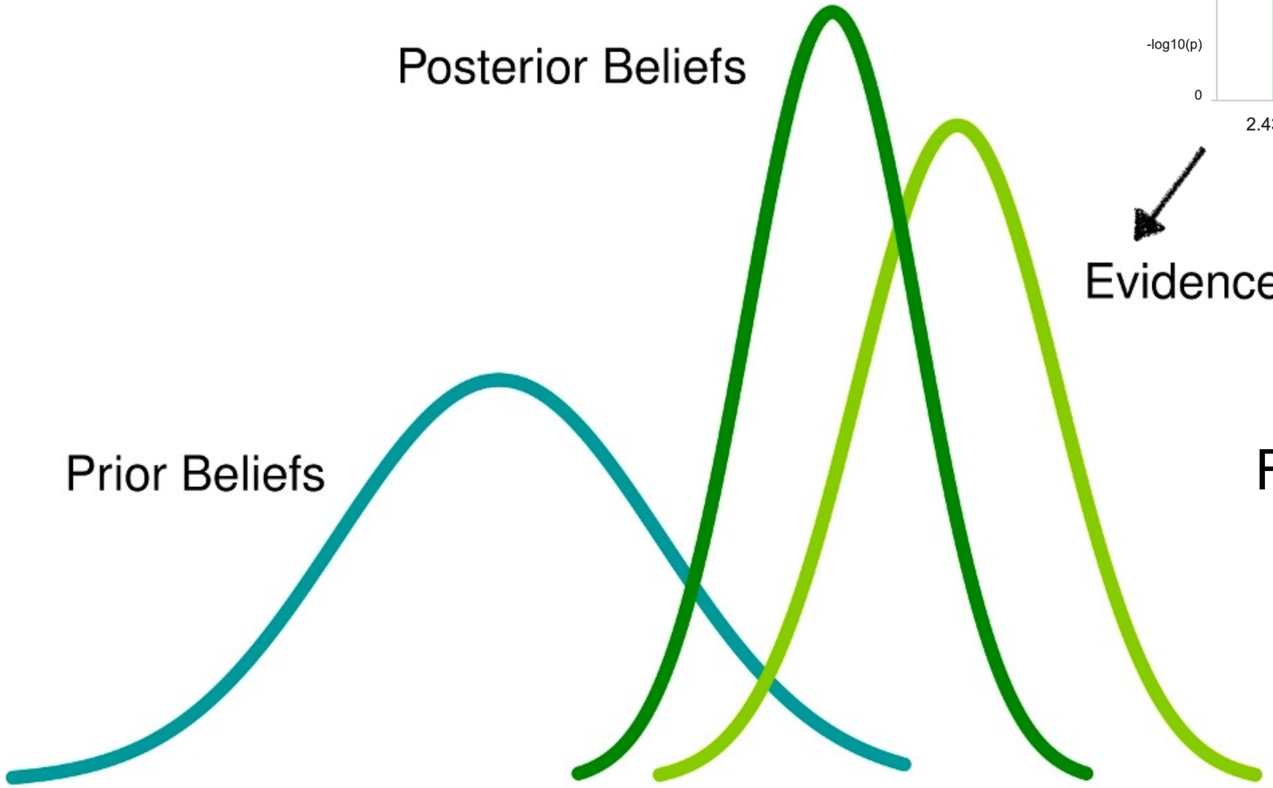
Example: election forecasting

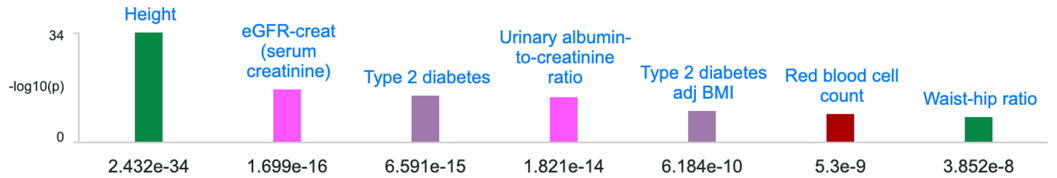




Applied to genetics

Use observed association strength to update prior belief in a gene





Evidence

Formally: $\frac{P(H_1|D)}{P(H_o|D)} = \frac{P(D|H_1)}{P(D|H_o)} \times \frac{P(H_1)}{P(H_o)}$

Bayes Factor (BF)

An important detail

What is the prior?

- **Conservative:** 1 / 20,000 (“we know there must be at least one”)
- **Unprincipled:** Investigator defined (“how much would you wager”)
- **Measured:** 5% (“1,000 T2D genes seems more reasonable than 100 or 10,000”)

Updating the prior

- **Nearby a GWAS association, we know:**
  - Genes with causal coding variants are almost certainly causal
  - Genes nearest to the signal are causal genes ~70% of the time
  - Genes with a coding GWAS association are causal ~50% of the time
  - At least one gene in a GWAS region is the causal gene
- **Given an observed rare variant association**

$$ABF = \sqrt{V/(V+\omega)} * \exp(\frac{\omega\beta^2}{2V(V+\omega)})$$

V = variance β = beta

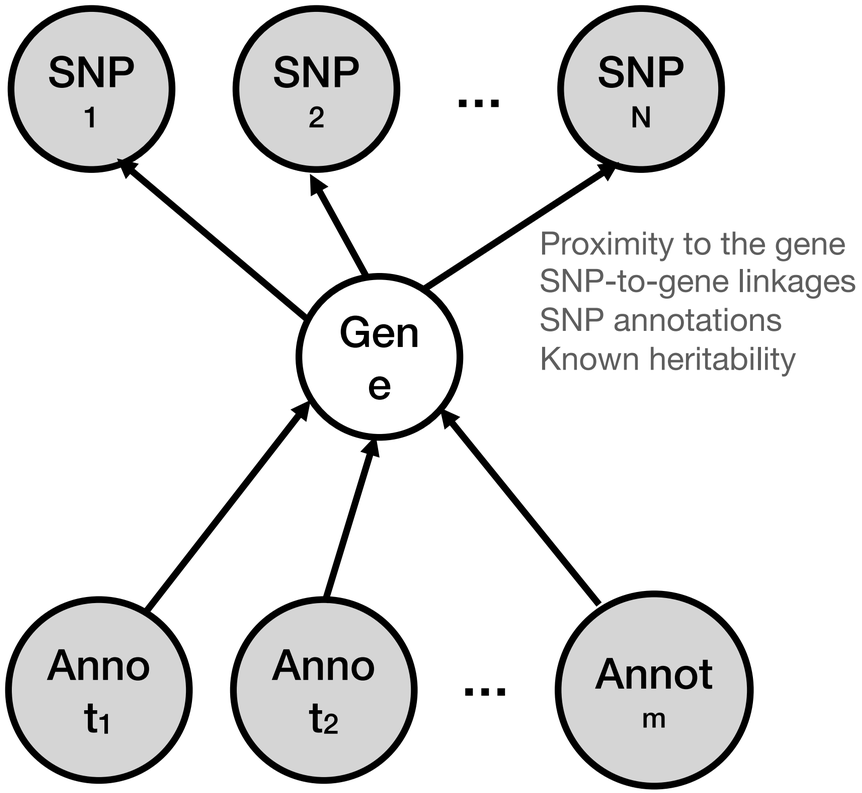
ω = prior allelic variance

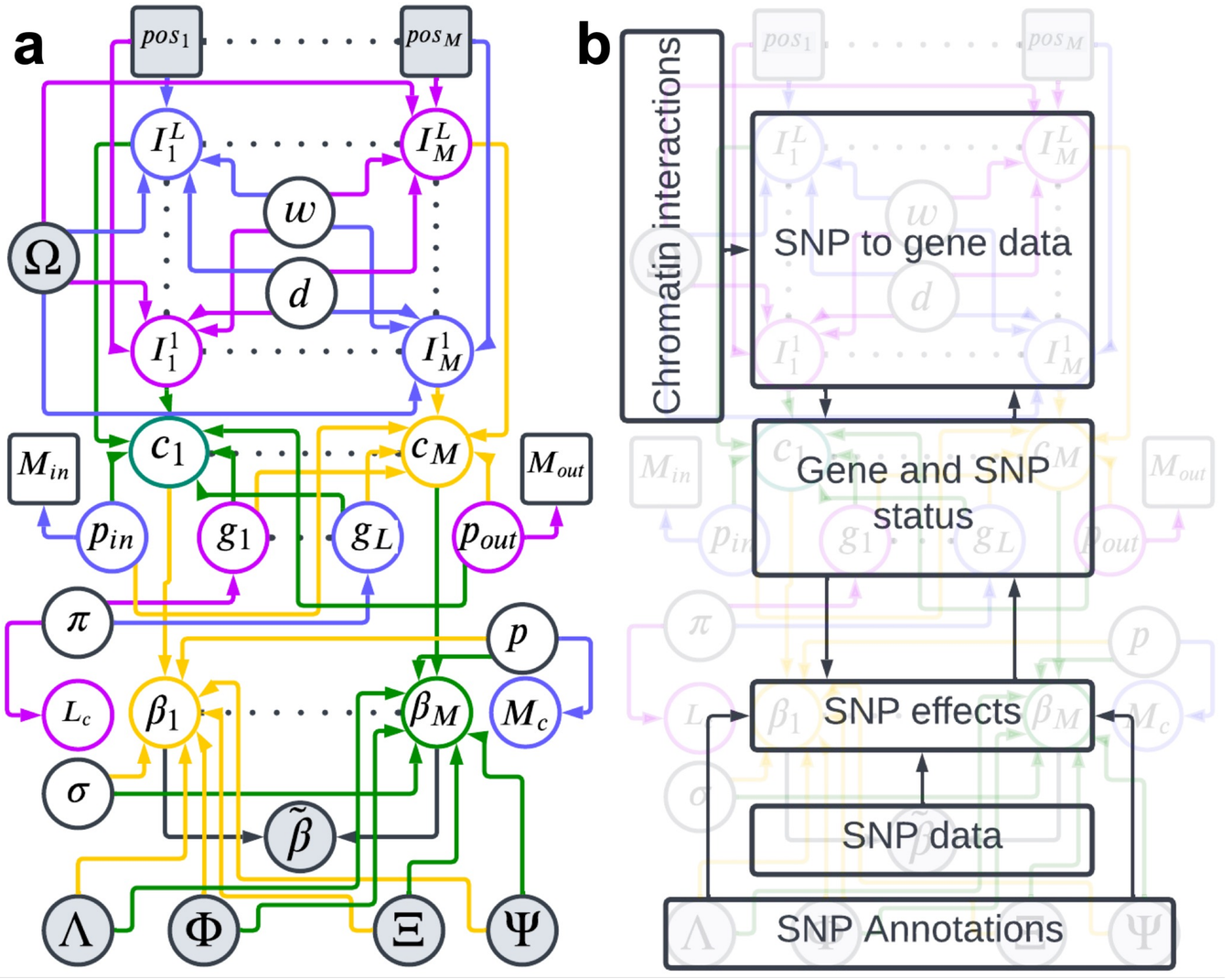
Simple rules for combining these data

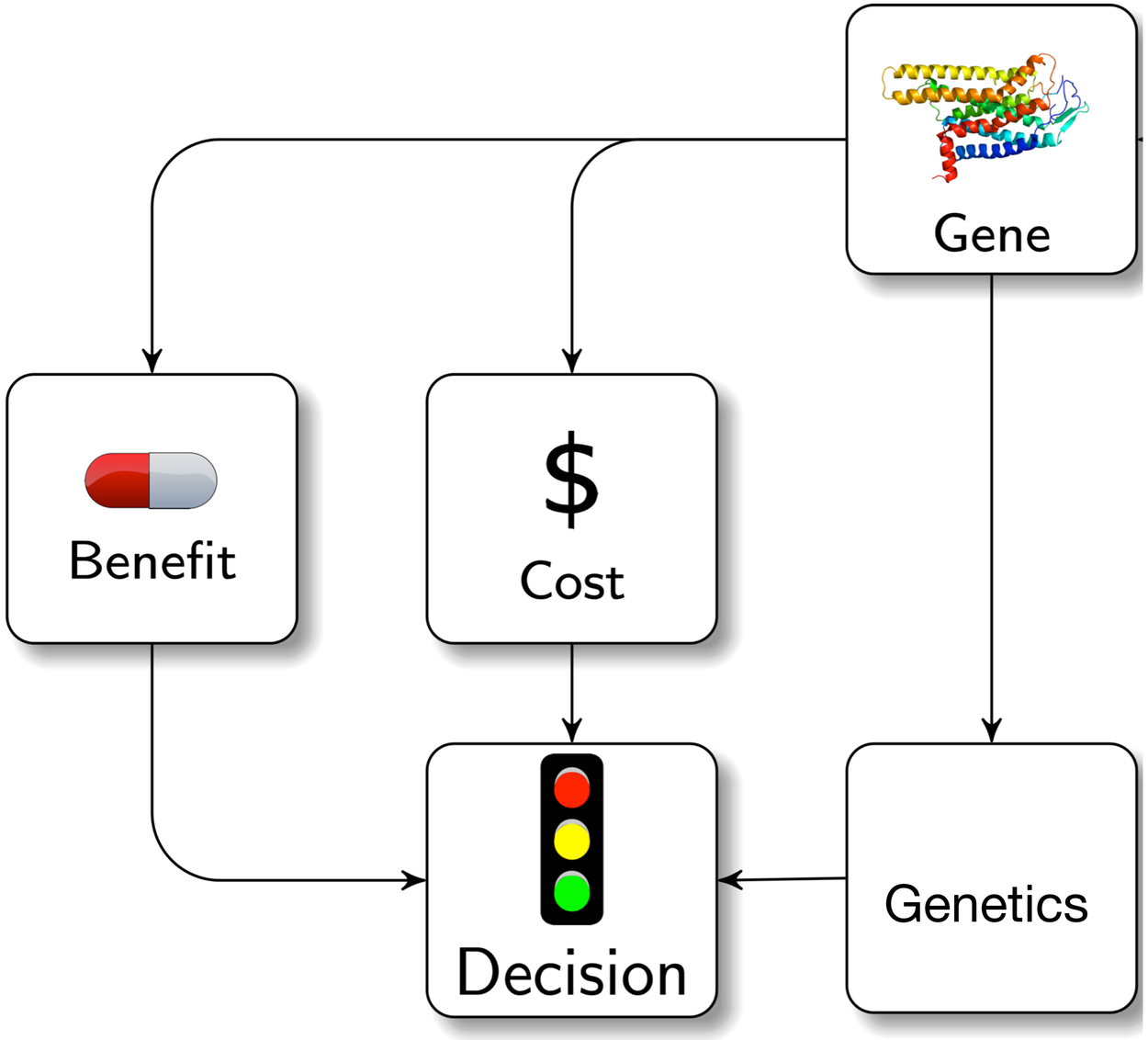
Common Variation	No evidence $p \geq 0.1$	Weak $p < 0.1$	Nominal $p < 0.05$	Strong $p < 1 \times 10^{-3}$	Exome-wide $p < 2.5 \times 10^{-6}$
	Rare Variation				
Causal coding variant	**Compelling** 95% | 99%	**Compelling** 95% | 99%	**Compelling** 99% | 9%	**Compelling** 99% | 99%	**Compelling** 99% | 99%
Nearest gene	**Very Strong** 70% | 90%	**Very Strong** 80% | 95%	**Extreme** 90% | 95%	**Compelling** 99% | 99%	**Compelling** 99% | 99%
Coding variant	**Strong** 50% | 85%	**Very Strong** 60% | 90%	**Very Strong** 75% | 95%	**Compelling** 95% | 99%	**Compelling** 99% | 99%
GWAS locus	**Moderate** 15% | 40%	**Moderate** 20% | 55%	**Moderate** 30% | 70%	**Very Strong** 75% | 95%	**Compelling** 99% | 99%
No evidence	**No evidence** 5% | 20%	**Anecdotal** 5% | 25%	**Moderate** 15% | 45%	**Strong** 50% | 85%	**Compelling** 95% | 99%

Dornbos et al, Cell Metabolism 2022

A peek at a more comprehensive model




a
b
Chromatin interactions
SNP to gene data
Gene and SNP status
SNP effects
SNP data
SNP Annotations


Gene
Benefit
$
Cost
Decision
Genetics

HuGE Calculator

Start by entering a gene name, then choose a phenotype. Browse a list of all available phenotypes here.

Build search criteria

Gene	Phenotype
Select a gene ...	Select a phenotype ...

Search Criterion SIN3A X Type 2 diabetes (T2D) X

HuGE Score (Combined Evidence) **45(Common variation BF) * 1(Rare variation BF) = 45**

*BF=Bayes Factor *HuGE Score(combined evidence) = BF of common variation X BF of rare variation

	45	<--Common Variation BF
X	1	<--Rare Variation BF (Reset Prior Allelic variance)
HuGE Score:	**45**	

HuGE score 45 falls in Very Strong evidence range

Compelling: HuGE Score >= 350 | Extreme: >=100 | Very Strong: >=30 | Strong: >=10 | Moderate: >=3 | Anecdotal: >1 | No Evidence: <=1

Build search criteria

Gene	Phenotype
Select a gene ...	Select a phenotype ...

