

Session 12: Mendelian Randomization



Drawback of observational studies





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 Boehm & Zhou, Comp Str Biotech J 2022



Possible effects of C-reactive protein (CRP) concentrations 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.



SCHOOL OF PUBLIC HEALTH

The three key assumptions in MR analyses

- **1. Relevance assumption**: G (SNP or a combination of SNPs) is robustly associated with X (exposure)
- 2. Independence assumption: 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. Exclusivity assumption**: G is related to Y only through its association with X—i.e., G is not associated with Y either directly or indirectly through other traits (i.e., no pleiotropy)

Pleiotropy: The potential for genetic variants to associate with multiple phenotypes MR Dictionary: https://mr-dictionary.mrcieu.ac.uk/





Assumption 1: Relevance assumption

> A "weak" instrument variable (IV) has been defined as having F<10, where

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

R² 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.





Assumption 2: Independence assumption

- > G is independent of factors (measured and unmeasured) that confound the X-Y relationship
- > Since G is randomized at birth and thus, independent of non-genetic confounders and not modified by the course of disease, the main concern is population stratification; i.e., if ancestry is related to G and Y.
- > If you have individual-level data, you can adjust for this (e.g., PCs)





Assumption 3: Exclusivity assumption

- > 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).





Scenarios invalidating assumption 3





BREAKOUT ACTIVITY

- > In which examples (a-f) below do the MR assumptions not hold for assessing the association between exposure (X) and outcome (Y)? Why? Why not?
- genetic a. → exposure → mediator − → outcome variants b. genetic → mediator → exposure − \rightarrow outcome variants с. genetic → exposure – → outcome Ψ variants related variable d. genetic \longrightarrow related variable \longrightarrow → outcome V variants exposure e. genetic ──→ outcome -→ exposure variants



Burgess, Wellcome Open Res 2019

BREAKOUT ACTIVITY

> In which examples (a-f) below do the MR assumptions not hold for assessing the association between exposure (X) and outcome (Y)? Why? Why not?



Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies¹

Philip C Haycock,²* Stephen Burgess,³ Kaitlin H Wade,² Jack Bowden,^{2,4} Caroline Relton,² and George Davey Smith²

Different design strategies for MR ⁴									
Study design	Test	Comments							
G-X + G-Y	Implies $X \rightarrow Y$	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	$X \rightarrow Y$ and $Y \rightarrow X$	Assesses causation in both directions							
Two-step MR	$X \rightarrow M \rightarrow Y$	Tests mediation in a causal pathway							
G×E	$X \rightarrow Y$	Able to detect direct effects (a violation of assumption							
	(relation is dependent on environment variable)	2 of MR)							

¹G×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

TABLE 2

One-sample MR

- > 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 IVs cause bias towards the confounded X-Y association



Two-sample MR

> GWAS summary statistics for G-X and the G-Y associations are estimated in two non-overlapping samples.

> Assumes the two populations are similar (ancestry, age, etc.)

- > Weak IVs cause bias towards the null
- > Note: The G-X and G-Y associations need to be coded using the same effect allele



Two-sample MR

Example IV criteria (can vary) based on G-X

- Genome-wide significant variants (P<5x10⁻⁸)
- Independent (e.g., r²<0.10)
- Common (MAF>1%, if sufficiently large N)
- Exclude palindromic variants (e.g., A/T, G/C) if MAF~50%
- Exclude incompatible alleles between G-X and G-Y (e.g., if a variant has A/G alleles for exposure but A/C for outcome

Extract these variants from G-X **and** G-Y GWAS summary statistics to conduct MR



Two-sample MR: Inverse-variance weighted (IVW) method to estimate the causal effect of X on Y

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

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

 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, $\sigma_{\beta_{2k}}^{-2}$ is the inverse variance for the G-Y association.



Hemani et al., eLife 2018

MR-base: An easy tool for Mendelian Randomization Analysis

- > <u>http://app.mrbase.org/</u>
- > A web-based platform (MR-Base) and an R-package "TwoSampleMR'.
- > Has catalogued thousands of genotype-phenotype associations and also allows manual file upload.





MR BASE

Welcome to MR Base

Acknowledgements

Data access agreement

TwoSampleMR R package

hongjie.chen41@gmail.com

Rerform MR analysis

Q Quick SNP lookup

Logged in as Hongjie Chen

i About





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 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/

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







Acknowledgements

Welcome to MR Base

i About

Data access agreement

TwoSampleMR R package

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

🗱 Perform MR analysis

account.

Current status

App version:

R version: 3.5.1 Host: e4ec2116cb55

0.4.18

Beta phase release

1.2.2 3a435d (31 January 2019)

R/TwoSampleMR version:

https://github.com/MRCIEU/TwoSampleMR

To use MR-Base using the TwoSampleMR R package directly:

See our sister website LD Hub for automated LD score regression:

Database version: 0.2.0 (17 December 2017)

http://ldsc.broadinstitute.org/

幸 Choose exposures
 幸 Choose outcomes

≢ Run MR

Q Quick SNP lookup

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

studies

Select the exposure (Instrumental variable),

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

✤ Get started

outcome and analysis scheme here.

