

# Multi-Instrument Mendelian Randomisation

*from genetic associations to generating hypotheses  
about causality*

Baptiste Couvy-Duchesne

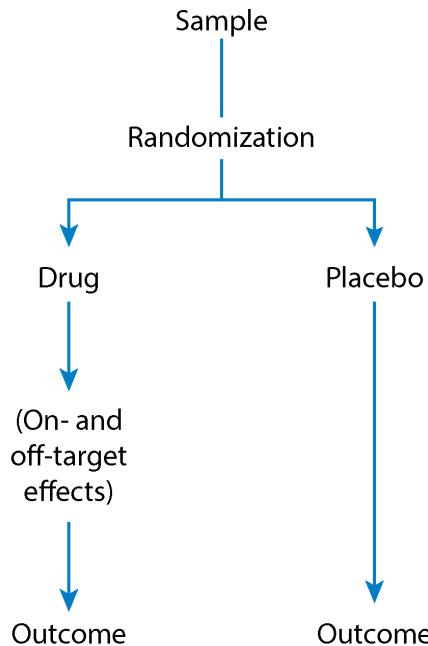
[b.couvyduchesne@uq.edu.au](mailto:b.couvyduchesne@uq.edu.au)

Program for Complex Trait Genetics (PCTG)

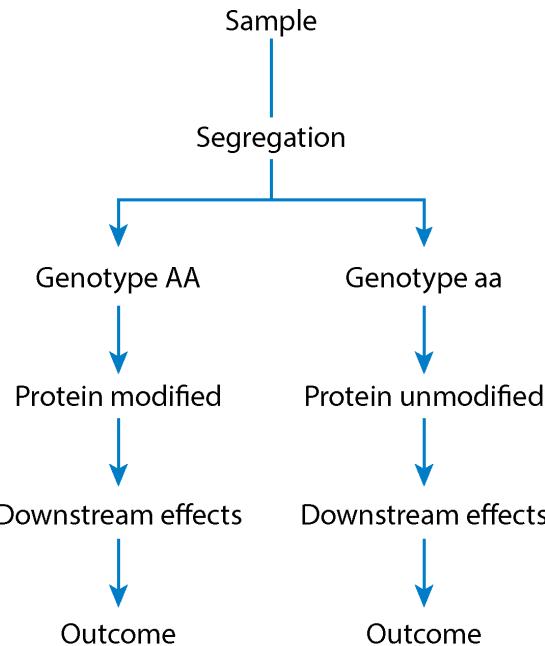
*International Congress of Twin Studies, Madrid, 16<sup>th</sup> November 2017*

# Mendelian Randomisation: rationale

Randomized controlled trial



Mendelian randomization



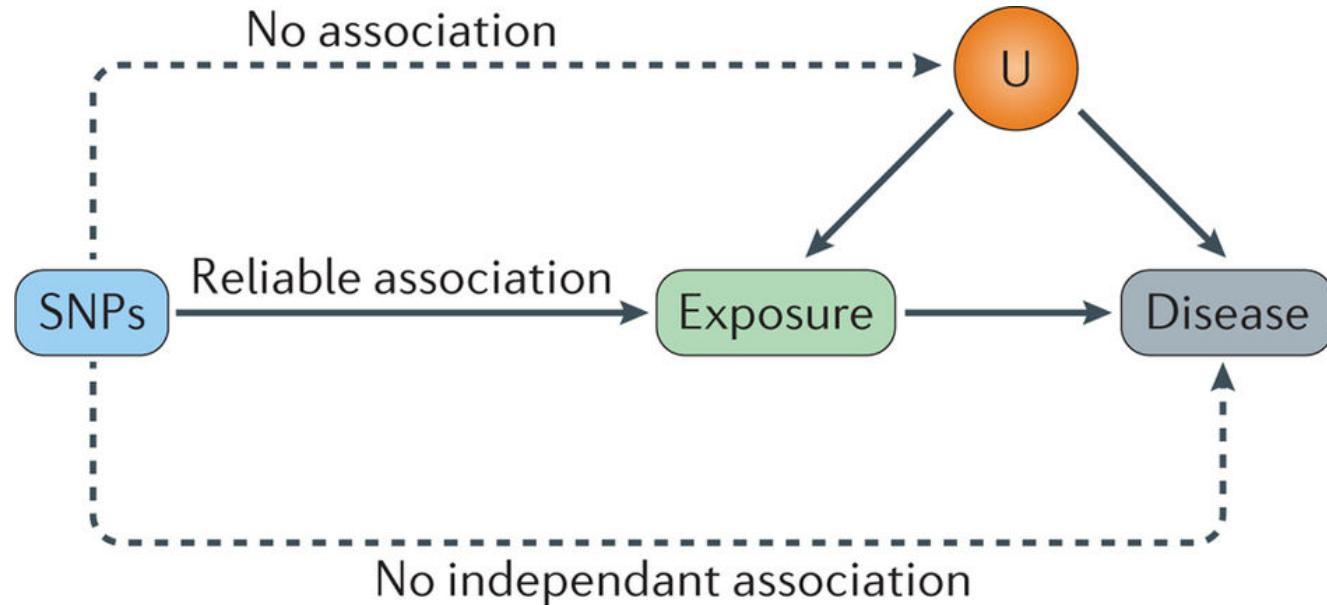
Mendelian  
Randomization: New  
Applications in the  
Coming Age of  
Hypothesis-Free  
Causality



Evans DM, Smith GD. 2015.

Annu. Rev. Genomics Hum. Genet. 16:327–50

# Mendelian Randomisation: the model



Holmes, M. V. et al. (2017) Mendelian randomization in  
cardiometabolic disease: challenges in evaluating causality  
*Nat. Rev. Cardiol.* doi:10.1038/nrccardio.2017.78

