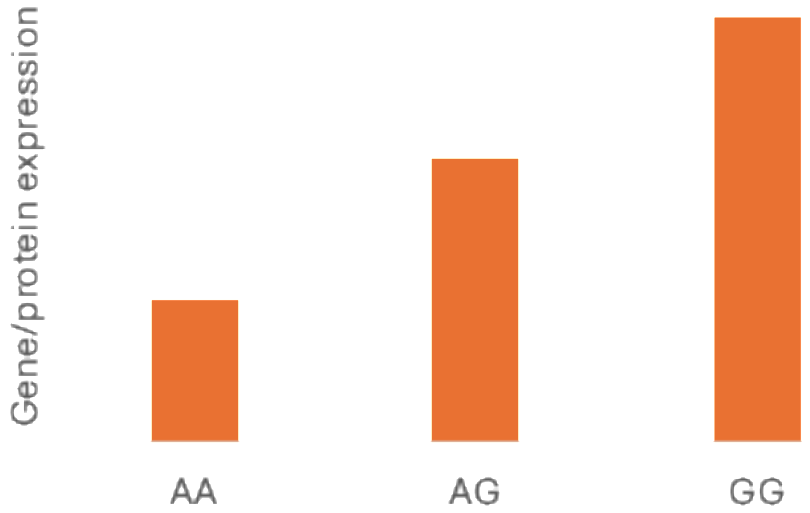
Using QTL Data to Infer Drug Effects

Dr Clara Jiang (IMB, UQ)

Quantitative trait locus

- A QTL is a genetic variant linked to variations in a measurable trait
- eQTL: gene/mRNA expression
- pQTL: protein expression



- Others
  - mQTL: DNA methylation
  - vQTL: variance of a trait
  - LOF mutations

Using QTLs to model drug effects on targets

- If a genetic variant alters the expression of a gene or the abundance of a protein, individuals carrying that variant effectively experience a lifelong "experiment" in altered target activity.
- For example:
  - An eQTL that reduces the expression of gene *X* can model the effect of a drug that inhibits gene *X*.
  - A pQTL that lowers circulating levels of protein *Y* can model the effect of an antibody that blocks protein *Y*.
  - **A common assumption**: The direction of gene/protein expression changes is assumed to model the direction of drug effects of the target (which may not be true)

eQTLs versus pQTLs

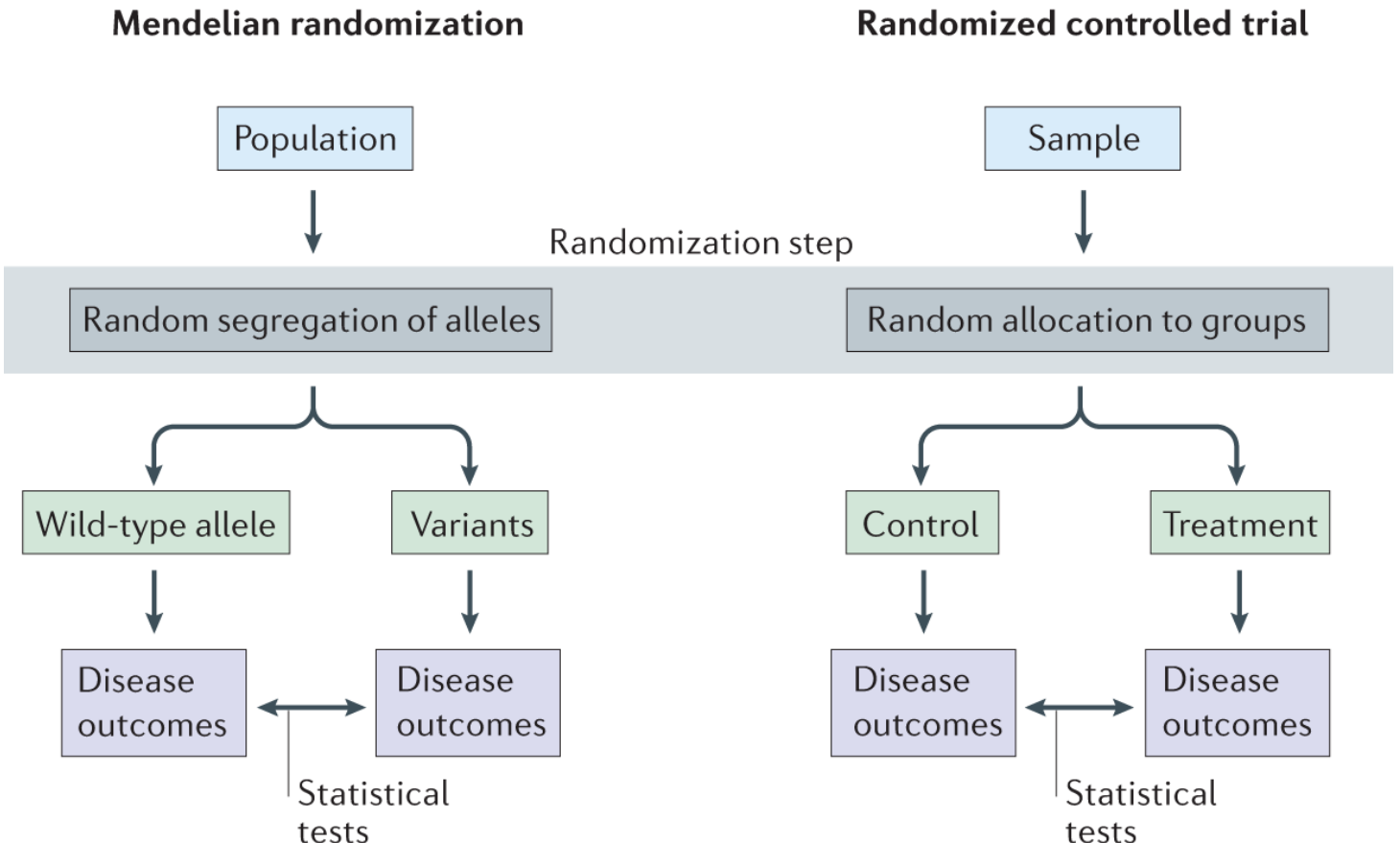
	eQTL	pQTL
Trait	Gene expression	Protein expression
Advantages	• Available in many tissues (e.g. GTEX) • Large sample sizes (e.g. ~31k in eQTLGen) • Useful when drugs act by changing the transcription level of a gene (e.g. mRNA-based therapeutics)	• Useful when drugs act by activating or inhibiting a drug
Limitations	• May not be useful when the drug acts on a protein (mRNAs levels may not correlate with protein abundance)	• Limited data availability (compared to eQTL) • Smaller sample size
Data resources	• GTEX • eQTLGen (whole blood) • …	• UKB-PPP (N ~54k) (plasma) • ONTIME (brain, CSF, plasma) • …

Correlating with disease/health outcomes

- QTLs can be used to test whether the genetically proxied modulation of a gene/protein is associated with changes in disease risk or adverse outcomes.
- Methods:
  - Summary statistics
    - Summary-data-based MR
    - Coloc
  - Individual-level data
    - Polygenic score