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

Choosing instruments for the exposure

Choose instruments

Select exposure source

MR Base GWAS catalog O Gene expression QTLs

O Protein level QTLs

O Manual file uploa O NHGRI-EBI GWAS catalog

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.

Welcome to MR Base i About

MR BASE

Acknowledgements

Data access agreement

TwoSampleMR R package

Logged in as **Hongjie Chen**

hongjie.chen41@gmail.c C Perform MR analysis

Choose exposures

Choose outcomes

🚊 Run MR

Q Quick SNP lookup

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.

Can either use the instruments provided by MR-base, O Metabolite level QTLs Or use the manually uploaded file. O Methylation level QTLs

	LD Rsq					🦯 Sel	ection c	riteria c	of SNPs							
	0.001															
	Clumping	g distance (kb)														
	10000															
	G First	t author			Sample size			Access		🗹 s	d					
	🗹 Con	sortium			Number of variants			Category		🗹 S	ex					
	🗹 Nun	nber of cases			Year			Population		🗹 S	ubcategory					
	🗹 Nun	nber of controls			PubmedID			Priority		U	nit					
ries															Search:	
l1 l1 Trait	Note	First author	Consortium	1 Number of case	of Number of es controls	Sample size	Number of variants	Year	PubmedID Access	Category	11 It Population	Priority	sd	↓† Sex	Subcategory	l li Unit
Adiponectin		Dastani Z	ADIPOGen			39883	2675209	2012	22479202 public	Risk factor	Mixed	1	0.57	Males and females	Protein	ln(mg/dL)
Crohn's disease		Jostins L	IIBDGC	14	763 15977	30740	13898	2012	23128233 public	Disease	European	1		Males and females	Autoimmune / inflammatory	log odds
Hip circumference	Adjusted for BMI	Randall JC	GIANT			60586	2725796	2013	23754948 public	Risk factor	European	15	8.4548	Males	Anthropometric	SD (cm)
Depressive symptoms		Okbay	SSGAC			161460	6524475	2016	27089181 public	Risk factor	European	1		Males and females	Psychiatric / neurological	SD
Years of schooling		Okbay	SSGAC			293723	8146841	2016	27225129 public	Risk factor	European	1	3.71	Males and females	Education	SD (years)
Leptin	Adjusted for BMI; effect all frequencies are missing	ele Kilpelainen				32161	2474010	2016	26833098 public	Risk factor	European	1		Males and females	Hormone	log ng/ml
Leptin	Effect allele frequencies ar missing	re Kilpelainen				32161	2473865	2016	26833098 public	Risk factor	European	2		Males and females	Hormone	log ng/ml
Age at menopause		Day	ReproGen			69360	2418696	2015	26414677 public	Risk factor	European	1	3.93	Females	Reproductive aging	years
Percent emphysema		Manichaikul A	MESA	Specify	the litera	ture 7667	2968584	2014	24383474 public	Risk factor	Mixed	1	0.708	Males and females	Lung disease	log (% emphysema + 1)
Mean platelet volume	Effect allele frequencies ar missing	re Gieger C	HaemGen	to be in	cluded h	ere 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
	ries Trait Adiponectin Crohn's disease Hip circumference Depressive symptoms Years of schooling Leptin Leptin Leptin Age at menopause Percent emphysema Mean platelet volume	LD Rsq 0.001 Clumpin 10000 Firs Con Nur Nur ries II II II Trait Note Adjonectin Crohn's disease Hip Adjusted for BMI circumference Depressive symptoms Vears of schooling Leptin Adjusted for BMI; effect all frequencies are missing Leptin Effect allele frequencies ar missing Age at menopause Percent emphysema Mean platelet volume All All	LD Rsq 0.001 Clumping distance (kb) 10000 I First author Consortium Number of cases Vumber of cases Vummer Adjusted for BMI Randall JC Okbay Schooling Leptin Adjusted for BMI; effect allele Kilpelainen frequencies are missing Vears of Schooling Leptin Adjusted for BMI; effect allele Kilpelainen frequencies are missing Percent emphysema Mean platelet Volume All	LD Rsq 0.001 Clumping distance (kb) 10000 First author Consortium Number of cases Number of controls ries Image: Trait imag	LD Rsq 0.001 Clumping distance (kb) 10000 Clumping distance (kb) 10000 First author Consortium Number of cases Number of cases Numbe	LD Rsq 0.001 Clumping distance (kb) 10000 10000 10000 2 First author 3 Sample size Consortium Number of cases 11 Mumber of cases 12 Number of controls 13 Number of controls 14 Number of [] Number