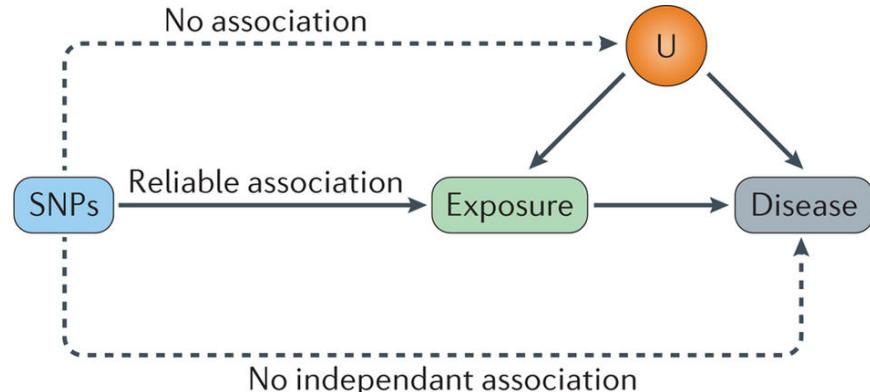
Nature Reviews | Cardiology

# Mendelian Randomisation: some strong assumptions

The SNP (instrumental variable) is associated with the exposure

The SNP is not associated with confounders, known or unknown

All the effect of the SNP on the disease is via the exposure



Nature Reviews | Cardiology

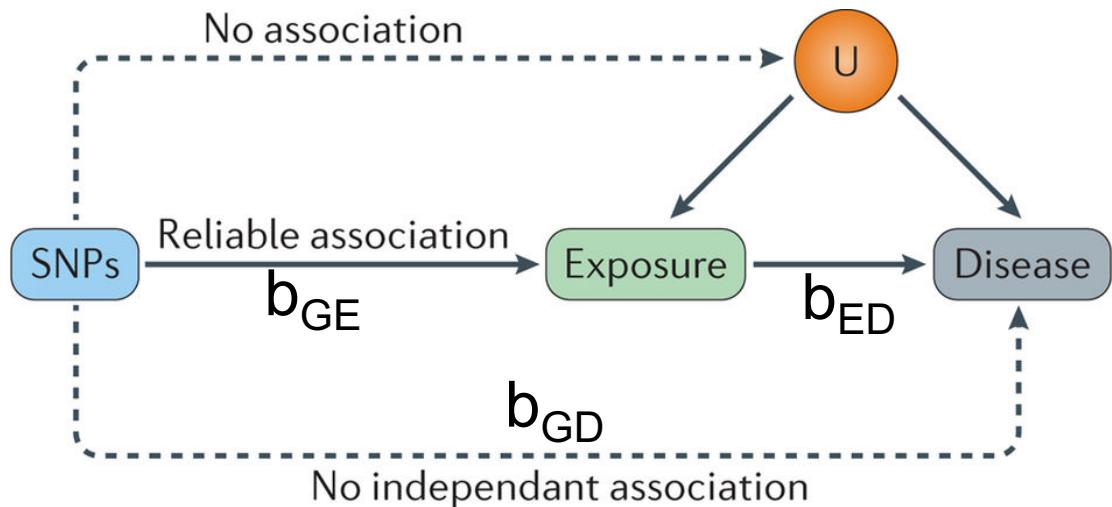
Holmes, M. V. et al. (2017) Mendelian randomization in cardiometabolic disease: challenges in evaluating causality  
Nat. Rev. Cardiol. doi:10.1038/nrccardio.2017.78

# Two-Sample MR

Often we do not have large GWAS studies with SNPs, exposure and outcome

- ⇒ Use summary statistics of GWAS of exposure and disease:  $b_{GE}$  and  $b_{GD}$
- ⇒ Extra assumption: same population ancestry
- ⇒ Allows bidirectional MR

Under MR assumptions:  
The effect of exposure  
on disease for 1 SNP is  
 $b_{ED} = b_{GD} / b_{GE}$



# Multi Instrument MR

Integrate all the SNPs associated with the exposure

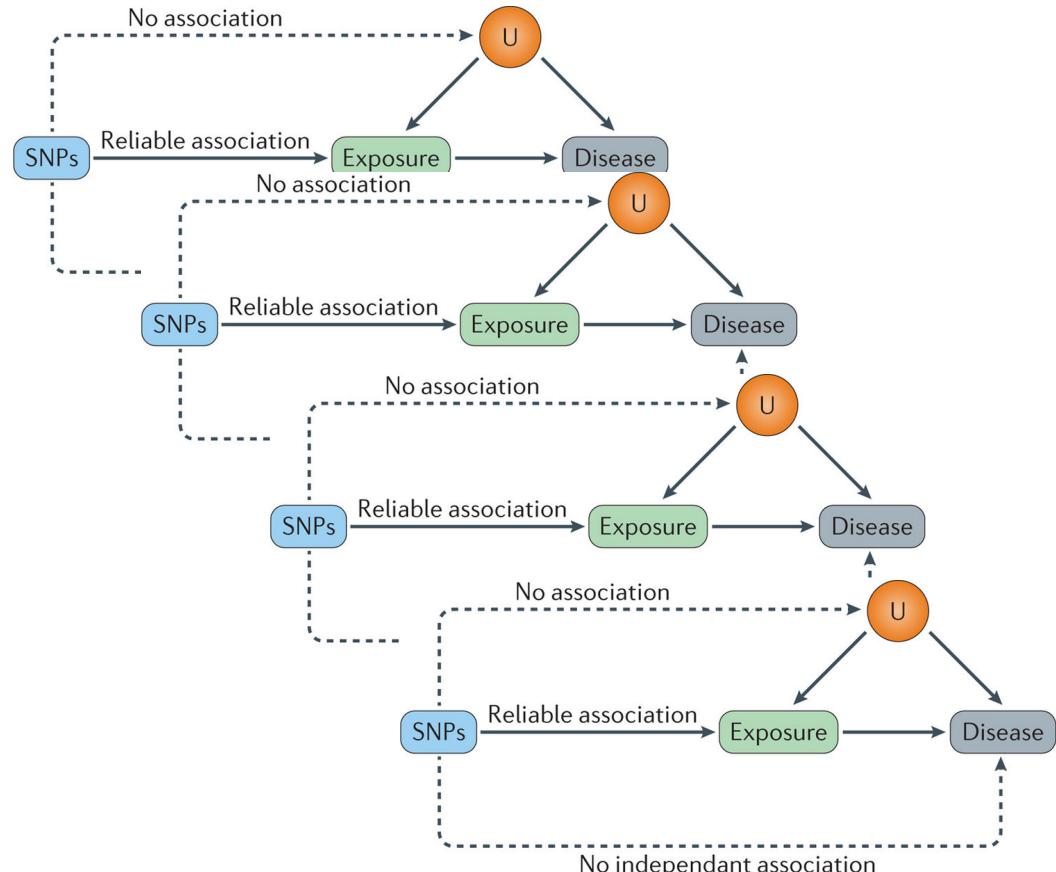
Meta-analysis approach

<http://www.mrbbase.org/>

Hemani et al., The MR-Base Collaboration.  
*MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations.* bioRxiv.

<https://mrcieu.github.io/TwoSampleMR/>

```
install.packages("devtools")
library(devtools)
install_github("MRCIEU/TwoSampleMR")
```



# Example: schizophrenia and population density



**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

New Results

## Higher genetic risk for schizophrenia is associated with living in urban and populated areas

Lucia Colodro-Conde, Baptiste Couvy-Duchesne, John B. Whitfield, Fabian Streit, Scott Gordon, Marcella Rietschel, John McGrath, Sarah E. Medland, Nicholas G. Martin

doi: <https://doi.org/10.1101/179422>

This article is a preprint and has not been peer-reviewed [what does this mean?].