Mendelian randomisation

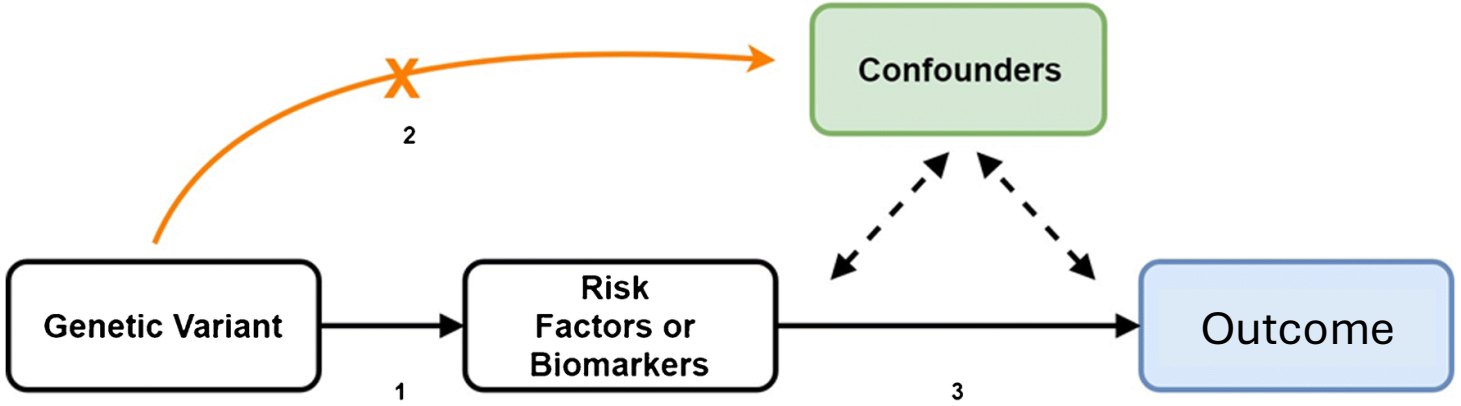
- Genetic variants are used as proxies for a modifiable exposure and determine whether the exposure causes a disease/health outcome.



Sanderson et al. Nature Reviews Methods Primers, 2016

Assumptions of MR

- In order to obtain unbiased estimates, three key assumptions of MR need to be met.



Assumptions of Mendelian Randomization Study:

Genetic variants are associated with the risk factor
Genetic variants are not associated with confounders
Genetic variants influence bone outcomes only through the risk factor

Trajanoska and Rivadeneira. Current Osteoporosis Reports. 2018

Summary-data-based MR (SMR)

Analysis | Published: 28 March 2016

Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets

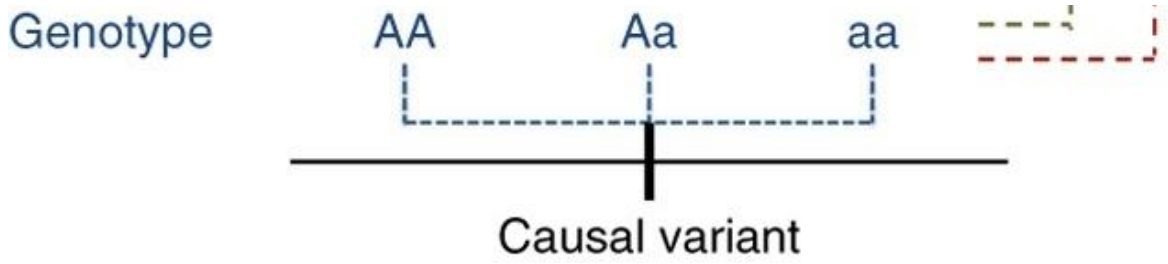
Zhihong Zhu, Futao Zhang, Han Hu, Andrew Bakshi, Matthew R Robinson, Joseph E Powell, Grant W Montgomery, Michael E Goddard, Naomi R Wray, Peter M Visscher & Jian Yang

Nature Genetics **48**, 481–487 (2016) | Cite this article

Download: https://yanglab.westlake.edu.cn/software/smr/
SMR portal: https://yanglab.westlake.edu.cn/smr-portal/

SMR

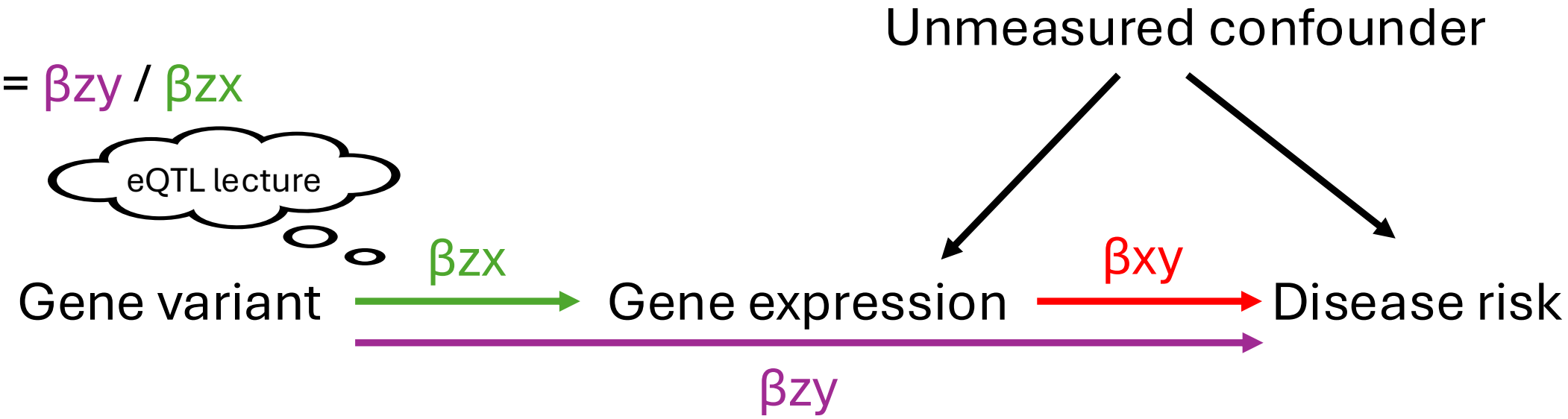
- To test if the effect of a SNP on a phenotype is mediated through gene expression.
- Assumptions:
  - Single causal variant
  - Linear relationship between the exposure and the outcome
- Can also be applied on other QTL data



