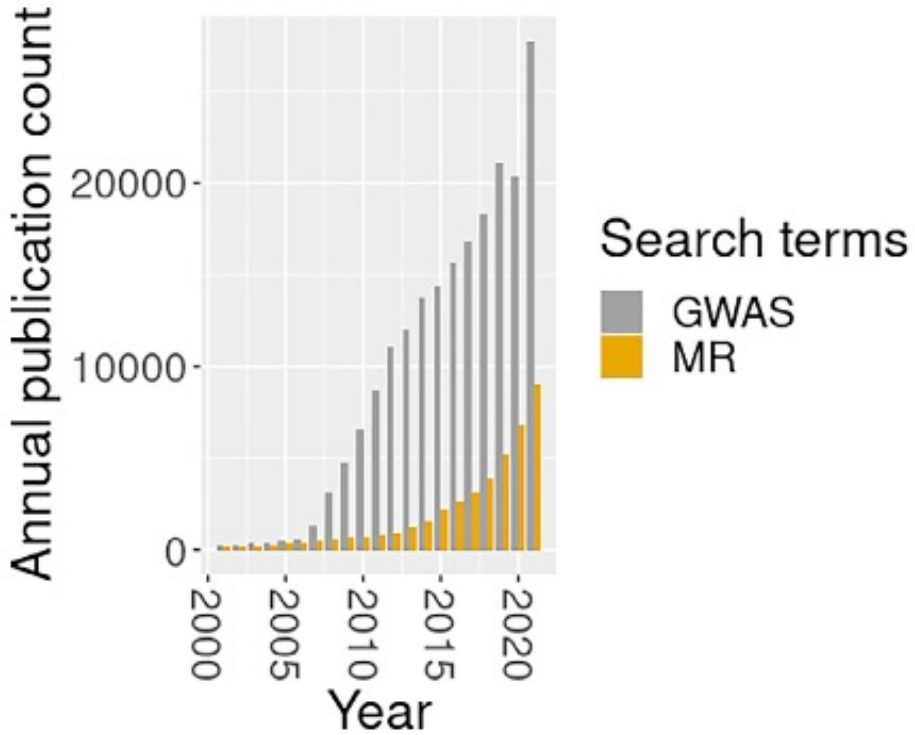
Leveraging Summary-data-based Mendelian Randomisation for Gene Target Discovery

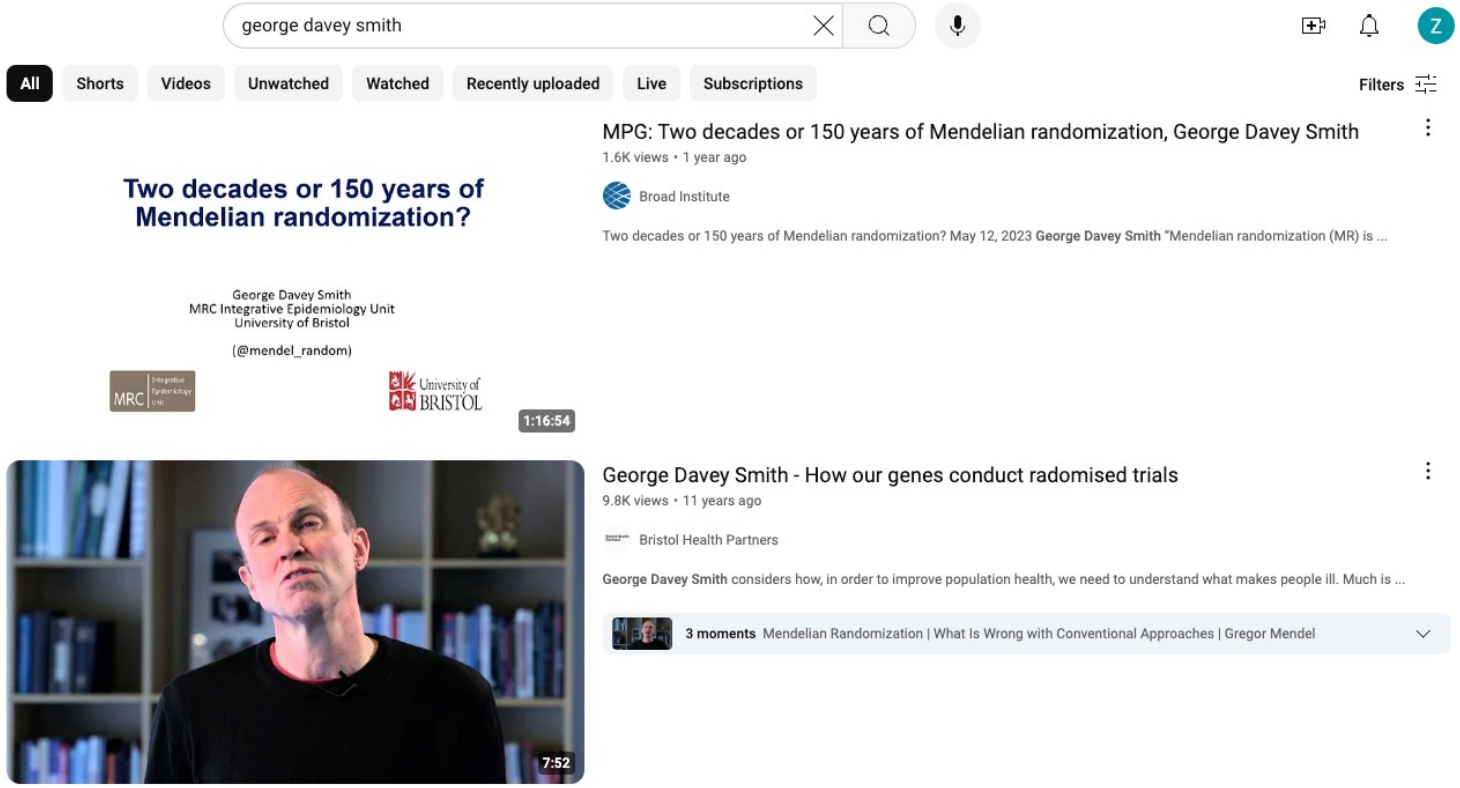
Zhihong Zhu, Ph.D

Senior Researcher, NCRR, Aarhus University
Visitor, PCTG, University of Queensland
z.zhu@econ.au.dk

