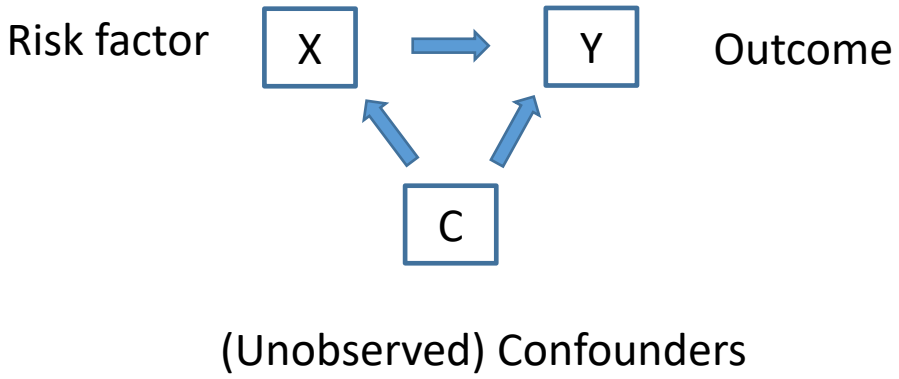
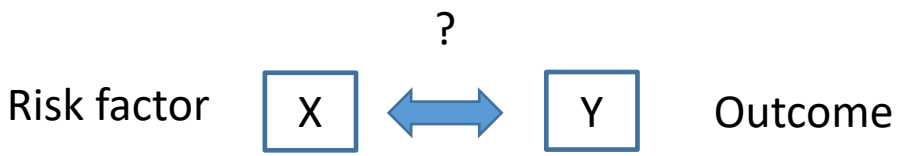
Mendelian Randomization

Drawback with observational studies

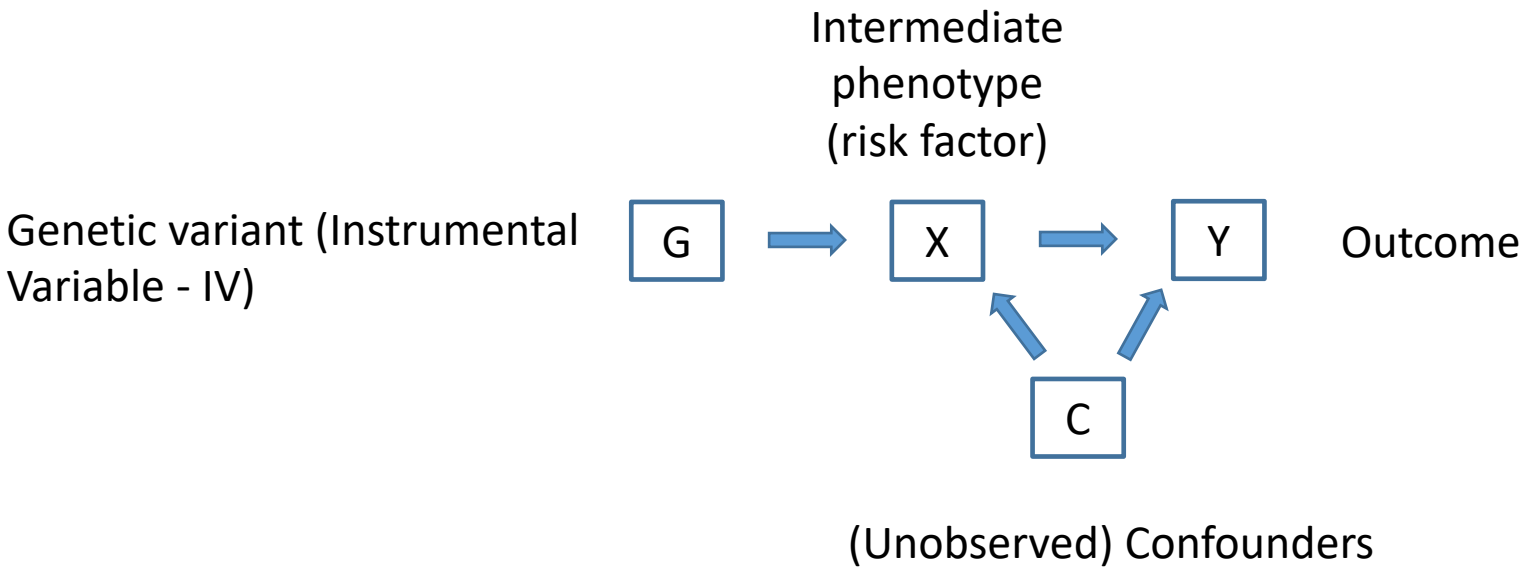
CONFOUNDING



REVERSE CAUSATION



We can leverage genetic variation to (partly) overcome these issues



Mendelian Randomization

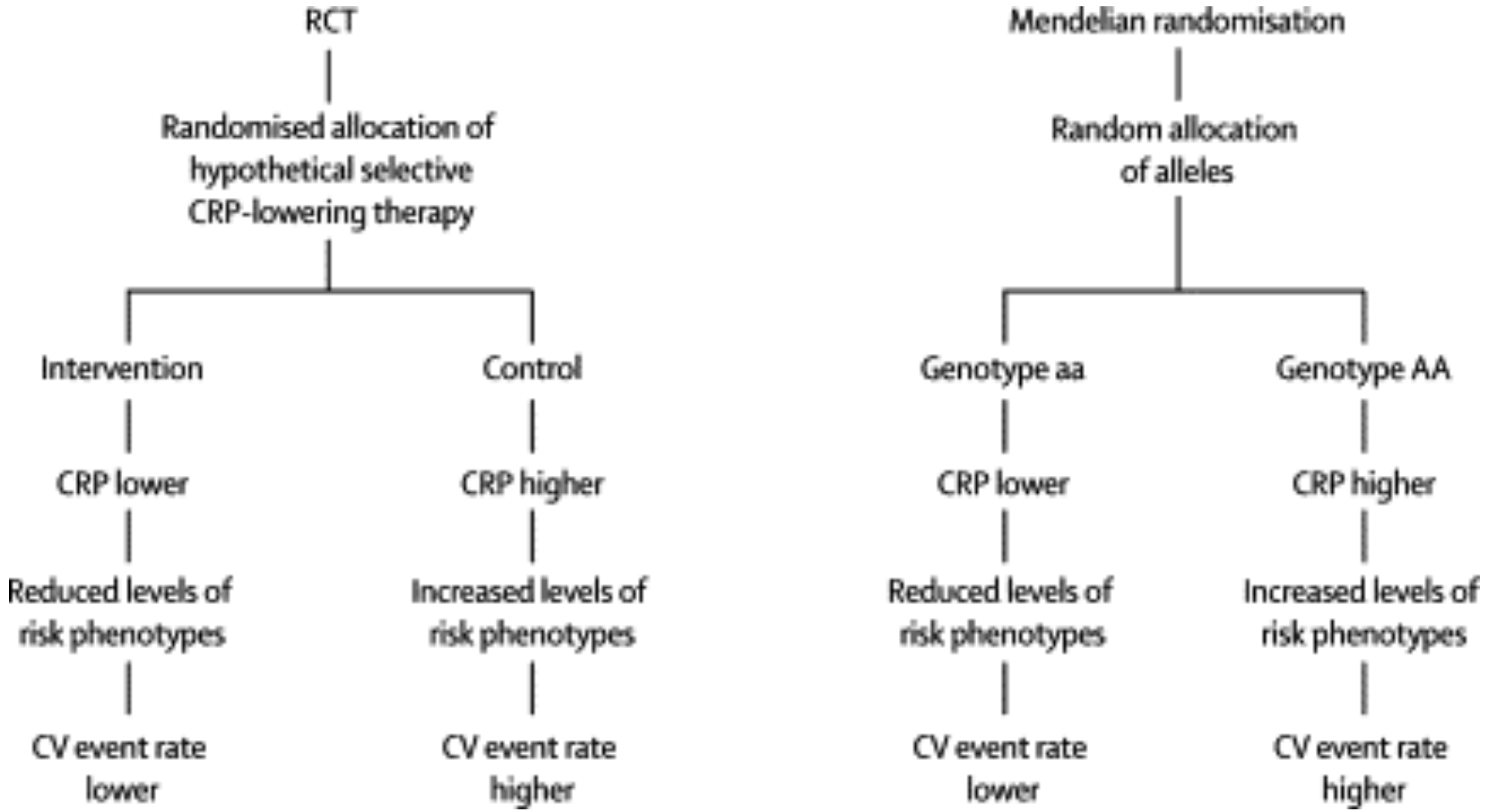
- *Basic principle: "genetic variants which mirror the biological effects of a modifiable environmental exposure and alters disease risk should be associated with disease risk to the extent predicted by their influence on exposure to the risk factor."*

- The random allocation of genetic variants from parents to offspring means these variants will generally be unrelated to other factors which affect the outcome.

- Furthermore, associations between the genotype and the outcome will not be affected by reverse causation because disease does not affect genotype

Ebrahim & Davey Smith, Hum Genet 2008
Davey Smith & Ebrahim, Int J Epi 2004

Possible effects of C-reactive protein (CRP) on cardiovascular (CV) events. Expected outcome from hypothetical randomized clinical trial of selective CRP-lowering intervention, and from Mendelian randomization analysis, if CRP were causal in developing CV.