of [] Number of [] 15 Adiponectin Dastani Z Adiponectin Dastani Z Adiponectin Depressive symptoms Okbay SSGAC Symptoms Okbay SSGAC Symptoms Okbay SSGAC Symptoms Okbay SSGAC Symptoms Okbay SSGAC Symptoms Okbay SSGAC Symptoms Okbay SSGAC Symptoms Okbay SSGAC Okbay SSGAC Okbay Strait Illelainen missing Age at menopause Day ReproGen	ID Rsq Seli 0.001 Clumping distance (kb) 10000 Sample size Sample size 0.001 First author I Sample size Number of variants Year Image: Consortium Image: Consortium Number of variants Year PubmediD Trait Note First author Consortium Consortium Sample size Adiponectin Dastani Z ADIPOGen 39883 Crohn's disease Jostins L IIBDGC 14763 15977 30740 Hip Adjusted for BMI Randall JC GIANT 60596 161460 Symptoms Okbay SSGAC 161460 293723 Leptin Adjusted for BMI; effect allele frequencies are missing Kilpelainen 32161 Leptin Effect allele frequencies are missing Day ReproGen 69300 Percent Manichaikul A MESA Specify the literature 7667 Mean platelet Effect allele frequencies are missing Gieger C HaemGen to be included here 66657	Selection c Clumping distance (kb) 10000 First author Consortium Number of sarants Sample size Number of sarants Sample size Number of sarants Sample size Samp	Selection criteria colspan="2">Selection criteria colspan="2" Cumping distance (kb) Solection criteria colspan="2" Selection criteria colspan="2" Selection criteria colspan="2" Cumping distance (kb) Solection criteria colspan="2" Selection criteria colspan="2" Selection criteria colspan="2" Cumping distance (kb) Selection criteria colspan="2" Selection criteria colspan="2" Selection criteria colspan="2" Cumping distance (kb) Selection criteria colspan="2" Toth colspan="2" Number of variants Number of variants <td>Selection criteria of SNPs Clumping distance (kb) 10000 9 Sample size 9 Access © Consortium 9 Sample size 9 Access © Consortium 9 Number of variants 9 Category Number of cases 9 Pubmedio 9 Poulution res 11 Number of 1 11 Mode First author Consortium 9 Poulution res 11 Number of 1 11 Number of 1 Adponectin Daxan Z ADPOGen 39853 3675209 2012 2247202 public Crohn's disease Jostan Z ADPOGen 39855 327576 2013 2375486 public Crohn's disease Jostan Z ADPOGen 5956 3275796 2013 23751846 public Pubmedio 6441 GiaAT 69556 3275796 2015 23708181 public Schooling Adjusted for BMI Randell JC GiAAT 69556 3275796 2015 22708181 public Leptin Adjusted for BMI Randell JC GiAAT 59562 2015 2433306 public</td> <td>Selection criteria of SNPs 0.001</td> <td>Image: Selection criteria of SNPs 0.001 0.002 0.001 0.002 0.002 0.001 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.0010 0.002 0.0010 0.0</td> <td>Selection criteria of SNPs Selection criteria of SNPs</td> <td>Image: Selection criteria of SNPs Image: Selection criteria of SNPs</td> <td>Selection criteria of SNPs Selection criteria of SNPs</td> <td>Selection criteria of SNPs Selection criteria of SNPs</td>	Selection criteria of SNPs Clumping distance (kb) 10000 9 Sample size 9 Access © Consortium 9 Sample size 9 Access © Consortium 9 Number of variants 9 Category Number of cases 9 Pubmedio 9 Poulution res 11 Number of 1 11 Mode First author Consortium 9 Poulution res 11 Number of 1 11 Number of 1 Adponectin Daxan Z ADPOGen 39853 3675209 2012 2247202 public Crohn's disease Jostan Z ADPOGen 39855 327576 2013 2375486 public Crohn's disease Jostan Z ADPOGen 5956 3275796 2013 23751846 public Pubmedio 6441 GiaAT 69556 3275796 2015 23708181 public Schooling Adjusted for BMI Randell JC GiAAT 69556 3275796 2015 22708181 public Leptin Adjusted for BMI Randell JC GiAAT 59562 2015 2433306 public	Selection criteria of SNPs 0.001	Image: Selection criteria of SNPs 0.001 0.002 0.001 0.002 0.002 0.001 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.002 0.0010 0.002 0.0010 0.0	Selection criteria of SNPs Selection criteria of SNPs	Image: Selection criteria of SNPs Image: Selection criteria of SNPs	Selection criteria of SNPs Selection criteria of SNPs	Selection criteria of SNPs Selection criteria of SNPs

Choosing instruments for the exposure

MR BASE

Welco
Abou
Ackno
Data

Logged i Hongjie hongjie Choo

Q Quick

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.