Lucia Colodro-Conde



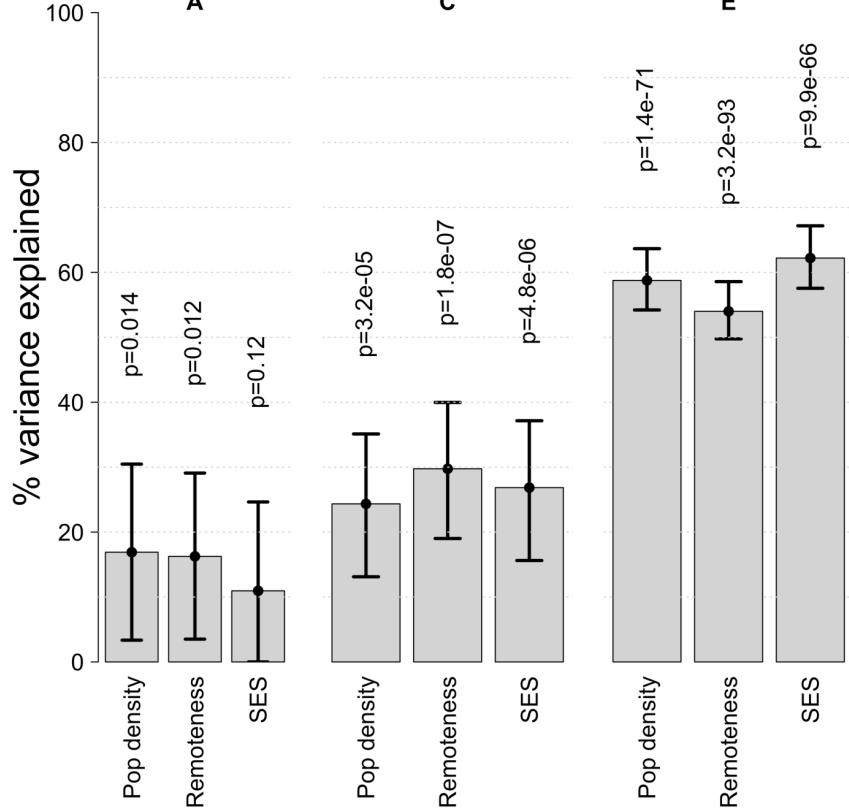
Sarah Medland



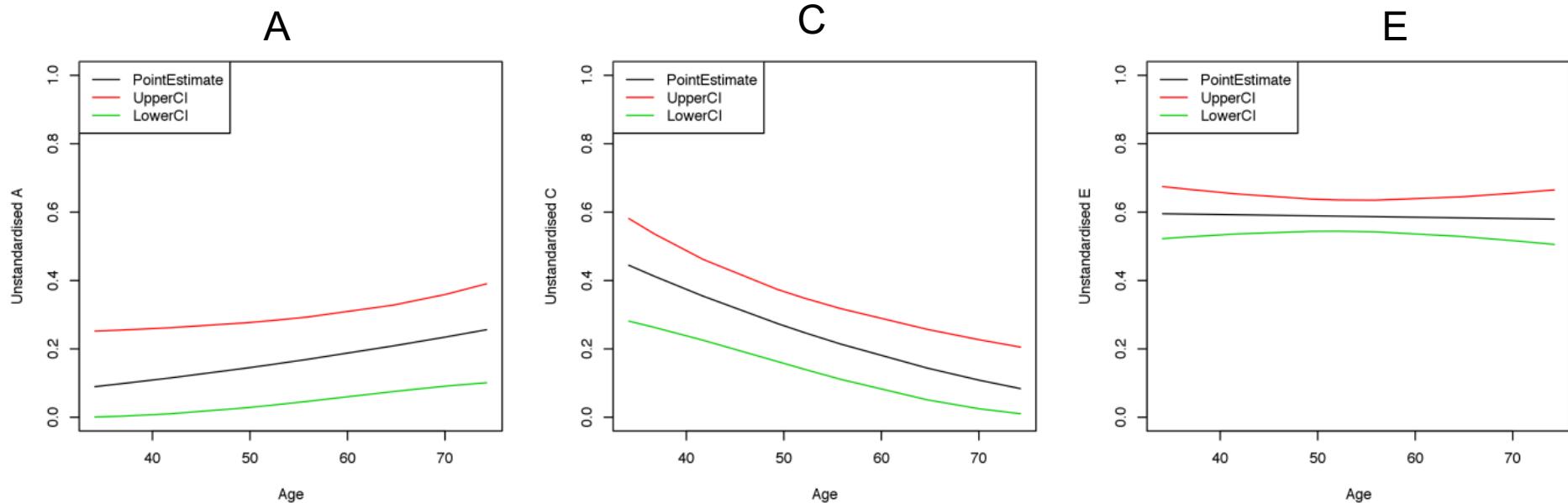
Nick Martin

HOME | AE

Search



# Example: schizophrenia and population density



GxE model in OpenMX

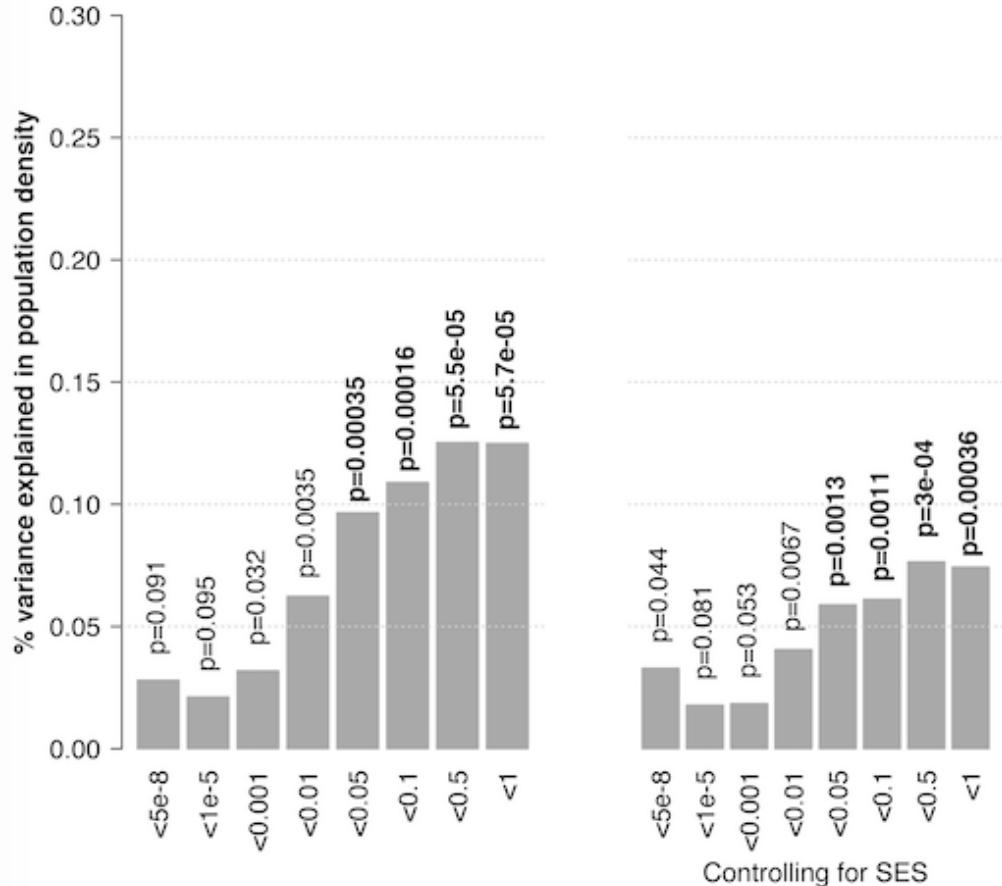
Purcell & Sham, 2002, Variance Components Models for Gene–Environment Interaction in Quantitative Trait Locus Linkage Analysis

# Example: schizophrenia and population density

PRS for schizophrenia  
associated with living in denser  
areas

Even when controlling for SES

**Association between SCZ and  
population density is partially  
genetic but what are the  
mechanisms?**



# MR base: “TwoSampleMR” package

```
scz_file <- read.table("scz2.snp.results_subset.txt")
colnames(scz_file)<-c("CHR", "SNP", "effect_allele",
"other_allele", "BP", "INFO", "OR", "se", "pval", "ngt")

outcome_dat <- read_outcome_data(snps = scz_file$SNP,
filename = "RMWResults_A_1000G_adu_popden_allchr_0K.txt",
sep = "\t",
snp_col = "SNP_dbSNP",
beta_col = "ALT_EFFECTSIZE",
se_col = "SE",
effect_allele_col = "ALT",
other_allele_col = "REF",
eaf_col = "ALL_AF",
pval_col = "PVALUE",
samplesize_col = "N_INFORMATIVE")
```