Zhu et al. Nat Genet. 2016

Effect estimate by SMR

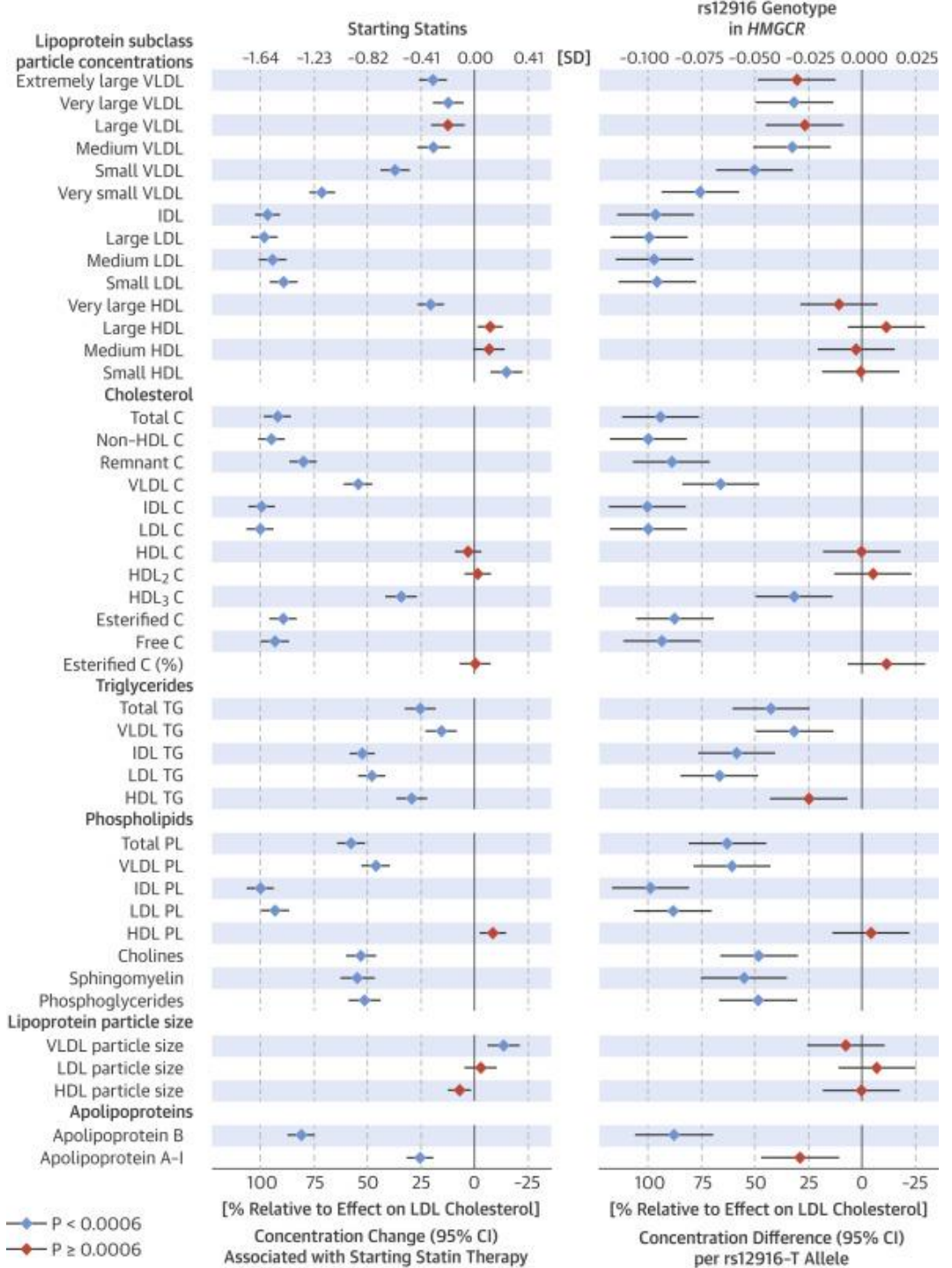
- βxy = βzy / βzx



- Interpretation: βxy represents in the change in disease risk per 1 unit (e.g. standard deviation) increase in genetically predicted gene expression

SMR calculations - example

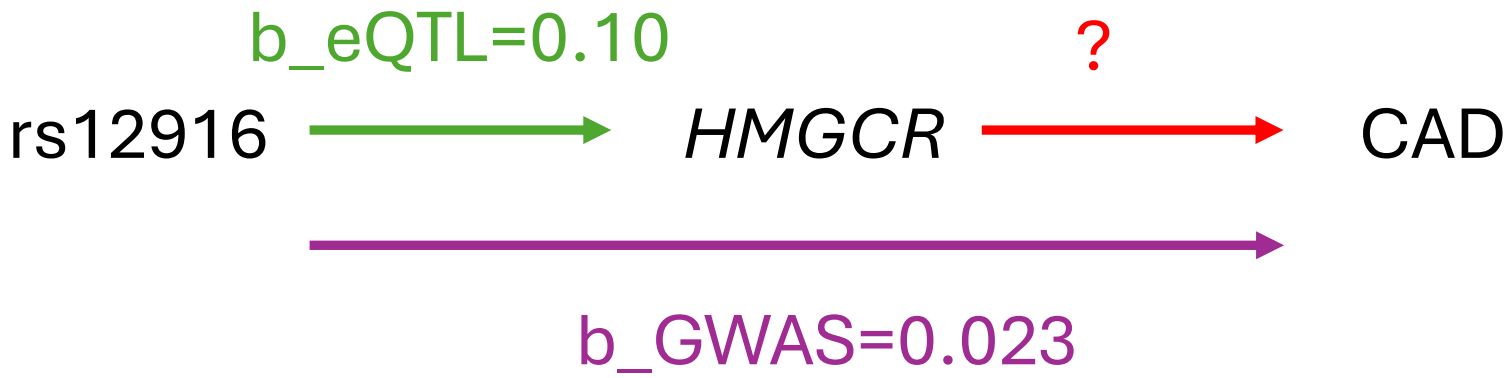
- Statins are widely used to treat coronary artery disease (CAD)
- Statins are HMGCR inhibitors
- rs12916 is commonly used to proxy for HMGCR inhibition and statin use



Wurtz et al. JACC. 2016

SMR calculations - example

βxy = βzy / βzx



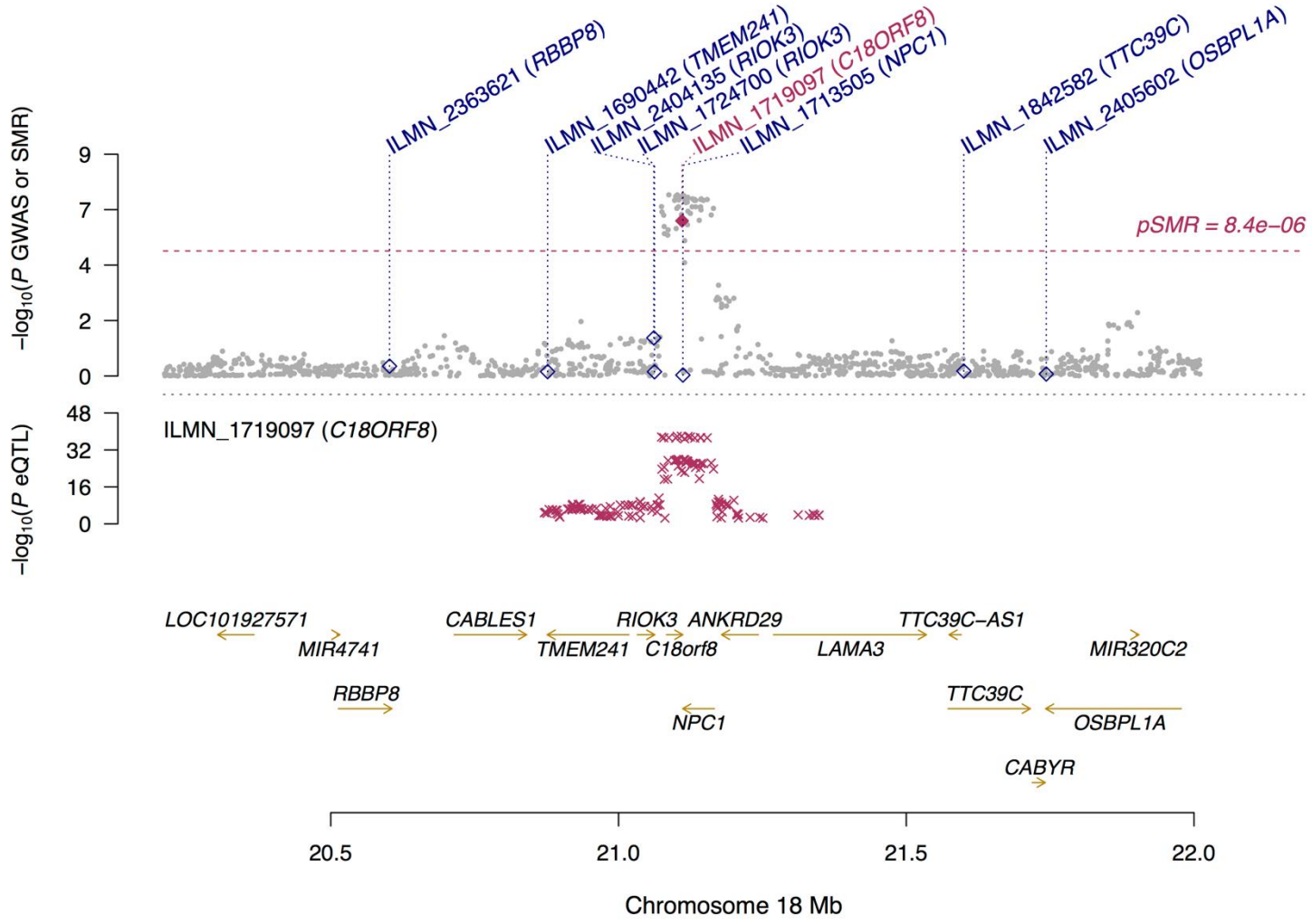
b_SMR = 0.023/0.10 = 0.23

$$\mathrm{var}(\hat{b}_{xy}) \approx \frac{b_{zy}^2}{\beta_{zx}^2}\left[\frac{\mathrm{var}(\hat{\beta}_{zx})}{\beta_{zx}^2}+\frac{\mathrm{var}(\hat{b}_{zy})}{b_{zy}^2}-\frac{2\,\mathrm{cov}(\hat{\beta}_{zx},\hat{b}_{zy})}{\beta_{zx}b_{zy}}\right] \quad (4)$$