Search Criterion SIN3A X Type 2 diabetes (T2D) X

HuGE Score (Combined Evidence) 45(Common variation BF) * 1(Rare variation BF) = 45

*BF=Bayes Factor *HuGE Score(combined evidence) = BF of common variation X BF of rare variation

	45	<--Common Variation BF
X	1	<--Rare Variation BF (Reset Prior Allelic variance)
HuGE Score:	**45**	

HuGE score 45 falls in Very Strong evidence range

Compelling: HuGE Score >= 350 | Extreme: >=100 | Very Strong: >=30 | Strong: >=10 | Moderate: >=3 | Anecdotal: >1 | No Evidence: <=1

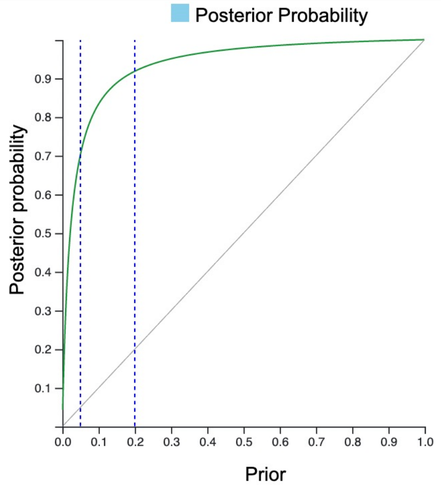
Compelling	Extreme	Very Strong	Strong	Moderate	Anecdotal	No Evidence

▶ Posterior probability

How is PPA calculated?

The posterior probability of association (PPA) is a measure of how different "prior" beliefs in a gene's disease-relevance are modified by the HuGE score. Priors may be defined using various criteria, such as an estimate of the total number of genes impacting a common disease, or experimental results bearing on the gene's role in a disease.

	Suggested prior	Posterior probability	
	0.2	0.9183673	✕
	0.05	0.703125	✕
Set prior			



Knowledge Portal Network

Neurodegenerative Disease Knowledge Portal

Learn more

Association to Function Knowledge Portal

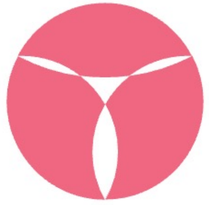
Learn more

Common Metabolic Diseases Knowledge Portal

Learn more

Cardiovascular Disease Knowledge Portal

Learn more



Reproductive System Knowledge Portal

Learn more

Ocular Knowledge Portal

Learn more

Autoimmune Disease Knowledge Portal

Learn more

Cerebrovascular Disease Knowledge Portal

Learn more

Sleep Disorder Knowledge Portal

Learn more

Type 1 Diabetes Knowledge Portal

Learn more

Type 2 Diabetes Knowledge Portal

Learn more

Musculoskeletal Knowledge Portal

Learn more

Lung Disease Knowledge Portal

Learn more

Lipid Droplet Knowledge Portal

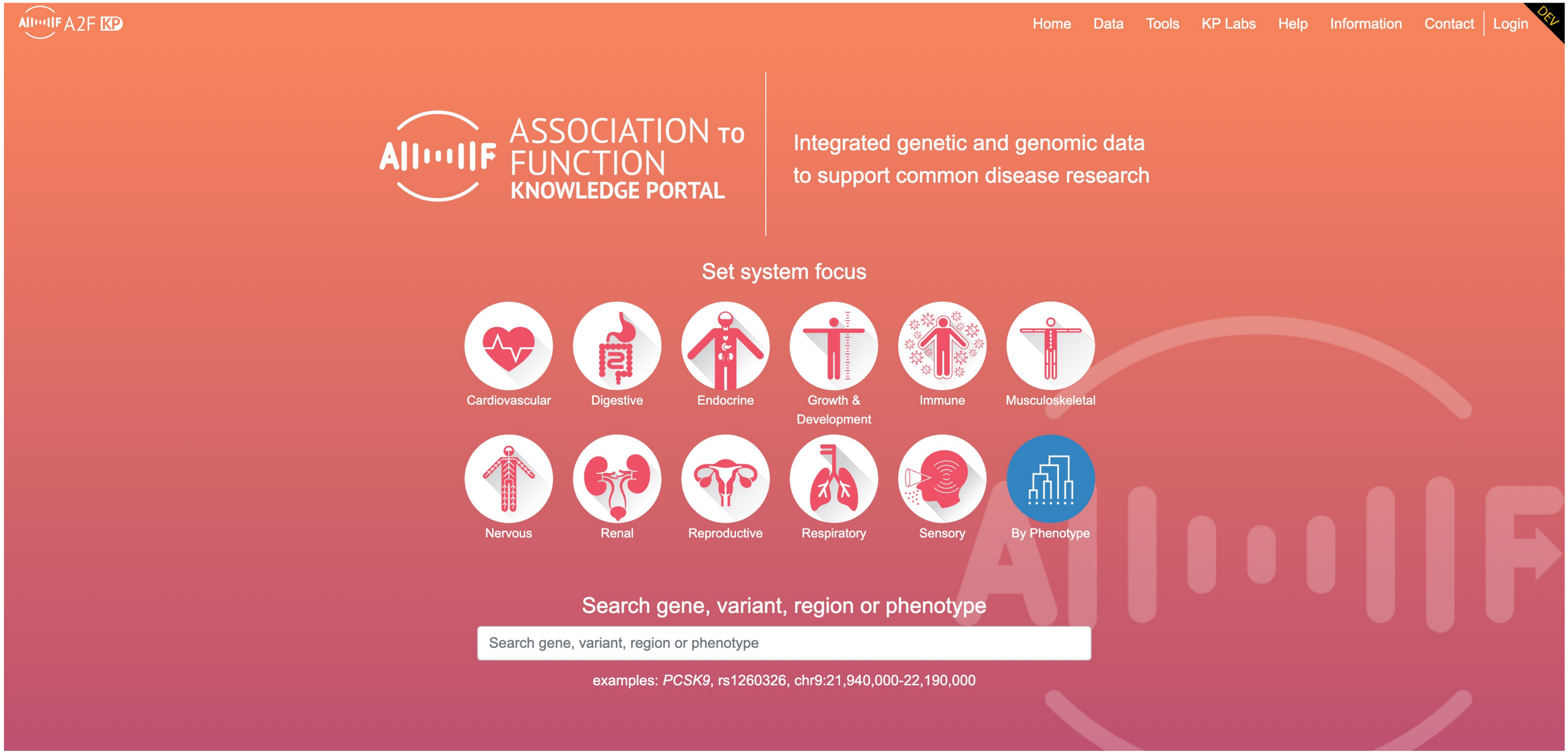
Learn more

Non-Additive Genetic Effects Knowledge Portal

Learn more

- The Association to Function portal combines all of these portals in one location

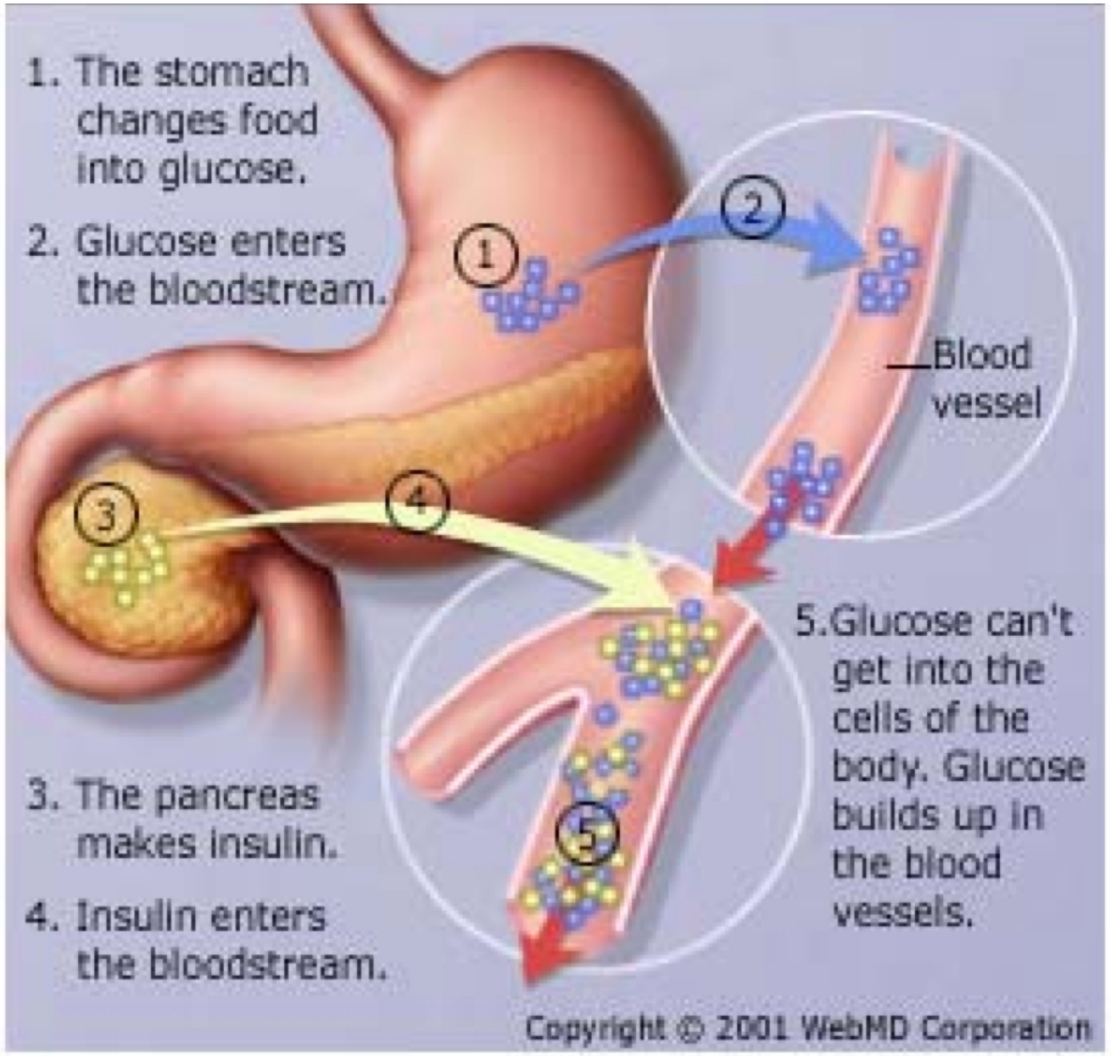
[a2fkp.org](a2fkp.org)



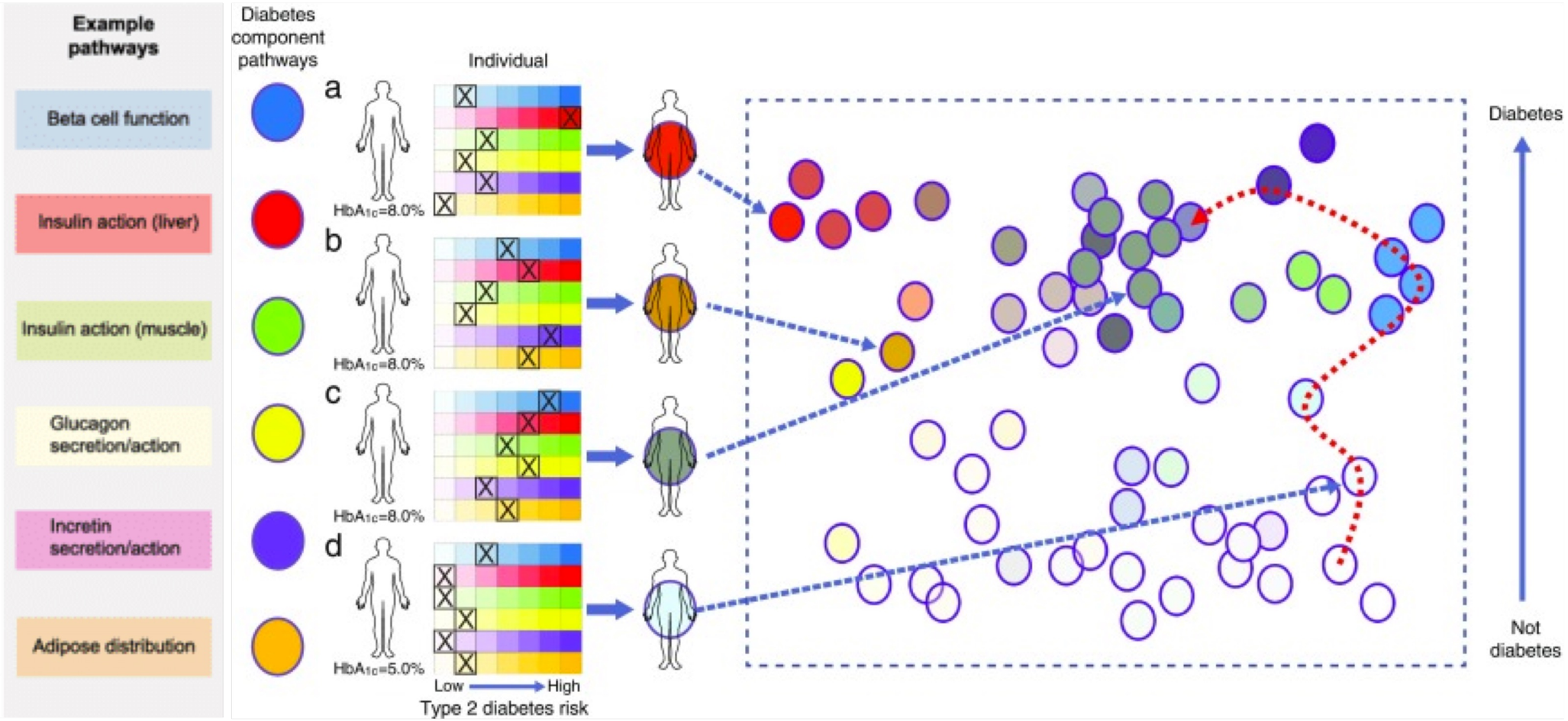
Problem:

What is the trait?

Type 2 diabetes…



…is not really a single disease



McCarthy, Diabetologia, 2017

RADIANT
Rare and Atypical Diabetes Network

Our Research | About Us | Information for Researchers | Join RADIANT

or Participant Portal Login

English | Español

If you've been diagnosed by your doctor with diabetes, but do not fit the usual pattern of either type 1 or type 2 diabetes, you may be eligible to join RADIANT.

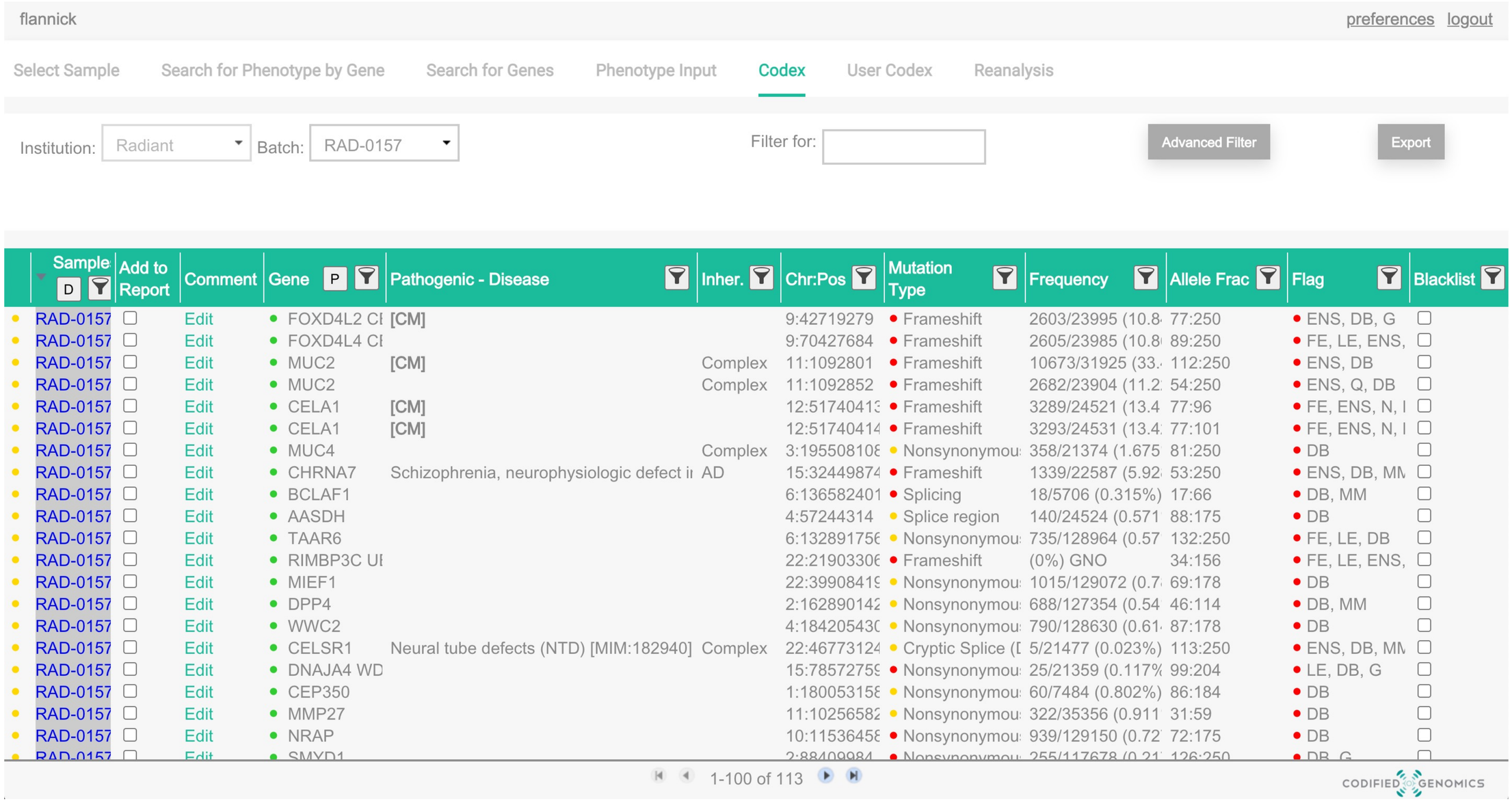
Learn More

There is currently little information and resources for atypical diabetes.
RADIANT wants to change this.