Open SCZ GWAS summary statistics with only genome-wide significant SNPs

Built in function: finds the same SNPs in GWAS of population density

Selects the appropriate columns

# MR base: “TwoSampleMR” package

```
sczMR<-format_data(scz_file, type ="exposure")
dat <- harmonise_data(exposure_dat = sczMR,
                      outcome = outcome_dat,
                      action = 2)
```

Formats and merges  
tables (built-in)

```
datClump <- clump_data(dat) # 94 SNPs kept
```

Clumps the SNPs for  
you! Needs internet  
connection, fast.

```
# Run all MR methods
res <- mr(datClump )
```

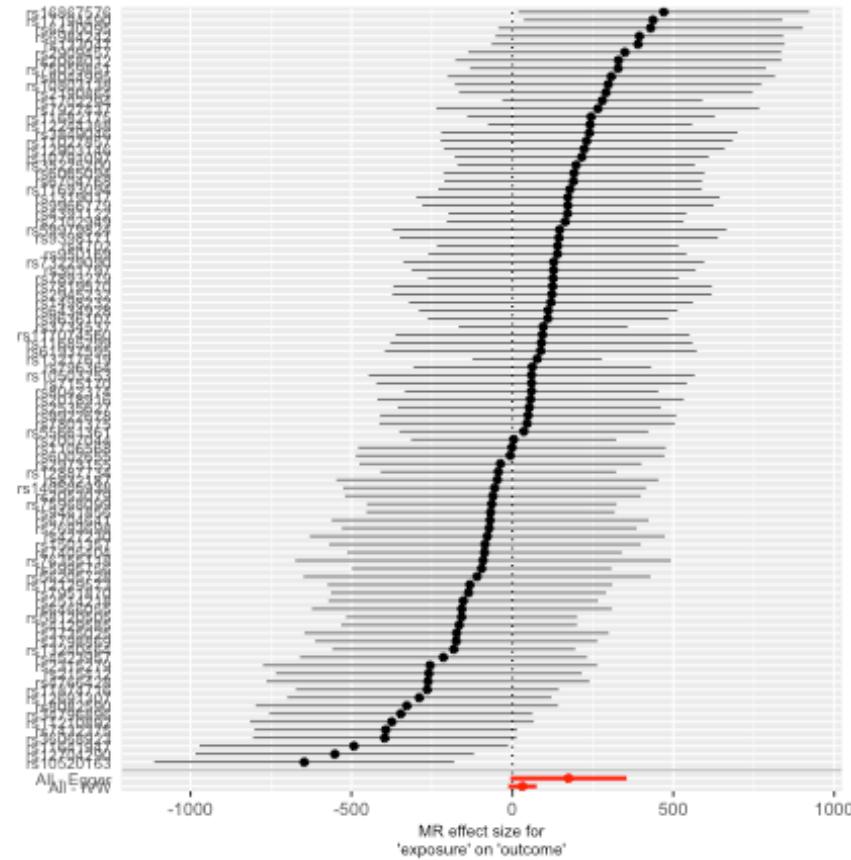
Run MR using a bunch  
of methods

```
# Creates a html report with all the plots and results
mr_report(datClump)
```

# MR base: “TwoSampleMR” outputs

Forest plot: effect of each SNP  
Meta-analysed effect at the bottom

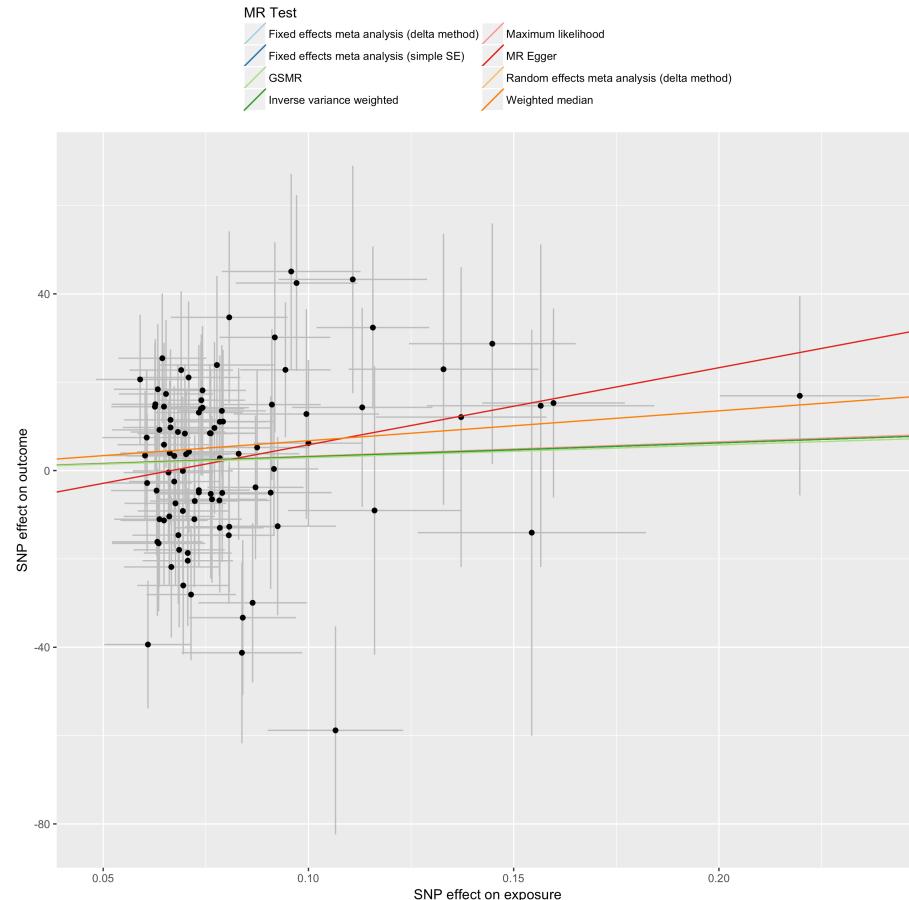
=> Check for outliers



# MR base: “TwoSampleMR” outputs

Inverse Variance Weighted (IVW) and MR-Egger considered the most robust methods in MR base

Method	Population density			
	N SNP	b	SE	p
Fixed effects meta analysis (simple SE)	94	31.4	22.1	0.15
Fixed effects meta analysis (delta method)	94	32.6	22.3	0.14
Random effects meta analysis (delta method)	94	32.6	22.3	0.14
Maximum likelihood	94	32.5	22.3	0.15
MR Egger	94	174.6	92.1	0.061
Weighted median	94	67.7	33.9	0.046
Inverse variance weighted	94	31.4	22.5	0.16
<b>GSMR</b>	<b>93</b>	<b>29.2</b>	<b>22.9</b>	<b>0.20</b>