Zhu et al. Nat Genet. 2016

SMR sensitivity analyses

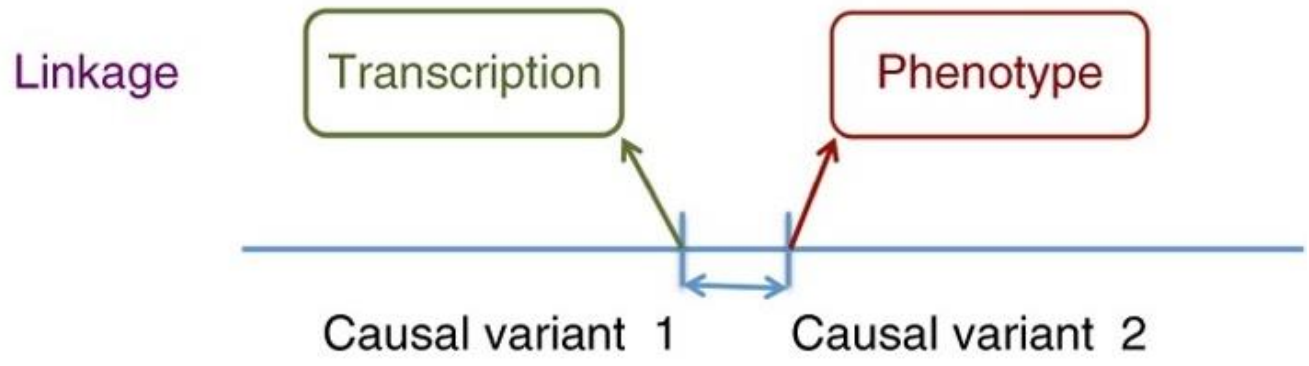
- Locus plot



GCTA

Heterogeneity in dependent instruments (HEIDI)

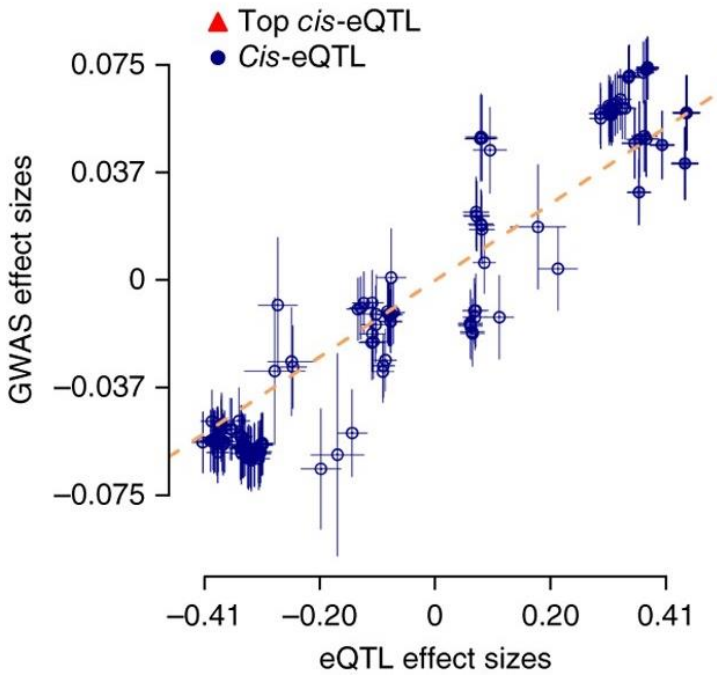
- The association observed in a SMR test could also be due to LD between two distinct causal variants, one affecting gene expression and the other affecting trait variation.
- If the SMR association is due to LD, we would expect heterogeneity in the bxy estimates calculated from other SNPs that are in LD with the SNP used in the SMR test.



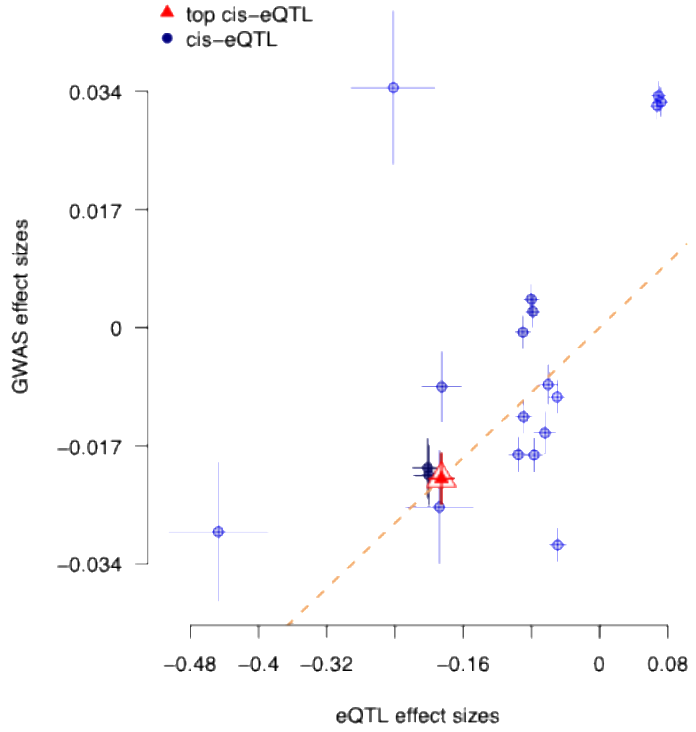
Zhu et al. Nat Genet. 2016

Heterogeneity in dependent instruments (HEIDI)

- The smaller the P_{HEIDI} value, the larger the probability of the observations being consistent with a model of linkage



p_HEIDI = 0.22



p_HEIDI = 2.28E-11

Heterogeneity in dependent instruments (HEIDI)

- The HEIDI test may be too conservative when
  - There are multiple associations signals, potentially through different mechanisms (e.g. gene vs protein levels), at a single locus
  - complex LD structure