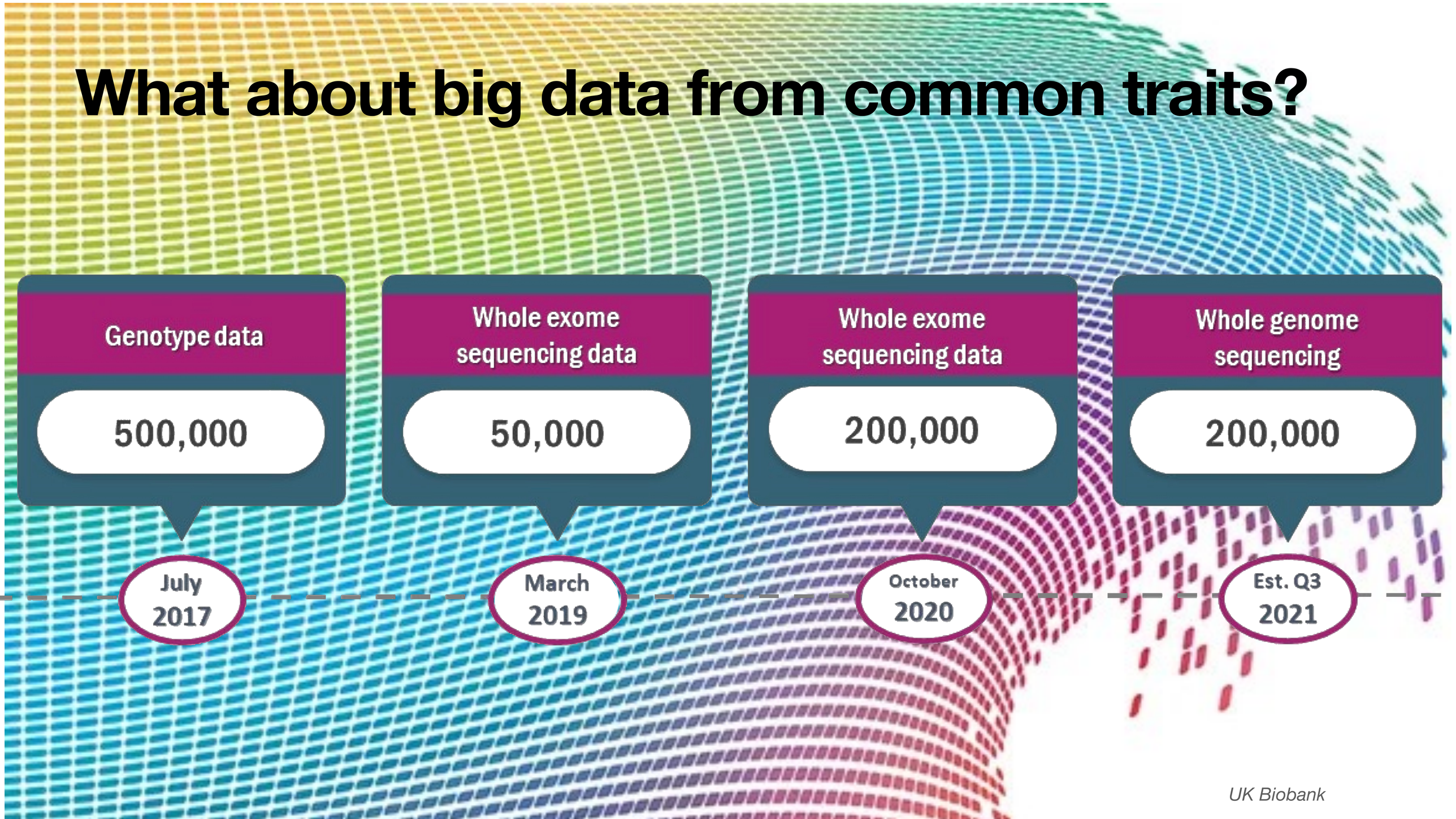
Example case

- “…diagnosed with diabetes at age 18…”
- “…highest HbA1C 8.7%…”
- “…history of ADHD, depression, tinnitus, and hyperlipidemia…”
- “…father diagnosed with diabetes at age 33 when not overweight; two grandparents and an uncle diagnosed with diabetes later in life while overweight…”

Sifting through genes and “post-hoc storytelling”

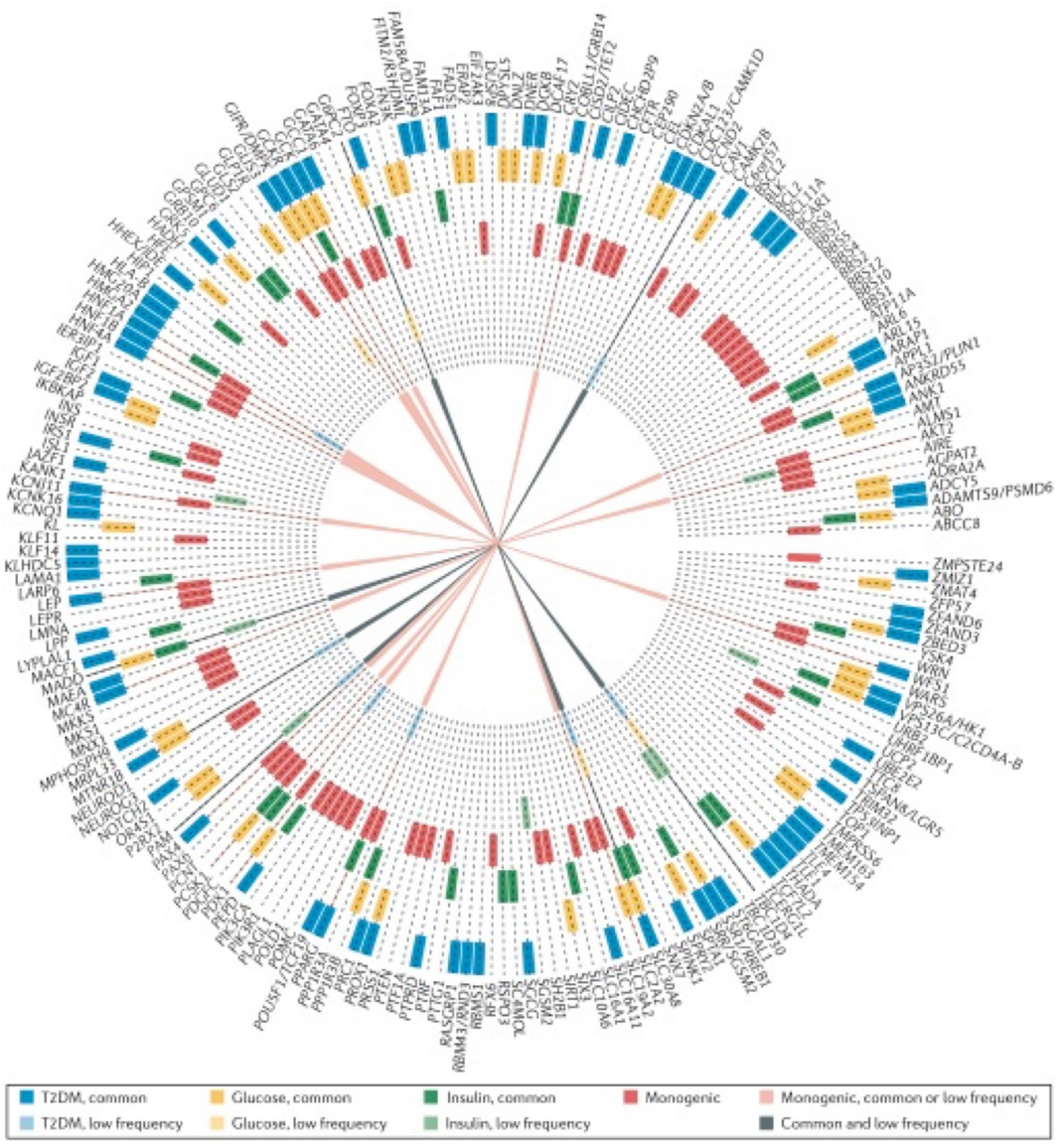


What about big data from common traits?