Mendelian randomisation



Boehm et al. 2022



From YouTube

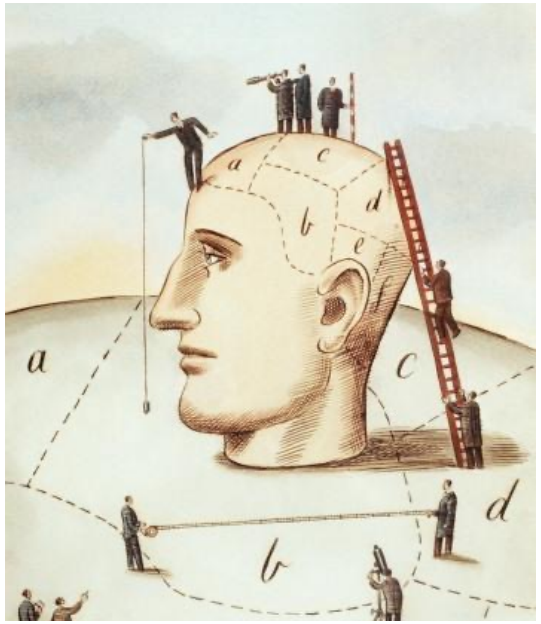
Applications of Mendelian randomisation

International Journal of Epidemiology 2003;**32**:1–22
DOI: 10.1093/ije/dyg070

30TH THOMAS FRANCIS JR MEMORIAL LECTURE

'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?*

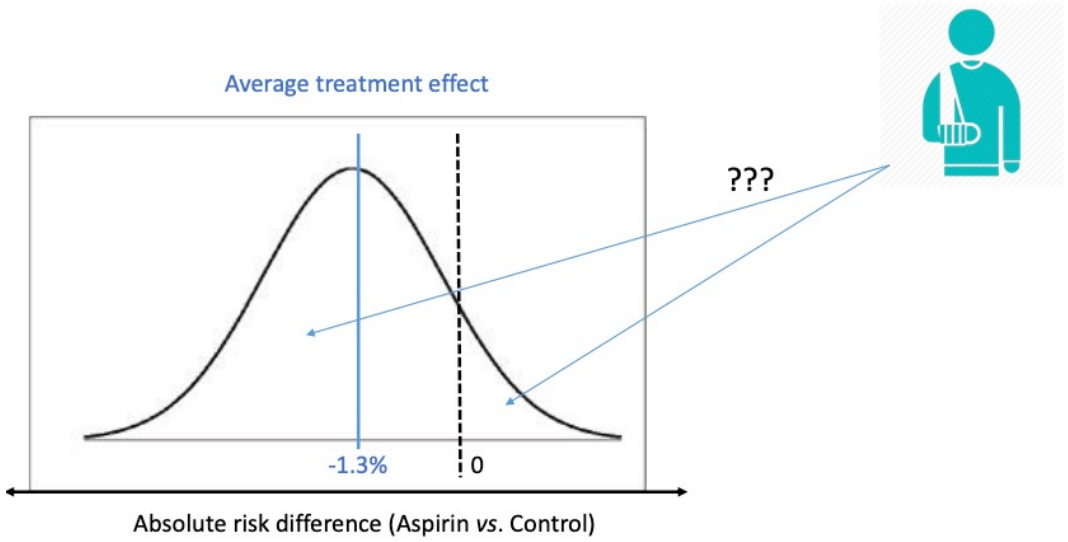
George Davey Smith and Shah Ebrahim

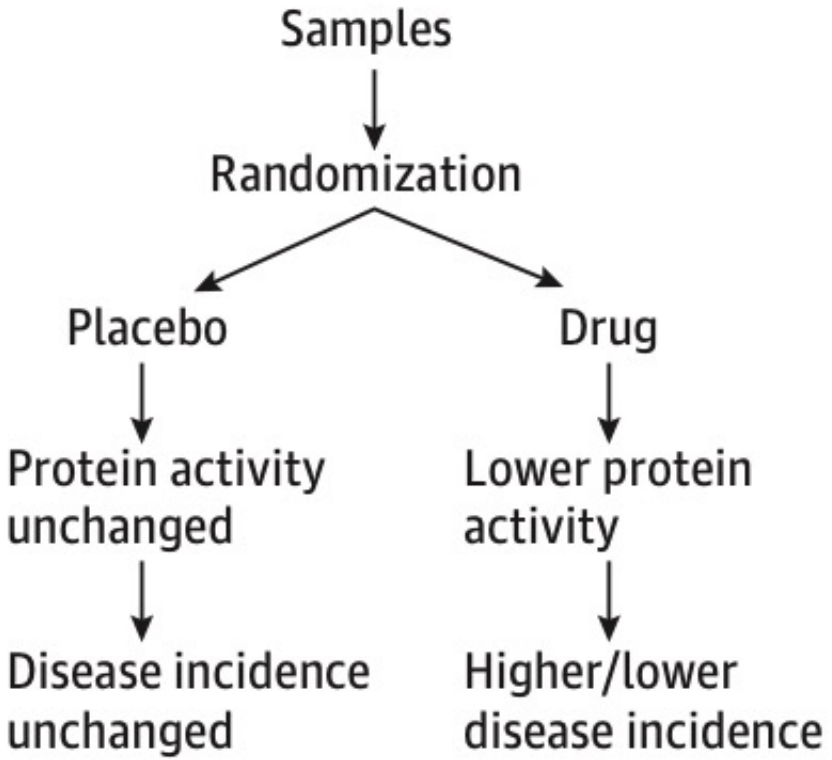


Outlines

- Causal inference
- Mendelian randomization
  - ❑ *CACNA2D4* as an example
- Summary-data-based Mendelian randomization
  - ❑ A two-sample Mendelian randomization method
- Software