Other sensitivity analyses

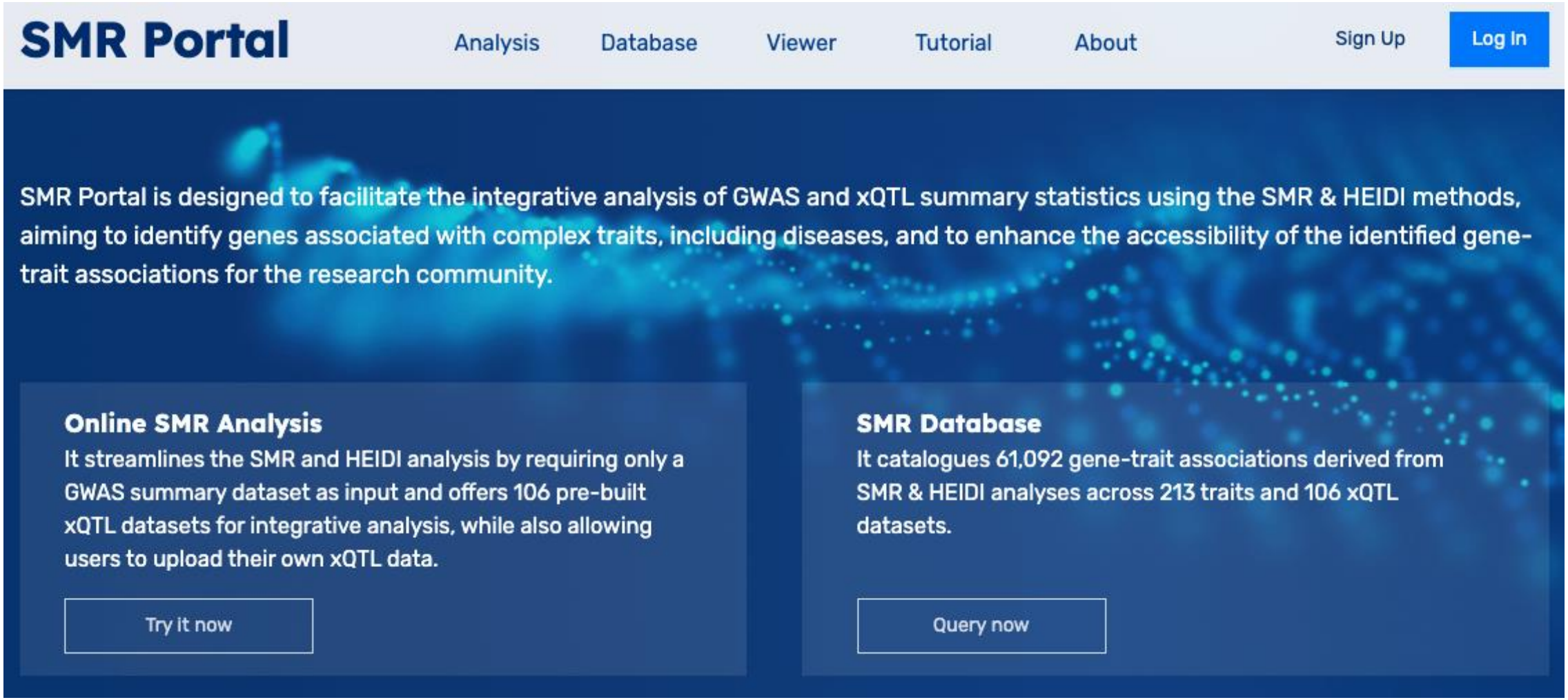
- eQTL pleiotropy:
  - Genes that are nearby should be tested to rule out pleiotropy

	SNP			Gene					Allele		
P-value	ID	Chr	Pos (hg19)	ID	Symbol	Chr	Pos (hg19)	Z-score	Assessed	Other	Nr Cohorts
2.4322e-79	rs6453133	5	74692776	ENSG00000152359	POC5	5	74991631	18.8602	G	A	36
1.2137e-50	rs6453133	5	74692776	ENSG00000113161	HMGCR	5	74645041	14.9667	G	A	37

P-value	ID	Chr	Pos	ID	Symbol	Chr	Pos	Z-score	Assessed	Other	Nr Cohorts

- Tissue specificity
- Replication in a different ancestry
- Replication with pQTL / LOF mutations

SMR portal


SMR Portal
Analysis
Database
Viewer
Tutorial
About
Sign Up
Log In
SMR Portal is designed to facilitate the integrative analysis of GWAS and xQTL summary statistics using the SMR & HEIDI methods, aiming to identify genes associated with complex traits, including diseases, and to enhance the accessibility of the identified gene-trait associations for the research community.
Online SMR Analysis
It streamlines the SMR and HEIDI analysis by requiring only a GWAS summary dataset as input and offers 106 pre-built xQTL datasets for integrative analysis, while also allowing users to upload their own xQTL data.
Try it now
SMR Database
It catalogues 61,092 gene-trait associations derived from SMR & HEIDI analyses across 213 traits and 106 xQTL datasets.
Query now

Coloc

PLOS Genetics

About Browse Publish

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics

Claudia Giambartolomei, Damjan Vukcevic, Eric E. Schadt, Lude Franke, Aroon D. Hingorani, Chris Wallace, Vincent Plagnol

Published: May 15, 2014 • https://doi.org/10.1371/journal.pgen.1004383

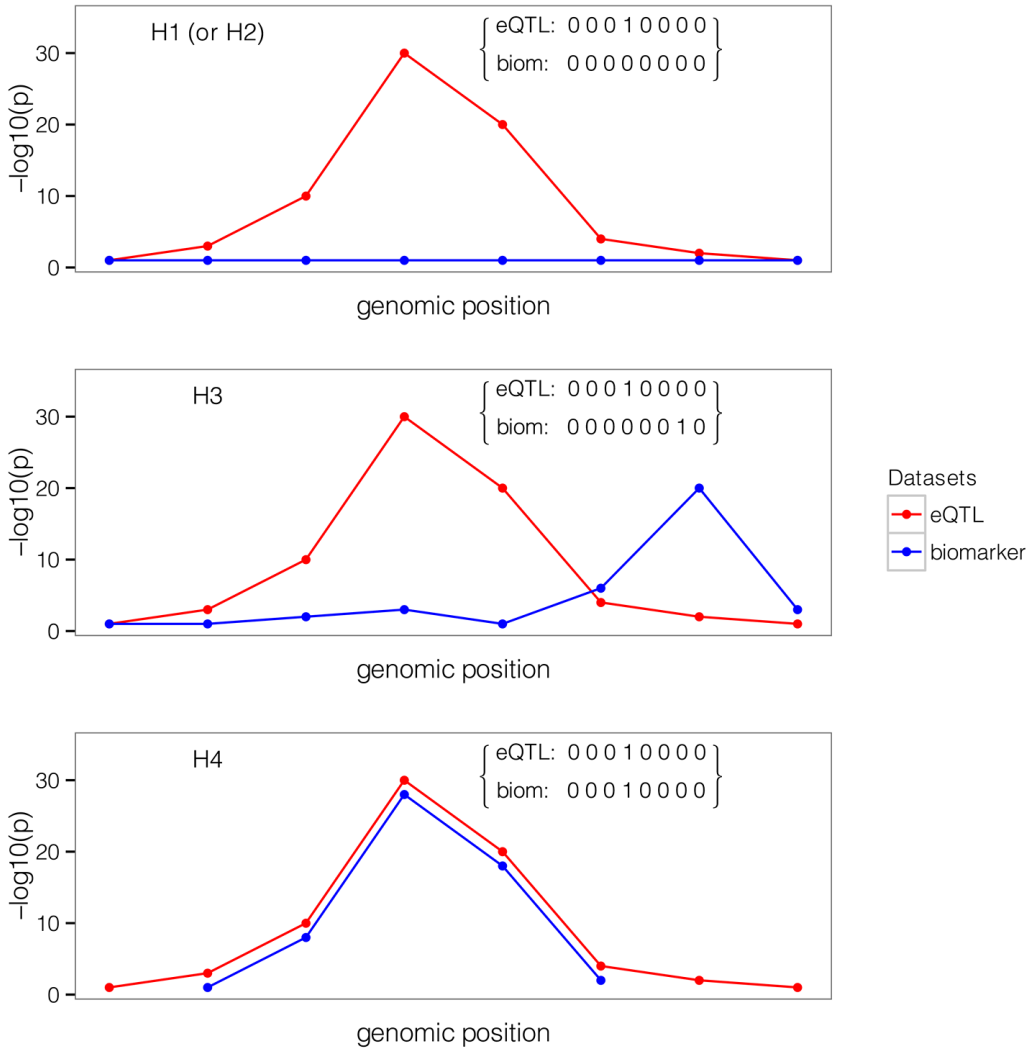
Coloc

- Colocalisation analysis aims to determine whether two traits share the same causal variant in a genomic region.
- *Coloc* uses the assumption of 0 or 1 causal variant in each trait, and tests for whether they share the same causal variant
  - if violated: SuSiE-Coloc, GCTA-COJO-Coloc
- Unlike SMR, *Coloc c*an't provide information on the directionality of the association
- Download R package:
  - install.packages("coloc")