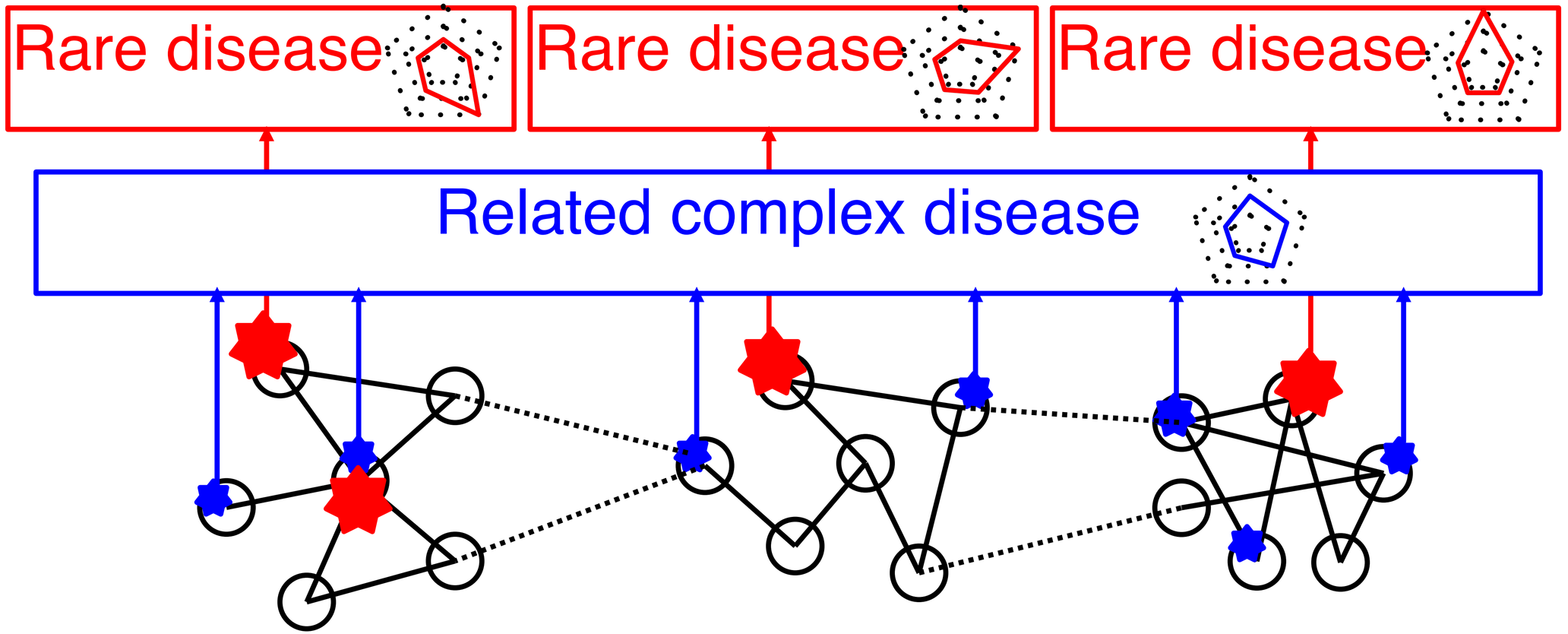
UK Biobank

Intuitively, T2D genes are logical candidates

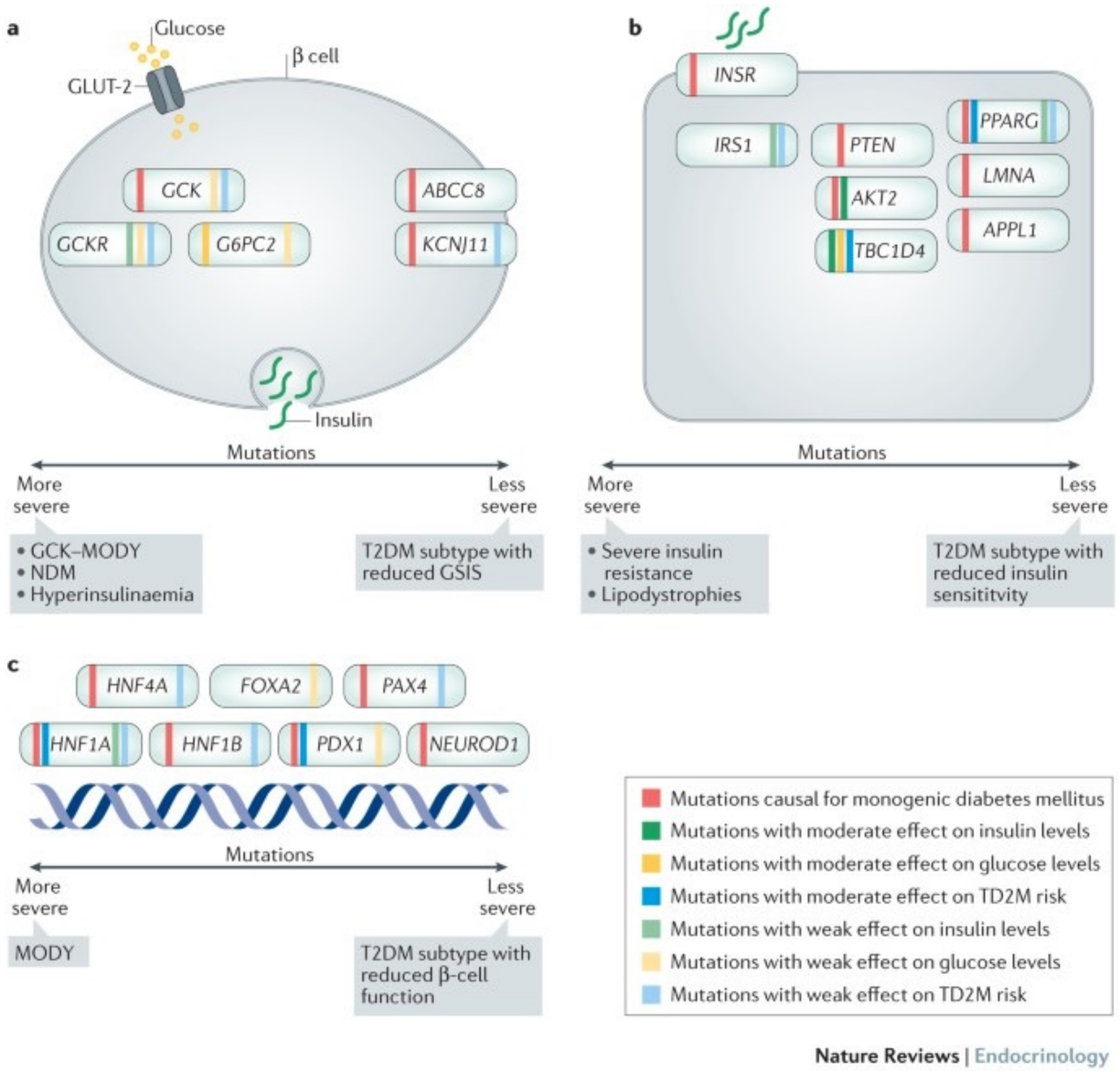


Flannick et al, 2016

Model: mutations in a pathway cause similar phenotypes



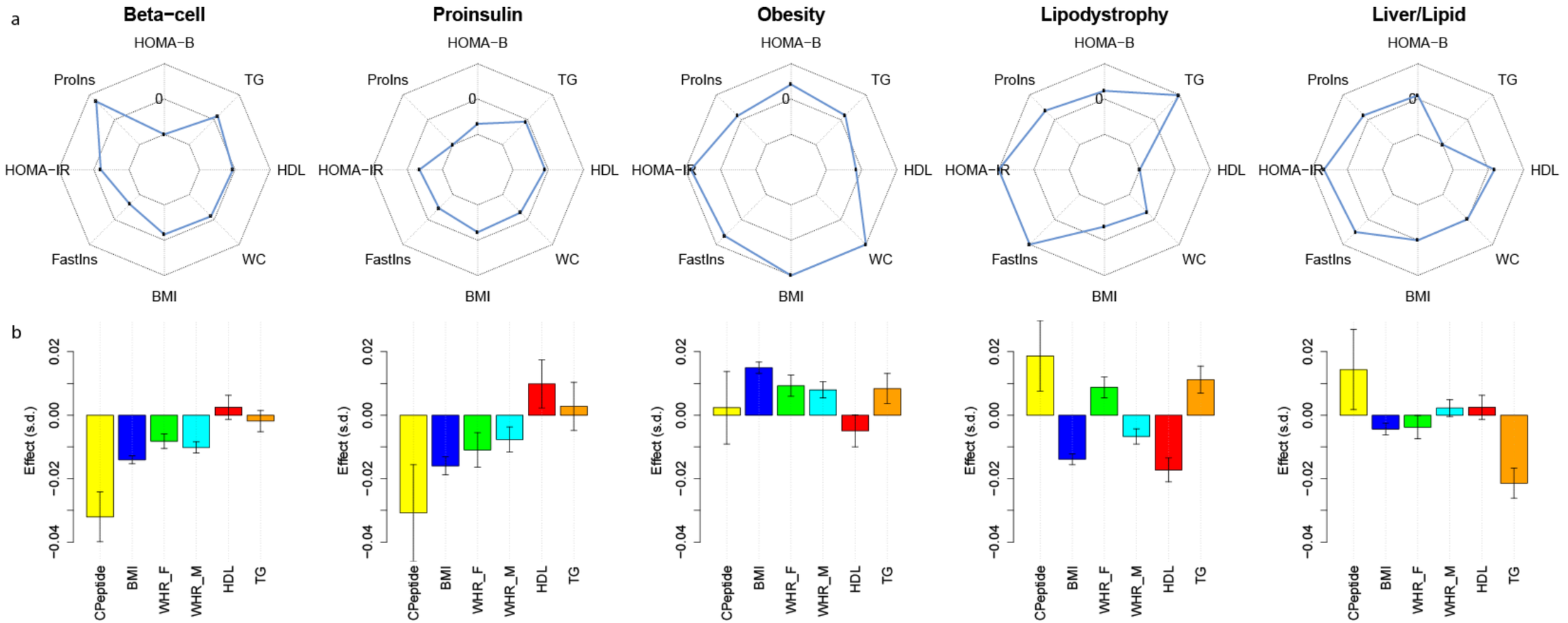
Supported anecdotally



Flannick et al, 2016

Supported anecdotally

- Patterns of trait associations for T2D GWAS SNPs



Udler et al, 2018

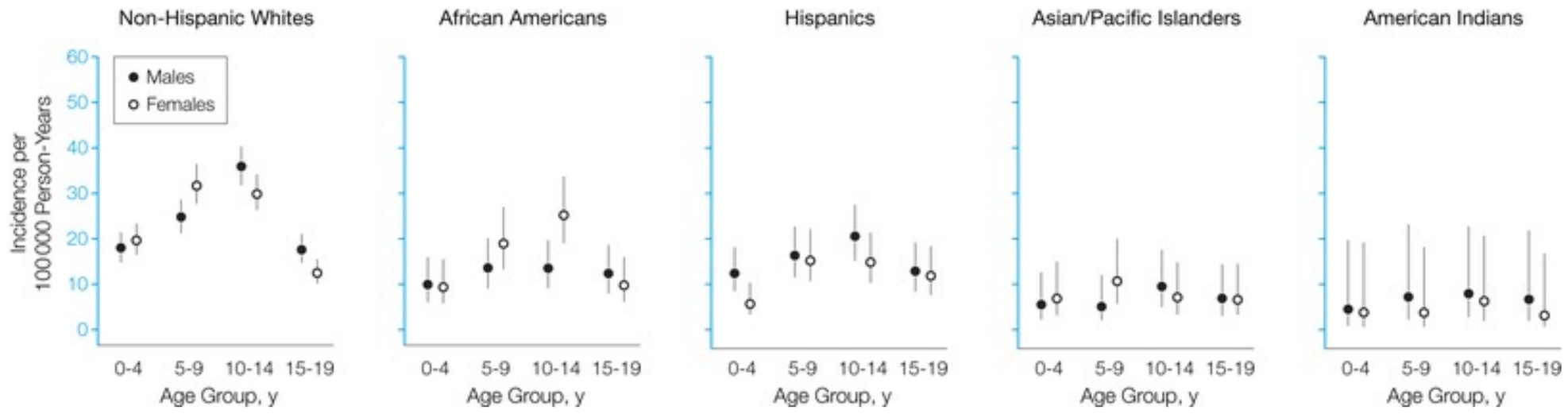
The extent to which we can “transfer” associations across similar diseases depends on the extent to which phenotypic similarity predicts genetic similarity

And the extent to which rare and common diseases lie on a “phenotypic continuum”

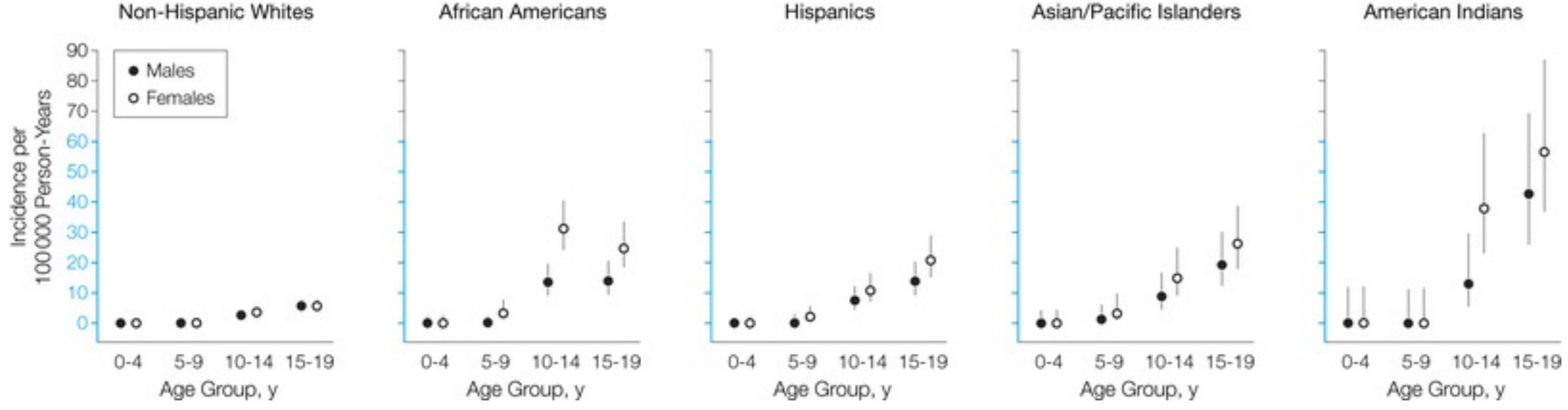
The ProDiGY study of T2D in youth

- Incidence of ~3,700 cases/year and increasing, particularly ages 10-19
  - 15% of new diabetes cases in whites, 46-86% in minorities

T1D



T2D



Samples

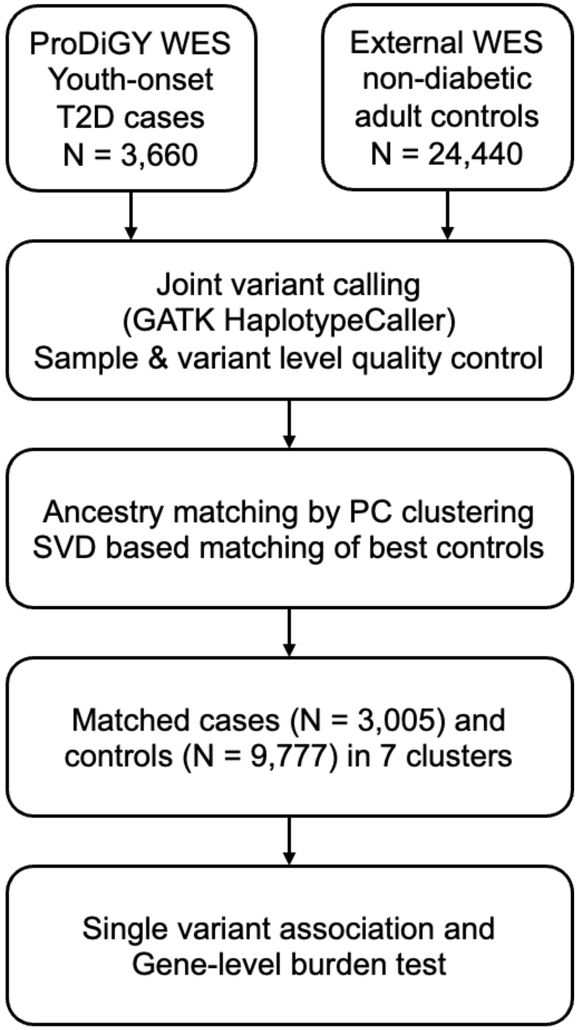
- SEARCH for Diabetes in Youth
  - Longitudinal follow up to assess natural history and complication risk factors
  - Active registry of youth diagnosed with diabetes at age < 20
- TODAY
  - Clinical trial of ages 10-17 to compare treatment efficacy of *Metformin vs Metformin+Lifestyle Intervention vs Metformin+Rosiglitazone*
  - BMI above 85th percentile
- Both studies are multi-ethnic

Ancestry	Samples
African-American	1,491 (40.8%)
East-Asian	62 (1.7%)
European	757 (20.7%)
Hispanic	1,306 (35.9%)
NA	34 (0.9%)
Total	3,650

	Total	Male	Female
N	3,650	1,294 (35.4%)	2,356 (64.6%)
Current Age	15.2±3.0	15.1±3.1	15.4±2.8
Age at Onset	13.6±2.3	13.3±2.3	14.1±2.2

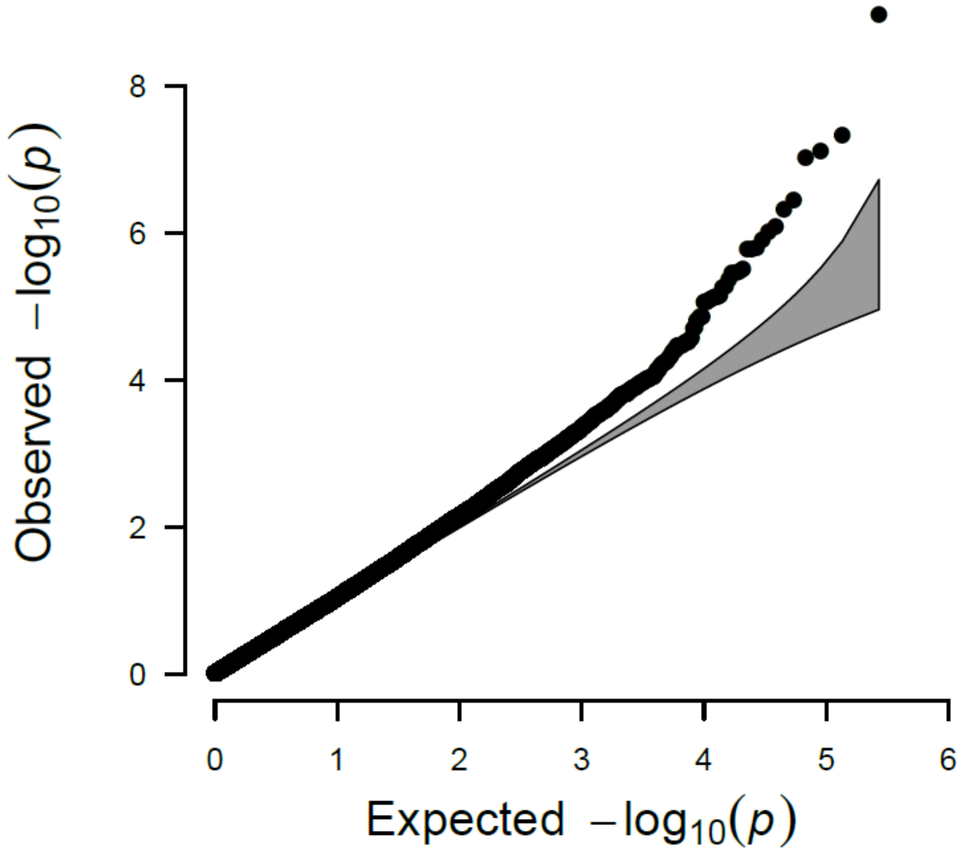
Analysis design

- Whole exome sequencing of 3,650 youth-onset T2D cases
- Match to controls from AMP-T2D exomes
  - Total analysis of 3,005 cases and 9,777 controls
- Single variant analysis (mostly for common variants)
- Gene-level analysis (for rare variants)