# GSMR: generalised summary based MR



**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

HOME | A

Search

New Results

## Causal associations between risk factors and common diseases inferred from GWAS summary data

Zhihong Zhu, Zhili Zheng, Futao Zhang, Yang Wu, Maciej Trzaskowski, Robert Maier, Matthew Robinson, John McGrath, Peter Visscher, Naomi Wray, Jian Yang

doi: <https://doi.org/10.1101/168674>



Zhihong  
Zhu



Naomi  
Wray



Peter  
Visscher

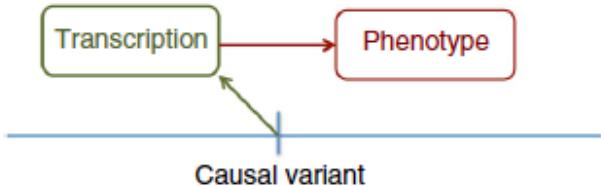


Jian  
Yang

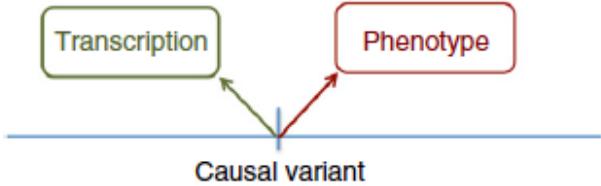
Models residual LD between SNPs  
Integrates standard error of the instruments  
HEIDI test of linkage

**b**

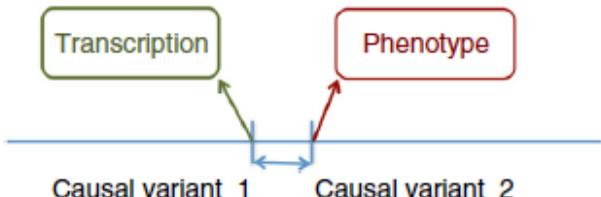
Causality



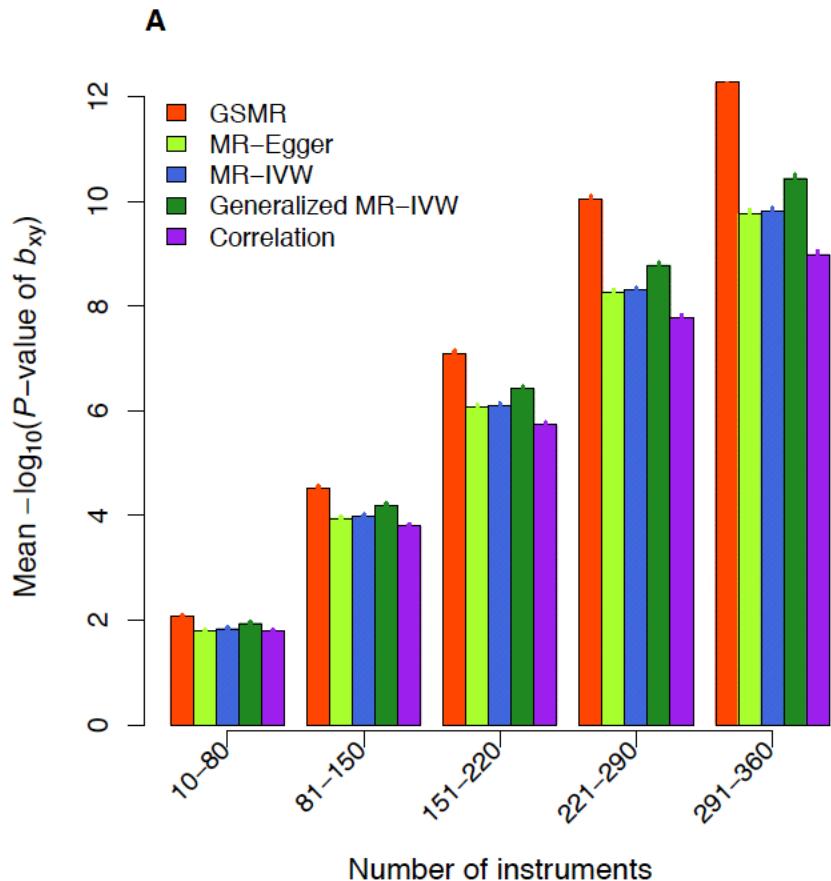
Pleiotropy



Linkage



# GSMR: more powerful than concurrent methods



**GSMR**  
Generalised Summary-data-based Mendelian Randomisation

GCTA GCTA forum SMR GSMR Program in CTG

**Overview**

Citation  
Installation  
Tutorial  
Package Document

## Overview

The `gsmr` R-package implements the GSMR (Generalised Summary-data-based Mendelian Randomisation) method to test for causal association between a risk factor and disease<sup>1</sup>. The R package is developed by [Zhihong Zhu](#), [Zhili Zheng](#), [Futao Zhang](#) and [Jian Yang](#) at Institute for Molecular Bioscience, the University of Queensland. Bug reports or questions: [z.zhu1@uq.edu.au](mailto:z.zhu1@uq.edu.au) or [jian.yang@uq.edu.au](mailto:jian.yang@uq.edu.au).

## Citation

Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *BioRxiv*, 168674.

## Installation

The `gsmr` requires R >= 2.15, you can install it in R by:

```
# gsmr requires the R-package survey  
install.packages("survey")  
# install gsmr  
install.packages("http://cnsgenomics.com/software/gsmr/static/gsmr_1.0.3.tar.gz", repos=NULL, type="source")
```



# GSMR: usage (<http://cnsgenomics.com/software/gsmr/>)

Requires additional matrix of LD between SNPs (from own sample or appropriate reference panel)

```
# Extract SNPs of interest
for iii in 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22;
do
mkdir ./chr${iii}
cd ./chr${iii}
gcta64 --bfile /data/bestguess/chr${iii}_thres0_cleaned --keep
./Ids.txt --extract ../gsmr_example_snps.allele --update-ref-allele
../gsmr_example_snps.allele --recodecp ./gcta.xmat.gz
../gcta.xmat${iii}.gz
cd ..
done

# Calculate the LD matrix
ldrho = cor(snp_coeff)
```