Coloc

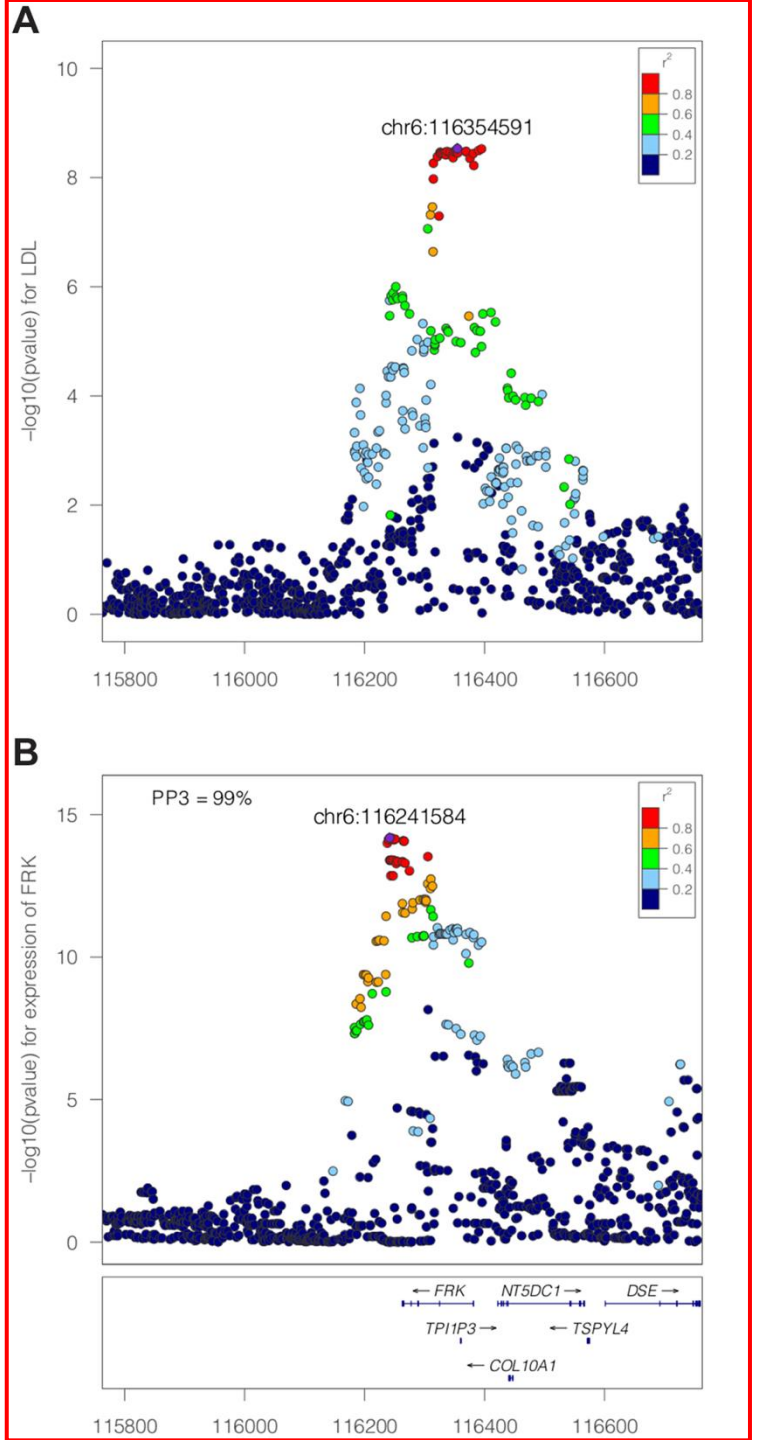
- H_0: neither trait has a genetic association in the region
- H_1: only trait 1 has a genetic association in the region
- H_2: only trait 2 has a genetic association in the region
- H_3: both traits are associated, but with different causal variants
- H_4: both traits are associated and share a single causal variant

`PP.H4.abf` is the posterior probability that two traits share a same causal variant.



