Integrating Multi-omics Summary Data Using Mendelian Randomization

Chong Jin
Assistant Professor, Department of Mathematical Sciences
New Jersey Institute of Technology

Apr 13, 2023

Outline

1. Mendelian randomization (MR)
2. MR methods for multi-omics summary statistics
3. Methods in *p*-value combinations
4. Simulations
5. Analyzing an Alzheimer's disease dataset

Outline

1. Mendelian randomization (MR)
2. MR methods for multi-omics summary statistics
3. Methods in *p*-value combinations
4. Simulations
5. Analyzing an Alzheimer's disease dataset

Finding a causal relationship

Instrumental variable analysis:

Natural experiment

“encouragement”

Example:

We have subjects with lung disease. Some subjects are randomly selected and encouraged to exercise.

We record if the exercise is performed and compute the effect of exercise on outcome.

Rosenbaum, Paul R. “Identification of Causal Effects Using Instrumental Variables: Comment.” *Journal of the American Statistical Association* **91**, no. 434 (1996): 465–68. https://doi.org/10.2307/2291633.

Finding a causal relationship

Mendelian randomization publications using UKBioBank:

ukbiobank.ac.uk/enable-your-research/publications?terms=causal&query=mendelian+randomization#articles

International Journal of Cancer, July 29th 2022
The causal association between smoking, alcohol consumption and risk of bladder cancer: A univariable and multivariable Mendelian randomization study
J Xiong et al

ESC Heart Failure, June 28th 2022
Causal associations between sleep traits and four cardiac diseases: a Mendelian randomization study
Y Yang et al

Frontiers in Nutrition, June 22nd 2022
Causal Relationship of Genetically Predicted Serum Micronutrients Levels With Sarcopenia: A Mendelian Randomization Study
T Sha et al

The Journal of Clinical Endocrinology and Metabolism, June 15th 2022
The Causal Effect of Systolic Blood Pressure Lowering on Vascular Outcomes in Diabetes: A Mendelian Randomization Study
T Hou et al

BMC Psychiatry , June 15th 2022
The causal relationship between sleep traits and the risk of schizophrenia: a two-sample bidirectional Mendelian randomization study
Z Wang et al

Search Publications:

Enter search term

Reset filters

Search Terms

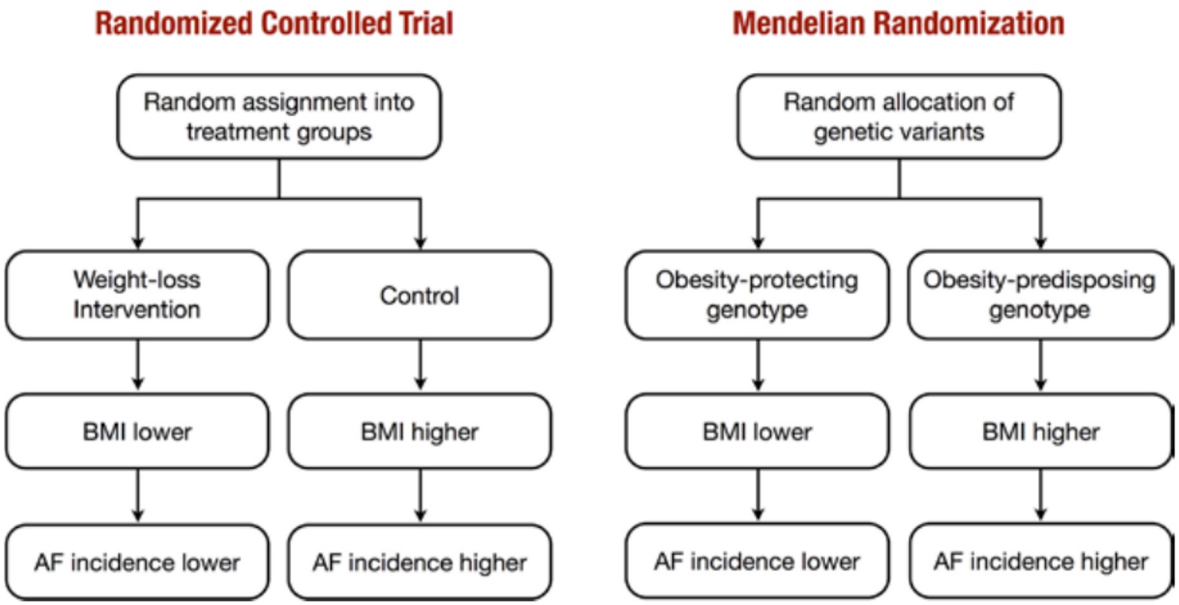
CAUSAL x | MENDELIAN RANDOMIZATION x

Year

2022 (24) | 2021 (29) | 2020 (14) | 2019 (7) | 2018 (3) | 2017 (1)

Finding a causal relationship

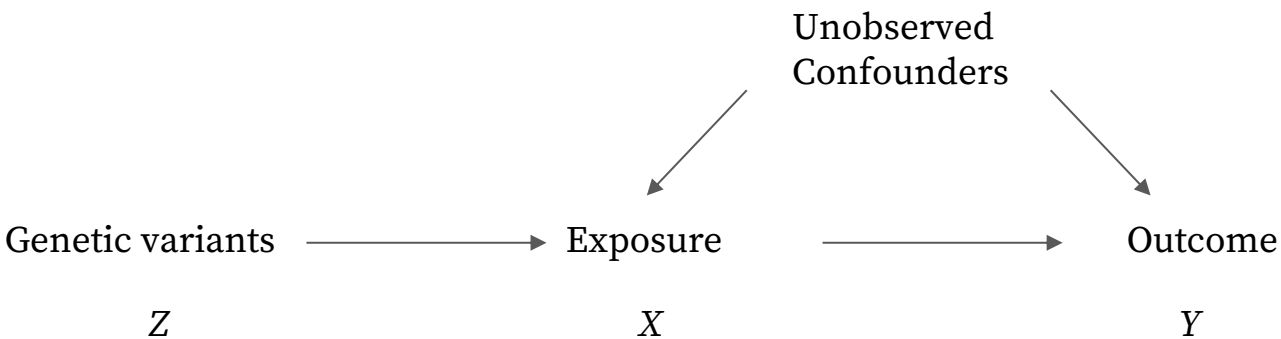
How to find if a higher BMI will cause a higher AF incidence?



AF, atrial fibrillation; BMI, body mass index.

Neeland, I. J. & Kozlitina, J. Mendelian Randomization. *Circulation* **135**, 755–758 (2017).

Finding a causal relationship

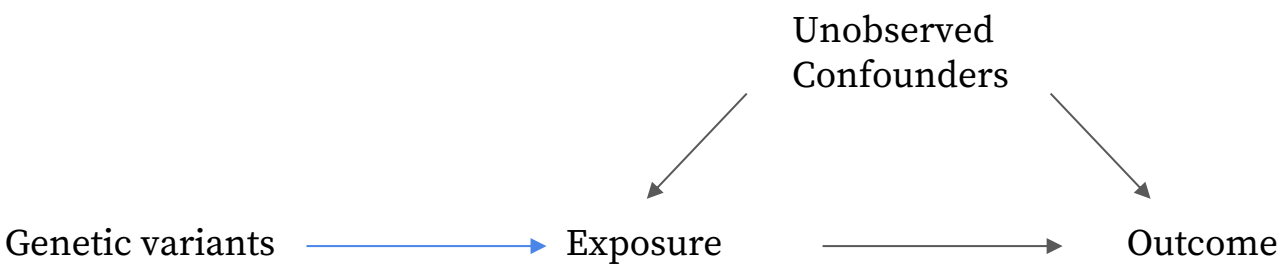


Goal:
Finding casual effect of the exposure on the outcome.

We only have observational data, and a regression of outcome on exposure will be biased since unobserved confounders are unadjusted for.