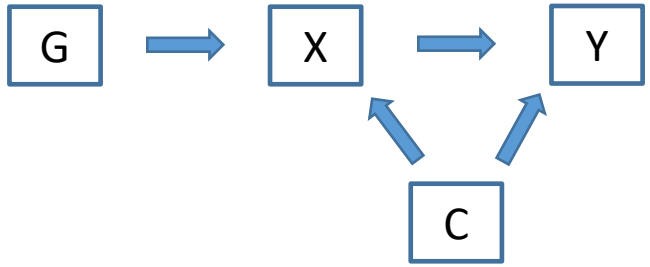
Hingorani & Humphries, Lancet 2005

Three key assumptions in MR analysis

1. G (SNP or a combination of multiple SNPs) is robustly associated with X (risk factor)

2. G is unrelated to any confounders C, that can bias the relationship between G and Y (outcome). In other words, there are no common causes of G and Y (e.g. population stratification)

3. G is related to Y only through its association with X (i.e. no pleiotropy)



Assumption 1: G is robustly associated with X

- Under certain conditions, the relative bias of the instrument variable (IV) estimate is ~1/F. A “weak” IV has been defined as having F<10, where

$$F = \frac{R^2(n-1-k)}{(1-R^2)k}$$

R^2 is variance in X explained by the IV(s),
n is sample size and k is number of IVs

- Weak IVs can lead to biased effect estimates (in the direction of the observed X-Y association) in the presence of confounding of the X–Y relationship.

Pierce, IJE 2011

Assumption 2: No confounding

- G is independent of factors (measured and unmeasured) that confound the X-Y relation

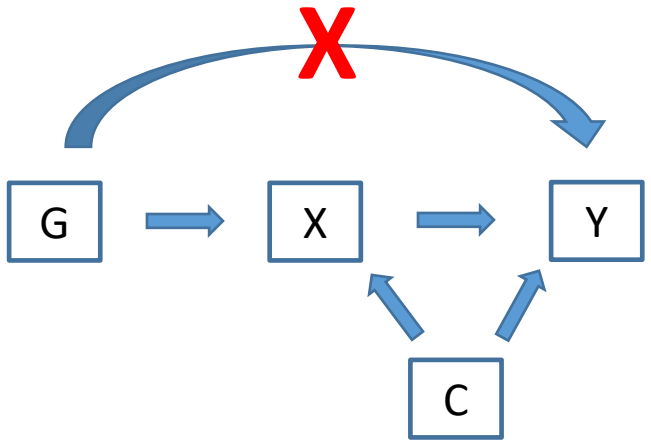
- Since G is randomized at birth and thus is independent of non-genetic confounders and is not modified by the course of disease, the one main concern here is population stratification – i.e. if ancestry is related both to G and Y.

- If you have individual-level data, you can test for this (e.g. PCs)

Assumption 3: No pleiotropy

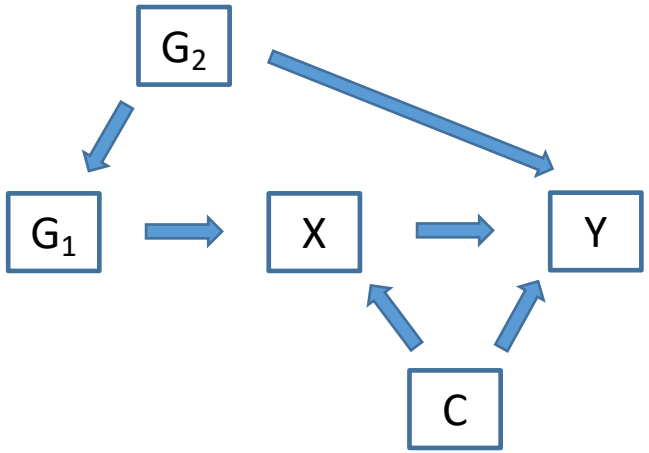
- This assumption is the trickiest

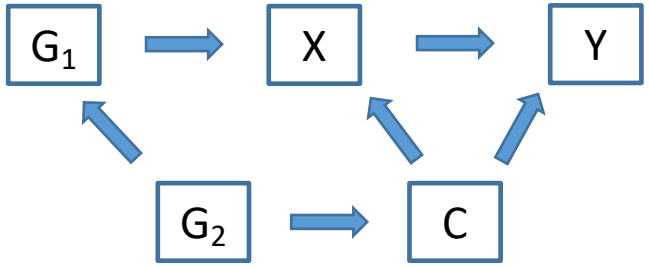
- Assumes that G is only associated with Y via X and thus the association between G and Y is fully mediated by X and not through any unmeasured factor(s). Needs to be true for SNPs in LD too



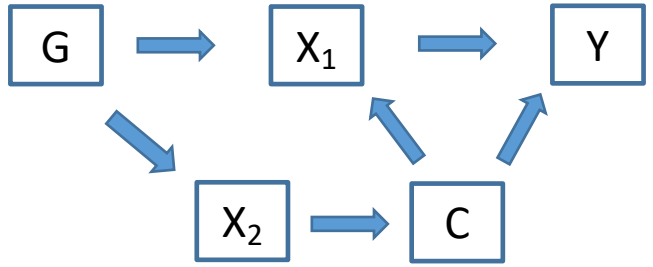
Scenarios invalidating assumption 3

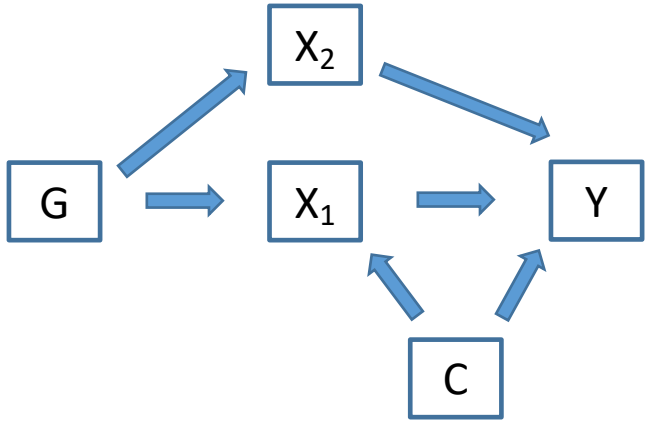
LD





Pleiotropy





Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies[1]

Philip C Haycock,[2] Stephen Burgess,[3] Kaitlin H Wade,[2] Jack Bowden,[2,4] Caroline Relton,[2] and George Davey Smith[2]*

TABLE 2
Different design strategies for MR[1]

Study design	Test	Comments
G-X + G-Y	Implies X→Y	No estimation of magnitude of causal effect
One-sample MR	Various hypotheses	Requires individual-level data; lower power; MR estimates are biased toward the confounded observational association by weak instruments
Two-sample MR	Various hypotheses	Individual-level or summary data; greater power (due to greater potential sample sizes); MR estimates are biased toward the null by weak instruments
Bidirectional MR	X→Y and Y→X	Assesses causation in both directions
Two-step MR	X→M→Y	Tests mediation in a causal pathway
G×E	X→Y (relation is dependent on environment variable)	Able to detect direct effects (a violation of assumption 2 of MR)

[1]G×E, gene-environment interaction; G-X, SNP-exposure association; G-Y, SNP-outcome association, M, mediator; MR, Mendelian randomization; SNP, single nucleotide polymorphism; X, hypothesized exposure; Y, outcome variable of interest.

Haycock et al, Am J Clin Nutr 2016

Individual-level data in one sample

- Access to SNPs, risk factor, and outcome for all participants

- The causal effect of X on Y can be estimated using 2-stage least-squares (2SLS) regression:

1. $X = a + \gamma G$
2. $Y = c + \beta X^*$, where X^* are the genetically predicted exposure levels as measured in (1)

- The causal estimate is given by β
- Can be implemented in R using the “ivpack” package
- Weak instruments cause bias towards the observed confounded association

Summary data from two samples

- The G-X and the G-Y associations are estimated in two different samples.
- Assumes no overlap among samples and that the two populations are similar (ethnicity, age, sex, etc.)
- Here, bias due to weak IVs will be towards the null
- Note: The G-X and G-Y associations need to be coded using the same effect allele

Summary data from two samples

$$\hat{\beta} = \frac{\sum_k \beta_{1k}\beta_{2k}\sigma_{\beta_{2k}}^{-2}}{\sum_k \beta_{1k}\sigma_{\beta_{2k}}^{-2}}$$

$$se(\hat{\beta}) = \sqrt{\frac{1}{\sum_k \beta_{1k}^2\sigma_{\beta_{2k}}^{-2}}}$$

β_{1k} is the mean change in X per allele for SNP k, β_{2k} is the mean change in Y per allele for SNP k, σ_{2k}^{-2} is the inverse variance for the G-Y association.