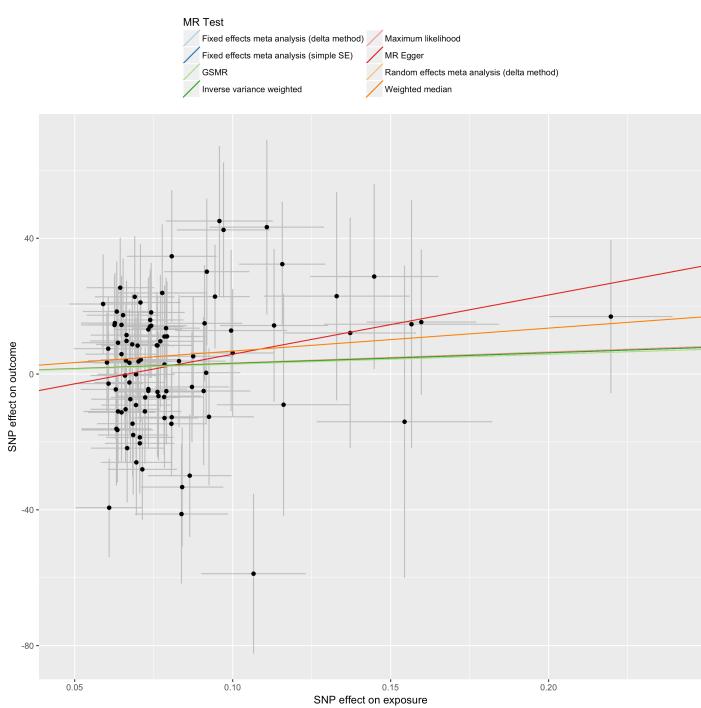
# GSMR: usage (<http://cnsgenomics.com/software/gsmr/>)

```
# Perform HEIDI-outlier analysis
filtered_index = heidi_outlier(bzx, bzx_se, bzx_pval, bzy, bzy_se,
ldrho, gwas_thresh, heidi_thresh)

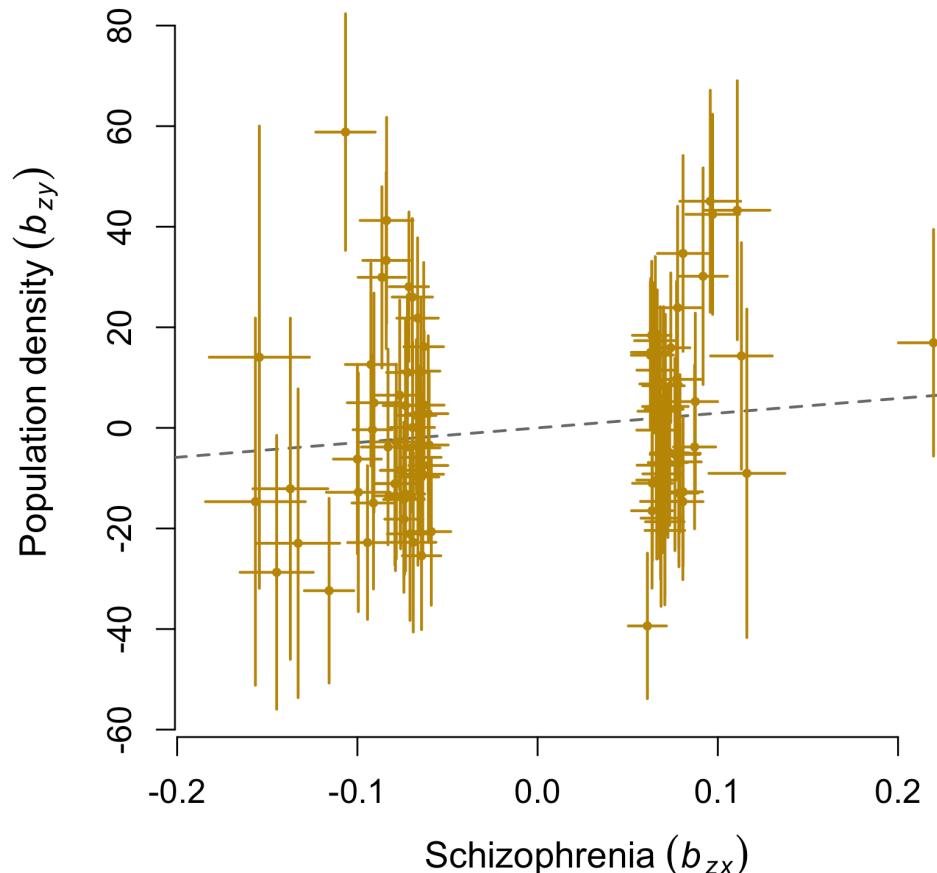
# select instruments that passed HEIDI-outlier filtering
filtered_gsmr_data = gsmr_data[filtered_index,]
filtered_ldrho = ldrho[filtered_gsmr_data$SNP,
filtered_gsmr_data$SNP]

# Run GSMR
gsmr_results = gsmr(bzx, bzx_se, bzx_pval, bzy, bzy_se,
filtered_ldrho)
```

# GSMR results



The plot difference is in the scale of the beta effects, easy to adapt



# Examples from the PGC-MDD2

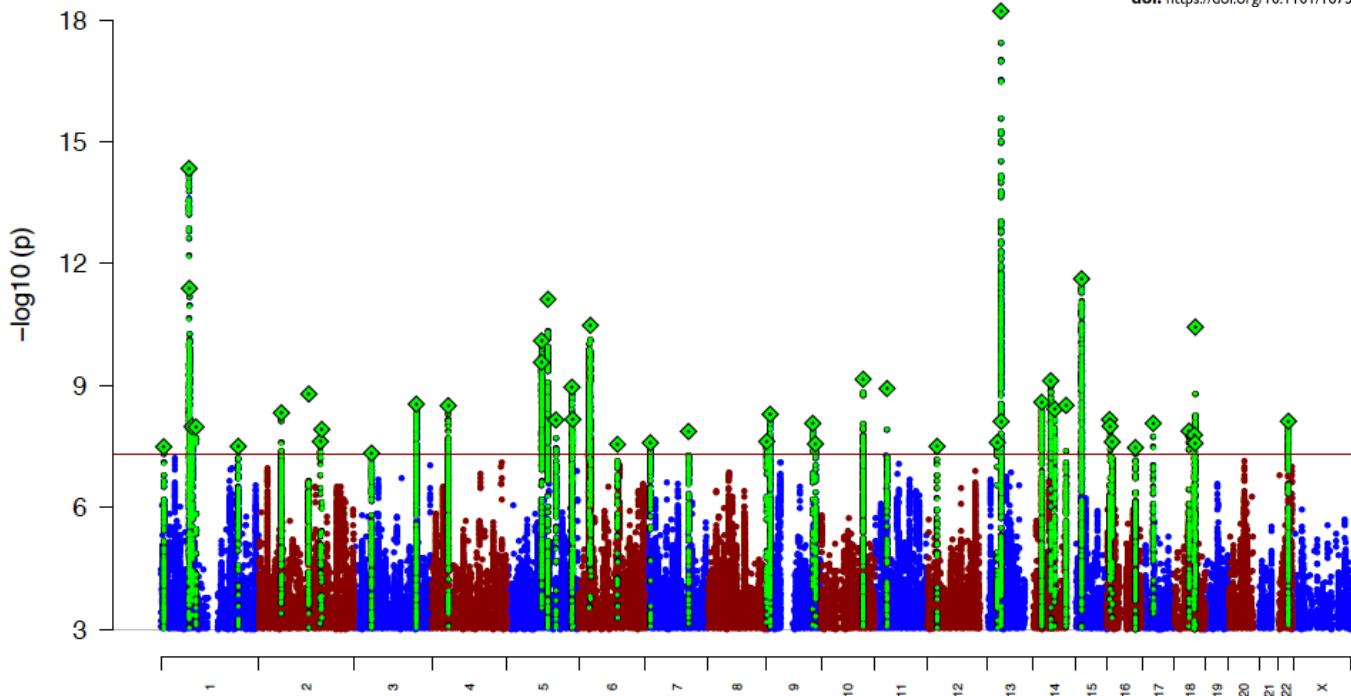


**bioRxiv**  
beta  
THE PREPRINT SERVER FOR BIOLOGY

HOME |

Search

(c)



New Results

**Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression**

- Major Depressive Disorder Working Group of the PGC, Naomi R. Wray, Patrick F Sullivan  
doi: <https://doi.org/10.1101/167577>