Causal inference

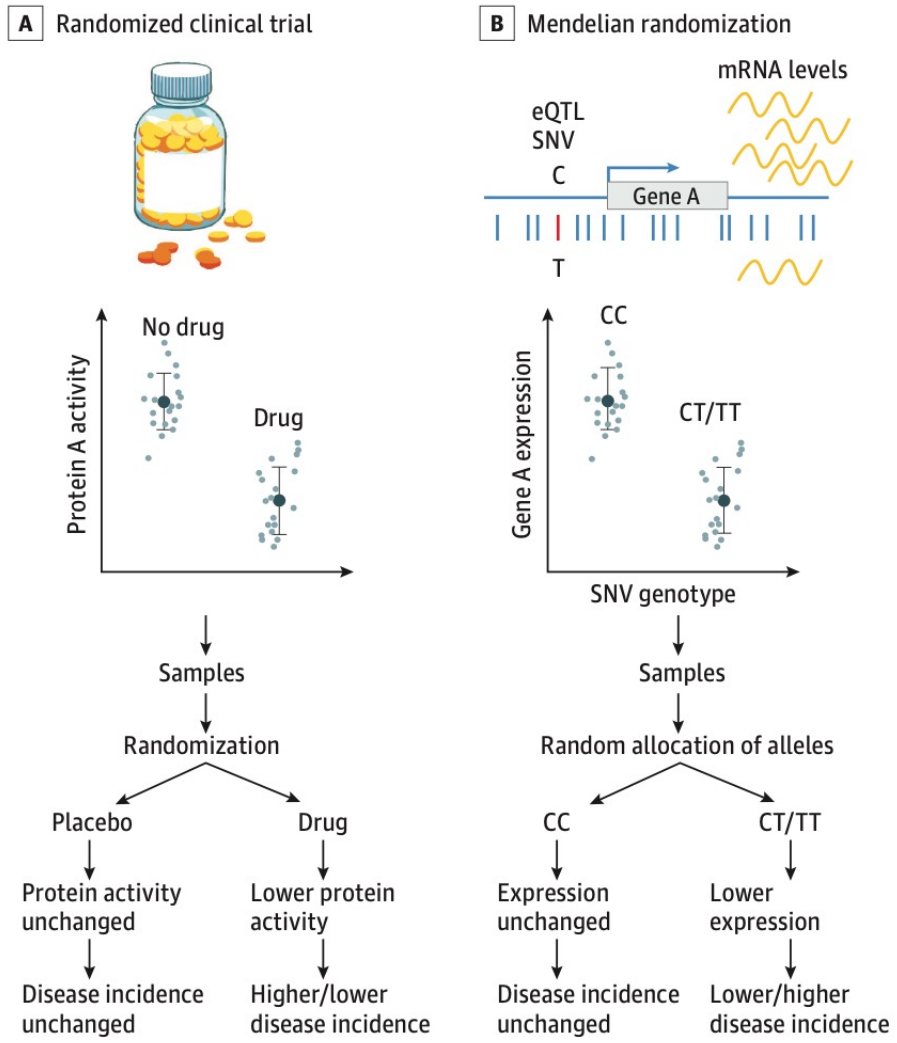




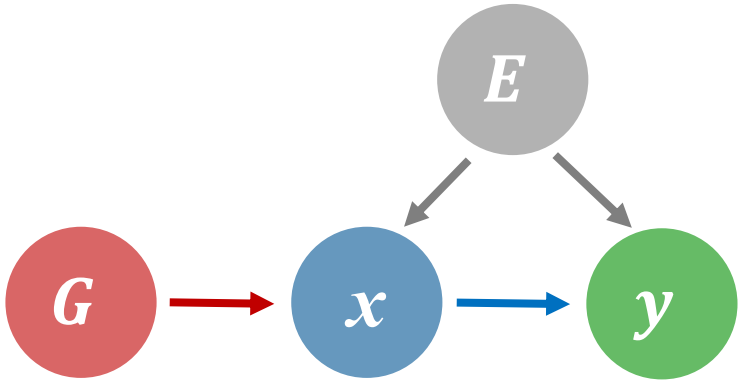
1. Randomisation
2. Instrument
3. Intervention
4. Difference

Independent of confounders

Causal inference by MR



Mendelian randomization



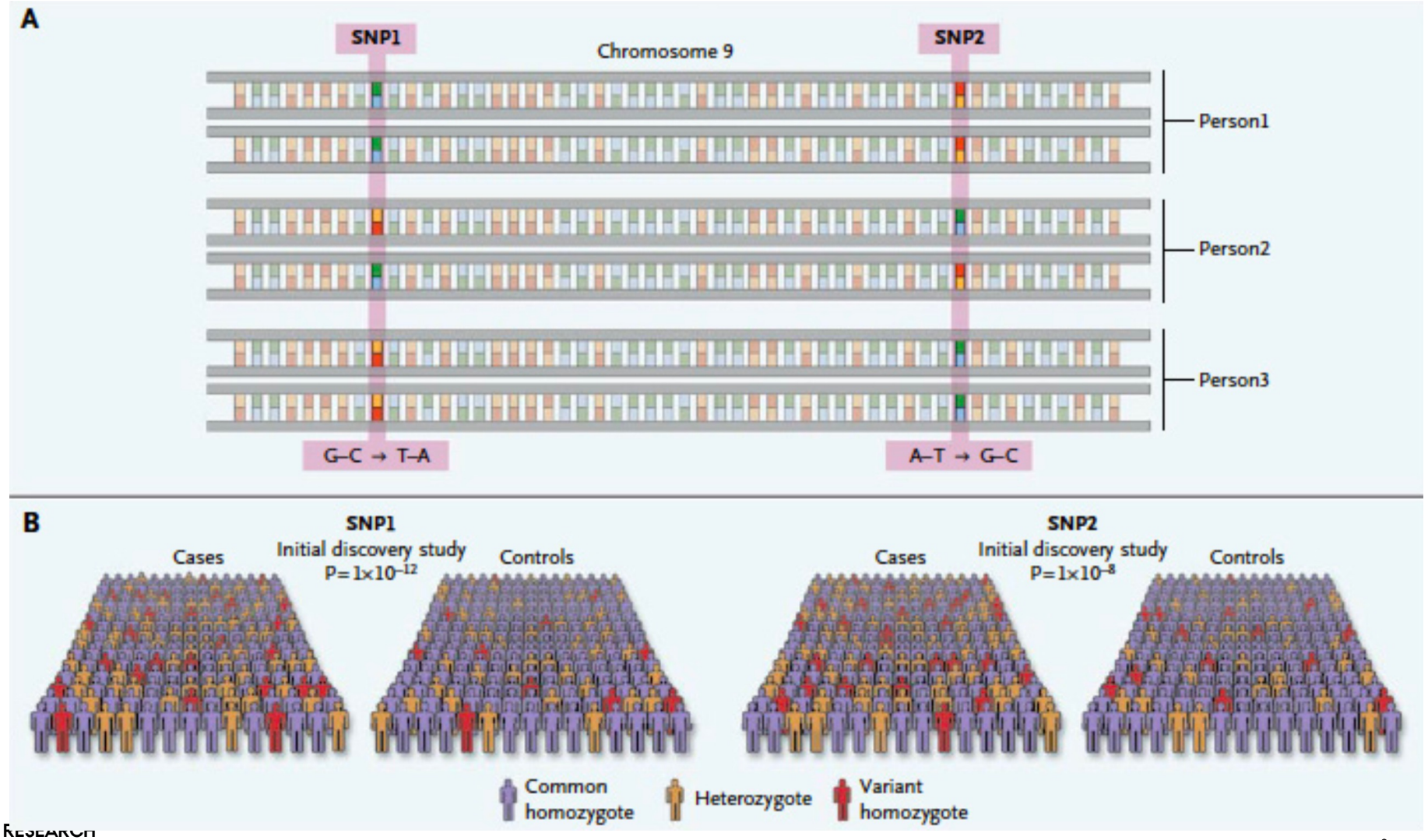
Independent of environmental confounders