Giambartolomei et al. PLOS Genet. 2014

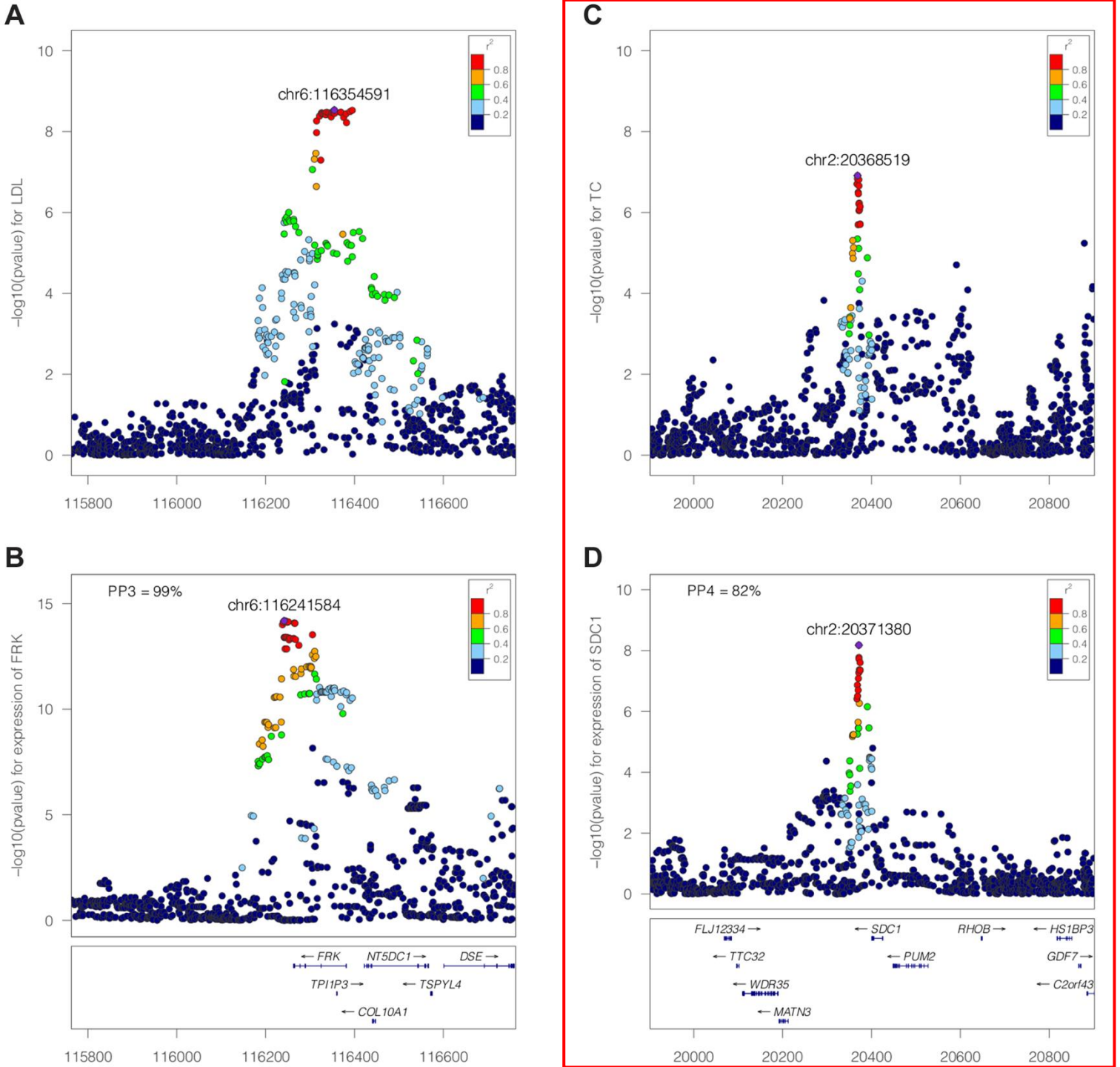
Coloc



Giambartolomei et al. PLOS Genet. 2014

PP.H3 > 90%

Coloc



Giambartolomei et al.
PLOS Genet. 2014

PP.H4 > 80%

Individual-level analysis

- A polygenic score (PGS) is a number that summarises the estimated effect of many genetic variants on an individual's phenotype (e.g. gene expression).
- A PGS can be generated from multiple independent QTLs, and used to explore associations with disease outcome
- Logistic regression, Cox proportional hazards regression

eQTL	Effect allele	Non-effect allele	Weight per copy of the effect allele
SNP 1	A	C	+1
SNP 2	G	A	+2
SNP 3	T	C	-1.5
…			

Individual-level analysis

- Advantages
  - Can use more refined definitions of diseases
  - Allows stratification
  - Can incorporate multiple genes in a pathway
- Cohorts
  - UK Biobank
  - All of US
  - China Kadoorie Biobank
  - Biobank Japan
  - …

OmicsPred

Genetic Score > **OPGS001144**

Score information

Score Name:	PCSK9.5231.79.3
Publication:	Xu Y *et al.* Nature (2023) (OPP000001)
Platform:	Somalogic (3.0) - Proteomics
Dataset:	INTERVAL SomaScan (OPD000001)
Tissue:	blood plasma (UBERON_0001969)
Method:	Bayesian Ridge regression
Reported Trait:	Proprotein convertase subtilisin/kexin type 9 (Q8NBP7)
Number of Variants:	3
Genome Build:	GRCh37
Terms & Licenses:	Creative Commons Attribution 4.0 International (CC BY 4.0)

Linked annotations

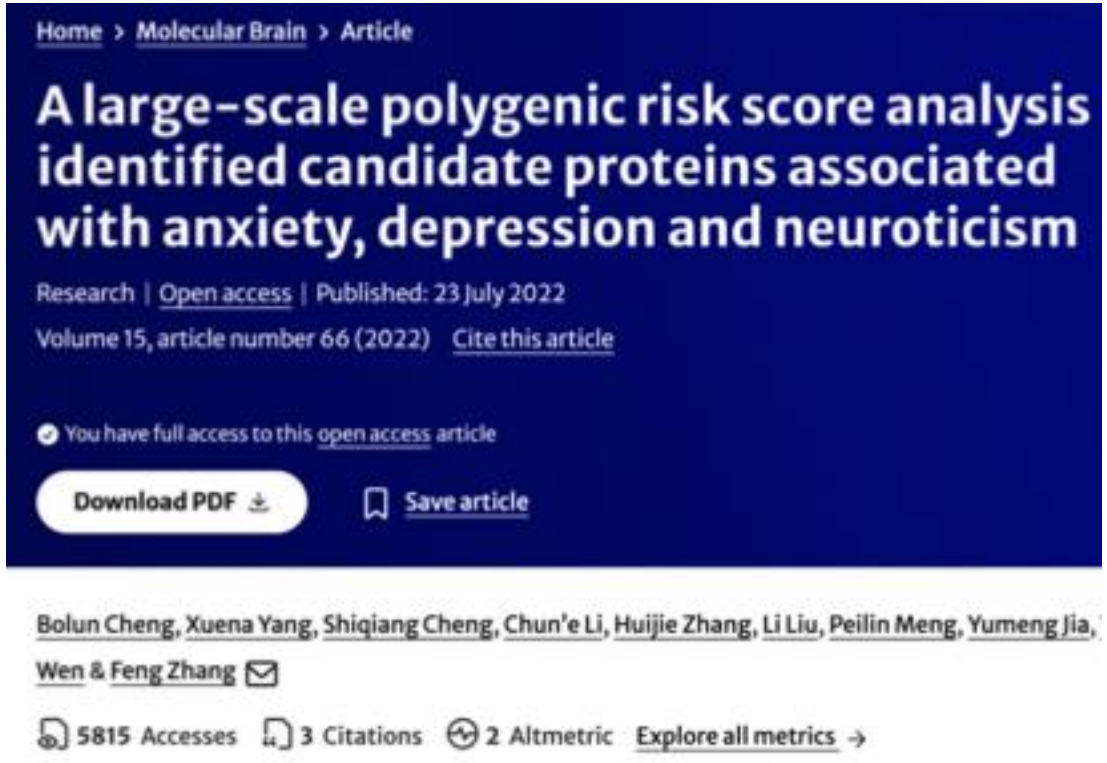
Gene:	PCSK9 (ENSG00000169174)
Protein:	Proprotein convertase subtilisin/kexin type 9 (Q8NBP7)
PheWAS:	**25** associated phenotype entries
Pathways:	**9** associated pathways

OmicsPred

Linked PheWAS data 29

Phenotype name	Phenotype ID	Reported trait	GWAS Catalog	PheWAS Publication	Ancestry	Sample	Cohort	Effect Size	HR / OR	Z-Score	P-Value	Adj P-Value	Method ...
hyperlipidemia	MONDO_0021187	Hyperlipidemia (PheCode 272.1)	-	OPP000001 Xu Y *et al.* Nature (2023)	European	360,241	UKB	0.0973	1.1 [1.08 - 1.13]	7.72	1.17e-14	1.37e-12	Cox regression (per 1 SD increase in score value)
disease, heart disorder, myocardial ischemia	MONDO_0000001, MONDO_0005267, MONDO_0024644	Other chronic ischemic heart disease. unspecified (PheCode 411.8)	-	OPP000001 Xu Y *et al.* Nature (2023)	European	385,050	UKB	0.0474	1.05 [1.03 - 1.06]	6.76	1.37e-11	1.42e-9	Cox regression (per 1 SD increase in score value)
coronary atherosclerosis	MONDO_0021661	Coronary atherosclerosis (PheCode 411.4)	-	OPP000001 Xu Y *et al.* Nature (2023)	European	390,159	UKB	0.0479	1.05 [1.04 - 1.06]	7.50	6.23e-14	5.77e-12	Cox regression (per 1 SD increase in score value)

Pathway PRS



Home › Molecular Brain › Article

A large-scale polygenic risk score analysis identified candidate proteins associated with anxiety, depression and neuroticism

Research | Open access | Published: 23 July 2022

Volume 15, article number 66 (2022) Cite this article

You have full access to this open access article

Download PDF Save article

Bolun Cheng, Xuena Yang, Shiqiang Cheng, Chun'e Li, Huijie Zhang, Li Liu, Peilin Meng, Yumeng Jia, Wen & Feng Zhang

5815 Accesses 3 Citations 2 Altmetric Explore all metrics →

Tissue	Phenotype	Protein	r (95%CI)	P
Brain	Neuroticism score	MSP		8.30×10^{-4}
	Anxiety score	TNF sR-II		1.75×10^{-3}
	Anxiety score	Ferritin		2.17×10^{-3}
	Neuroticism score	CNTN2		2.63×10^{-3}
	Anxiety self	cGMP-stimulated PDE		5.79×10^{-3}
CSF	Neuroticism score	C1-Esterase Inhibitor *		3.09×10^{-8}
	Neuroticism score	NADPH-P450 *		2.51×10^{-5}
	Depression self	ERBB1 *		4.65×10^{-5}
	Neuroticism score	IL-1F6 *		5.59×10^{-5}
	Depression self	granzyme A*		6.93×10^{-5}
Plasma	Neuroticism score	C1-Esterase Inhibitor *		2.56×10^{-9}
	Neuroticism score	IL-6 sRa		4.72×10^{-4}
	Depression self	Apo L1		5.09×10^{-4}
	Depression self	Siglec-6		6.87×10^{-4}
	Anxiety self	IL-1 R AcP		8.80×10^{-4}

-0.02 -0.01 0 0.01 0.02

OpenGWAS

home datasets about api

OpenGWAS

Empowering decision makers to maximise improvements to public health

The aim of **OpenGWAS** is to bring together disjointed sources of GWAS summary data alongside the suite of analytical tools that make value from those data. Building this coherence has drastically increased the scope of phenome-wide causal inference, alongside its reliability.