Ch	Choose instruments				MR Base GWAS catalog													
Sel O	Select exposure source O Manual file upload				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.													
0	NHGRI-EBI GWAS catalog MR Base GWAS catalog				To use a tra in the MR a	ait as an exposure, analysis.	nighlight the relevant ro	w in the table below	multiple traits can	be selected). All SNP	's with p values below the spe	ecified threshold will be	extracted, and clumpi	ing will be used to remo	ve SNPs in LD with sentina	il SNPs. These SNPs will be	used as instruments	
0	Gene expre	ession QTLs																
0	Metabolite	level QTLs																
0	Methylatior	n level QTLs																
p-v	alue thresh	hold		LD Rsq														
3	e-06			0.001														
	Perform clu	umping		Clumping dis	tance (kb)													
D'e					baan													
	ID	ins		Consorti	ium			 Sample size Number of variants 			Category		So Se	x				
	Trait			Number	of cases			Vear			Population		Su	bcategory				
	Note			Number	of controls			PubmedID			Priority		🗹 Un	lit				
																	Search	
Sho	0W 10	✓ entries															ocarem	
Sho	10 ID	↓† Trait	11 Note	11	First author	Consortium	1 Number cas	of Number of es controls	Sample size	Number of variants	11 Year	PubmedID Access	11 Category	11 11 Population	11 Priority	Sd Sex	11 Subcategory	Lt It
Sho 1	10 10 1	Trait Adiponectin	11 Note	11	First author Dastani Z	Consortium ADIPOGen	1 Number cas	of Number of tes	I† Sample size 39883	Number of variants 2675209	Vear 2012	PubmedID Access 22479202 public	Category Risk factor	L1 L1 Population Mixed	Priority 1	Sd Sex 0.57 Males an females	d Protein	Lt Lt Unit In(mg/dL)
Sho 1 10	ID 1 10	Trait Adiponectin Crohn's disea	11 Note	11	First author Dastani Z Jostins L	Consortium ADIPOGen IIBDGC	Lî Number cas	of Number of controls	11 Sample size 39883 30740	Number of 1 variants 2675209 13898	11 Year 2012 2012	PubmedID Access 22479202 public 23128233 public	Lit Category Risk factor Disease	Lit Population Lit Mixed European	Priority 1	Sd Sex 0.57 Males an females Males an females	id Protein d Autoimmune / inflammatory	Lit Unit Lit In(mg/dL) log odds
Sho 1 10	ID 1 10 0 100	entries If Trait Adiponectin Crohn's dises Hip circumference	Note Note	11	First author Dastani Z Jostins L Randall JC	Consortium ADIPOGen IIBDGC GIANT	1 Number cas	of Number of the controls	Sample size 39883 30740 60586	Number of variants 2675209 13898 2725796	11 Year 2012 2012 2013	PubmedID Access 22479202 public 23128233 public 23754948 public	Category Risk factor Disease Risk factor	Image: Population Image: Population Mixed European European European	Priority 1 1 15	Sd Sex 0.57 Males an females Males an females 8.4548 Males	Id Protein Id Autoimmune / Inflammatory Anthropometric	It Unit In(mg/dL) log odds c SD (cm)
Sha 1 10 100	ID ID 1 10 00 100 00 1000	 entries Trait Adiponectin Crohn's disea Hip circumference Depressive symptoms 	Adjusted for BN	Jî 1	First author Dastani Z Jostins L Randall JC Okbay	Consortium ADIPOGen IIBDGC GIANT SSGAC	If Number cas	of Number of es controls 4763 15977 he instrur	Sample size 39883 30740 60586 ments1460	Number of ************************************	11 Year 2012 2012 2013 2016	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public	It Category Risk factor Disease Disk factor Risk factor Risk factor Risk factor	It Population It Mixed European European European European European	Priority 1 15 1	Sd Sex 0.57 Males an females Males an females 8.4548 Males Males an females	id Protein id Autoimmune / inflammatory Anthropometric d Psychiatric / neurological	Lt Unit Lt In(mg/dL) log odds c SD (cm) SD
Sho 1 10 100 100	ID ID ID 1 1 0 0 100 00 1000 01 1001	It Trait Adiponectin Crohn's disea Hip circumferenc Depressive symptoms Years of schooling	Adjusted for BN	11	First author Dastani Z Jostins L Randall JC Okbay Okbay	Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC	If Number cas	Number of controls	11 Sample size 39883 30740 60586 ments1460 293723	Number of variants 2675209 13898 2725796 6524475 8146841	11 Year 2012 2013 2016 2016	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public	It Category Risk factor Disease Risk factor Risk factor Risk factor	It Population It Mixed European European European	Priority 1 1 15 1 1 1	sd Sex 0.57 Males an females Males an females 8.4548 Males Males an females 3.71 Males an females	It Subcategory nd Protein nd Autoimmune / inflammatory Anthropometric d Psychiatric / neurological d Education	In (mg/dL) log odds SD (cm) SD (years)