SNP (DNA variant)

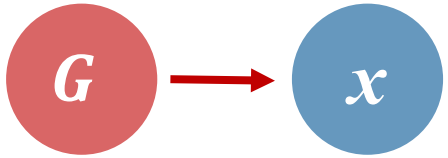
Risk alleles

Pathways

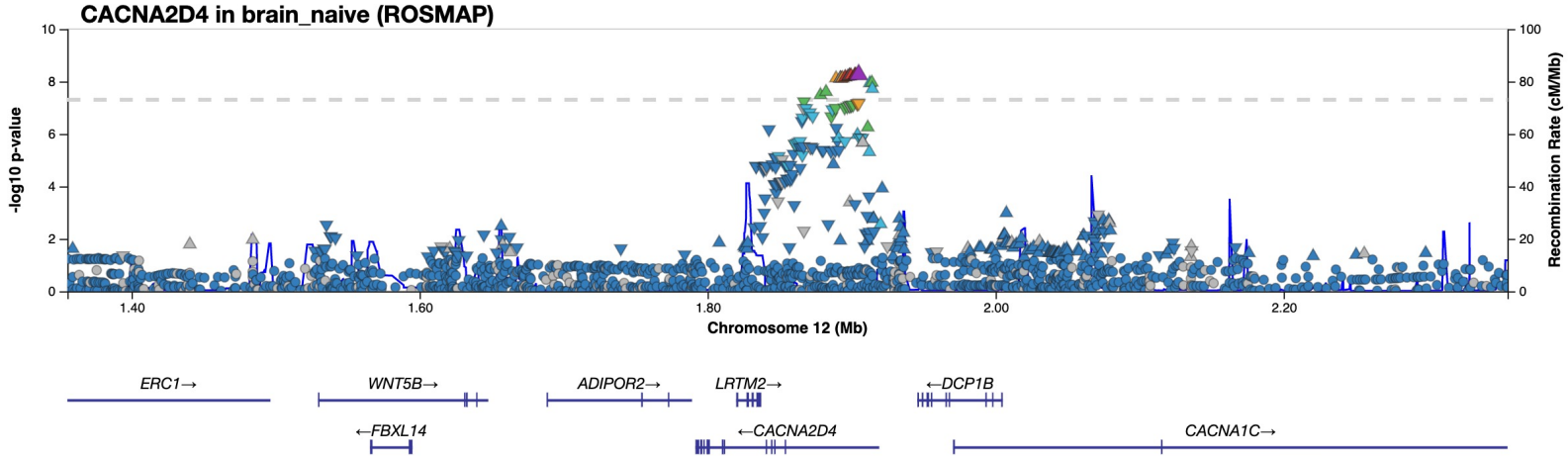
Genetic risks



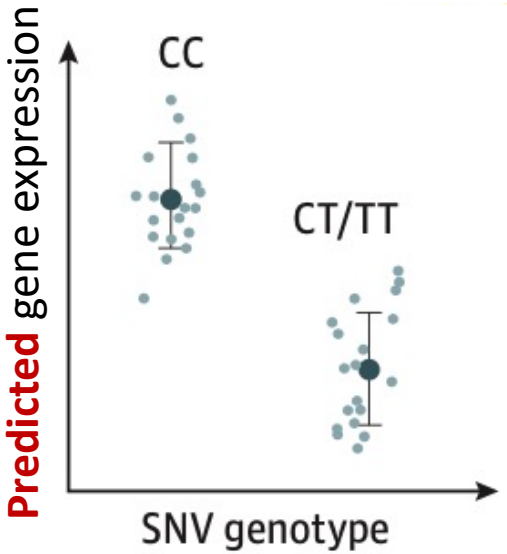
Genetically predicted exposure



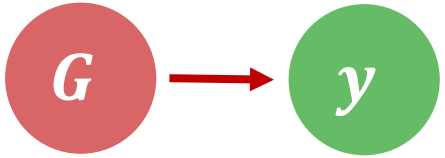
- *cis*-eQTL data
  - ❑ Association between SNP and exposure



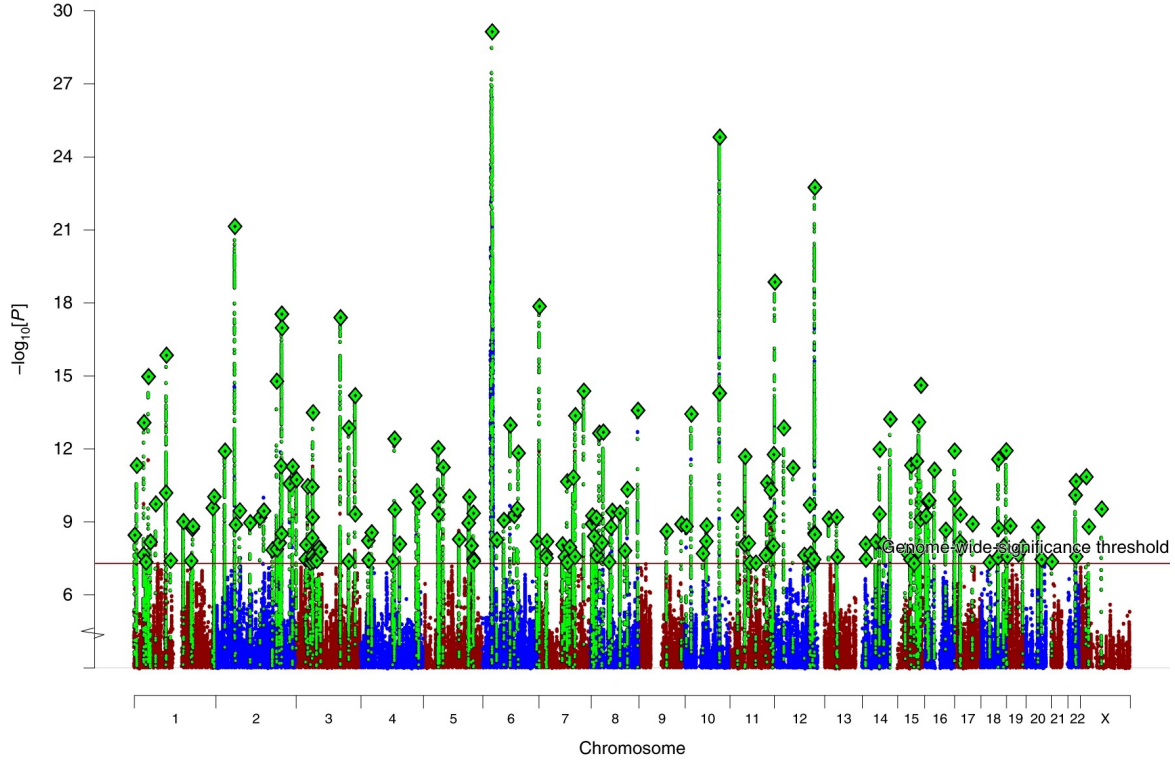
- Assumption
  - ❑ Strong association between SNP and exposure
  - ❑ Linear relationship between SNP and exposure
- Genetically predicted exposure
  - ❑ $\hat{x}_j = z_j\hat{\delta} + e_{x(j)}$