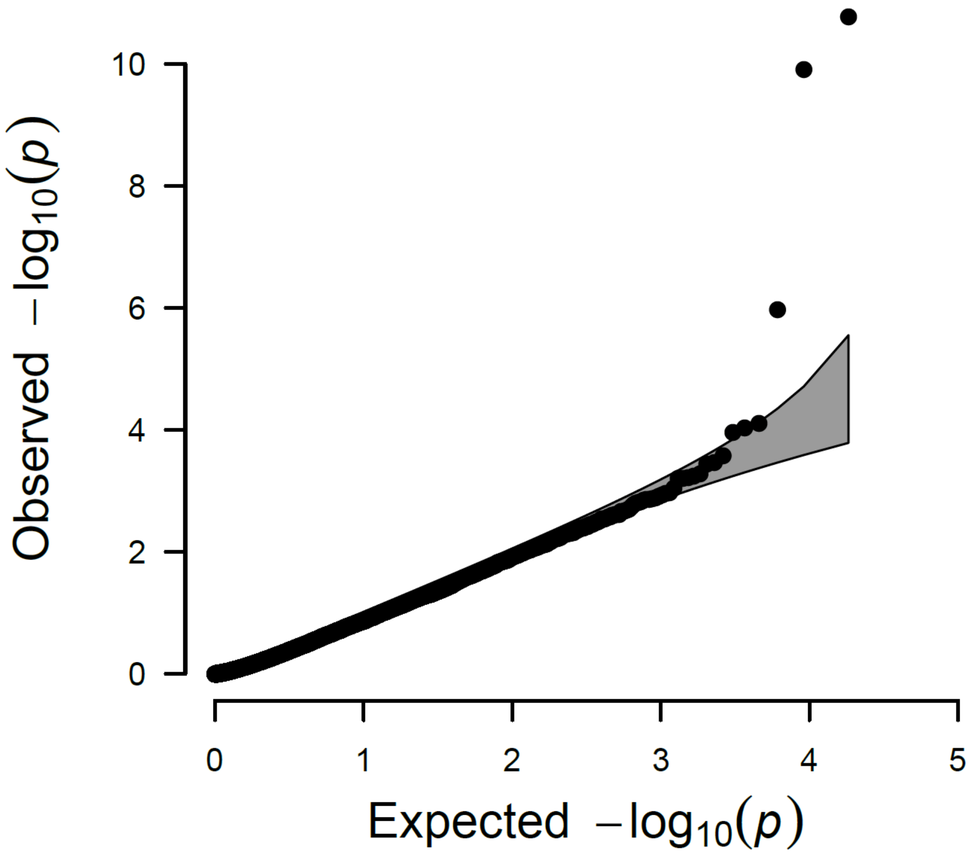


Statistics are well-calibrated

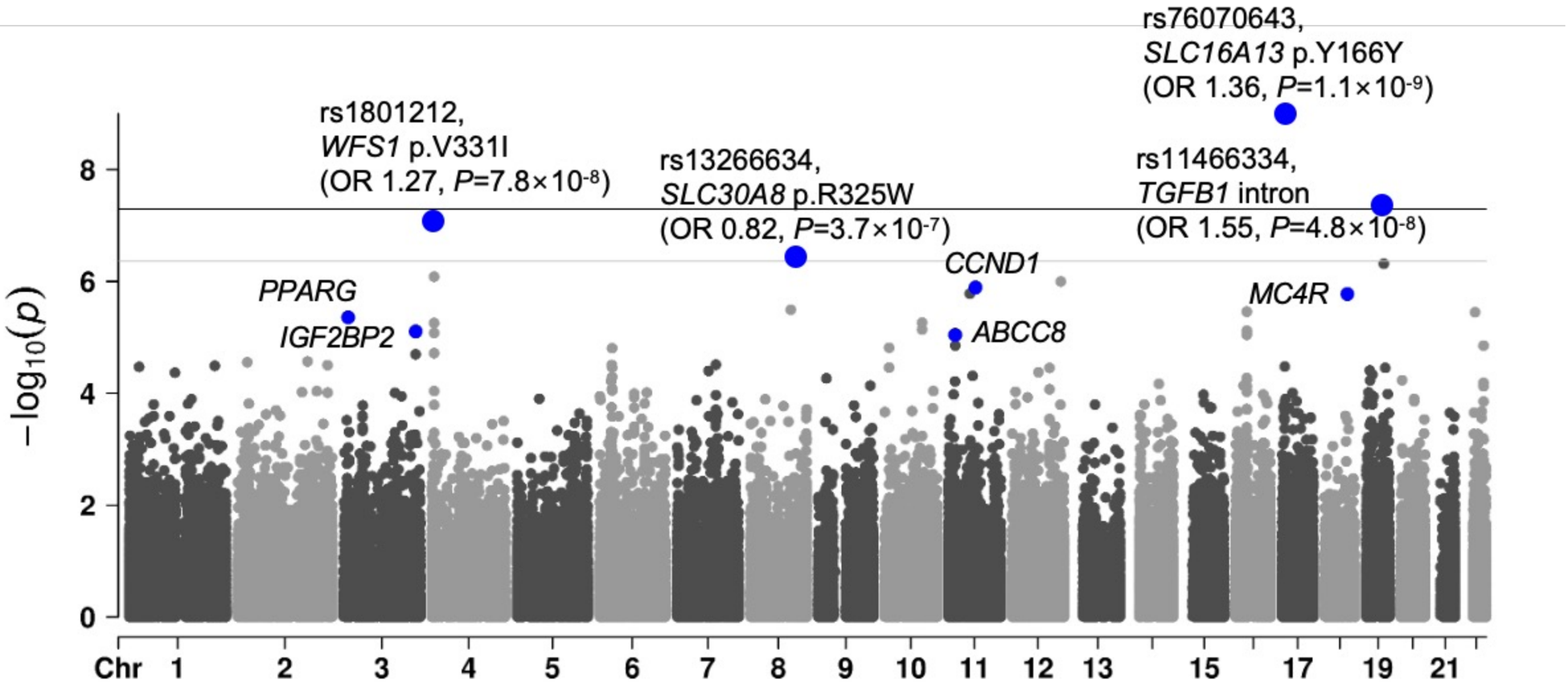
Single variant results



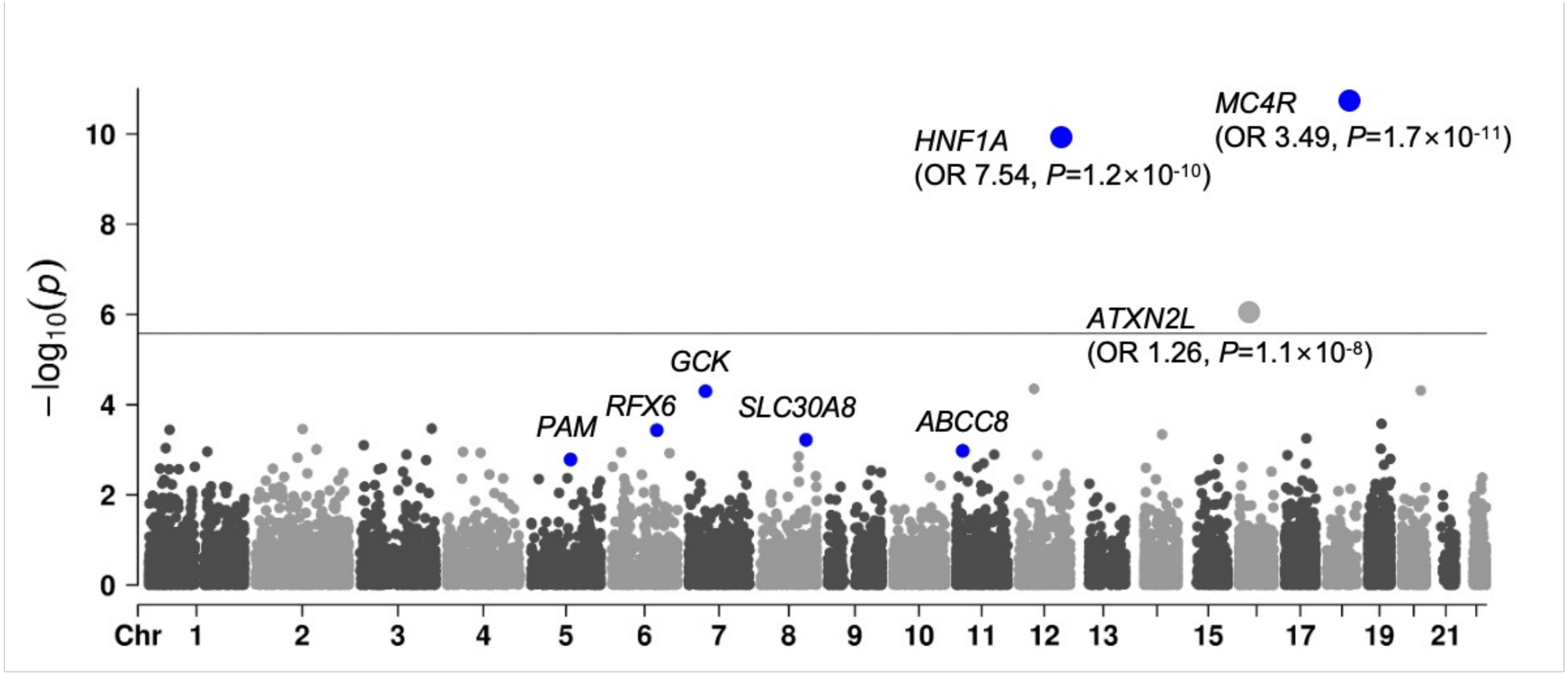
Gene-level results



Four exome-wide significant associations



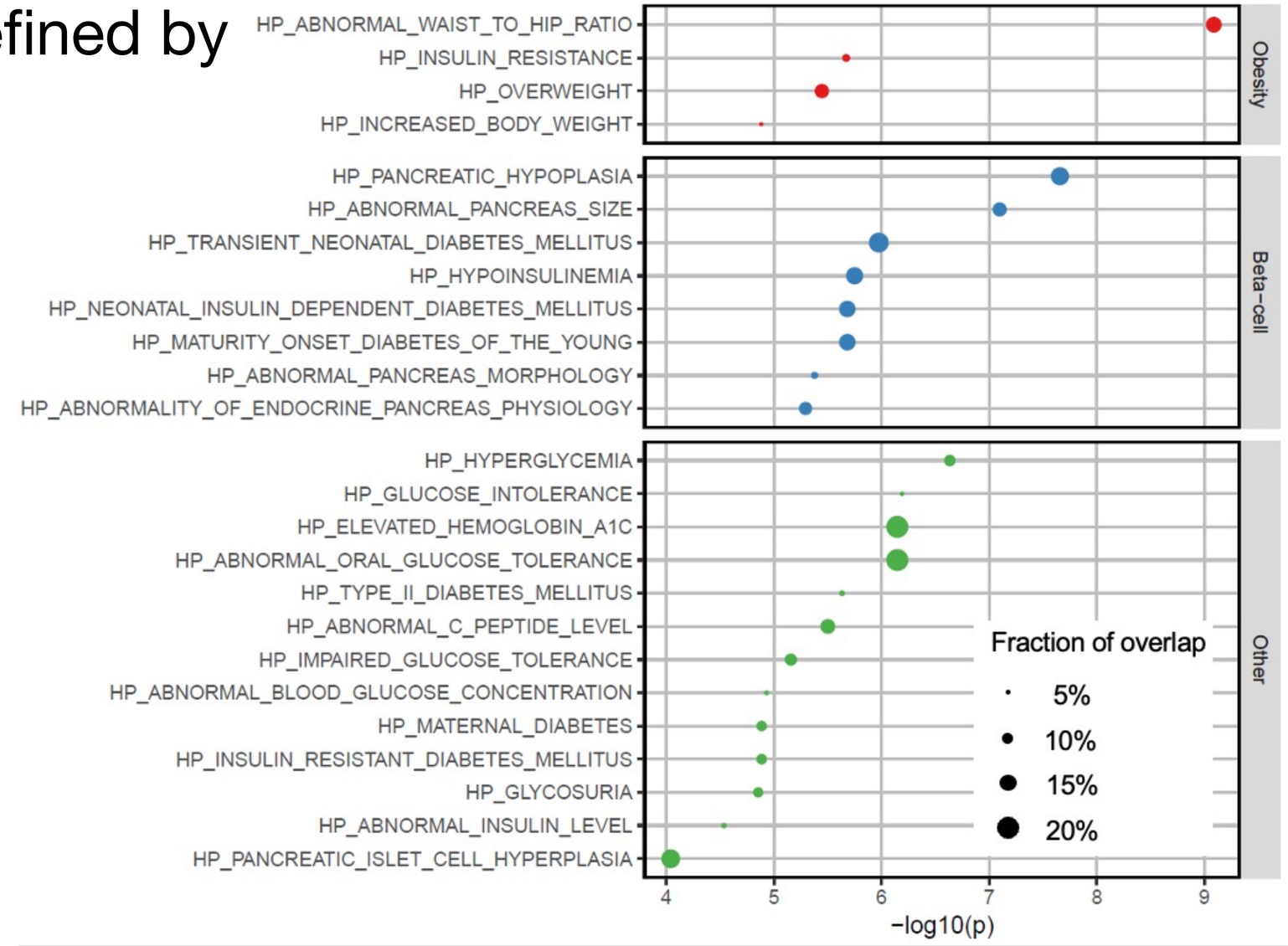
Three exome-wide significant gene-level associations



- **Additionally:** 2.1% of cases carry a monogenic diabetes causing variant

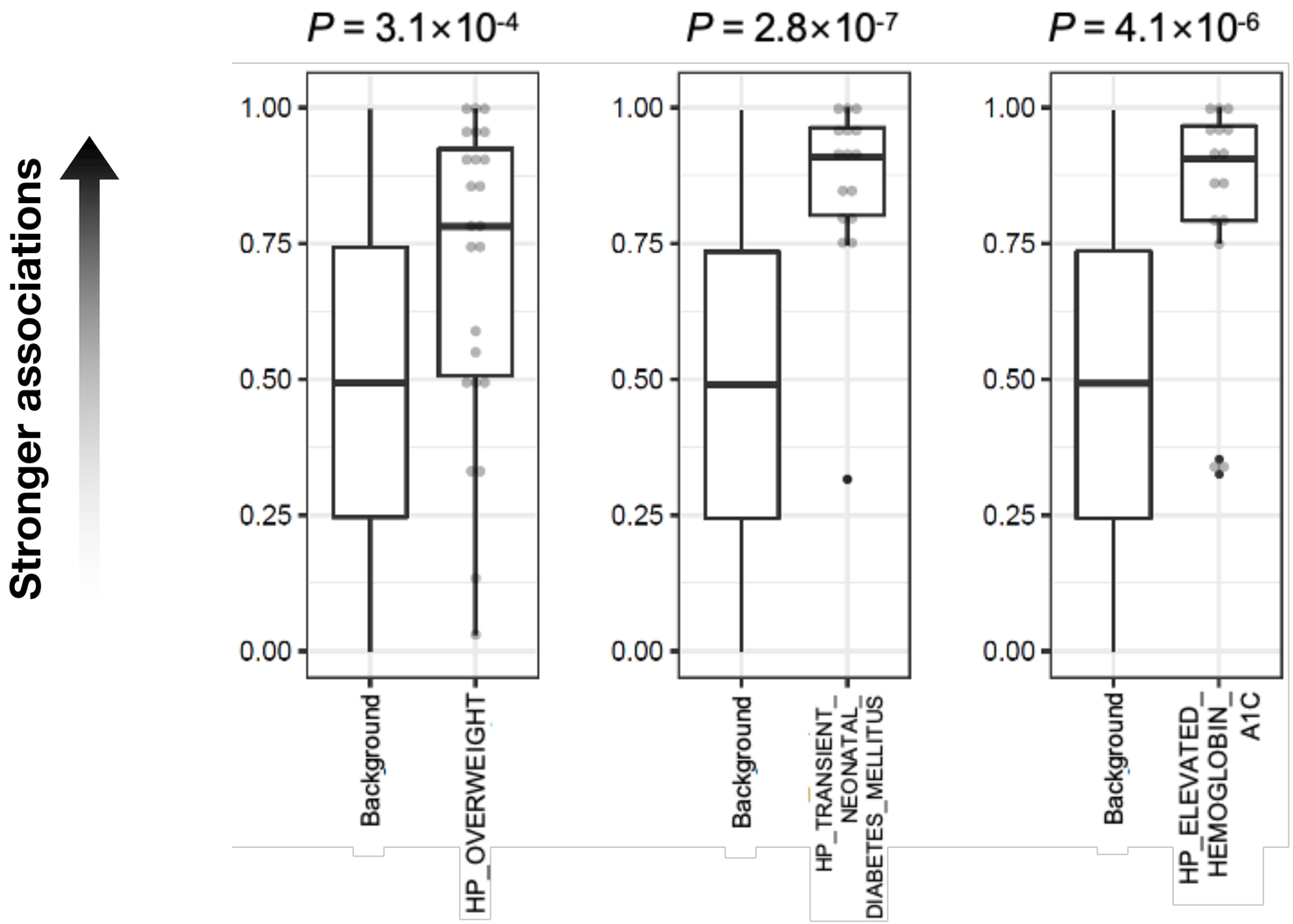
Substantial enrichment in diabetes-relevant gene sets

- Gene sets defined by HPO terms



Three categories

Enrichments are due to many genes

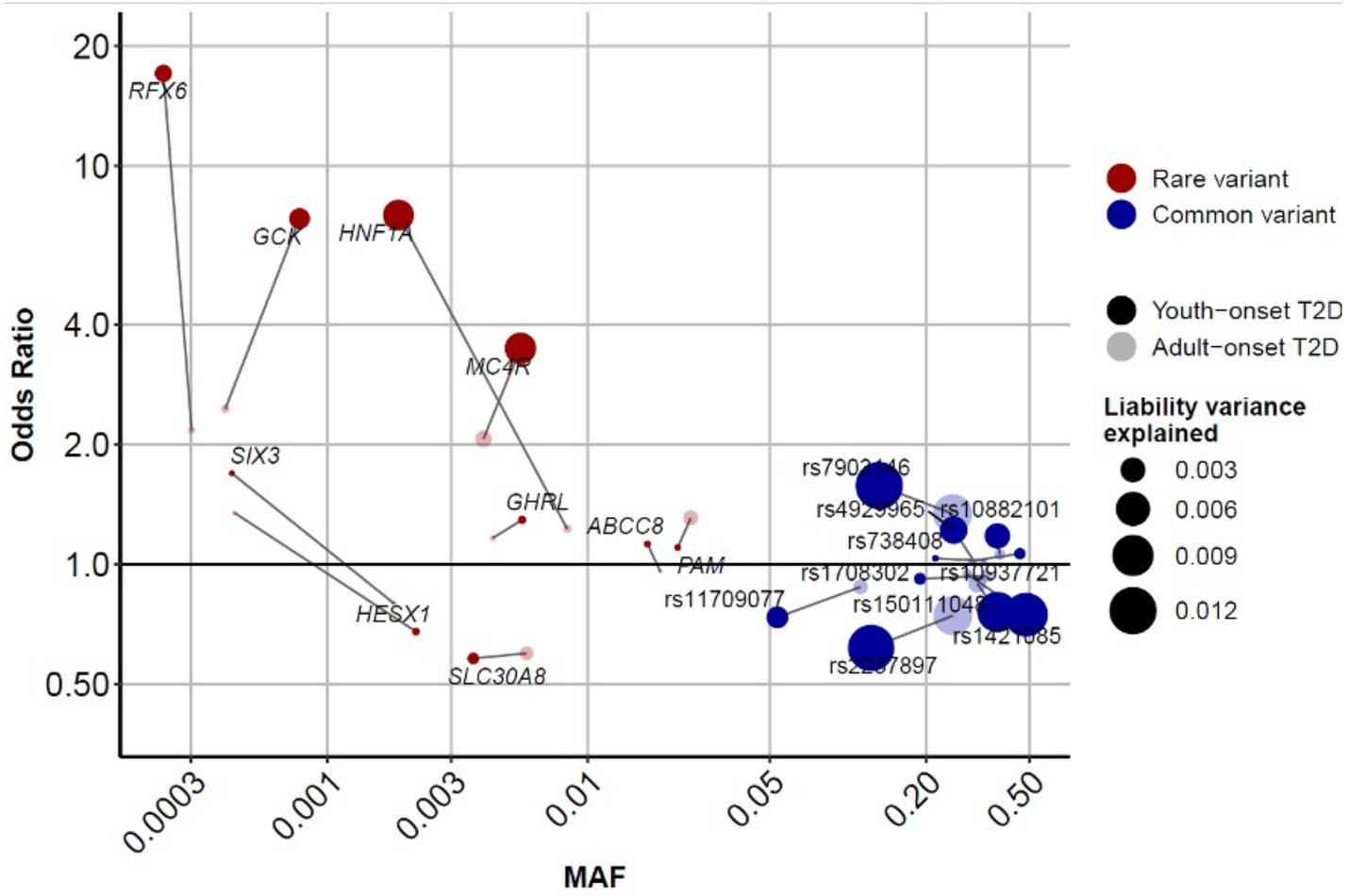


Tiers of candidate genes

- **Tier 1:** Exome-wide significant genes (*MC4R*, *HNF1A*, *ATXN2L*)
- **Tier 2:** among top 50 and causal for monogenic diabetes or T2D (*GCK*, *SLC30A8*, *ABCC8*, *PAM*)
- **Tier 3:** among the top 50 and in an enriched HPO gene set (*RFX6*, *GHRL*, *HESX1*, *SIX3*)
- **Tier 4:** $p<0.05$ and in a diabetes-relevant gene set (38 additional genes)