Sho 1 10 100 100 100	ID ID ID 1 10 10 00 1000 00 1000 01 1001 02 1002	It Trait Adiponectin Crohn's disea Hip circumference Depressive symptoms Years of schooling Leptin	Adjusted for BN frequencies are	LT II II: effect allele e missing	First author Dastani Z Jostins L Randall JC Okbay Okbay	Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC	It Number cas	es Number of (controls 1763 15977 he instrur cluded MR analy	11 Sample size 39883 30740 60586 ments1460 293723 293723 22515 32161	Number of 1 variants 2675209 13898 2725796 6524475 8146841 2474010	Vear 2012 2012 2013 2016 2016 2016	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public 26833098 public	It Category Risk factor Disease Risk factor Risk factor Risk factor Risk factor	It Population It Mixed European It European European It European European It	Priority 1 1 15 1 1 1 1	sd Sex 0.57 Males an females Males an females 8.4548 Males Males an females 3.71 Males an females Males an females	It Subcategory It Subcategory Ind Protein Id Autoimmune / Inflammatory Anthropometric Id Psychiatric / neurological Id Education Id Hormone	In Unit In(mg/dL) log odds SD (cm) SD SD (years) log ng/ml
Sho 1 10 100 100 100 100	IO IO ID I I I IO IOO O IOO OO IOOO	It Trait Adiponectin Crohn's disea Hip circumference Depressive symptoms Years of schooling Leptin	Adjusted for BA Adjusted for BA Frequencies are Effect allele fre missing	LT III III: III: III: III: III: III: III	First author Dastani Z Jostins L Randall JC Okbay Okbay Kilpelainen	Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC	I Number cas	es Number of (controls 15977 he instrur cluded MR analy	11 Sample size 39883 30740 60586 ments1460 293723 293723 293723 293723 293723 293723 293723	Number of 1 variants 2675209 13898 2725796 6524475 8146841 2474010 2473865	11 Year 2012 2013 2016 2016 2016 2016	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public 26833098 public	It Category Risk factor Disease Risk factor Risk factor Risk factor Risk factor Risk factor Risk factor Risk factor Risk factor	It Population It Mixed It European It	Priority 1 1 15 1 1 1 1 2	sd Sex 0.57 Males an females Males an females 8.4548 Males an females 3.71 Males an females Males an females Males an females Males an females Males an females	It Subcategory It Subcategory Ind Protein Ind Autoimmune / Inflammatory Anthropometriv Ind Psychiatric / neurological Ind Education Ind Hormone Ind Hormone	It Unit It In(mg/dL) Iog odds It Iog odds SD (cm) SD SD SD (years) It Iog ng/ml It It
Sho 1 10 100 100 100 100 100	IO IO ID I I I IO IOO OI IOO IO IOOI IO IOOI IO IOOI IO IOOI IO IOOI	 entries If Trait Adiponectin Crohn's disea Crohn's disea Hip circumference Depressive symptoms Years of schooling Leptin Leptin Age at menopause 	It Note See Adjusted for BN frequencies are Effect allele fre missing	Lî 11 11: effect allele missing quencies are	First author Dastani Z Jostins L Randall JC Gkbay Gkbay Kilpelainen Kilpelainen	Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC SSGAC	I Number cas	of: Number of: controls 15977 he instrur cluded MR analy	11 Sample size 39883 30740 60586 ments1460 293723 293723 293723 32161 32161 69360	Number of 1 variants 2675209 13898 2725796 6524475 8146841 2474010 2473865 2418696	11 Year 2012 2013 2016 2016 2016 2016 2015	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public 26833098 public 26833098 public 26414677 public	It Category Risk factor Disease Risk factor Risk factor	It Population It Mixed It European It	Priority 1 1 15 1 1 1 1 2 2	Sd Sex 0.57 Males an females 8.4548 Males an females 3.71 Males an females 3.71 Males an females Males an females 3.71 Males an females Males an females 3.71 Males an females Males an females 3.71 Males an females 3.71 Males an females Males an females 3.71 Males an females Males an females 3.71 Males an females 3.71 Males an females Males an females 3.71 Males an females 3.71 Males an females 3.71 Males an females 3.71 Males an females 3.71 Males an females Males an females 3.71 Males an females Males an females 3.93 Females	It Subcategory It Subcategory Ind Protein Ind Autoimmune / Inflammatory Anthropometriv Ind Psychiatric / neurological Ind Education Ind Hormone Ind Hormone Reproductive aging	It Unit In(mg/dL) log odds SD (cm) SD SD (years) log ng/ml log ng/ml years