Datasets

A database of more than 50,000 GWAS summary datasets, for querying or download.

Explore datasets >

API

The fast, programmatic way to query batches, associations, phewas, variants are more.

API account & docs >

Analysis integrations

Use TwoSampleMR for MR and gwasglue2 for fine mapping, colocalisation, etc.

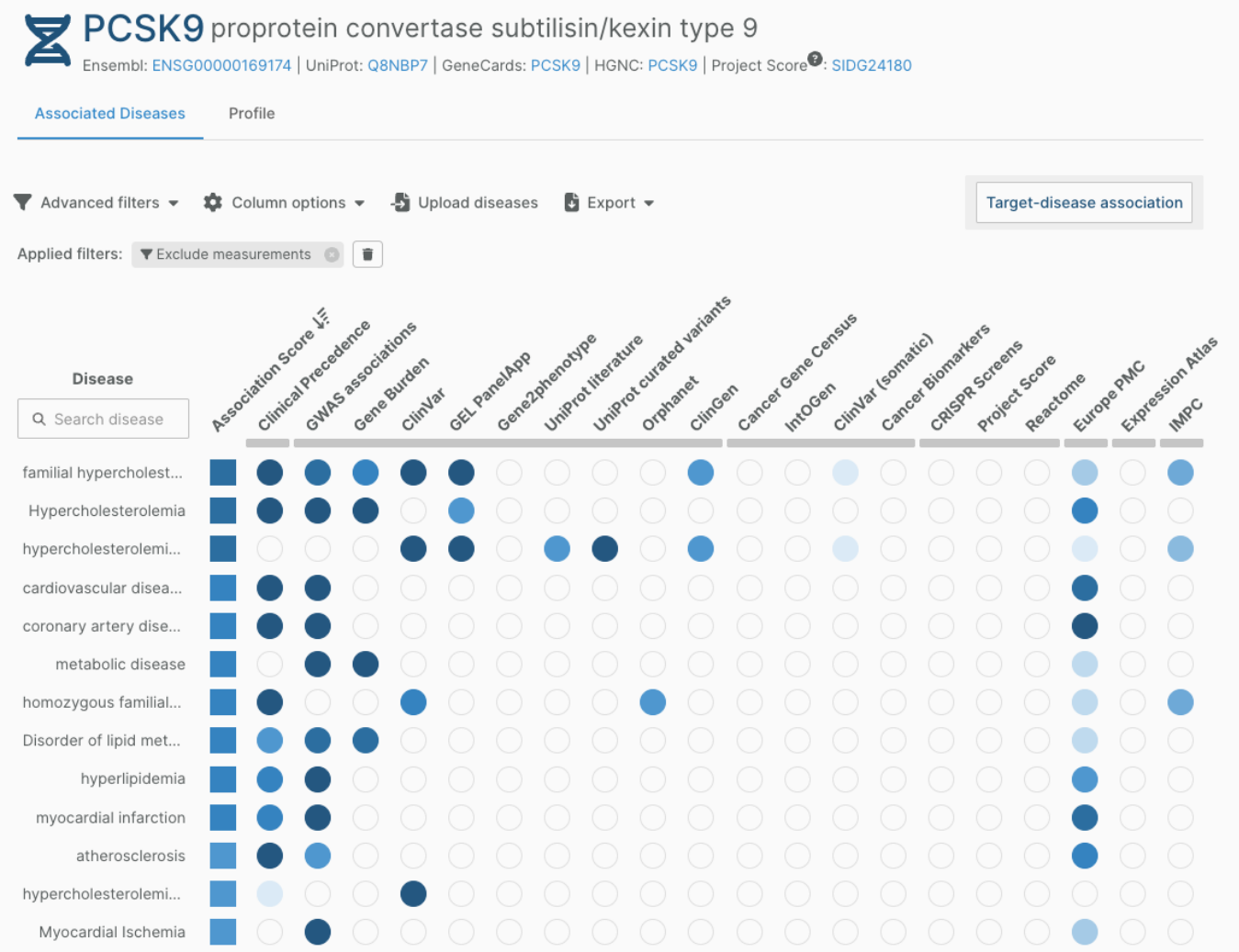
MRC IEU R-universe >

Genotype-phenotype map

Systematic finemapping and colocalization clustering across datasets with web application, API and R package.

Genotype-phenotype map >

Open Targets



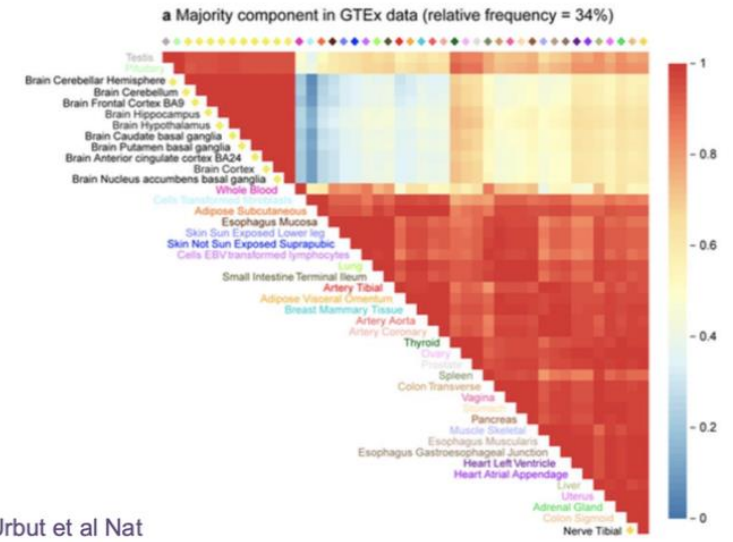
Things to consider when designing a study

- Cohorts used to generate the QTL and GWAS data (e.g. ancestry)
- Check your GWAS summary statistics (e.g. allele frequency, QQ plots, genome build)
- Biological context (e.g. relevant tissues)
- Positive control outcome
- PheWAS (potential side effects)
- MR assumptions

Limitations

- Need to know the drug target genes
- Rely on the availability of QTL data (ancestry, tissue etc.)
- Off-target effects
- Other considerations:
  - Brain-blood barrier

Many eQTLs are shared across tissues



Correlation of eQTL effect estimates for 16,069 (genes expressed and have effect estimates in all 44 tissues)

(1) effects are **positively** correlated among all tissues;

(2) the brain tissues—and, to a lesser extent, testis and pituitary—are particularly strongly correlated with one another, and less correlated with other tissues;

(3) effects in whole blood less well correlated with other tissues

Urbut et al Nat Genetics 2018

Summary

- Molecular QTLs, such as eQTLs and pQTLs, can be used as genetic instruments to infer drug effects
- SMR, *Coloc* and polygenic scores can be used to interrogate the effect of genetically predicted gene/protein modulation on disease outcomes