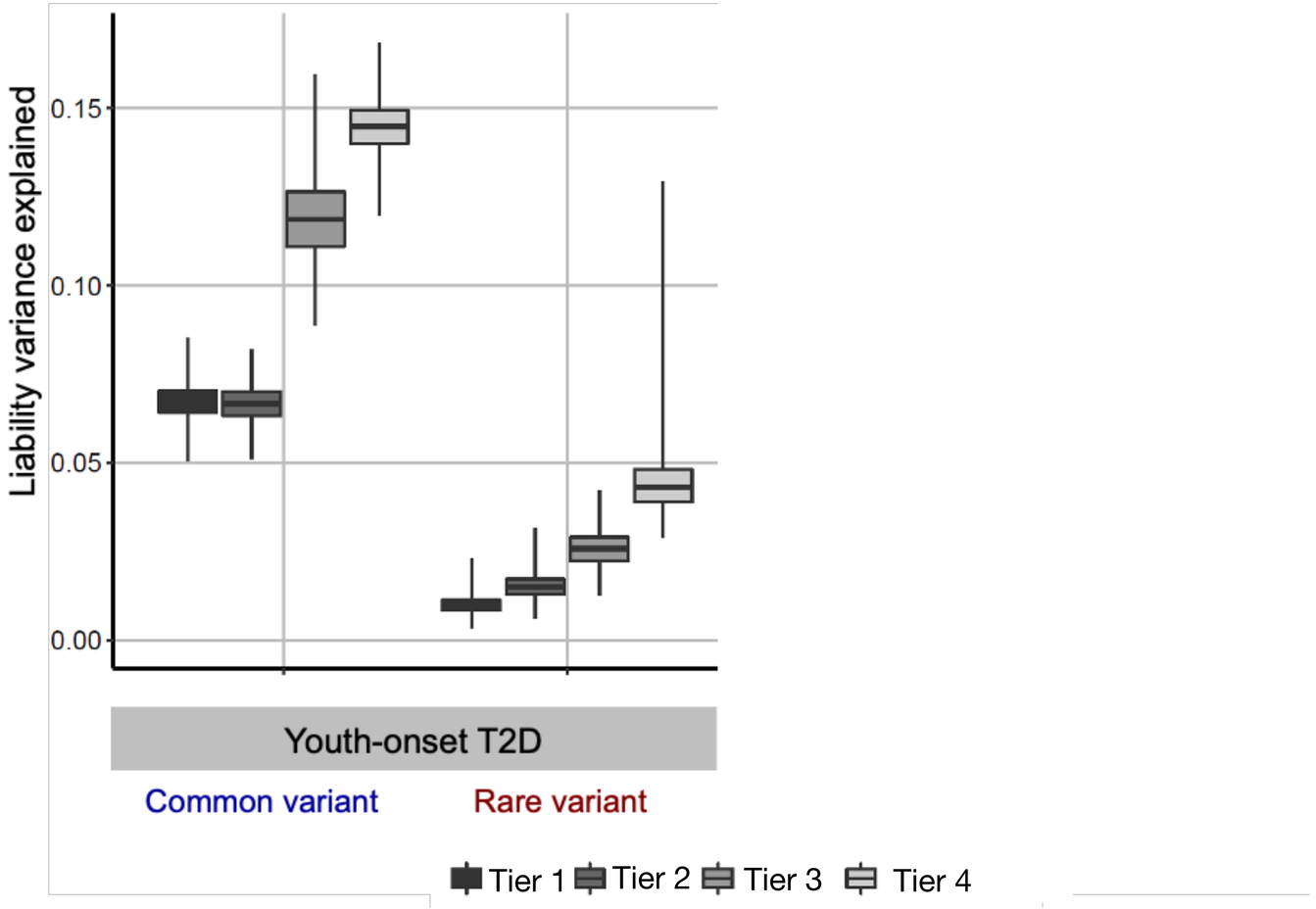
Both common and rare variants are enriched in ProDiGY

(relative to adult-onset T2D cases)



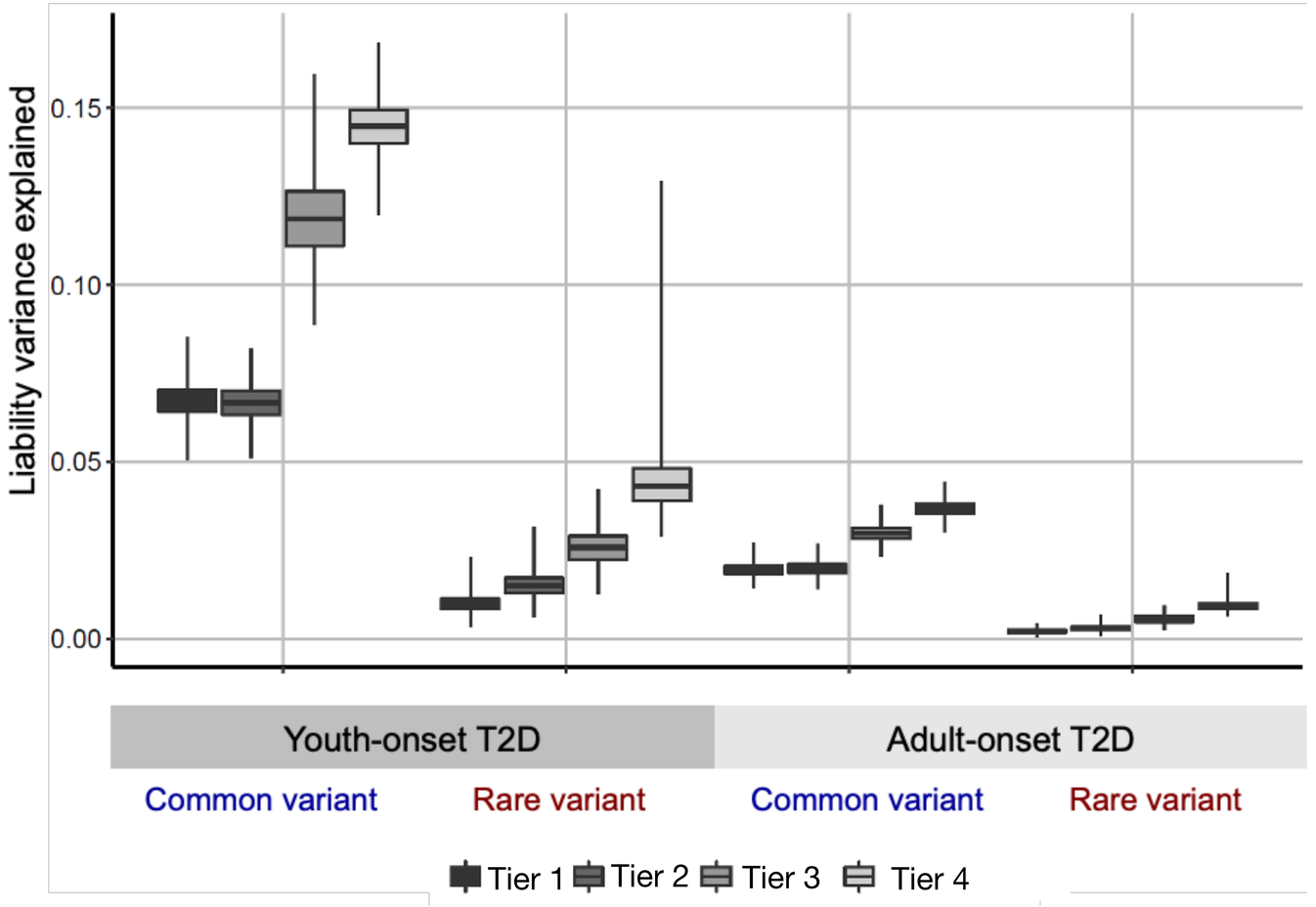
Both common and rare variants explain more heritability

(relative to adult-onset T2D cases)



Both common and rare variants explain more heritability

(relative to adult-onset T2D cases)

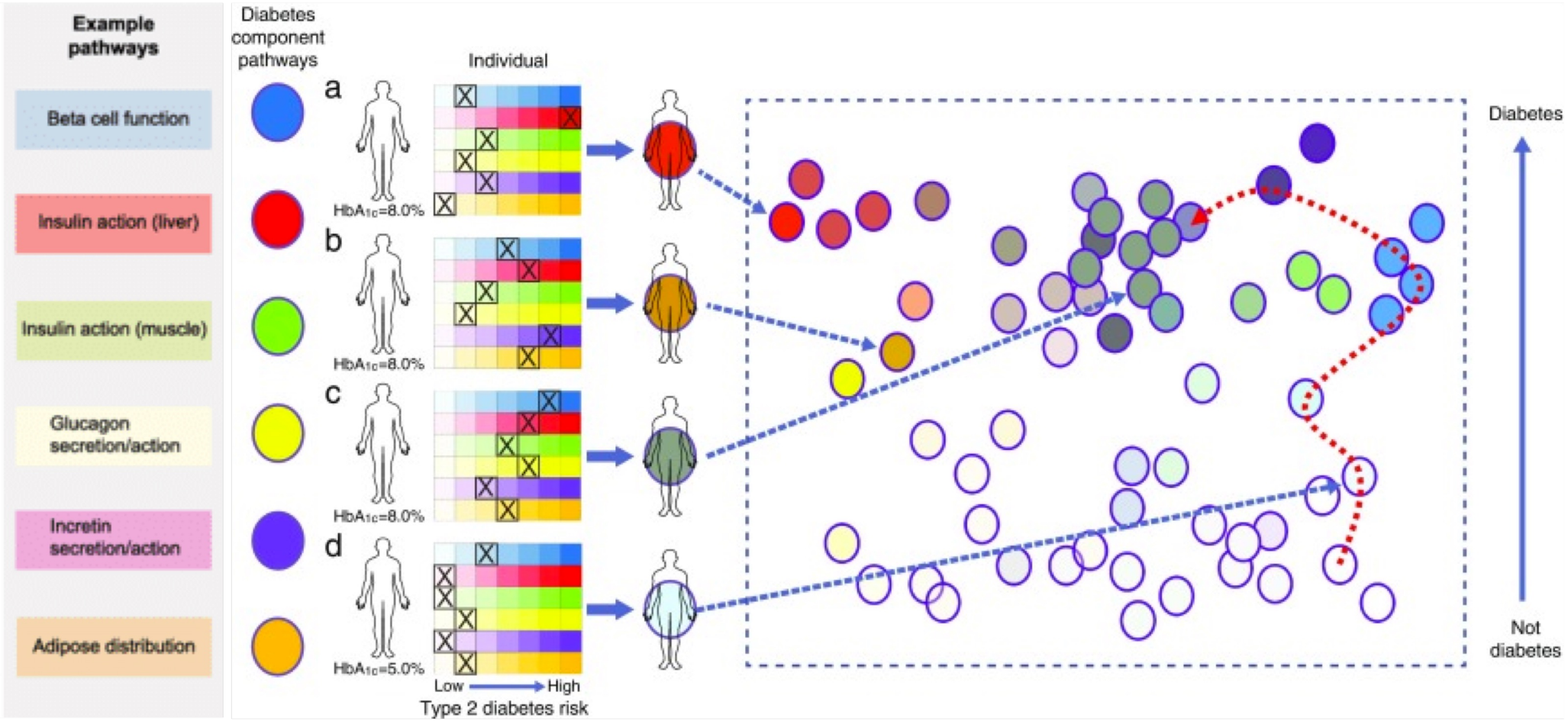


As a **population**, youth-onset T2D cases are enriched for all types of genetic risk factors

2.1% carry monogenic variants (MODY cases)
5.0-fold more rare variants than adult-onset cases
3.4-fold more common variants than adult-onset cases

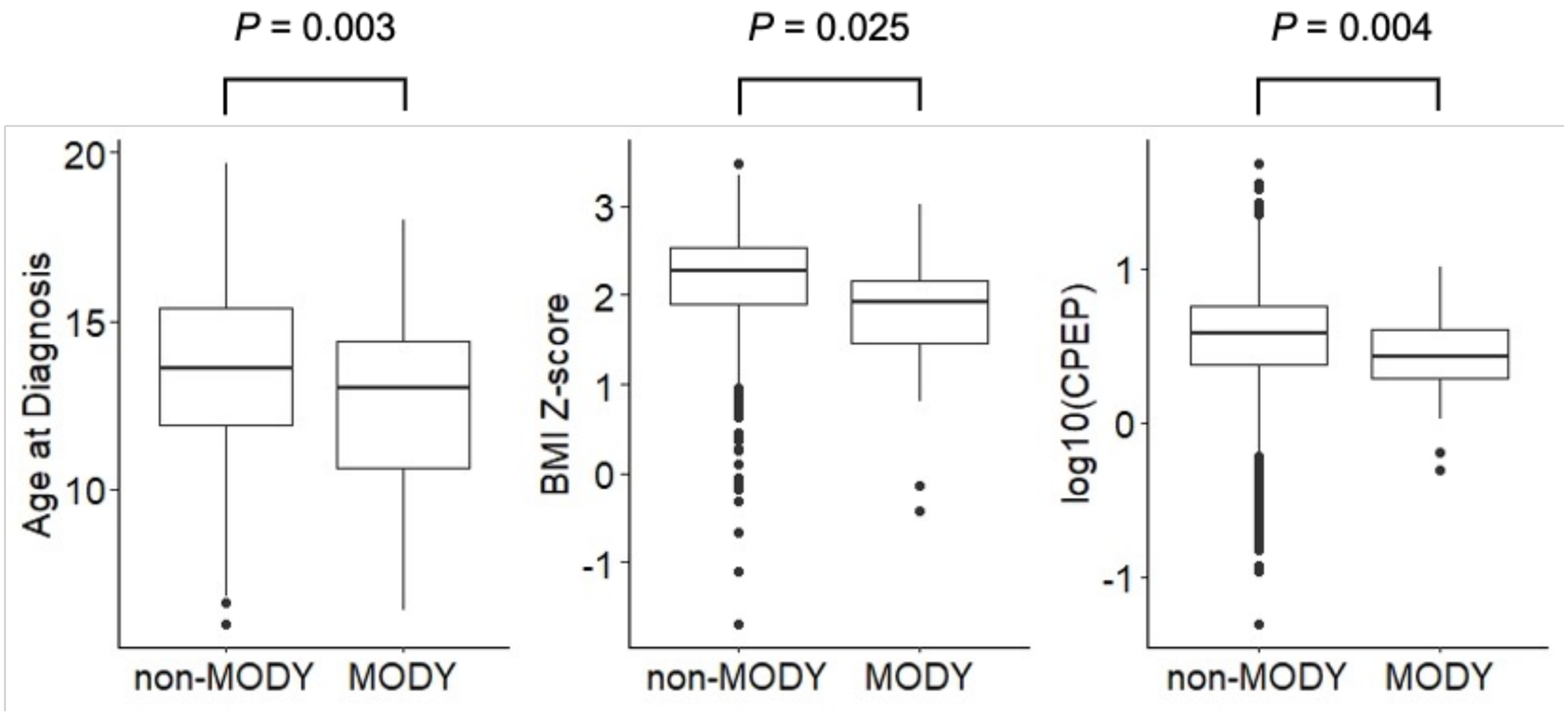
skew towards **common variants in absolute** terms
skew towards **rare variants relative** to adult-onset T2D

What about individually?