Sho 1 10 100 100 100 100 100 100	IO IO ID I I I IO IO IO IOO IO IOOI IO IOOI <td> entries It Trait Adiponectin Crohn's disea Hip circumference Depressive symptoms Years of schooling Leptin Leptin Age at menopause Percent emphysema </td> <td>It Note se Adjusted for BM frequencies are Effect allele fre missing</td> <td>Lî 11 11: effect allele missing quencies are</td> <td>First author Dastani Z Jostins L Randall JC Gkbay Gkbay Kilpelainen Kilpelainen Day</td> <td>Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC SSGAC ReproGen MESA</td> <td>I Number cas</td> <td>of: Number of: controls 4763 15977 he instrur cluded MR analy</td> <td>11 Sample size 39883 30740 60586 ments1460 293723 223723 32161 32161 69360 7667</td> <td>Number of 1 variants 2675209 13898 2725796 6524475 8146841 2474010 2473865 2418696 2968584</td> <td>11 Year 2012 2013 2016 2016 2016 2016 2015 2014</td> <td>PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public 26833098 public 26833098 public 26414677 public 24383474 public</td> <td>It Category Risk factor Disease Risk factor Risk factor Risk factor Risk factor</td> <td>It Population It Mixed It European It</td> <td>Priority 1 1 15 1 1 1 1 2 1 1 1 1 1</td> <td>Sd Sex Alles an females 8.4548 Males an females 3.71 Males an females Males an females Males an females 3.93 Females 0.708 Males an females</td> <td>It Subcategory It Subcategory Ind Protein Ind Autoimmune / Inflammatory Anthropometriu Ind Psychiatric / neurological Ind Education Ind Hormone Ind Hormone Ind Hormone Ind Hormone Ind Hormone</td> <td>In unit In In(mg/dL) Iog odds Iog odds Iog SD (cm) Iog SD (years) Iog ng/ml Iog ng/ml Iog ng/ml</td>	 entries It Trait Adiponectin Crohn's disea Hip circumference Depressive symptoms Years of schooling Leptin Leptin Age at menopause Percent emphysema 	It Note se Adjusted for BM frequencies are Effect allele fre missing	Lî 11 11: effect allele missing quencies are	First author Dastani Z Jostins L Randall JC Gkbay Gkbay Kilpelainen Kilpelainen Day	Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC SSGAC ReproGen MESA	I Number cas	of: Number of: controls 4763 15977 he instrur cluded MR analy	11 Sample size 39883 30740 60586 ments1460 293723 223723 32161 32161 69360 7667	Number of 1 variants 2675209 13898 2725796 6524475 8146841 2474010 2473865 2418696 2968584	11 Year 2012 2013 2016 2016 2016 2016 2015 2014	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public 26833098 public 26833098 public 26414677 public 24383474 public	It Category Risk factor Disease Risk factor Risk factor	It Population It Mixed It European It	Priority 1 1 15 1 1 1 1 2 1 1 1 1 1	Sd Sex Alles an females 8.4548 Males an females 3.71 Males an females Males an females Males an females 3.93 Females 0.708 Males an females	It Subcategory It Subcategory Ind Protein Ind Autoimmune / Inflammatory Anthropometriu Ind Psychiatric / neurological Ind Education Ind Hormone Ind Hormone Ind Hormone Ind Hormone Ind Hormone	In unit In In(mg/dL) Iog odds Iog odds Iog SD (cm) Iog SD (years) Iog ng/ml Iog ng/ml Iog ng/ml
Sho 1 10 100 100 100 100 100 100 100	IO IO ID I I I IO IO IO IOO IO IOO IO IOO IO	 entries It Trait Adiponectin Crohn's disea Hip circumference Depressive symptoms Years of schooling Leptin Leptin Leptin Percent emphysema Mean platele volume 	It Note See Adjusted for BM frequencies are Effect allele fre missing Effect allele fre missing	Lt II II: effect allele missing quencies are	First author Dastani Z Jostins L Randall JC Gkbay Gkbay Kilpelainen Kilpelainen Day Day	Consortium ADIPOGen IIBDGC GIANT SSGAC SSGAC SSGAC SSGAC ReproGen MESA HaemGen	I Number cas	of: Number of: controls	11 Sample size 39883 30740 60586 Ments1460 293723 2293723 32161 32161 69360 7667 66867	Number of 1 variants 2675209 13898 2725796 6524475 8146841 2474010 2473865 2418696 2968584 2690859	It Year 2012 2012 2013 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 2016 2017 2014 2011	PubmedID Access 22479202 public 23128233 public 23754948 public 27089181 public 27225129 public 26833098 public 26833098 public 26434677 public 24383474 public 2139419 public	It Category Risk factor Disease Disk factor Risk factor Risk factor Risk factor	It Population It Mixed It European It	Priority 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Sd Sex All Sex All Sex All Sea All	It Subcategory It Subcategory Ind Protein Ind Autoimmune / Inflammatory Anthropometriv Ind Psychiatric / neurological Ind Education Ind Hormone Ind Hormone Ind Hormone Ind Hormone Ind Hormone Ind Hormone Ind Hormone Ind Hormone	In unit In (ung/dL) Iog odds Iog odds SD (cm) SD SD (years) Iog ng/ml Iog ng/ml Iog ng/ml