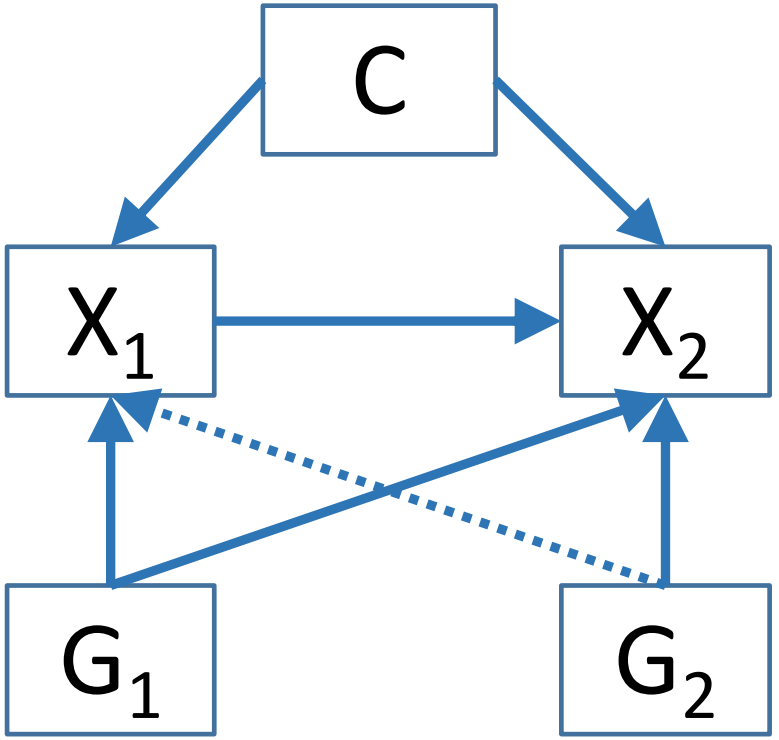
Bidirectional MR analysis

- Approach to overcome reverse causation

- IVs for both X_1 and X_2 are used to assess the causal association in both directions

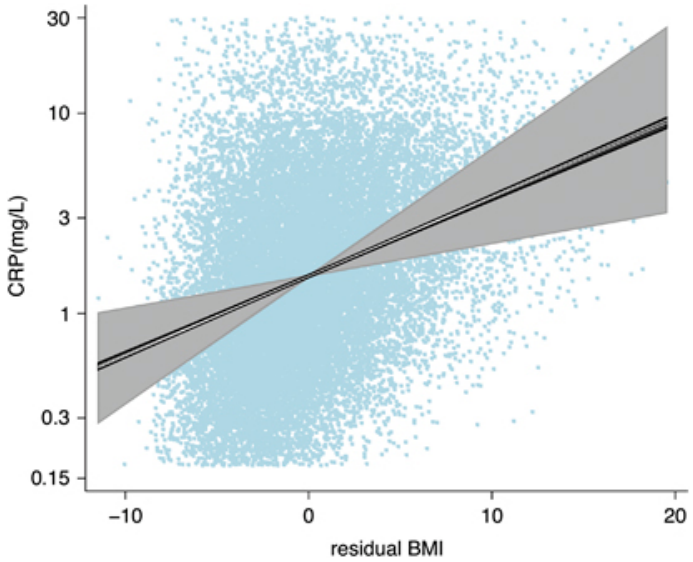
1. Is G_1 associated with X_2?
2. Is G_2 associated with X_1?

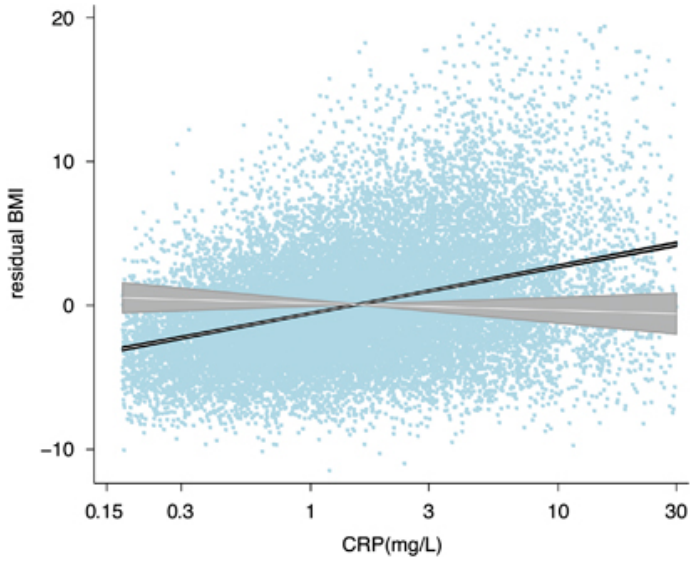
(Also confirm that G_1 is associated with X_1 and that G_2 is associated with X_2



BMI and CRP – what causes what?

- There is a consistent observed association between high BMI and high CRP levels





Light grey points represent a scatter plot of the correlation between circulating CRP and residual BMI. Gray areas represent 95% confidence regions around IV estimates. Black area represents 95% confidence regions around simple linear regression estimates.

Timpson et al, Int J Obesity 2011

Table 5. Observational and instrumental variable derived relationships between BMI and circulating CRP.

Previous table | Figures and tables index

	Effect estimates				
Outcome/explanatory variable	***Observational***	***Instrumental variable***	P_{IV}	P_{diff}	F_{first}
CRP/BMI	1.46 (1.44, 1.48)	1.41 (1.10, 1.80)	0.006	0.8	31.1
BMI/CRP	1.03 (1.00, 1.07)	−0.24 (−0.58, 0.11)	0.2	<0.0001	57.3

These data suggest that the observed association between circulating CRP and measured BMI is likely to be driven by BMI, with CRP being a marker of elevated adiposity.

Timpson et al, Int J Obesity 2011

Drawbacks with MR analysis

- Large sample sizes are needed
  - As genetic effects on risk factors are typically small, MR estimates of association have much wider confidence intervals than conventional epidemiological estimates.

- Make sure that the three key assumptions hold
  - In practice, this is very difficult, especially for the third assumption of no pleiotropy.

TABLE 4
Practical strategies for enhancing causal inference[1]

Strategy	Addresses	Rationale/explanation	Potential limitation
Pleiotropy analyses	Genetic confounding	Test association between instrument(s) and wide range of potential confounders	Does not test for association with unknown confounders
Exclusion of nonspecific SNPs	Genetic confounding	SNPs associated with multiple exposures may introduce pleiotropy	Power may be limited to detect nonspecific associations; exclusion of nonspecific SNPs can also introduce bias into the analysis
Weighted median estimator	Violation of all MR assumptions	Sensitivity analysis allowing 50% of the instruments to be invalid	At least 50% of the genetic proxies must be valid instruments
MR-Egger regression	Direct effects/horizontal pleiotropy	Sensitivity analysis allowing all instruments to be subject to direct effects (i.e., horizontal pleiotropy)	The InSIDE assumption is required: strength of the gene-exposure association must not correlate with the strength of bias due to horizontal pleiotropy
Gene-environment interactions	Genetic confounding	Association should only be observed in certain exposure subgroups (e.g., smoking instruments in ever- compared with never-smokers)	Limited number of available gene-environment interactions; can introduce collider bias
Multiple independent instruments	Genetic confounding	Association across multiple independent genomic regions should be robust to confounding	Power likely to be limited for individual genetic variants
Two-sample approaches	Weak instrument bias and low power	Allows larger sample sizes because measurement of the exposure is not required in all samples; bias from weak instruments is toward the null, rather than the confounded observational association	Samples must be independent and representative of the same population; less flexible than 2SLS
Multi-SNP instruments	Weak instrument bias and low power	Instrument will explain more of the variance in the exposure, reducing impact of weak instruments bias and increasing power	Requires multiple GWAS significant hits; increases chance of pleiotropy
External weights for 2SLS	Weak instrument bias	Weight the first stage by SNP-exposure effect from an external study	Precisely estimated external weights must be unavailable

[1]GWAS, genome-wide association study; InSIDE, Instrument Strength Independent of Direct Effect; MR, Mendelian randomization; SNP, single nucleotide polymorphism; 2SLS, 2-stage least squares.

Haycock et al, Am J Clin Nutr 2016

MR-base: An easy tool for Mendelian Randomization Analysis

Overview:

- Collaboratively developed by the University of Bristol, University of Cambridge and Translational Research Institute of Australia.
- A web-based platform (MR-Base) and an R-package "*TwoSampleMR*'.
- http://app.mrbase.org/
- With billions of genotype-phenotype associations and thousands of traits available on MR-Base; and allows manual file upload.

Hemani, Gibran, et al. "The MR-Base platform supports systematic causal inference across the human phenome." *Elife* 7 (2018): e34408.

MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

University of BRISTOL | MRC Integrative Epidemiology Unit | CANCER RESEARCH UK

MR BASE

A platform for Mendelian randomisation using summary data from genome-wide association studies

To begin analysis in the web application please review the data access agreement and accept by logging in with your google account.

Get started

All analyses, data extraction and more can be performed using the TwoSampleMR R package. Additionally, you can use the R package to analyse your own outcome datasets.

Get the R package

Accept the data access agreement;
Log in with your Google ID

Current status

Beta phase release

App version:
1.2.2 3a435d (31 January 2019)

R version:
3.5.1

Host:
e4ec2116cb55

R/TwoSampleMR version:
0.4.18

Database version:
0.2.0 (17 December 2017)

To use MR-Base using the TwoSampleMR R package directly:
https://github.com/MRCIEU/TwoSampleMR

See our sister website LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/

MRBASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
- Quick SNP lookup

University of BRISTOL | MRC Integrative Epidemiology Unit | CANCER RESEARCH UK

MRBASE

A platform for Mendelian randomisation using summary data from genome-wide association studies

Your name and Email address
should appear here after the log-in

To begin analysis in the web application please review the data access agreement and accept by logging in with your google account.

Get started

All analyses, data extraction and more can be performed using the TwoSampleMR R package. Additionally, you can use the R package to analyse your own outcome datasets.

Get the R package

Current status

Beta phase release

App version:
1.2.2 3a435d (31 January 2019)

R version:
3.5.1

Host:
e4ec2116cb55

R/TwoSampleMR version:
0.4.18

Database version:
0.2.0 (17 December 2017)

To use MR-Base using the TwoSampleMR R package directly:
https://github.com/MRCIEU/TwoSampleMR

See our sister website LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/

MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
- Quick SNP lookup

University of BRISTOL | MRC Integrative Epidemiology Unit | CANCER RESEARCH UK

MR BASE

A platform for Mendelian randomisation using summary data from genome-wide association studies

Click on 'Perform MR analysis'

To begin analysis in the web application please review the data access agreement and accept by logging in with your google account.