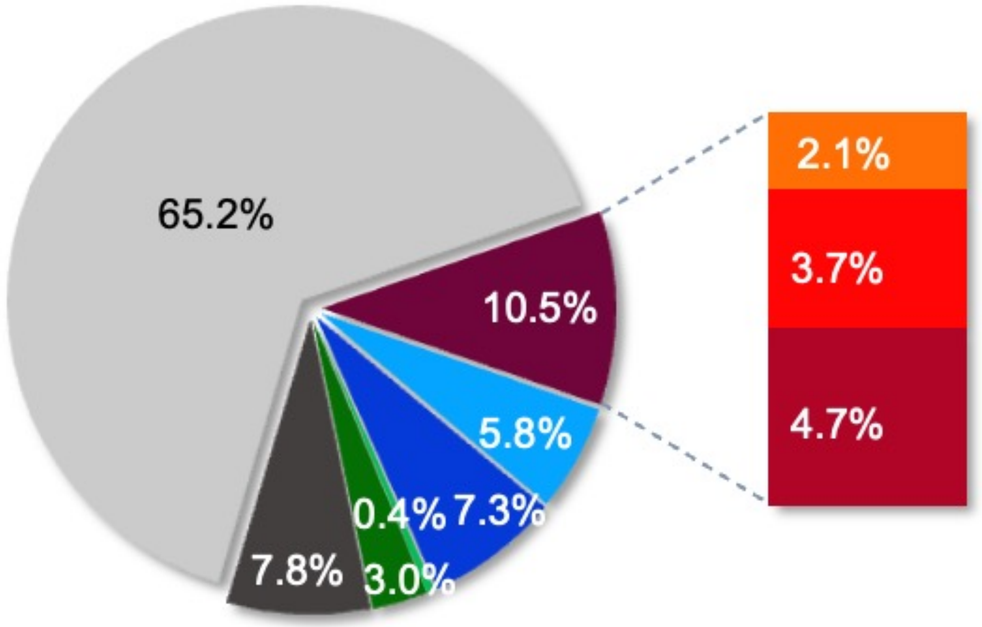


McCarthy, Diabetologia, 2017

Cases due to MODY mutations are phenotypically different



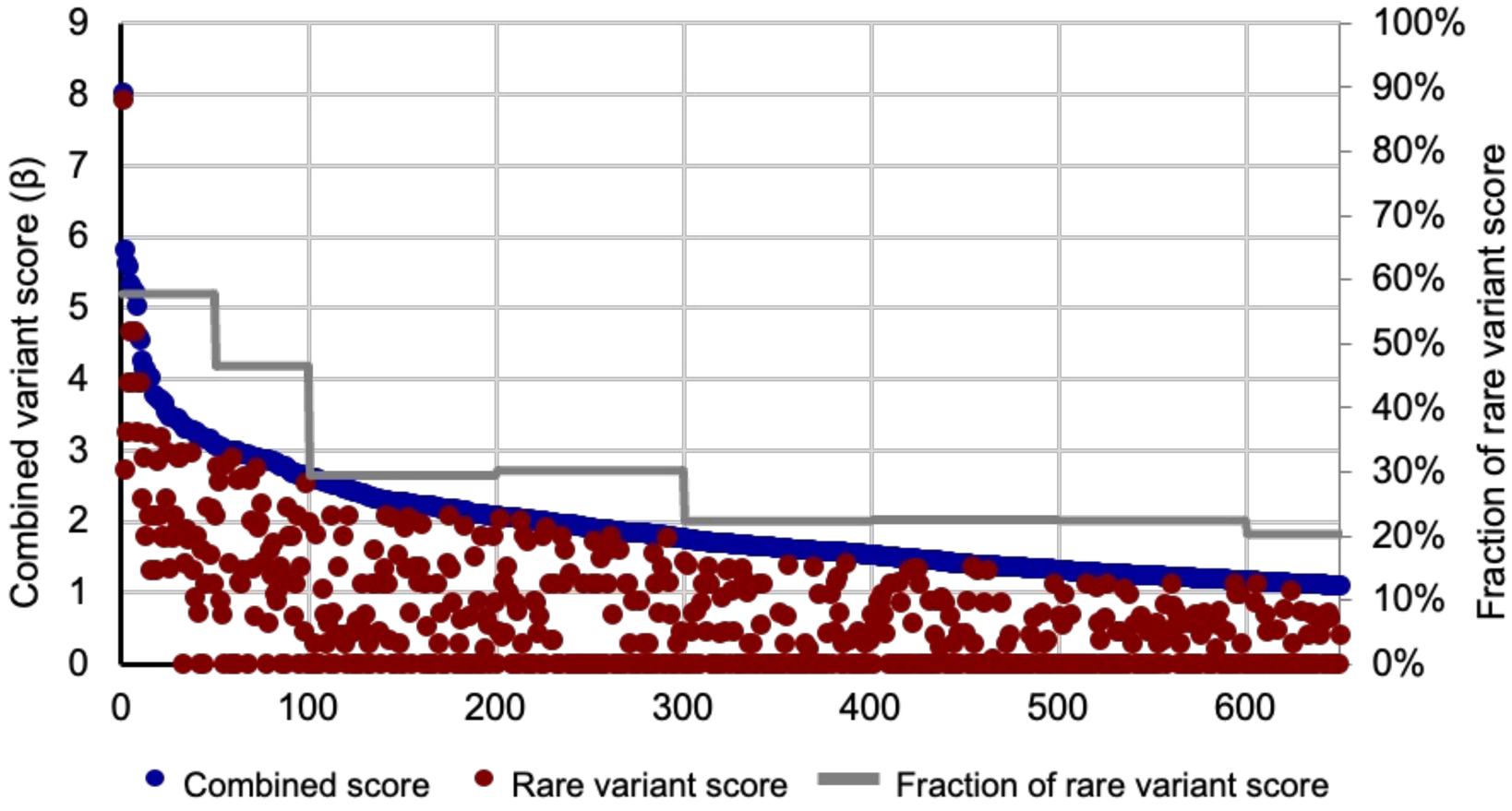
Start with cases “explained” by rare or common variants



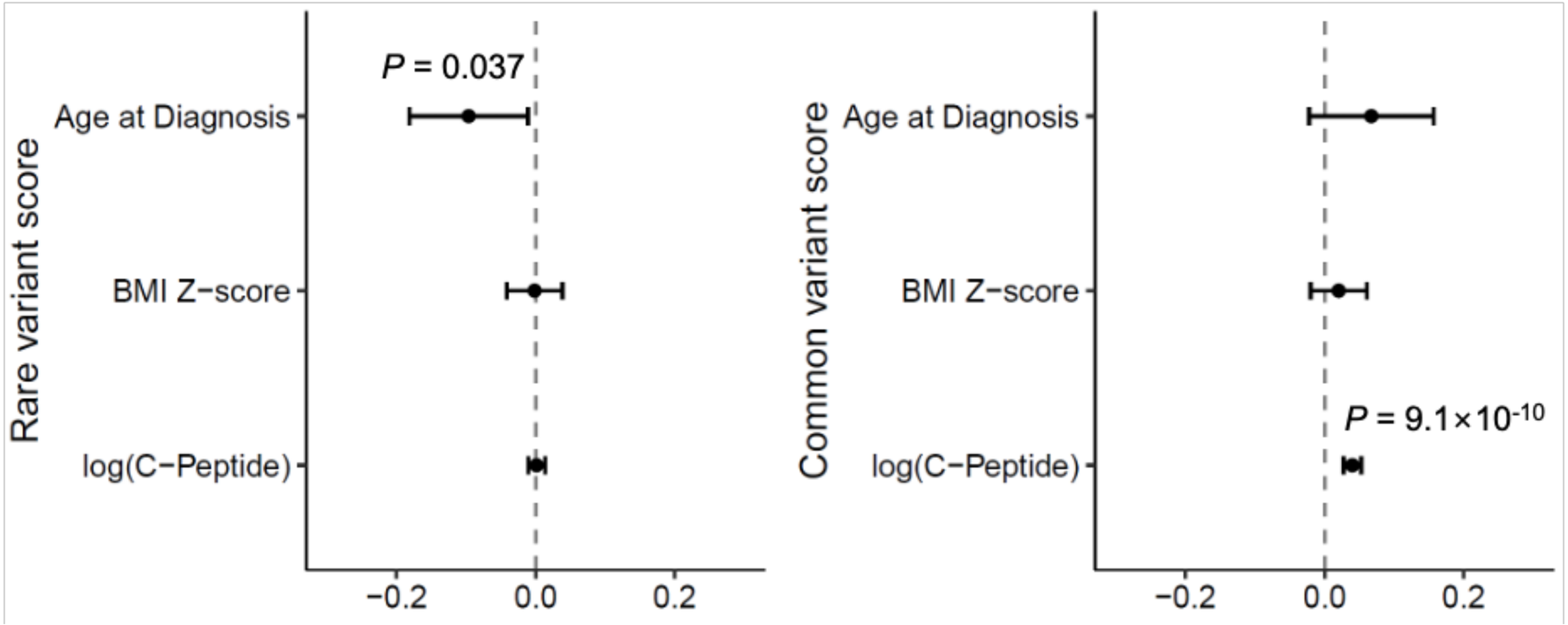
- MODY
- Rare variant score OR ≥ 5
- Rare variant score OR 3-5
- Common variant score OR ≥ 5
- Common variant score OR 3-5
- Combined score OR ≥ 5
- Combined score OR ≥ 3-5

No clear dividing line between cases due to rare vs. common variants

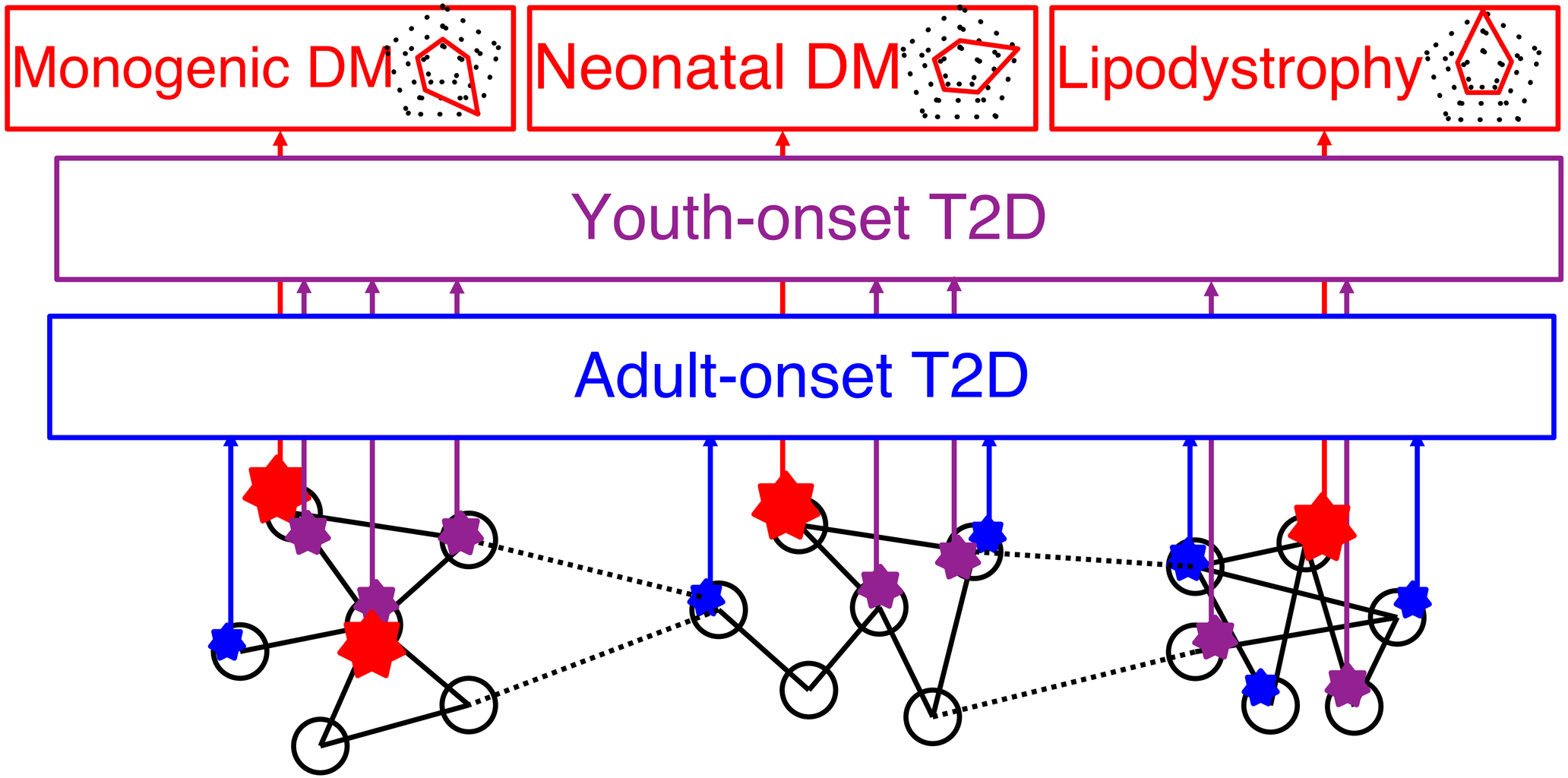
- But, a substantial amount of heterogeneity across cases



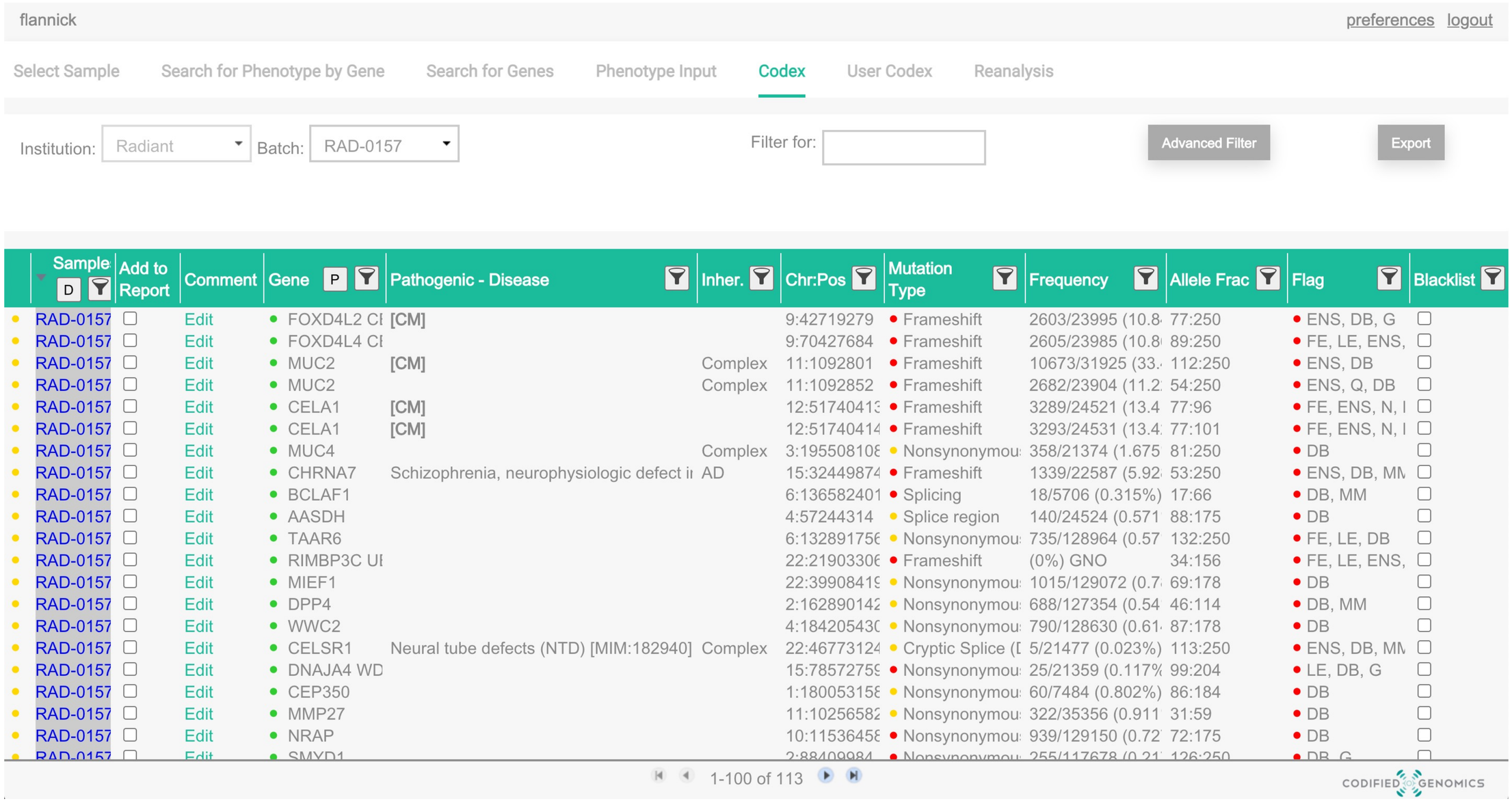
Cases due to rare vs. common variants are phenotypically different