Genetically predicted outcome



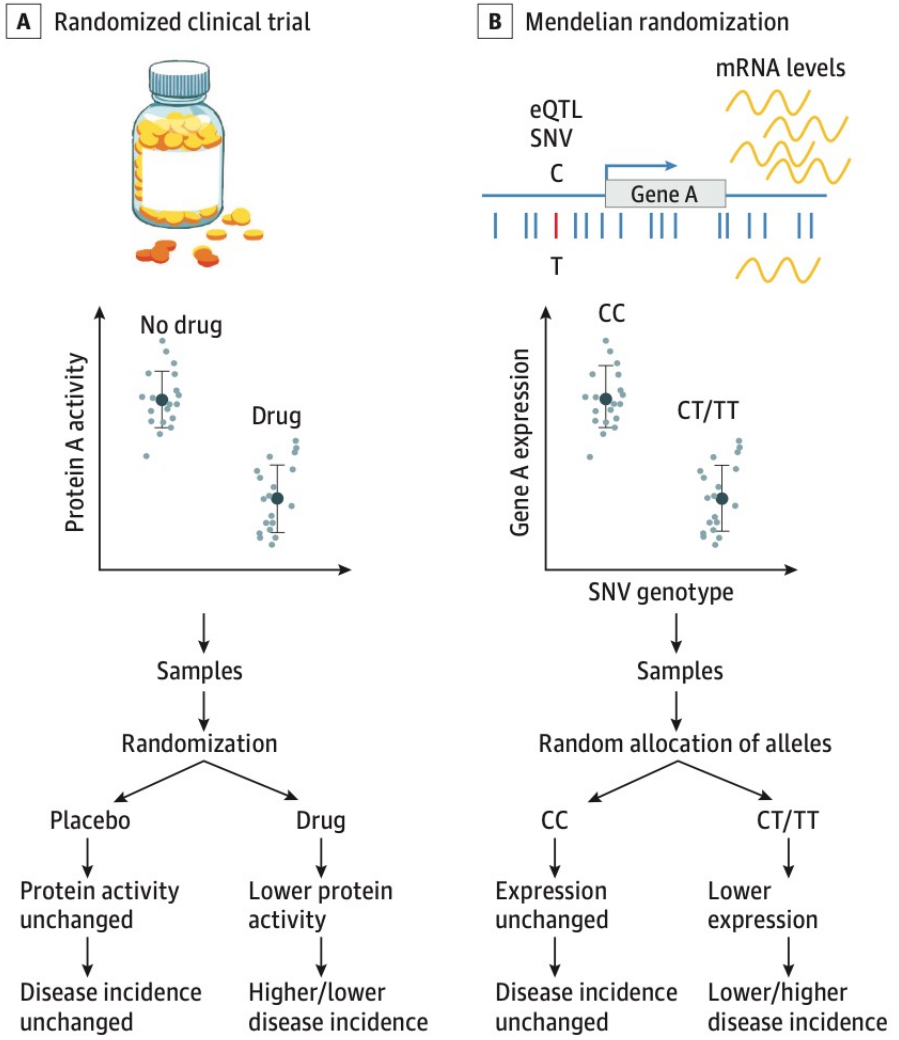
- GWAS data
  - ❑ Association between SNP and outcome
- Assumption
  - ❑ Linear relationship between SNP and outcome
- Genetically predicted outcome
  - ❑ $\hat{y}_j = z_j\hat{\gamma} + e_{x(j)}$



Lam et al 2019 Nature Genetics

Summary-data-based MR

- SMR estimate
  - ❑ Assumption: Single causal variant, linear relationship between exposure and outcome
  - ❑ Estimate: $\hat{\beta}_{\mathrm{SMR}} = \hat{\gamma}/\hat{\delta}$
  - ❑ Interpretation: a x mg/L (1SD unit) higher exposure increase/decrease y% genetic risk of disease outcome
  - ❑ Equivalent to estimate from regression if individual-level data are available.



Proof of MR estimate

- $\mathrm{E}(\hat{\beta}_{\mathrm{SMR}}) = \hat{\gamma}/\hat{\delta} = (\hat{\gamma}\times\hat{\delta})/(\hat{\delta}\times\hat{\delta})$
- SNP-exposure association: $\hat{\delta} = (z^T z)^{-1} z^T x$
  SNP-outcome association: $\hat{\gamma} = (z^T z)^{-1} z^T y$
- $\mathrm{E}(\hat{\beta}_{\mathrm{SMR}}) = \hat{\gamma}/\hat{\delta} = (\hat{\gamma}\times\hat{\delta})/(\hat{\delta}\times\hat{\delta}) = \frac{x^T P_z y}{x^T P_z x} = \beta + \frac{x^T P_z e}{x^T P_z x}$, where $P_z = z(z^T z)^{-1} z^T$
- DNA variants are independent of environmental factors, $z^T e = 0$

  $\mathrm{E}(\hat{\beta}_{\mathrm{SMR}}) = \beta$

SMR – *CACNA2D4*

Gene	SNP	A1 / A2	Data	*b*	SE	*P*-value
CACNA2D4	rs1044825	G / T	eQTL (blood)	0.447	0.0186	4.1E-128
			GWAS (schizophrenia)	-0.0377	0.0087	1.3E-5