Get started

All analyses, data extraction and more can be performed using the TwoSampleMR R package. Additionally, you can use the R package to analyse your own outcome datasets.

Get the R package

Current status

Beta phase release

App version:
1.2.2 3a435d (31 January 2019)

R version:
3.5.1

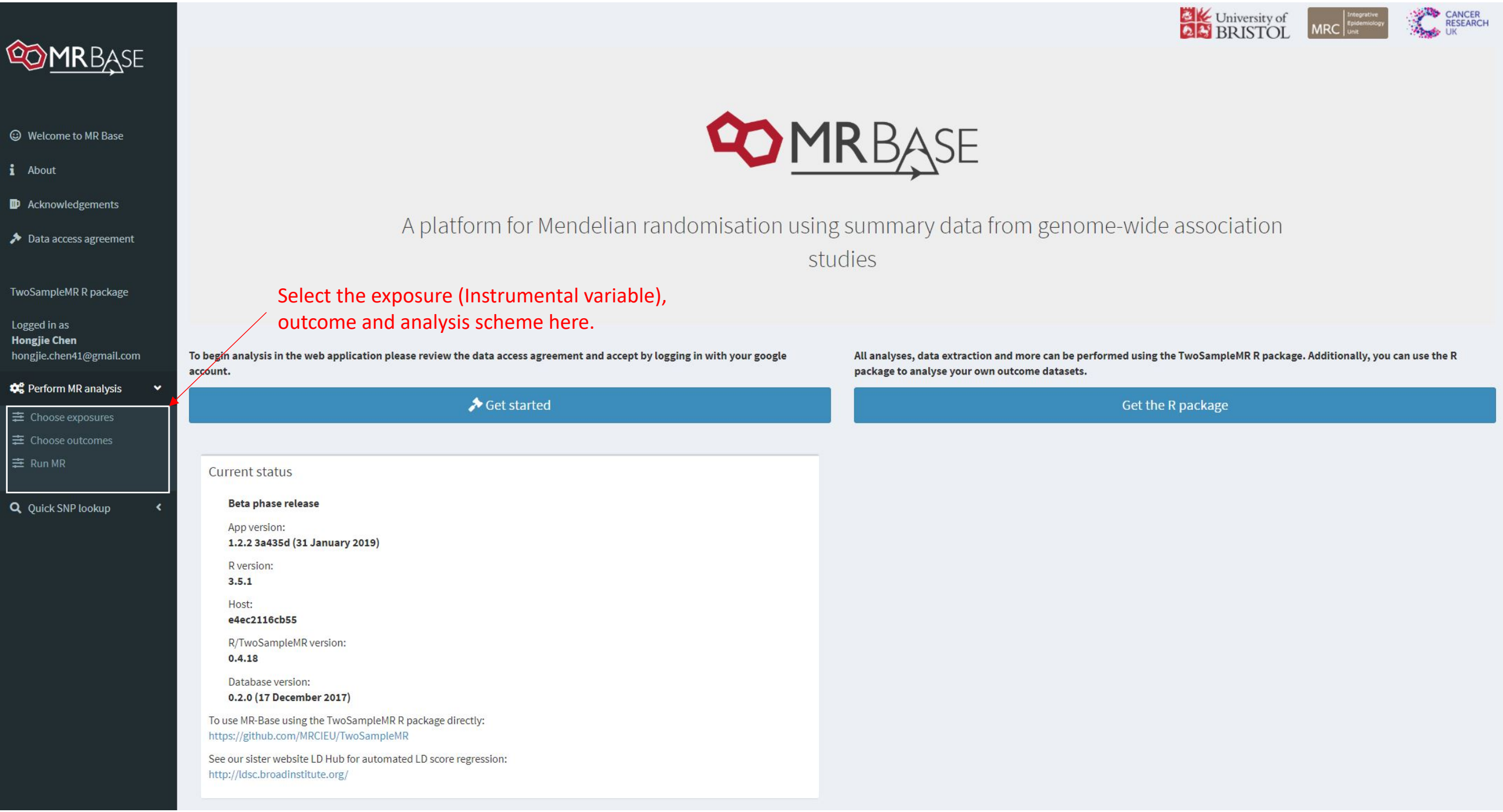
Host:
e4ec2116cb55

R/TwoSampleMR version:
0.4.18

Database version:
0.2.0 (17 December 2017)

To use MR-Base using the TwoSampleMR R package directly:
https://github.com/MRCIEU/TwoSampleMR

See our sister website LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/


MR BASE
Welcome to MR Base
About
Acknowledgements
Data access agreement
TwoSampleMR R package
Logged in as
Hongjie Chen
hongjie.chen41@gmail.com
Perform MR analysis
Choose exposures
Choose outcomes
Run MR
Quick SNP lookup
University of BRISTOL
MRC Integrative Epidemiology Unit
CANCER RESEARCH UK
MR BASE
A platform for Mendelian randomisation using summary data from genome-wide association studies
Select the exposure (Instrumental variable), outcome and analysis scheme here.
To begin analysis in the web application please review the data access agreement and accept by logging in with your google account.
Get started
All analyses, data extraction and more can be performed using the TwoSampleMR R package. Additionally, you can use the R package to analyse your own outcome datasets.
Get the R package
Current status
Beta phase release
App version:
1.2.2 3a435d (31 January 2019)
R version:
3.5.1
Host:
e4ec2116cb55
R/TwoSampleMR version:
0.4.18
Database version:
0.2.0 (17 December 2017)
To use MR-Base using the TwoSampleMR R package directly:
https://github.com/MRCIEU/TwoSampleMR
See our sister website LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/

Choosing instruments for the exposure

To use two sample MR to estimate the causal effect of an exposure on an outcome, the first step is to identify SNPs that are robustly associated with the exposure. These summary statistics for these SNPs can be taken from a sample from which there is no data on the outcome.

Please provide instruments by choosing from one of the data sources below, or by uploading your own data. You can choose multiple exposures to be analysed, and multiple instruments per exposure.

Choose instruments

Select exposure source

- ○ Manual file upload
- ○ NHGRI-EBI GWAS catalog
- ◉ MR Base GWAS catalog
- ○ Gene expression QTLs
- ○ Protein level QTLs
- ○ Metabolite level QTLs
- ○ Methylation level QTLs

Can either use the instruments provided by MR-base,
Or use the manually uploaded file.

MR Base GWAS catalog

The MR Base database holds a collection of the summary statistics from a large number of GWASs. It is possible to use this resource to manually identify instruments, and to therefore use these traits as exposures by finding the independent GWAS significant hits from these summary associations.

To use a trait as an exposure, highlight the relevant row in the table below (multiple traits can be selected). All SNPs with p values below the specified threshold will be extracted, and clumping will be used to remove SNPs in LD with sentinal SNPs. These SNPs will be used as instruments in the MR analysis.

p-value threshold: 5e-08

LD Rsq: 0.001

Selection criteria of SNPs

☑ Perform clumping

Clumping distance (kb): 10000

Display columns

☑ ID ☑ Trait ☑ Note ☑ First author ☑ Consortium ☑ Number of cases ☑ Number of controls ☑ Sample size ☑ Number of variants ☑ Year ☑ PubmedID ☑ Access ☑ Category ☑ Population ☑ Priority ☑ Sd ☑ Sex ☑ Subcategory ☑ Unit

Show 10 entries Search:

	ID	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	PubmedID	Access	Category	Population	Priority	Sd	Sex	Subcategory	Unit
1	1	Adiponectin		Dastani Z	ADIPOGen			39883	2675209	2012	22479202	public	Risk factor	Mixed	1	0.57	Males and females	Protein	ln(mg/dL)
10	10	Crohn's disease		Jostins L	IIBDGC	14763	15977	30740	13898	2012	23128233	public	Disease	European	1		Males and females	Autoimmune / inflammatory	log odds
100	100	Hip circumference	Adjusted for BMI	Randall JC	GIANT			60586	2725796	2013	23754948	public	Risk factor	European	15	8.4548	Males	Anthropometric	SD (cm)
1000	1000	Depressive symptoms		Okbay	SSGAC			161460	6524475	2016	27089181	public	Risk factor	European	1		Males and females	Psychiatric / neurological	SD
1001	1001	Years of schooling		Okbay	SSGAC			293723	8146841	2016	27225129	public	Risk factor	European	1	3.71	Males and females	Education	SD (years)
1002	1002	Leptin	Adjusted for BMI; effect allele frequencies are missing	Kilpelainen				32161	2474010	2016	26833098	public	Risk factor	European	1		Males and females	Hormone	log ng/ml
1003	1003	Leptin	Effect allele frequencies are missing	Kilpelainen				32161	2473865	2016	26833098	public	Risk factor	European	2		Males and females	Hormone	log ng/ml
1004	1004	Age at menopause		Day	ReproGen			69360	2418696	2015	26414677	public	Risk factor	European	1	3.93	Females	Reproductive aging	years
1005	1005	Percent emphysema		Manichaikul A	MESA			7667	2968584	2014	24383474	public	Risk factor	Mixed	1	0.708	Males and females	Lung disease	log (% emphysema + 1)
1006	1006	Mean platelet volume	Effect allele frequencies are missing	Gieger C	HaemGen			66867	2690859	2011	22139419	public	Risk factor	European	1	0.109	Males and females	Haemotological	log fl
	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All

Specify the literatures to be included here

Search for the exposure of interest here.