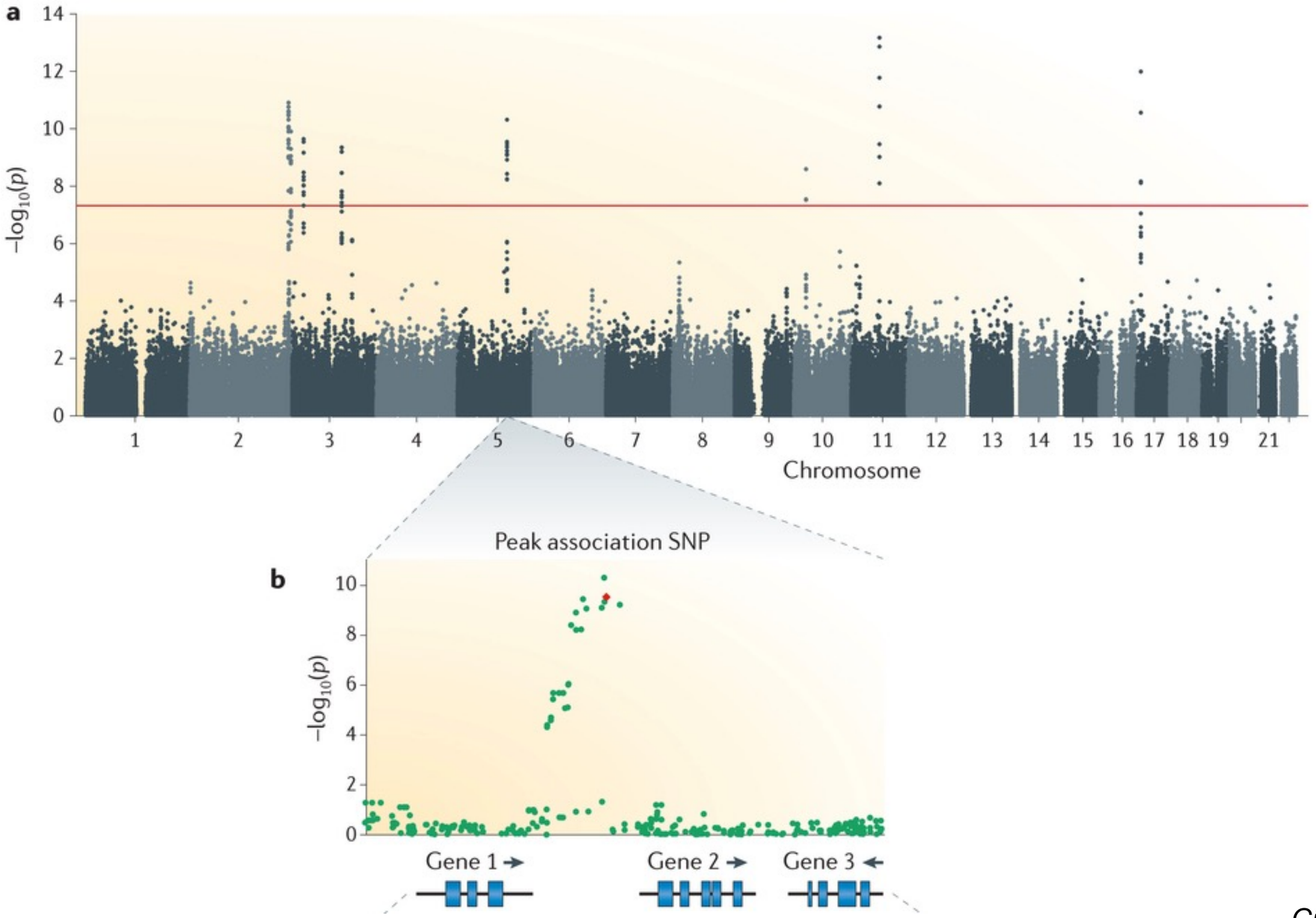
Summary: bridging the gap between monogenic and polygenic diabetes



But...

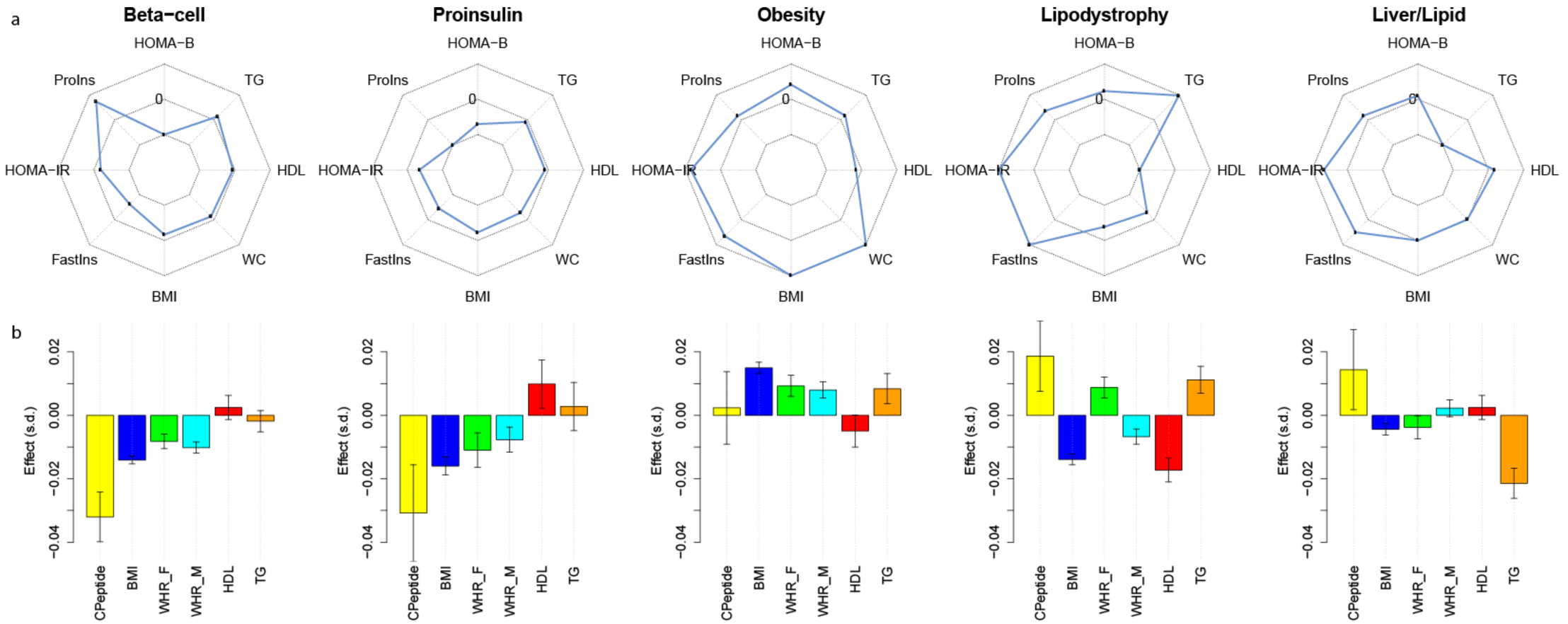

flannick
preferences
logout
Select Sample
Search for Phenotype by Gene
Search for Genes
Phenotype Input
Codex
User Codex
Reanalysis
Institution:
Radiant
Batch:
RAD-0157
Filter for:
Advanced Filter
Export
Sample
D
Add to Report
Comment
Gene
P
Pathogenic - Disease
Inher.
Chr:Pos
Mutation Type
Frequency
Allele Frac
Flag
Blacklist
RAD-0157 Edit FOXD4L2 CI [CM] 9:42719279 Frameshift 2603/23995 (10.8 77:250 ENS, DB, G
RAD-0157 Edit FOXD4L4 CI 9:70427684 Frameshift 2605/23985 (10.8 89:250 FE, LE, ENS,
RAD-0157 Edit MUC2 [CM] Complex 11:1092801 Frameshift 10673/31925 (33. 112:250 ENS, DB
RAD-0157 Edit MUC2 Complex 11:1092852 Frameshift 2682/23904 (11.2 54:250 ENS, Q, DB
RAD-0157 Edit CELA1 [CM] 12:51740413 Frameshift 3289/24521 (13.4 77:96 FE, ENS, N, I
RAD-0157 Edit CELA1 [CM] 12:51740414 Frameshift 3293/24531 (13.4 77:101 FE, ENS, N, I
RAD-0157 Edit MUC4 Complex 3:195508108 Nonsynonymou 358/21374 (1.675 81:250 DB
RAD-0157 Edit CHRNA7 Schizophrenia, neurophysiologic defect i AD 15:32449874 Frameshift 1339/22587 (5.92 53:250 ENS, DB, MM
RAD-0157 Edit BCLAF1 6:136582401 Splicing 18/5706 (0.315%) 17:66 DB, MM
RAD-0157 Edit AASDH 4:57244314 Splice region 140/24524 (0.571 88:175 DB
RAD-0157 Edit TAAR6 6:132891756 Nonsynonymou 735/128964 (0.57 132:250 FE, LE, DB
RAD-0157 Edit RIMBP3C UI 22:21903306 Frameshift (0%) GNO 34:156 FE, LE, ENS,
RAD-0157 Edit MIEF1 22:39908419 Nonsynonymou 1015/129072 (0.7 69:178 DB
RAD-0157 Edit DPP4 2:162890142 Nonsynonymou 688/127354 (0.54 46:114 DB, MM
RAD-0157 Edit WWC2 4:184205430 Nonsynonymou 790/128630 (0.61 87:178 DB
RAD-0157 Edit CELSR1 Neural tube defects (NTD) [MIM:182940] Complex 22:46773124 Cryptic Splice (5/21477 (0.023%) 113:250 ENS, DB, MM
RAD-0157 Edit DNAJA4 WD 15:78572759 Nonsynonymou 25/21359 (0.117% 99:204 LE, DB, G
RAD-0157 Edit CEP350 1:180053158 Nonsynonymou 60/7484 (0.802%) 86:184 DB
RAD-0157 Edit MMP27 11:10256582 Nonsynonymou 322/35356 (0.911 31:59 DB
RAD-0157 Edit NRAP 10:11536458 Nonsynonymou 939/129150 (0.72 72:175 DB
1-100 of 113
CODIFIED GENOMICS

Versus…

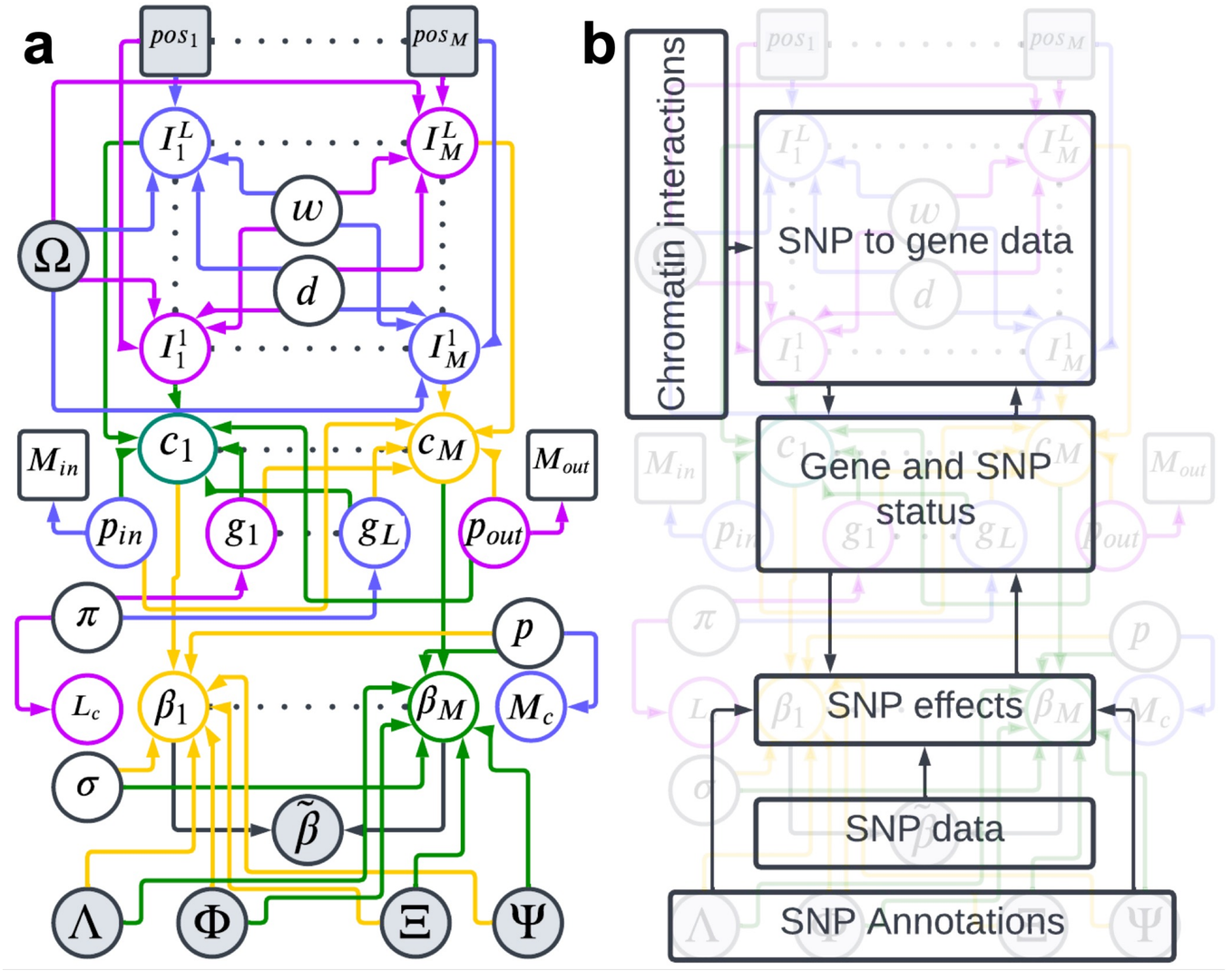


Civelek and Lusis, 2014

A future vision



Udler et al, 2018


a
b
Chromatin interactions
SNP to gene data
Gene and SNP status
SNP effects
SNP data
SNP Annotations

Build search criteria

Gene	Phenotype
Select a gene ...	Select a phenotype ...

Search Criterion SIN3A X Type 2 diabetes (T2D) X

HuGE Score (Combined Evidence) 45(Common variation BF) * 1(Rare variation BF) = 45

*BF=Bayes Factor *HuGE Score(combined evidence) = BF of common variation X BF of rare variation

	45	<--Common Variation BF
X	1	<--Rare Variation BF (Reset Prior Allelic variance)
HuGE Score:	**45**	

HuGE score 45 falls in Very Strong evidence range

Compelling: HuGE Score >= 350 | Extreme: >=100 | Very Strong: >=30 | Strong: >=10 | Moderate: >=3 | Anecdotal: >1 | No Evidence: <=1

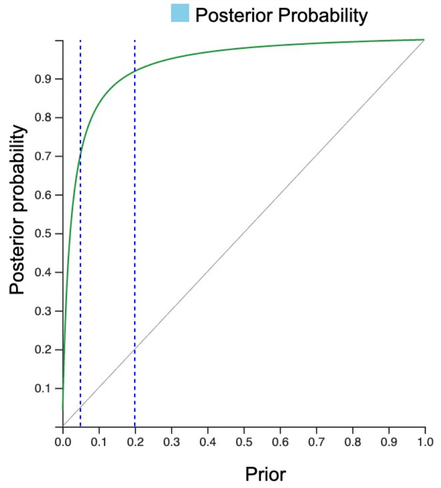
Compelling	Extreme	Very Strong	Strong	Moderate	Anecdotal	No Evidence

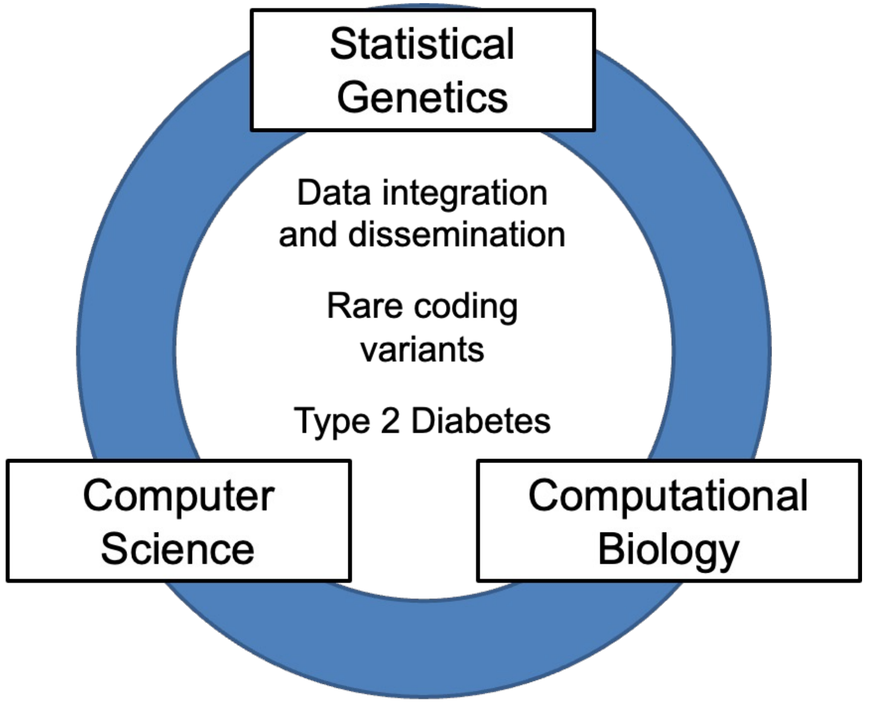
▶ Posterior probability

How is PPA calculated?

The posterior probability of association (PPA) is a measure of how different "prior" beliefs in a gene's disease-relevance are modified by the HuGE score. Priors may be defined using various criteria, such as an estimate of the total number of genes impacting a common disease, or experimental results bearing on the gene's role in a disease.

	Suggested prior	Posterior probability	
	0.2	0.9183673	✕
	0.05	0.703125	✕
Set prior			





FlannickLab

We are always seeking collaborators and motivated new members!

Contact flannick@broadinstitute.org or http://flannicklab.org