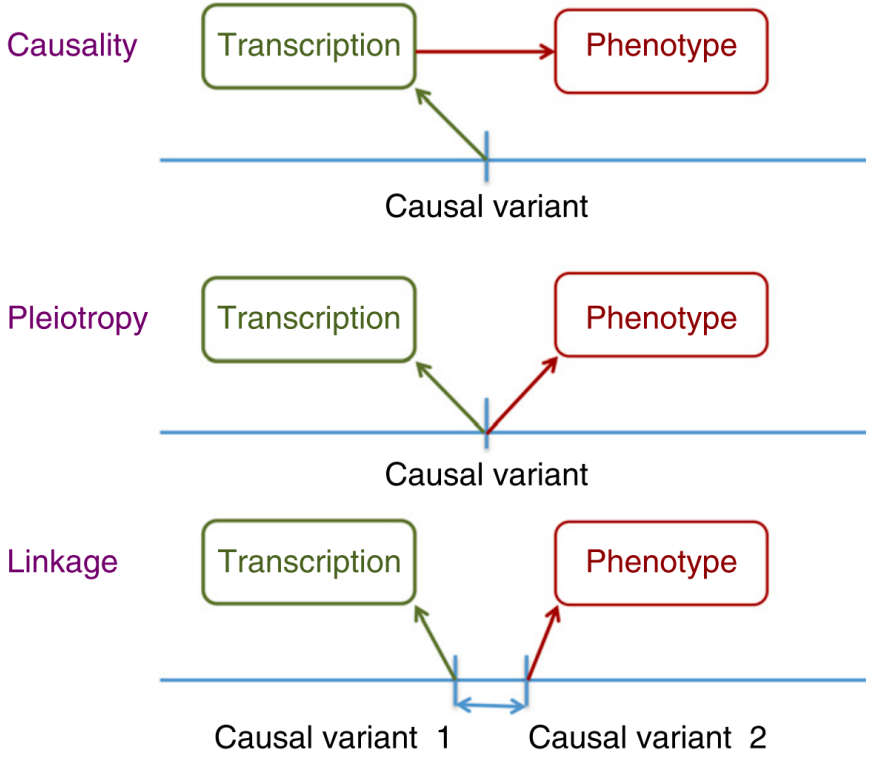
$$\hat{\beta} \approx \frac{\hat{\gamma}}{\hat{\delta}} = -\frac{0.0377}{0.447} = -0.084$$

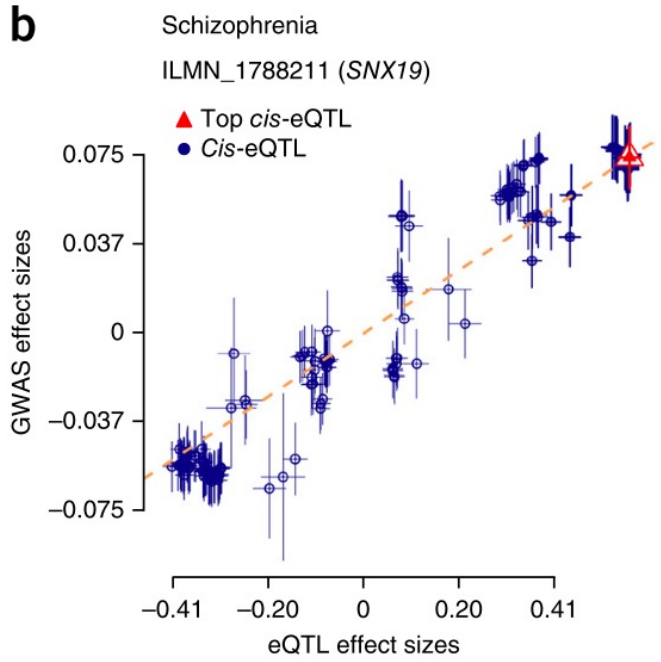
→ *P*-value = 2.0E-5

$$\mathrm{SE}(\hat{\beta}) \approx \sqrt{\left(\frac{\gamma}{\delta}\right)^2 \left[\frac{var(\delta)}{\delta^2} + \frac{var(\gamma)}{\gamma^2}\right]} = \sqrt{\left(\frac{0.45}{-0.04}\right)^2 \left[\frac{0.02^2}{0.45^2} + \frac{0.01^2}{(-0.04)^2}\right]} = 0.020$$