Showing 1 to 10 of 21,266 entries Previous 1 2 3 4 5 … 2127 Next

Choosing instruments for the exposure

To use two sample MR to estimate the causal effect of an exposure on an outcome, the first step is to identify SNPs that are robustly associated with the exposure. These summary statistics for these SNPs can be taken from a sample from which there is no data on the outcome.

Please provide instruments by choosing from one of the data sources below, or by uploading your own data. You can choose multiple exposures to be analysed, and multiple instruments per exposure.

Choose instruments

Select exposure source

- ○ Manual file upload
- ○ NHGRI-EBI GWAS catalog
- ◉ MR Base GWAS catalog
- ○ Gene expression QTLs
- ○ Protein level QTLs
- ○ Metabolite level QTLs
- ○ Methylation level QTLs

MR Base GWAS catalog

The MR Base database holds a collection of the summary statistics from a large number of GWASs. It is possible to use this resource to manually identify instruments, and to therefore use these traits as exposures by finding the independent GWAS significant hits from these summary associations.

To use a trait as an exposure, highlight the relevant row in the table below (multiple traits can be selected). All SNPs with p values below the specified threshold will be extracted, and clumping will be used to remove SNPs in LD with sentinal SNPs. These SNPs will be used as instruments in the MR analysis.

p-value threshold 5e-08

LD Rsq 0.001

☑ Perform clumping

Clumping distance (kb) 10000

Display columns

☑ ID ☑ Trait ☑ Note ☑ First author ☑ Consortium ☑ Number of cases ☑ Number of controls ☑ Sample size ☑ Number of variants ☑ Year ☑ PubmedID ☑ Access ☑ Category ☑ Population ☑ Priority ☑ Sd ☑ Sex ☑ Subcategory ☑ Unit

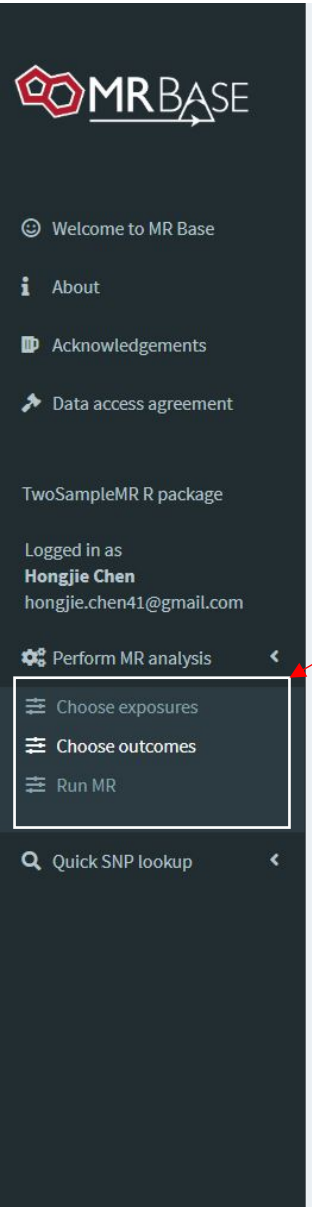
Show 10 entries Search:

	ID	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	PubmedID	Access	Category	Population	Priority	Sd	Sex	Subcategory	Unit
1	1	Adiponectin		Dastani Z	ADIPOGen			39883	2675209	2012	22479202	public	Risk factor	Mixed	1	0.57	Males and females	Protein	ln(mg/dL)
10	10	Crohn's disease		Jostins L	IIBDGC	14763	15977	30740	13898	2012	23128233	public	Disease	European	1		Males and females	Autoimmune / inflammatory	log odds
100	100	Hip circumference	Adjusted for BMI	Randall JC	GIANT			60586	2725796	2013	23754948	public	Risk factor	European	15	8.4548	Males	Anthropometric	SD (cm)
1000	1000	Depressive symptoms		Okbay	SSGAC			51460	6524475	2016	27089181	public	Risk factor	European	1		Males and females	Psychiatric / neurological	SD
1001	1001	Years of schooling		Okbay	SSGAC			293723	8146841	2016	27225129	public	Risk factor	European	1	3.71	Males and females	Education	SD (years)
1002	1002	Leptin	Adjusted for BMI; effect allele frequencies are missing	Kilpelainen				32161	2474010	2016	26833098	public	Risk factor	European	1		Males and females	Hormone	log ng/ml
1003	1003	Leptin	Effect allele frequencies are missing	Kilpelainen				32161	2473865	2016	26833098	public	Risk factor	European	2		Males and females	Hormone	log ng/ml
1004	1004	Age at menopause		Day	ReproGen			69360	2418696	2015	26414677	public	Risk factor	European	1	3.93	Females	Reproductive aging	years
1005	1005	Percent emphysema		Manichaikul A	MESA			7667	2968584	2014	24383474	public	Risk factor	Mixed	1	0.708	Males and females	Lung disease	log (% emphysema + 1)
1006	1006	Mean platelet volume	Effect allele frequencies are missing	Gieger C	HaemGen			66867	2690859	2011	22139419	public	Risk factor	European	1	0.109	Males and females	Haemotological	log fl

Select the instruments
to be included
In your MR analysis

Showing 1 to 10 of 21,266 entries

Previous 1 2 3 4 5 … 2127 Next



Select outcomes for analysis

The MR Base database houses a large collection of summary statistic data from hundreds of GWAS studies. In order to perform two sample MR, the SNPs that were selected for the exposures will be extracted from the outcomes that you select here.

Please select the outcomes that you want to test for being causally influenced by the exposures.

Studies available in MR base

Display columns

- [ ] ID
- [x] Trait
- [x] Note
- [x] First author
- [x] Consortium
- [x] Number of cases
- [x] Number of controls
- [x] Sample size
- [x] Number of variants
- [x] Year
- [ ] PubmedID
- [ ] Access
- [x] Category
- [ ] Population
- [ ] Priority
- [ ] Sd
- [ ] Sex
- [x] Subcategory
- [ ] Unit

Show 10 entries

Search:

Choose the outcome of the MR analysis

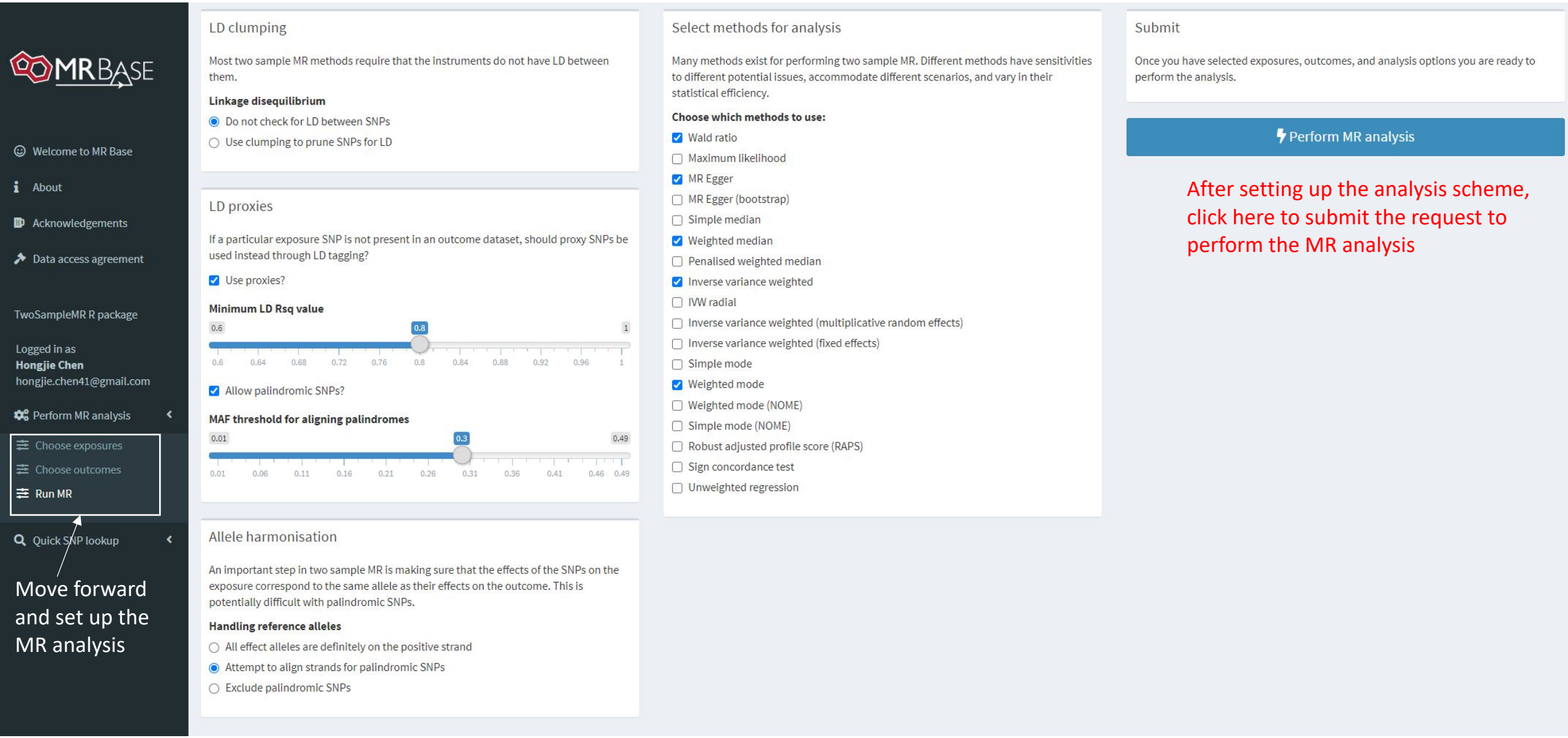
Select the outcome GWAS data to be used

	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	Category	Subcategory
1126	Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)		Michailidou K	BCAC	122977	105974	228951	10680257	2017	Disease	Cancer
1127	ER+ Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)		Michailidou K	BCAC	69501	105974	175475	10680257	2017	Disease	Cancer
1128	ER- Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)		Michailidou K	BCAC	21468	105974	127442	10680257	2017	Disease	Cancer
1129	Breast cancer (Oncoarray)		Michailidou K	BCAC	61282	45494	106776	10680257	2017	Disease	Cancer
1130	Breast cancer (iCOGS)		Michailidou K	BCAC	46785	42892	89677	10680257	2017	Disease	Cancer
1131	Breast cancer (GWAS)		Michailidou K	BCAC	14910	17588	32498	10680257	2017	Disease	Cancer
1132	ER+ Breast cancer (Oncoarray)		Michailidou K	BCAC	38197	45494	83691	10680257	2017	Disease	Cancer
1133	ER+ Breast cancer (iCOGS)		Michailidou K	BCAC	27078	42892	69970	10680257	2017	Disease	Cancer
1134	ER+ Breast cancer (GWAS)		Michailidou K	BCAC	4226	17588	21814	10680257	2017	Disease	Cancer
1135	ER- Breast cancer (Oncoarray)		Michailidou K	BCAC	9655	45494	55149	10680257	2017	Disease	Cancer
	Breast Cancer	All	All	All	All	All	All	All	All	All	All

Showing 1 to 10 of 28 entries (filtered from 21,266 total entries)

Search for the outcome of interest

Previous 1 2 3 Next


MR BASE
Welcome to MR Base
About
Acknowledgements
Data access agreement
TwoSampleMR R package
Logged in as
Hongjie Chen
hongjie.chen41@gmail.com
Perform MR analysis
Choose exposures
Choose outcomes
Run MR
Quick SNP lookup
Move forward and set up the MR analysis
LD clumping
Most two sample MR methods require that the instruments do not have LD between them.
Linkage disequilibrium
Do not check for LD between SNPs
Use clumping to prune SNPs for LD
LD proxies
If a particular exposure SNP is not present in an outcome dataset, should proxy SNPs be used instead through LD tagging?
Use proxies?
Minimum LD Rsq value
0.6 0.8 1
0.6 0.64 0.68 0.72 0.76 0.8 0.84 0.88 0.92 0.96 1
Allow palindromic SNPs?
MAF threshold for aligning palindromes
0.01 0.3 0.49
0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41 0.46 0.49
Allele harmonisation
An important step in two sample MR is making sure that the effects of the SNPs on the exposure correspond to the same allele as their effects on the outcome. This is potentially difficult with palindromic SNPs.
Handling reference alleles
All effect alleles are definitely on the positive strand
Attempt to align strands for palindromic SNPs
Exclude palindromic SNPs
Select methods for analysis
Many methods exist for performing two sample MR. Different methods have sensitivities to different potential issues, accommodate different scenarios, and vary in their statistical efficiency.
Choose which methods to use:
Wald ratio
Maximum likelihood
MR Egger
MR Egger (bootstrap)
Simple median
Weighted median
Penalised weighted median
Inverse variance weighted
IVW radial
Inverse variance weighted (multiplicative random effects)
Inverse variance weighted (fixed effects)
Simple mode
Weighted mode
Weighted mode (NOME)
Simple mode (NOME)
Robust adjusted profile score (RAPS)
Sign concordance test
Unweighted regression
Submit
Once you have selected exposures, outcomes, and analysis options you are ready to perform the analysis.
Perform MR analysis
After setting up the analysis scheme, click here to submit the request to perform the MR analysis

MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
  - MR Results
- Quick SNP lookup

Results appear after the analysis is done

Select analysis

Exposure

Age at menopause || id:1004

Outcome

Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis) || id:1126

Generate HTML report

Exposure details

Name: Age at menopause
ID: 1004
Number of instruments used: 42
Units: years
Number of cases: NaN
Number of controls: NaN
Sample size: 69360
PubmedID: 26414677
First author: Day
Consortium: ReproGen
Year: 2015

Outcome details

Name: Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)
ID: 1126
SNPs in GWAS: 10680257
Number of instruments identified: 35
... of which are LD proxies: 0
Units: log odds
Number of cases: 122977
Number of controls: 105974
Sample size: 228951
PubmedID: 29059683
First author: Michailidou K
Consortium: BCAC
Year: 2017

# of instruments were found in the outcome GWAS, which were used in the MR analysis.

Downloads for all analyses

Download harmonised summary statistics

Download MR results

Download leave-one-out sensitivity analysis

Download single SNP MR results

Download the generated datasets or MR analysis results here.

MR results | Heterogeneity statistics | Causal direction test | Horizontal pleiotropy

Tables

This table shows the MR estimates from each method of the causal effect of the exposure on the outcome. The effects are reported in the units that were used to estimate the SNP effects.

method	nsnp	b	se	pval
MR Egger	35	0.06926	0.02329	0.005452
Weighted median	35	0.05319	0.01036	2.815e-7
Inverse variance weighted	35	0.04993	0.01036	0.000001446
Weighted mode	35	0.06599	0.01204	0.000004092

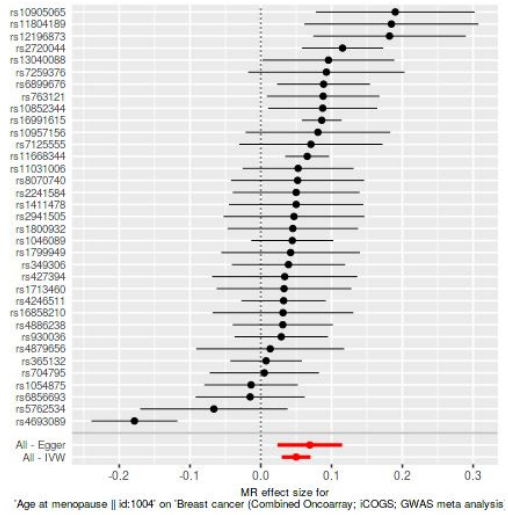
Test statistics corresponding to MR analysis approaches selected.

Single SNP analysis | Method comparison plot | Leave-one-out analysis | Funnel plot

Graphs

The causal effect of exposure on outcome is estimated using each SNP singly using the Wald ratio, and represented in a forest plot. The MR estimate using all SNPs using the MR Egger and IVW methods are also shown. Formal estimates of heterogeneity are shown in the tables below.

Download PDF of this graph

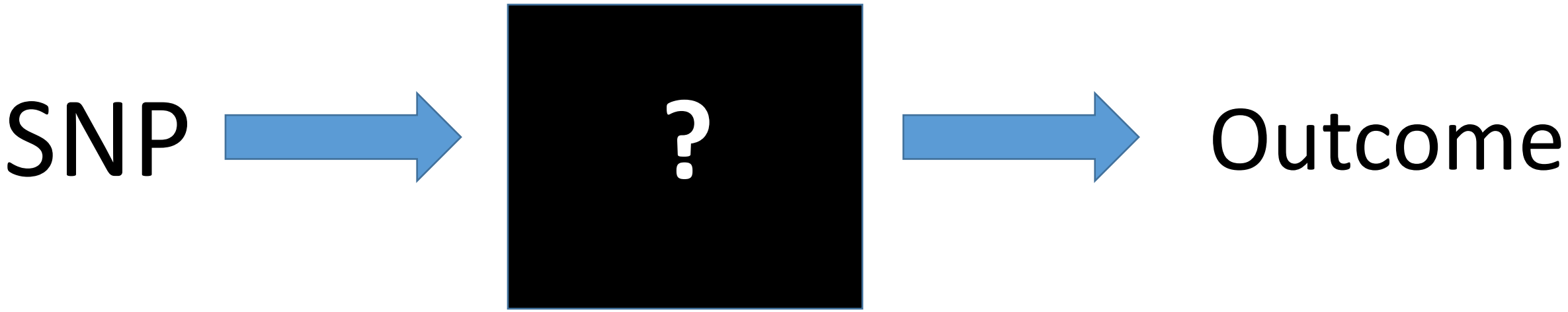


Causal effect of exposure on outcome, by SNP

Exercise

- Explore MR-Base (http://www.mrbase.org) to conduct your own MR study.
- Run an MR study of age at menopause and breast cancer risk following the example we showed in class.
- Check out UK Biobank (https://www.ukbiobank.ac.uk), an amazing resource for genetic epidemiological studies (~500,000 participants with GWAS and comprehensive phenotype data) .