Showing 1 to 10 of 21,266 entries

Previous 1 2 3 4 5 ... 2127 Next

Select outcomes for analysis

Trait

Note

Show 10 v entries



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.

Welcome to MR Base

	About
1	ADOUL

Acknowledgements

Data access agreement

TwoSampleMR R package

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

🗱 Perform MR analysis

🔁 Choose exposures

🛱 Run MR

Q Quick SNP lookup

Display columns	Image: A start and a start	First author
		Consortium

Number of cases

Number of controls

Choose the outcome of the MR analysis

•	Sample size
	Number of variants
•	Year
	PubmedID

	Access
<	Category
	Population
	Priority

Select the outcome GWAS data to be used

Sd

Sex

Unit 🗌

Subcategory

Search:

Lî Trait Note	↓↑ First author	L1 Consortium	11 It Number of cases	Number of controls	↓↑ Sample size	Number of variants	↓↑ Year Ca	tegory Subcategory
1126 Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)	Michailidou K	BCAC	122977	105974	228951	10680257	2017 Dis	ease Cancer
1127 ER+ Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)	Michailidou K	BCAC	69501	105974	175475	10680257	2017 Dis	ease Cancer
1128 ER- Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)	Michailidou K	BCAC	21468	105974	127442	10680257	2017 Dis	ease Cancer
1129 Breast cancer (Oncoarray)	Michailidou K	BCAC	61282	45494	106776	10680257	2017 Dis	ease Cancer
1130 Breast cancer (iCOGS)	Michailidou K	BCAC	46785	42892	89677	10680257	2017 Dis	ease Cancer
1131 Breast cancer (GWAS)	Michailidou K	BCAC	14910	17588	32498	10680257	2017 Dis	ease Cancer
1132 ER+ Breast cancer (Oncoarray)	Michailidou K	BCAC	38197	45494	83691	10680257	2017 Dis	ease Cancer
1133 ER+ Breast cancer (iCOGS)	Michailidou K	BCAC	27078	42892	69970	10680257	2017 Dis	ease Cancer
1134 ER+ Breast cancer (GWAS)	Michailidou K	BCAC	4226	17588	21814	10680257	2017 Dis	ease Cancer
1135 ER- Breast cancer (Oncoarray)	Michailidou K	BCAC	9655	45494	55149	10680257	2017 Dis	ease Cancer
Breast Cancer 🛞 🔻 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

MRBASE

- Welcome to MR Base
- i About
- Acknowledgements
- Data access agreement
- TwoSampleMR R package
- Logged in as Hongjie Chen hongjie.chen41@gmail.com
- 🗱 Perform MR analysis

- 韋 Run MR
- Q 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



0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41

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

0.46 0.49

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

	Select analysis	MR results	Heterogeneity statistics Causal	direction test Horizon	tal pleiotropy				Tables
MR BASE	Exposure	This table show the causal effec	us the MR estimates from each method o	f method	11	nsnp11	b∥↑	se†	pval
~ <u> </u>	Age at menopadae 10.1004	effects are repo	orted in the units that were used to	MR Egger		35	0.06926	0.02329	0.005452
	Outcome	estimate the SN	VP effects.	Weighted median		35	0.05319	0.01036	2.815e-7
S	Breast cancer (Combined Oncoarray; ICOGS; GWAS meta analysis) Id:1126	Test statistics	corresponding to	Inverse variance we	ighted	35	0.04993	0.01036	0.000001446
Welcome to MR Base	➡ Generate HTML report	MR analysis a	nnroaches selected	Weighted mode	5	35	0.06599	0.01204	0.00004092
About			pproderies selected.						
Acknowledgements	Exposure details	Single SNP and	alysis Method comparison plot	Leave-one-out analysis	Funnel plot				Graphs
Data access agreement	Name: Age at menopause	The second offer							
bata access agreement	ID: 1004	using each SNP	ct of exposure on outcome is estimated	rs10905065 rs11804189 rs12196873					
	Units: vears	represented in	a forest plot. The MR estimate using all	rs2720044 rs13040088 rs7259376					
woSampleMR R package	Number of cases: NaN	SNPs using the	MR Egger and IVW methods are also	rs6899676 rs763121		_			
	Number of controls: NaN	shown. Formal	estimates of heterogeneity are shown in	rs10852344 rs16991615 rs10957156		_			
ogged in as	Sample size: 69360	the tables beto		rs7125555 rs11668344 rs11031006		-			
nonglie.chen41@gmail.com	First author: Day	🛓 Download	PDF of this graph	rs8070740 rs2241584			C		
	Consortium: ReproGen			rs2941505 rs1800932			Causa	l effect of	exposure
🞗 Perform MR analysis 🛛 🔇	Year: 2015			rs1046089 rs1799949 rs349306			on out	tcome, by	/ SNP
き Choose exposures				rs42/394 rs1713460 rs4246511				-	
	Outcome details			rs18858210 rs4886238 rs930036					
≠ Dun MD	Name: Proact concor (Combined Operation (COCC) (WAS mote applying)			rs4879656 rs365132 rs704795					
	ID: 1126			rs1054875 rs6856693 rs5762534					
MR Results	SNPs in GWAS: 10680257			rs4693089					
Q Quick SNP lookup <	Number of instruments identified: 35		to the subscript OM	All - IVW	.01 00 01	0'2 0'3			
	Units: log odds	ts were found	In the outcome GW	AS, 'Age at menopause id:1	MR effect size for 004' on 'Breast cancer (Combined Oncoarray; i	COGS; GWAS meta anal	ysis		
	Number of cases: 122977 which were use	ed in the MR a	inalysis.		· · · · · · · · · · · · · · · · · · ·				
tesults appear	Number of controls: 105974		•		1				
thee the applycie	Sample size: 228951								
nuel une analysis -	First author: Michailidou K		Evidence	of heterogene	ity of association	ons (con	firmed i	n "Hetero	geneity
sidone	Consortium: BCAC								Seriety
	Year: 2017		statistics	tab) suggests	some SNPs exr	non fidir	izontal p	leiotropy	•
			but MP	Eggor is uphis	and avan if hat	orogono	ity accu	motion is	violated
	Downloads for all analyses		but Ivin-	Leger is unbid		erogene	ity assui	inpuonis	violateu,
			and MR-E	gger is signific	ant here.				
	Lownload harmonised summary statistics								
	La Download MR results	d the generate	ed datasets or MR a	nalysis results	here.				
	L Download leave-one-out sensitivity analysis								
	➡ Download single SNP MR results								