Maciej  
Trzaskowski



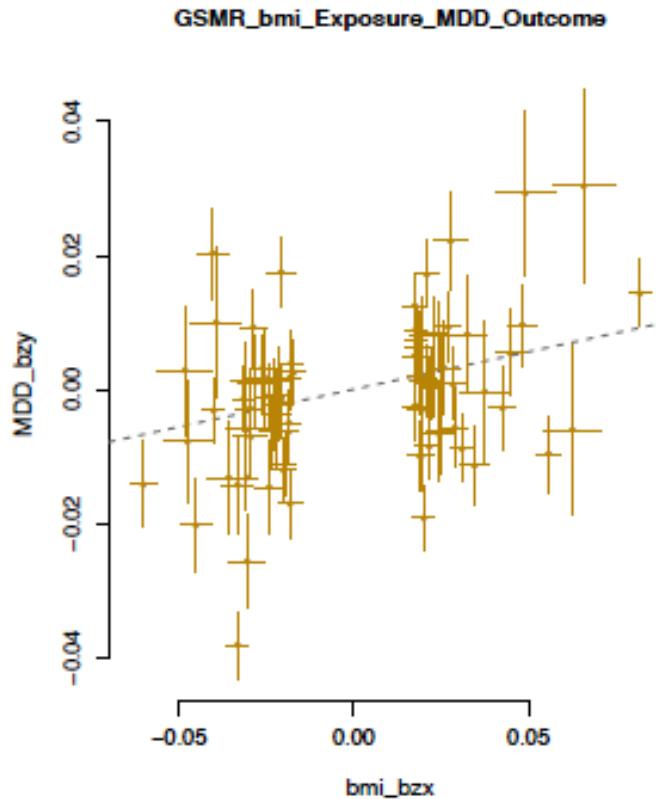
Naomi  
Wray

# Examples from the PGC-MDD2 paper

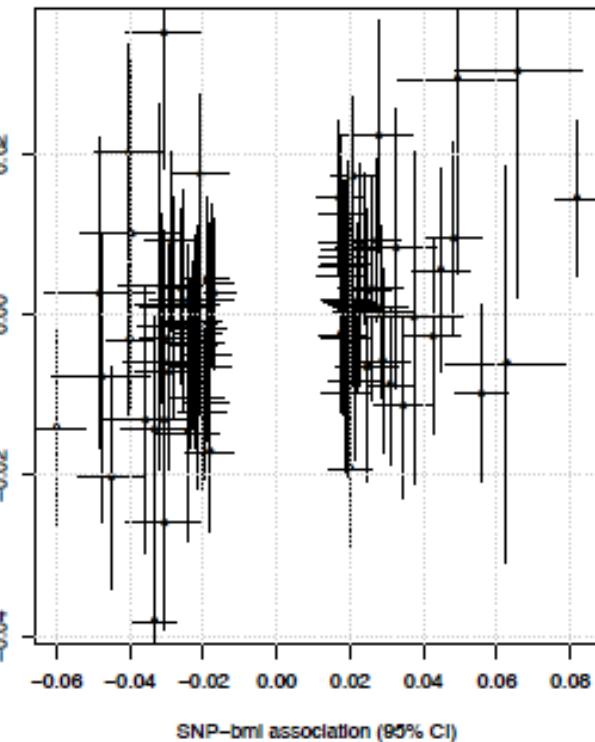
Effect of BMI on MDD

$$P_{\text{GSMR}} = 2.7 \times 10^{-7}$$

No evidence of reverse causality ( $P_{\text{GSMR}} = 0.81$ )



**MR-EGGER bmi Exposure – MDD Outcome**



Enda Byrne

# Using MR to identify target genes from GWAS

nature  
genetics

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

Zhihong Zhu<sup>1</sup>, Futao Zhang<sup>1</sup>, Han Hu<sup>2</sup>, Andrew Bakshi<sup>1</sup>, Matthew R Robinson<sup>1</sup>, Joseph E Powell<sup>1,3</sup>,  
Grant W Montgomery<sup>4</sup>, Michael E Goddard<sup>5,6</sup>, Naomi R Wray<sup>1</sup>, Peter M Visscher<sup>1,7</sup> & Jian Yang<sup>1,7</sup>

**SMR**  
Summary-data-based Mendelian Randomization

GCTA   SMR   GSMR   OSCA   Program in PCTG   CTG forum

**Overview**

- About
- Credits
- Questions and Help Requests
- Citations

**Download**

- Basic options
- Data Management
- SMR locus plot
- Query QTL Results
- MeCS
- Options Reference

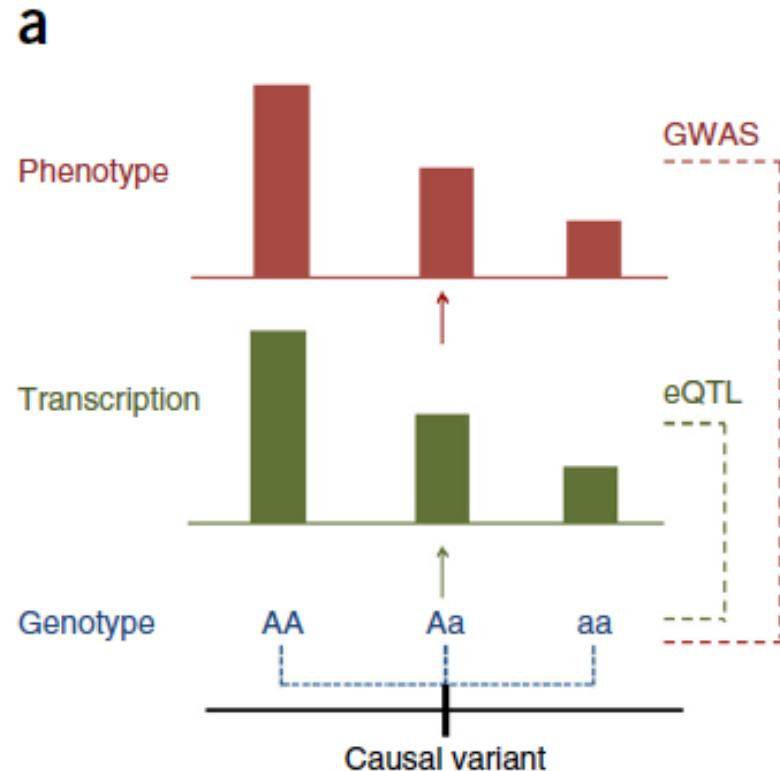
**About**

This software tool implements the SMR & HEIDI methods to test for pleiotropic association between the expression level of a gene and a complex trait of interest using summary-level data from GWAS and expression quantitative trait loci (eQTL) studies (Zhu et al. 2016 *Nat Genet*). The methodology can be interpreted as an analysis to test if the effect size of a SNP on the phenotype is mediated by gene expression. This tool can therefore be used to prioritize genes underlying GWAS hits for follow-up functional studies.

**Credits**

Futao Zhang developed the software tool and webpages with supports from Zhihong Zhu, Zhihong Zhu, Ting Qi and Jian Yang.  
Zhihong Zhu and Jian Yang developed the SMR and HEIDI methods.  
Ting Qi and Jian Yang developed the MeCS method.