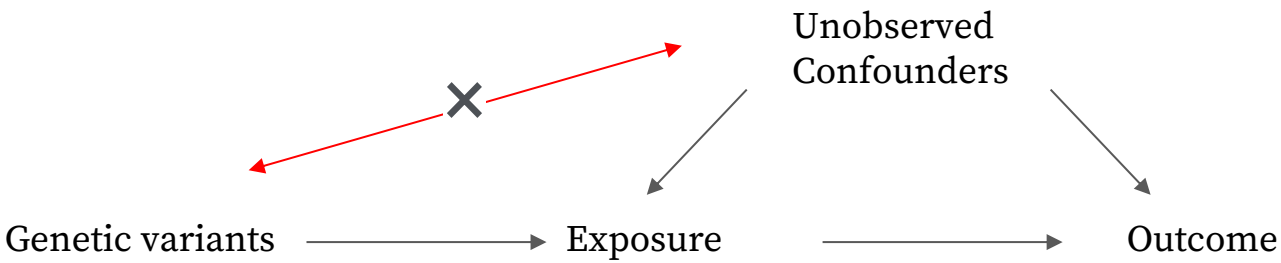
In Mendelian randomization, we will use genetic variants as proxies for the exposure.

Mendelian randomization



Assumptions:
(i) The genetic variant is associated with the exposure.

Mendelian randomization

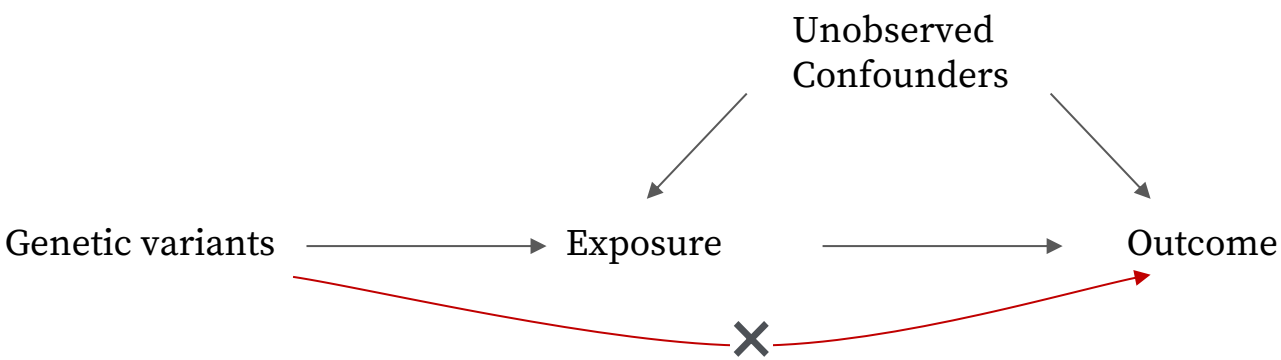


Assumptions:

(i) The genetic variant is associated with the exposure.

(ii) The genetic variant cannot be associated with any confounder that lies within the exposure-outcome relationship.

Mendelian randomization



Assumptions:

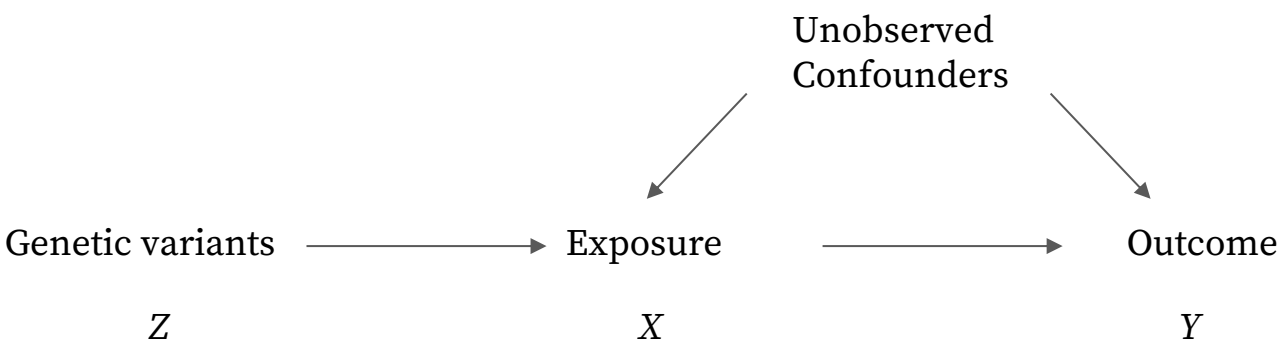
(i) The genetic variant is associated with the exposure.