BREAKOUT ACTIVITY

> (Explore MR-Base (<u>http://www.mrbase.org</u>) to conduct your own MR study. Run an MR study of smoking pack years and lung cancer risk following the example in class.)



Details about MR-Base: Hemani et al., eLife 2018 (https://elifesciences.org/articles/34408)

Bidirectional MR analysis

- > Approach to overcome reverse causation ($X \leftarrow \rightarrow Y$)
- > IVs for both X and Y are used to assess the causal association in both directions
- 1. Is G_X associated with Y?
- 2. Is G_Y associated with X?

(Also confirm that G_X is associated with X and that G_Y is associated with Y





BMI and CRP – what causes what?

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



Gray: 95% CI for CRP locus rs3091244 as IV for CRP.



Blue dots 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 **EPIDEMIOLOGY** SCHOOL OF PUBLIC HEALTH The observed association between circulating CRP and measured BMI is suggested to be driven by BMI, with adiposity causally influencing circulating CRP levels and not vice-versa

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

 Previous table 		 Figures and tables index 								
		Effec								
Outcome /ex	Outcome /explanatory variable		Observational Instrumental variable			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				



Considerations for MR analyses

- > 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.
- Consider the three core assumptions and how they apply in your study
 In practice, this is very difficult, especially for the third assumption of no pleiotropy.
- > Look for consistency across MR approaches
 - Tells you how robust your results are given the different assumptions
- > In the end, it is a helpful tool to complement observational findings
 - You can never assume causality because you can never be 100% sure that all assumptions are met



 Table 1. Summary of some methods proposed for Mendelian randomization: inverse-variance weighted method and robust methods.

Method	Consistency assumption	Strengths and weaknesses	Reference	Software
Inverse-variance weighted	All variants valid or balanced pleiotropy	Most efficient (greatest statistical power), biased if average pleiotropic effect differs from zero	18	*†
MR-Egger	InSIDE	Sensitive to outliers, sensitive to violations of InSIDE assumption, InSIDE assumption often not plausible, often less efficient	19	*†
Weighted median	Majority valid	Robust to outliers, sensitive to addition/removal of genetic variants	20	*†
Mode-based estimation	Plurality valid	Robust to outliers, sensitive to bandwidth parameter and addition/ removal of genetic variants, generally conservative	21	*†
MR-PRESSO	Outlier-robust	Removes outliers, efficient with valid IVs, very high false positive rate with several invalid IVs	22	+
MR-Robust	Outlier-robust	Downweights outliers, efficient with valid IVs, high false positive rate with several invalid IVs	23	*
MR-Lasso	Outlier-robust	Removes outliers, efficient with valid IVs, high false positive rate with several invalid IVs	23	
MR-RAPS	Balanced pleiotropy (except outliers)	Downweights outliers, sensitive to violations of balanced pleiotropy assumption	24	‡
Contamination Mixture	Plurality valid	Robust to outliers, sensitive to variance parameter and addition/ removal of genetic variants	25	*
MR-Mix	Plurality valid	Robust to outliers, requires large numbers of genetic variants, very high false positive rate in several scenarios	26	‡

Each of the methods in the table can be implemented using summarized data. False positive rates refer to the simulation study by Slob and Burgess²⁷. InSIDE is the Instrument Strength Independent of Direct Effect assumption.



Mendelian Randomization in R

- > Encodes several methods for performing Mendelian randomization analyses with summarized data. Summarized data on genetic associations with the exposure and with the outcome can be obtained from large consortia. These data can be used for obtaining causal estimates using instrumental variable methods.
- > <u>https://cran.r-project.org/web/packages/MendelianRandomization/index.html</u>
- > <u>https://www.youtube.com/channel/UCHjMrVSqOu1rcrYQPAD_bNA</u>