Zhihong Zhu   Jian Yang   Futao Zhang   Zhihong Zhu   Ting Qi



# Example: PGC-MDD2

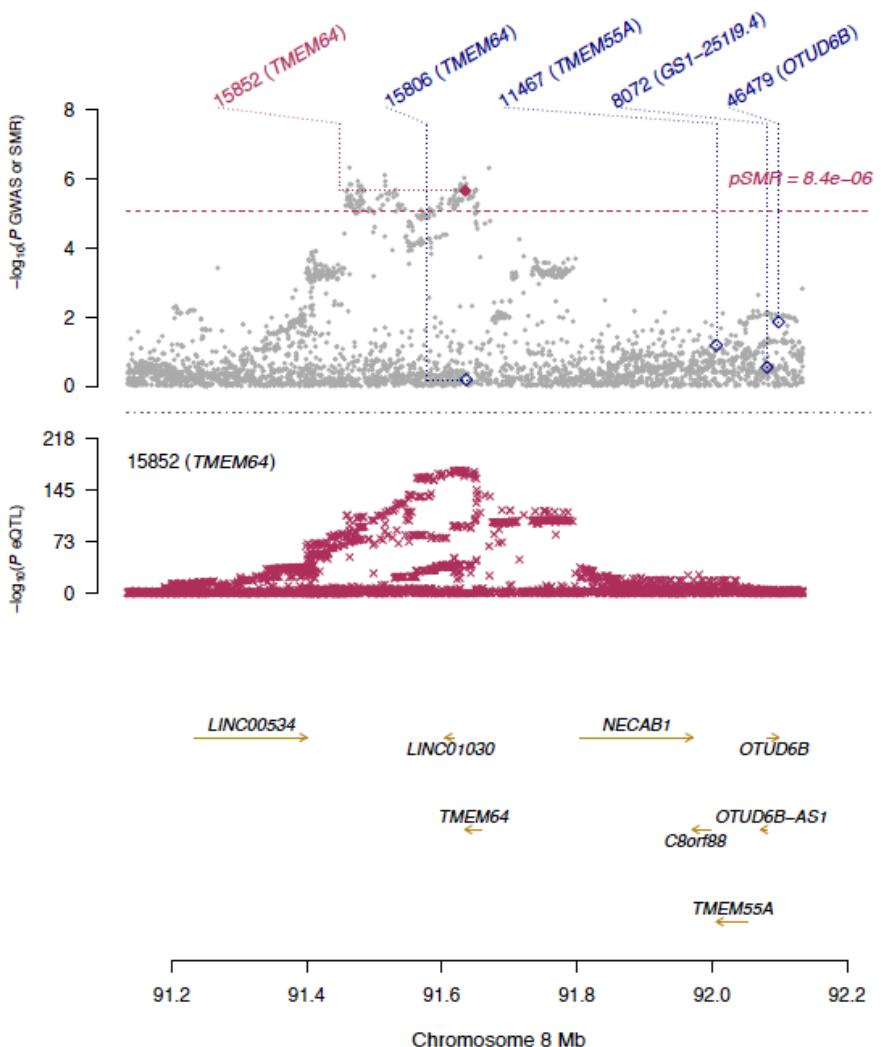
Exposure is gene expression  
Outcome is MDD

MR tests for each gene in tissue

=> Transcriptome wide association study using GTEx, CMC, eQTLGen Consortium



Enda Byrne

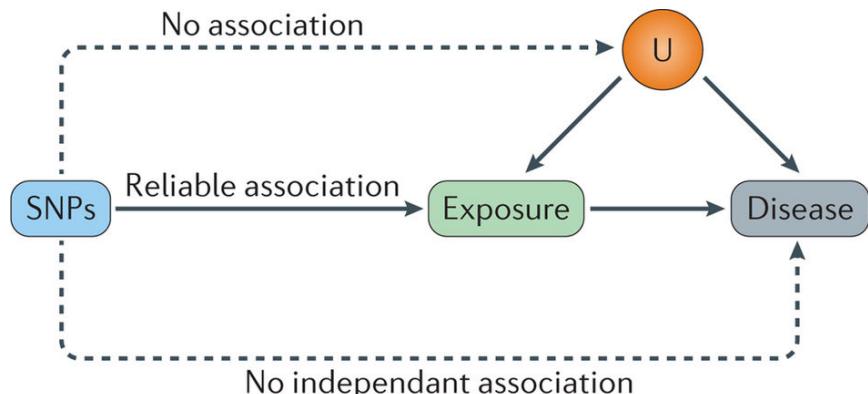


# Summary

**MR relies on strong (unverifiable?) assumptions about the data:**

- *The SNP is not associated with confounders, known or unknown*
- *All the effect of the SNP on the disease is via the exposure*

⇒ Should be cautious about using MR to conclude about causality  
⇒ Useful to generate hypotheses (that may be tested in randomised trials or in the lab)



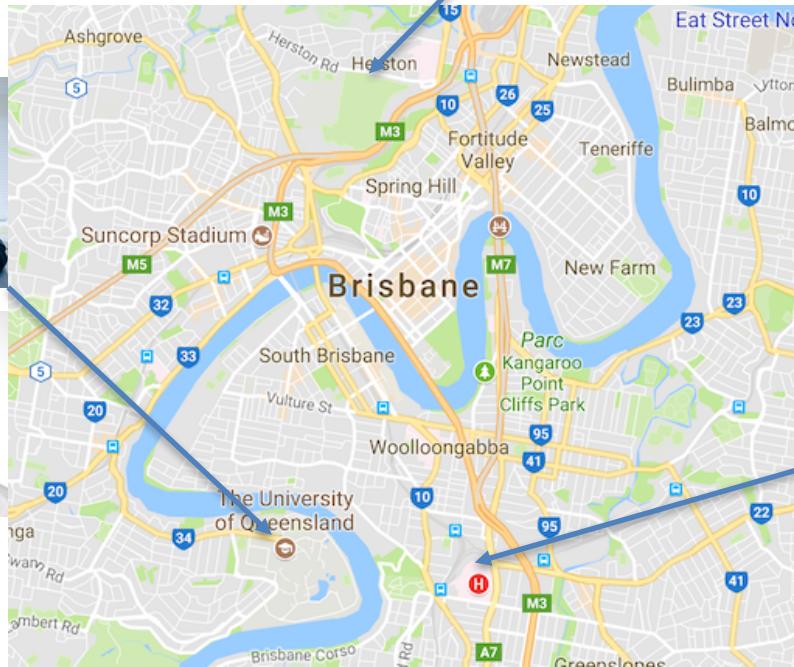
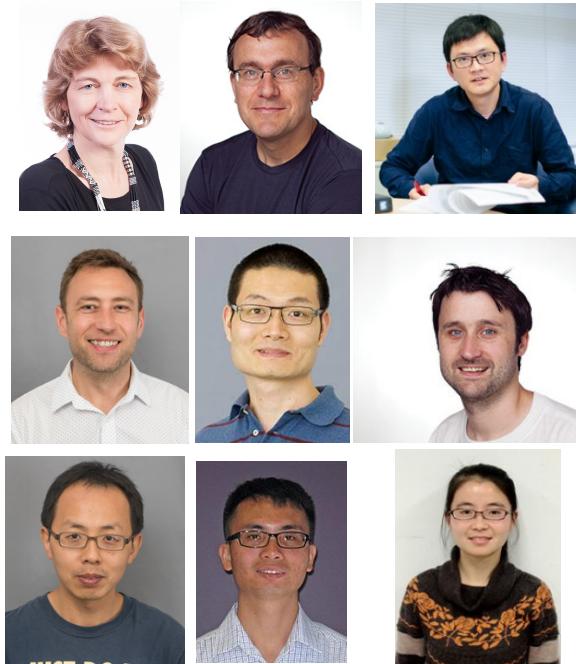
Nature Reviews | Cardiology

Holmes, M. V. et al. (2017) Mendelian randomization in cardiometabolic disease: challenges in evaluating causality  
*Nat. Rev. Cardiol.* doi:10.1038/nrcardio.2017.78

Interpretation limited to genetic signal  
Overall limited by GWAS power and number of instruments

# Thank you

Program for Complex Trait Genetics – the University of Queensland



QIMR Berghofer Medical Research Institute



Translational Research Institute