(ii) The genetic variant cannot be associated with any confounder that lies within the exposure-outcome relationship.

(iii) The genetic variant cannot be associated with the outcome through any pathway other than through the exposure in question.

(Usually, a fourth assumption of linearity or monotonicity is needed for identifiability)

Individual level data vs. summary statistics



Individual-level data:

Individual-level genetic variants Z_i, exposure X_i, and outcome Y_i are available.

Summary statistics:

Summary statistics of regressing exposure on genetic variants $X \sim Z$ and regressing outcome on genetic variants $Y \sim Z$ are available.

General guidelines

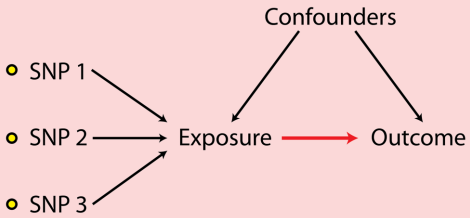
STROBE-MR Guidelines (BMJ 2021)

Guidelines for performing Mendelian randomization investigations (Wellcome Open Research 2020)

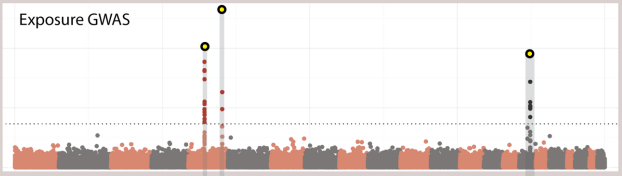
Mendelian randomization (Nature Reviews Methods Primers 2022)

Using the TwoSampleMR package

Objective: Infer the causal effect of the exposure on the outcome



1.



Description

Define instruments: Obtain SNPs that are GWAS significant for the exposure. Ensure that they are independent.

Instruments can be defined from a variety of different sources.

R commands

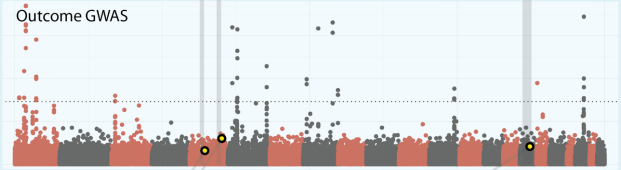
```
extract_instruments()

read_exposure_data()

library(MRInstruments);
  data(gwas_catalog);
  data(aries_mqtl);
  data(gtex_eqtl);
  data(proteomic_qtls);
  data(metab_qtls);

clump_data()
```

2.



Get effects on outcome: Extract the instrument SNPs from the outcome GWAS. If they are not available, use LD proxies instead.

MR Base contains a large database of entire GWAS summary statistics.

```
extract_outcome_data()

read_outcome_data()
```


Gibran Hemani, https://mrcieu.github.io/TwoSampleMR/articles/introduction.html

Using the TwoSampleMR package

3.

	Exposure GWAS				Outcome GWAS			
SNP	*Effect*	*Effect allele*	*Other allele*	*Effect allele frequency*	*Effect*	*Effect allele*	*Other allele*	*Effect allele frequency*
rs123456	0.132	A	G	0.28	0.022	A	G	0.26
rs234567	-0.485	G	T	0.41	**0.056**	**T**	**G**	**0.61**
rs345678	0.203	G	C	0.11	**-0.046**	**G**	**C**	**0.88**

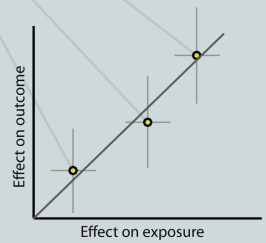
Harmonize effects

	Exposure GWAS				Outcome GWAS			
SNP	*Effect*	*Effect allele*	*Other allele*	*Effect allele frequency*	*Effect*	*Effect allele*	*Other allele*	*Effect allele frequency*
rs123456	0.132	A	G	0.28	0.022	A	G	0.26
rs234567	-0.485	G	T	0.41	**-0.056**	**G**	**T**	**0.39**
rs345678	0.203	G	C	0.11	**0.046**	**G**	**C**	**0.12**

Harmonize effects: Ensure that the effect of the SNP on the exposure and the effect of the SNP on the outcome correspond to the same allele.

```
harmonise_data()
```

4.