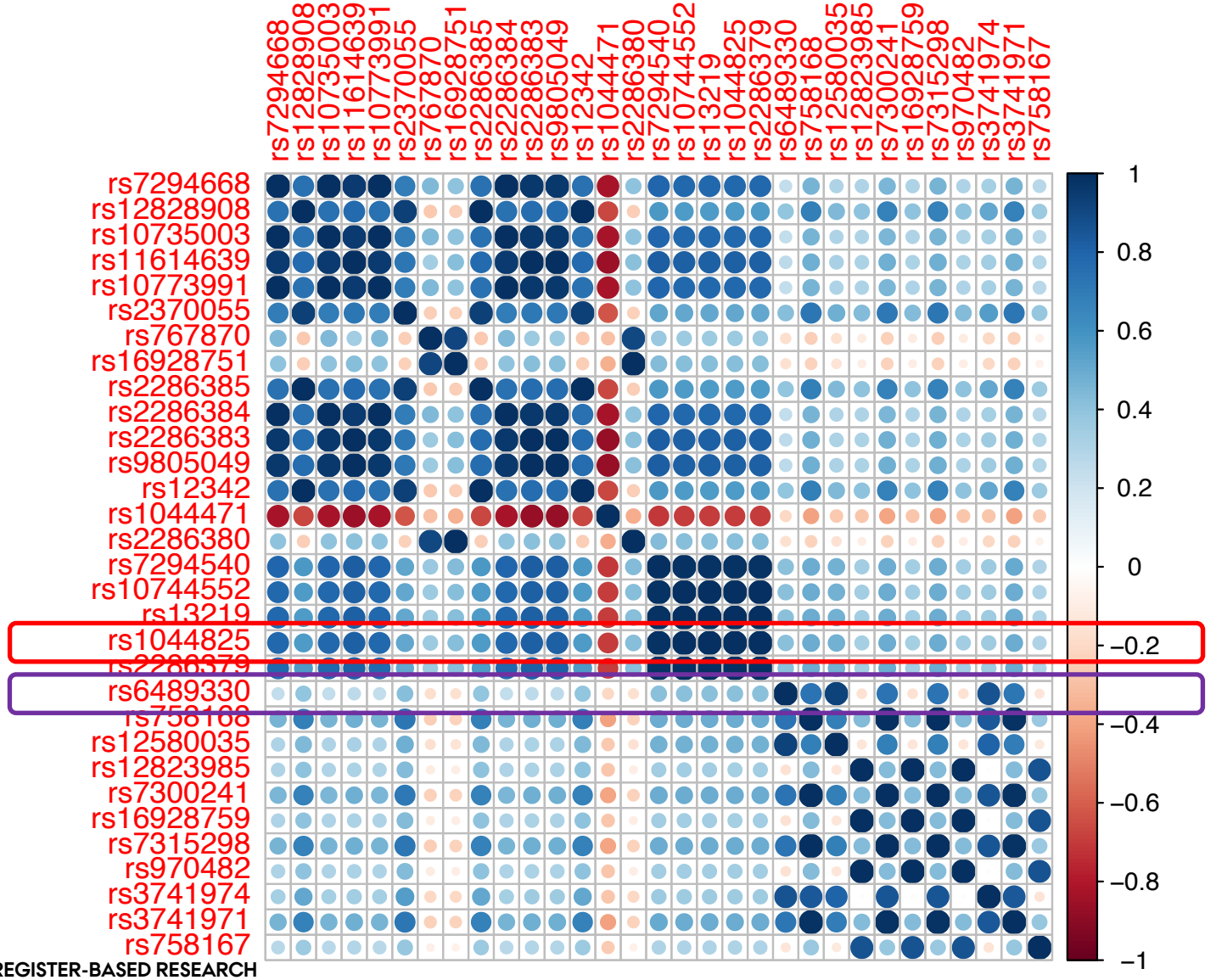
Causality/pleiotropy versus linkage

- Causality/pleiotropy
  - One causal variant
  - No significant difference
- Linkage
  - Multiple causal variants
  - Significant difference
- HEIDI
  - Heterogeneity in Dependent Instruments)





SMR/HEIDI – *CACNA2D4*



The top-associated SNPt

The SNP to test difference

SMR/HEIDI – *CACNA2D4*

SNP	A1 / A2	Data	*b*	SE	*P*-value
rs1044825	G / T	eQTL (blood)	0.447	0.0186	4.1E-128
		GWAS (schizophrenia)	-0.0377	0.0087	1.3E-5
rs6489330	A / G	eQTL (blood)	0.211	0.02384	9.5E-19
LD r = 0.413		GWAS (schizophrenia)	-0.0378	0.0108	4.7E-4

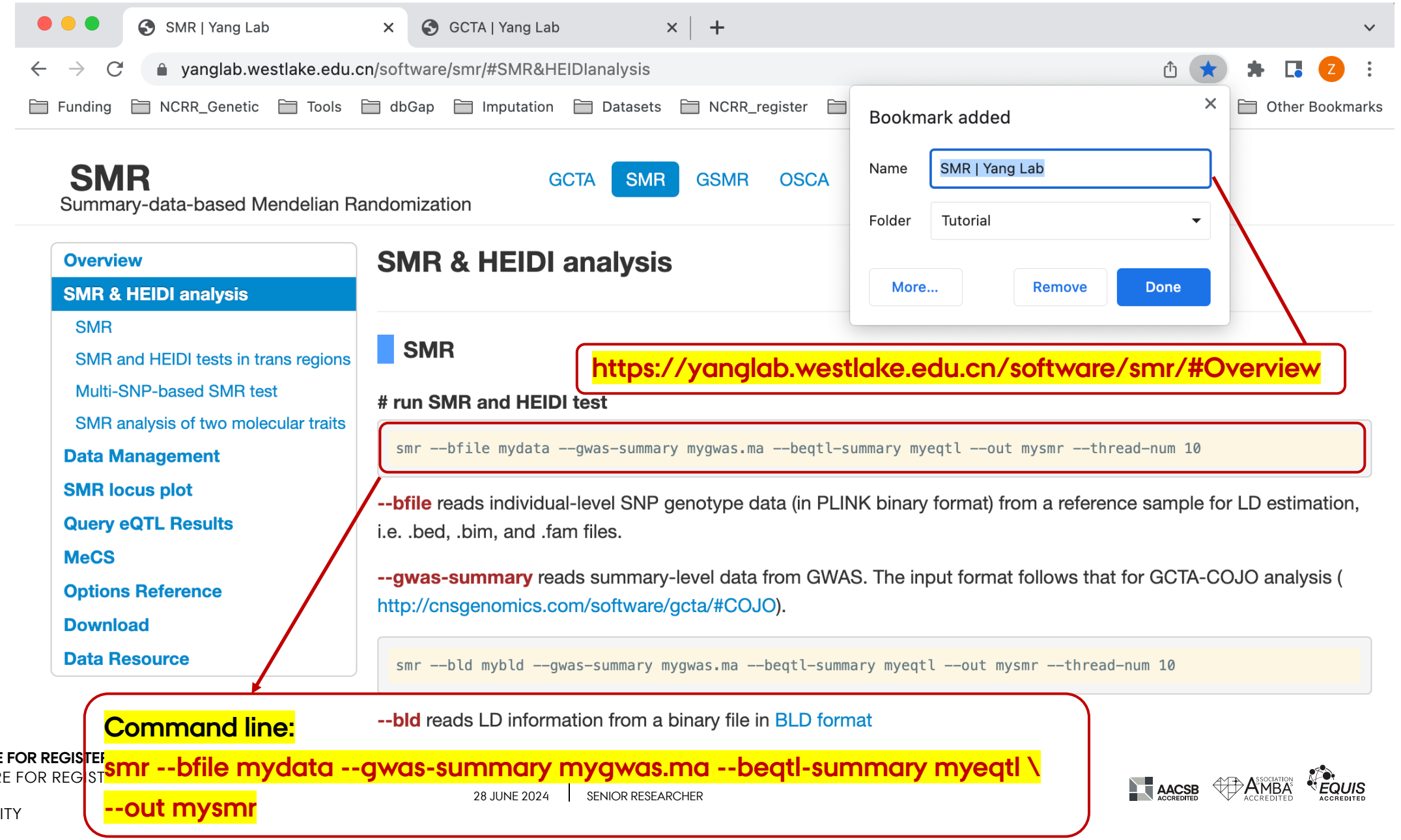
rs1044825, $\hat{\beta}_1 = -0.084$, $\mathrm{SE}(\hat{\beta}_1) \approx 0.020$ rs6489330, $\hat{\beta}_2 = -0.179$, $\mathrm{SE}(\hat{\beta}_2) \approx 0.055$