Transcriptome-wide association analysis (TWAS)

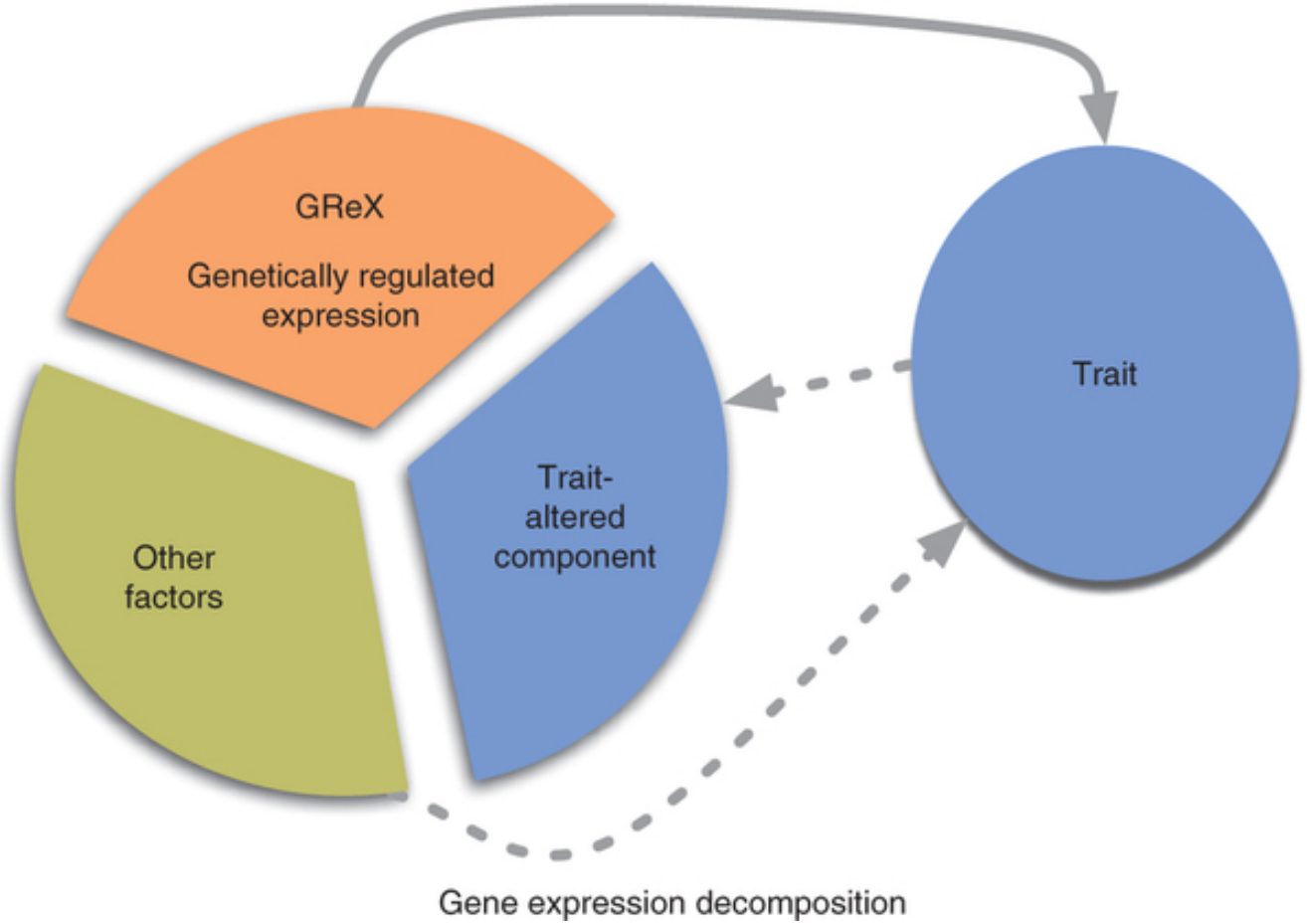


Transcriptome-wide association analysis (TWAS)

Transcriptome-wide association analysis (TWAS)

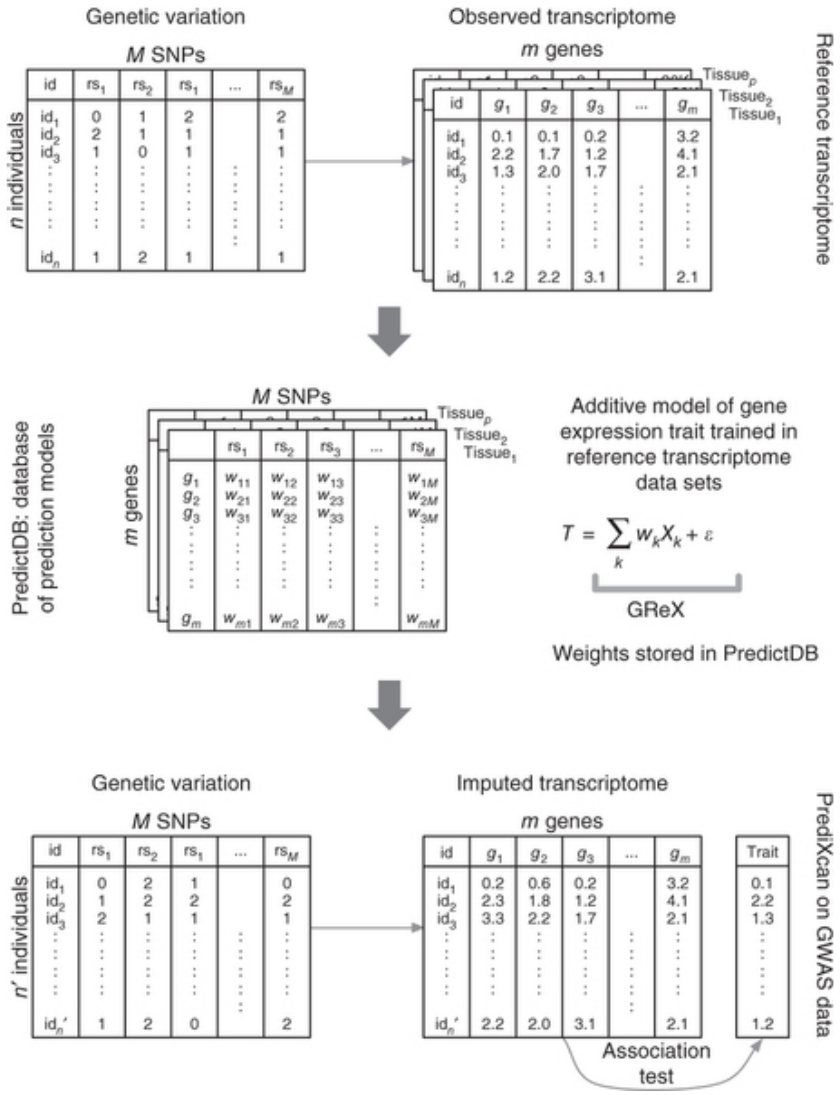
Studies examining the association between gene expression and outcome face multiple issues including reverse causation, confounding, access to specimens and cost of generating genome-wide gene expression profiles

Transcriptome-wide association analysis (TWAS)



Gamazon, Nat Genet 2015

TWAS using individual-level data (PrediXcan)



Gamazon, Nat Genet 2015

TWAS methods

Include only SNPs within 1 Mb of the gene start or end

Range of approaches for model building:

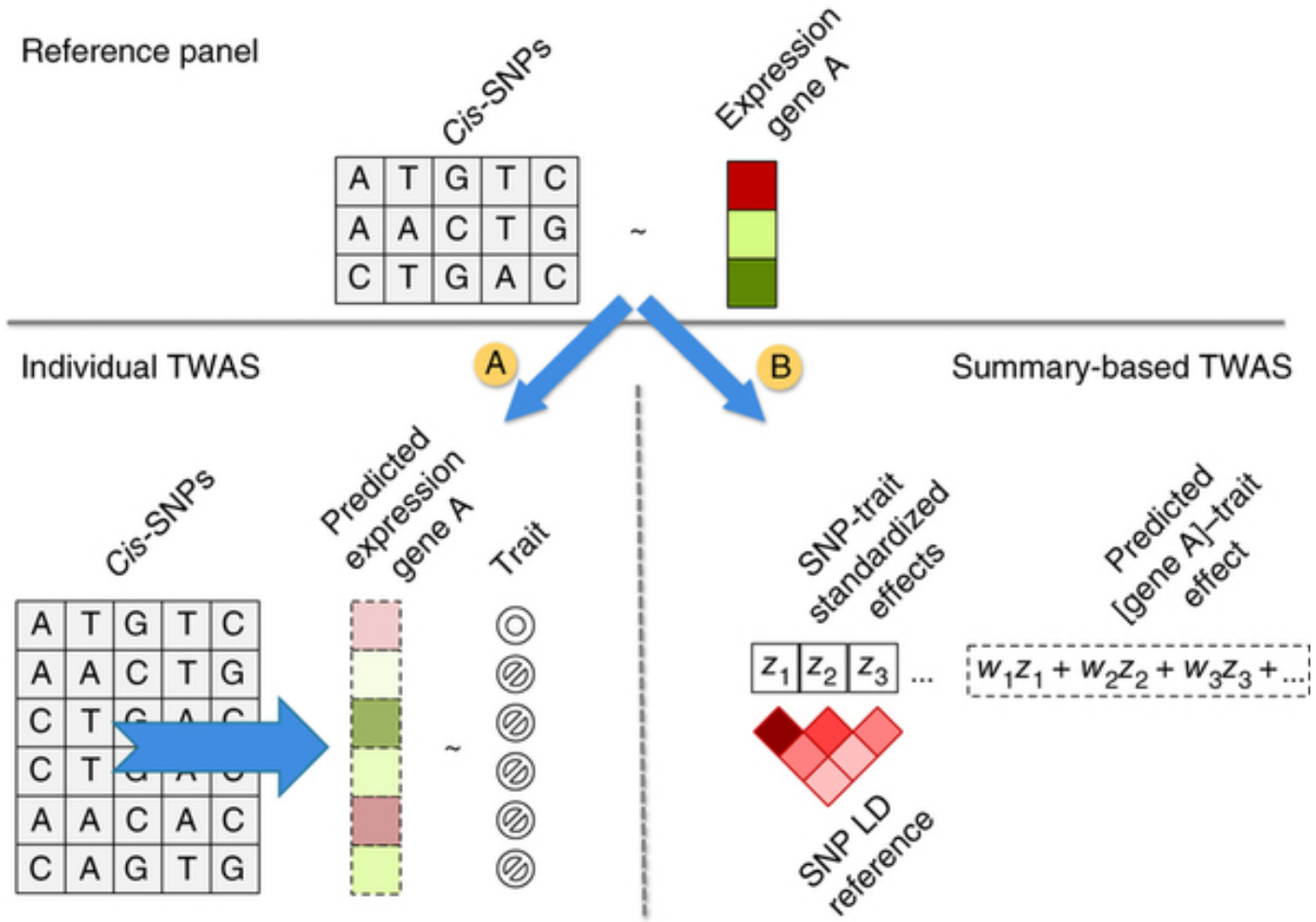
Lasso

Elastic Net

Best eQTL SNP

BLUP

BLSMM

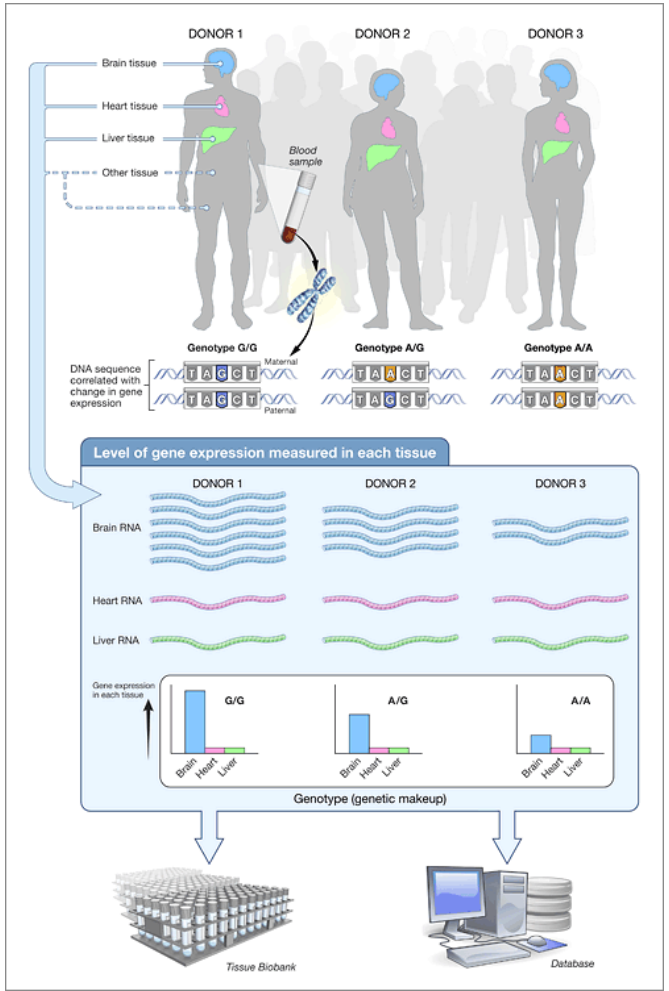


Gusev, Nat Genet 2016

The Genotype-Tissue Expression (GTEx) project

The GTEx Consortium*

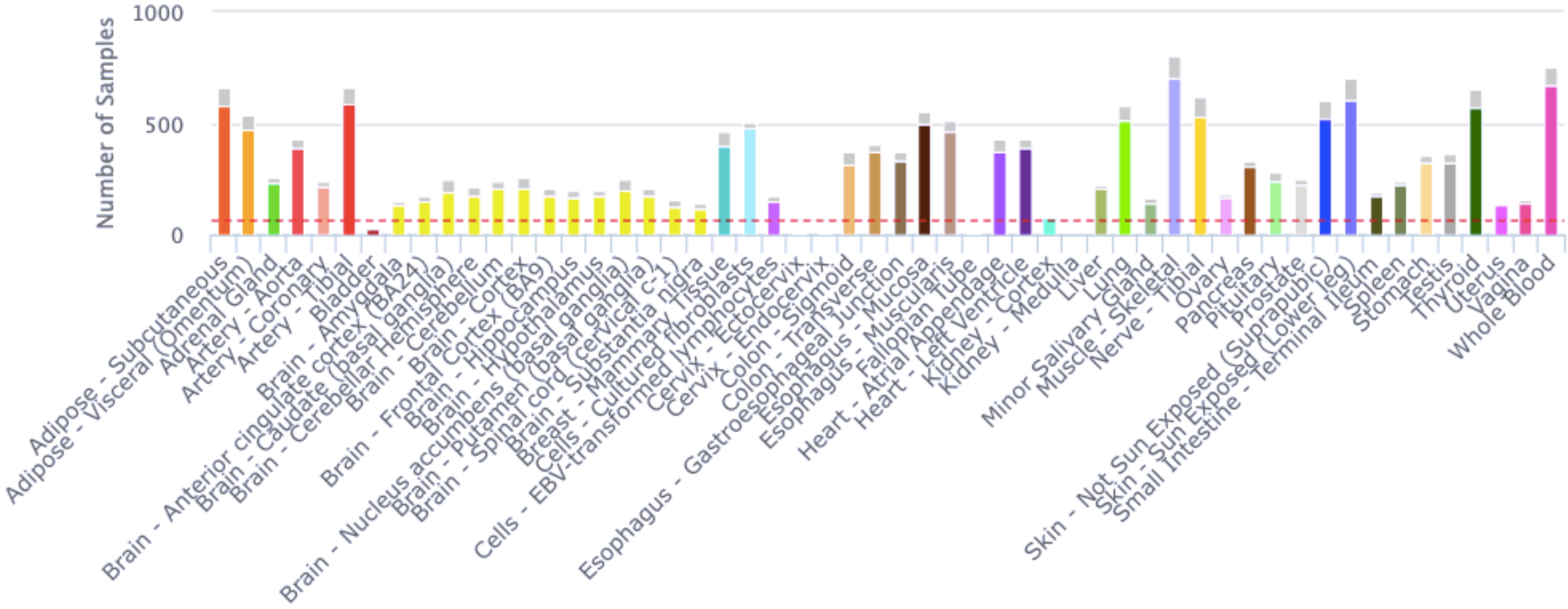
Genome-wide association studies have identified thousands of loci for common diseases, but, for the majority of these, the mechanisms underlying disease susceptibility remain unknown. Most associated variants are not correlated with protein-coding changes, suggesting that polymorphisms in regulatory regions probably contribute to many disease phenotypes. Here we describe the Genotype-Tissue Expression (GTEx) project, which will establish a resource database and associated tissue bank for the scientific community to study the relationship between genetic variation and gene expression in human tissues.



GTEx Consortium, Nat Genet, 2013

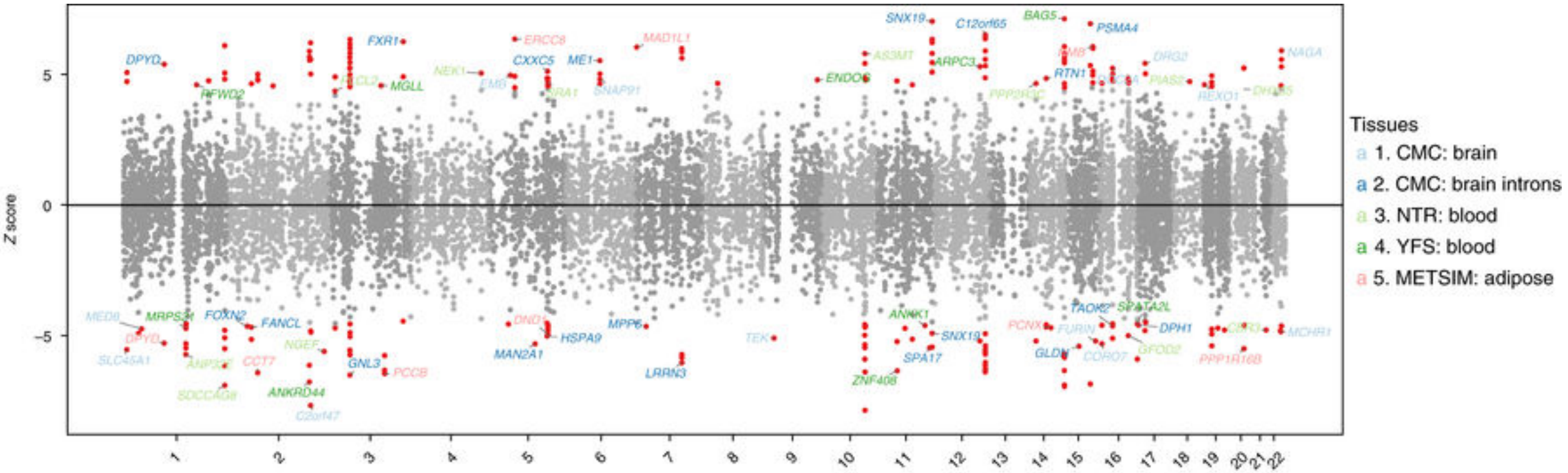
GTEx (https://gtexportal.org/home/)

Tissues available in GTex



https://www.gtexportal.org/home/

TWAS of schizophrenia across 5 tissues



10,819 genes
34,300 cases and 45,511 controls

157 significant TWAS signals of which 35 did not overlap with previous GWAS signals.

Gusev, Nat Genet 2018

Module evaluation

- Please complete the online evaluation available by logging into your SISG account.

- After you complete the evaluation, you will be able to download your Certificate of Completion through the account.

- http://uwsurvey.qualtrics.com/jfe/form/SV_6A3uZcCSyPxMwfz