Difference, $\hat{d} = \hat{\beta}_2 - \hat{\beta}_1 = -0.179 + 0.084 = -0.095$

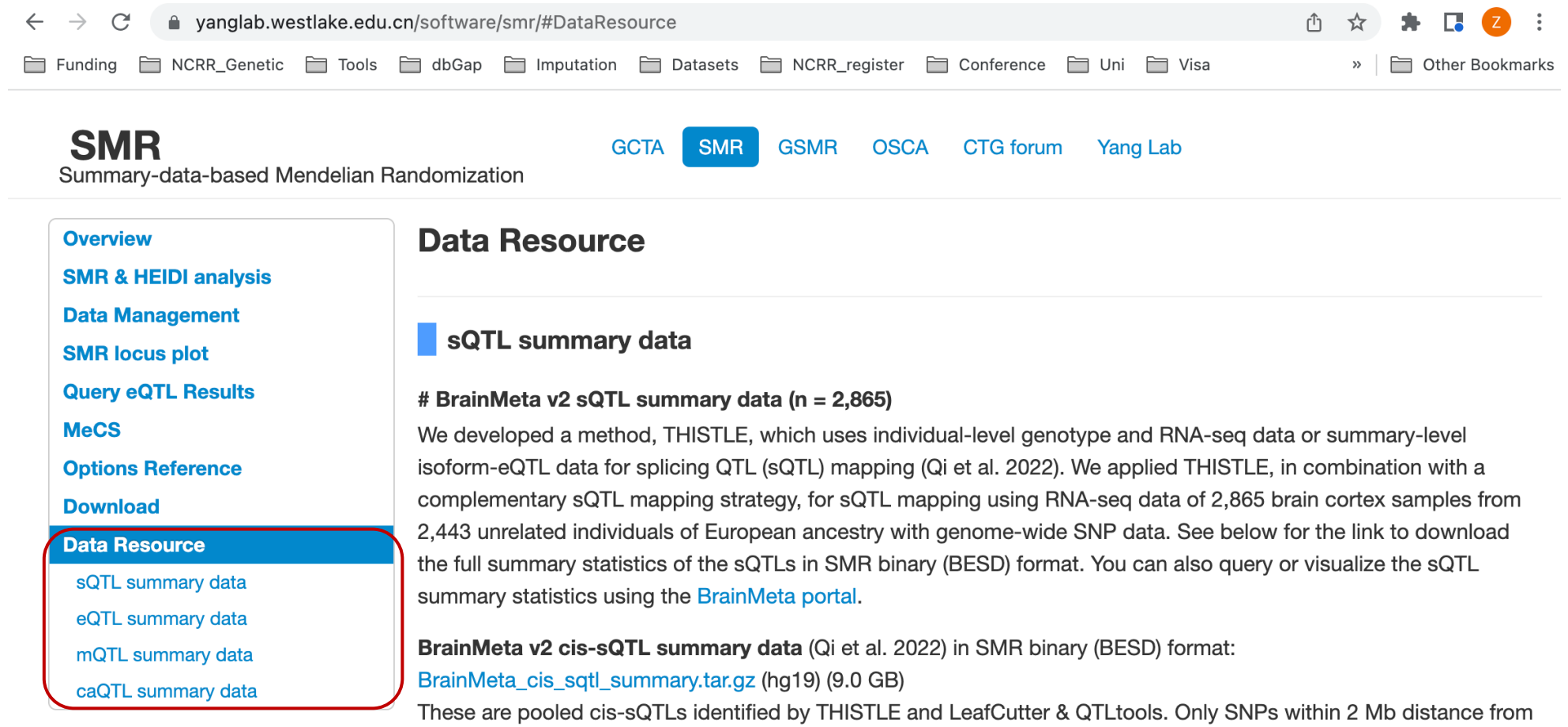
$$\mathrm{SE}(\hat{d}) = \sqrt{\mathrm{var}(\hat{\beta}_2 - \hat{\beta}_1)} = \sqrt{var(\hat{\beta}_2) + var(\hat{\beta}_1) - 2 \times cov(\hat{\beta}_1, \hat{\beta}_2)} = 0.050$$

→ *P*-value = 0.06

Software


SMR | Yang Lab
GCTA | Yang Lab
yanglab.westlake.edu.cn/software/smr/#SMR&HEIDIanalysis
Funding
NCRR_Genetic
Tools
dbGap
Imputation
Datasets
NCRR_register
Other Bookmarks
Bookmark added
Name
SMR | Yang Lab
Folder
Tutorial
More...
Remove
Done
SMR
Summary-data-based Mendelian Randomization
GCTA
SMR
GSMR
OSCA
Overview
SMR & HEIDI analysis
SMR
SMR and HEIDI tests in trans regions
Multi-SNP-based SMR test
SMR analysis of two molecular traits
Data Management
SMR locus plot
Query eQTL Results
MeCS
Options Reference
Download
Data Resource
SMR & HEIDI analysis
SMR
https://yanglab.westlake.edu.cn/software/smr/#Overview
run SMR and HEIDI test
smr --bfile mydata --gwas-summary mygwas.ma --beqtl-summary myeqtl --out mysmr --thread-num 10
--bfile reads individual-level SNP genotype data (in PLINK binary format) from a reference sample for LD estimation, i.e. .bed, .bim, and .fam files.
--gwas-summary reads summary-level data from GWAS. The input format follows that for GCTA-COJO analysis (http://cnsgenomics.com/software/gcta/#COJO).
smr --bld mybld --gwas-summary mygwas.ma --beqtl-summary myeqtl --out mysmr --thread-num 10
--bld reads LD information from a binary file in BLD format
Command line:
smr --bfile mydata --gwas-summary mygwas.ma --beqtl-summary myeqtl \
--out mysmr

Data resource



sQTL – Summary statistics of splicing QTLs

eQTL – Summary statistics from associations of gene expression

mQTL – Summary statistics from associations of methylation

…

Summary

- Mendelian randomisation is a method for causal inference.
- Mendelian randomisation uses DNA variants as instrumental variables.
- SMR assumes a single underlying causal variant (e.g. gene expression etc.).
- HEIDI can identify potential linkage SNPs.
- SMR&HEIDI can be used to identify gene targets.
- Software - SMR

NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

AARHUS UNIVERSITY