Perform analysis: Using the harmonized data, perform Mendelian randomization analyses and related sensitivity analyses.

The slope of the regression line corresponds to the causal effect of the exposure on the outcome

```
mr()
mr_singlesnp()
mr_leaveoneout()
mr_heterogeneity()
mr_steiger()
mr_pleiotropy_test()
```


Gibran Hemani, https://mrcieu.github.io/TwoSampleMR/articles/introduction.html

Association between gene expression and phenotype through genotypes

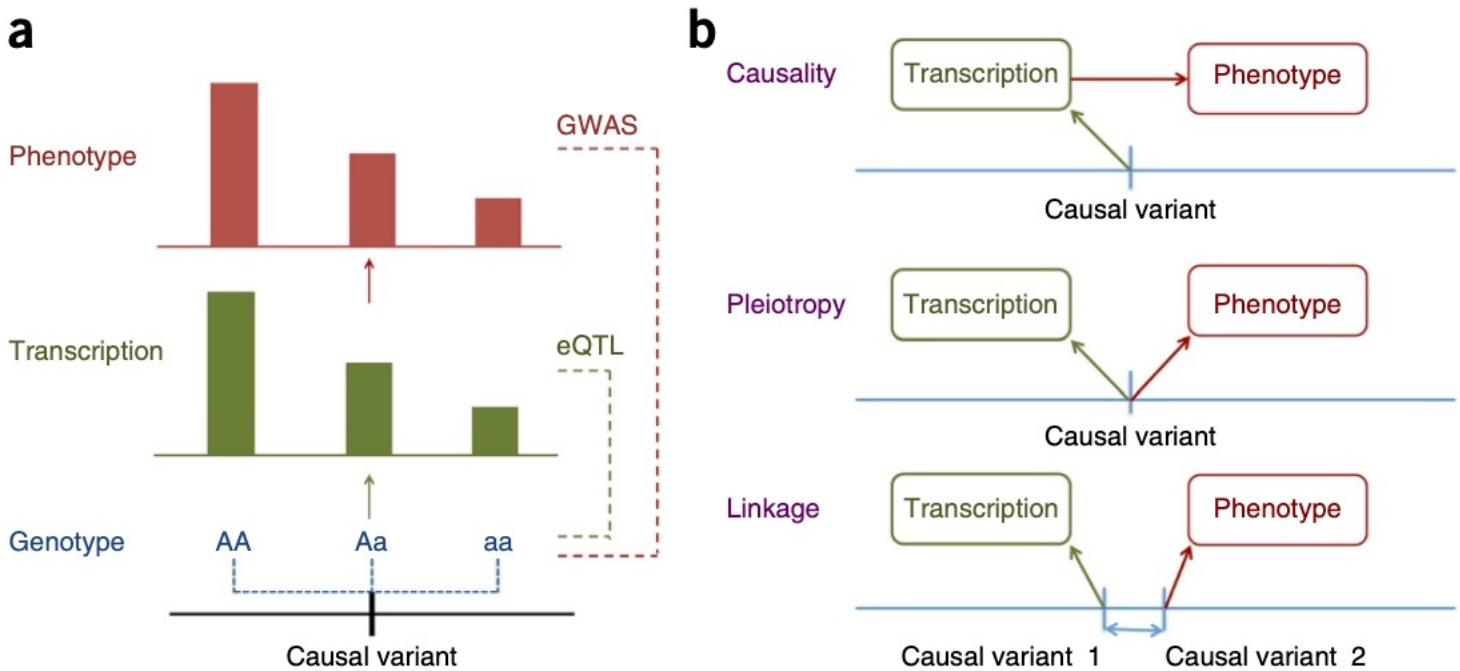


Figure 1 of Zhu, Zhihong, et al. "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets." Nature genetics 48.5 (2016): 481-487.

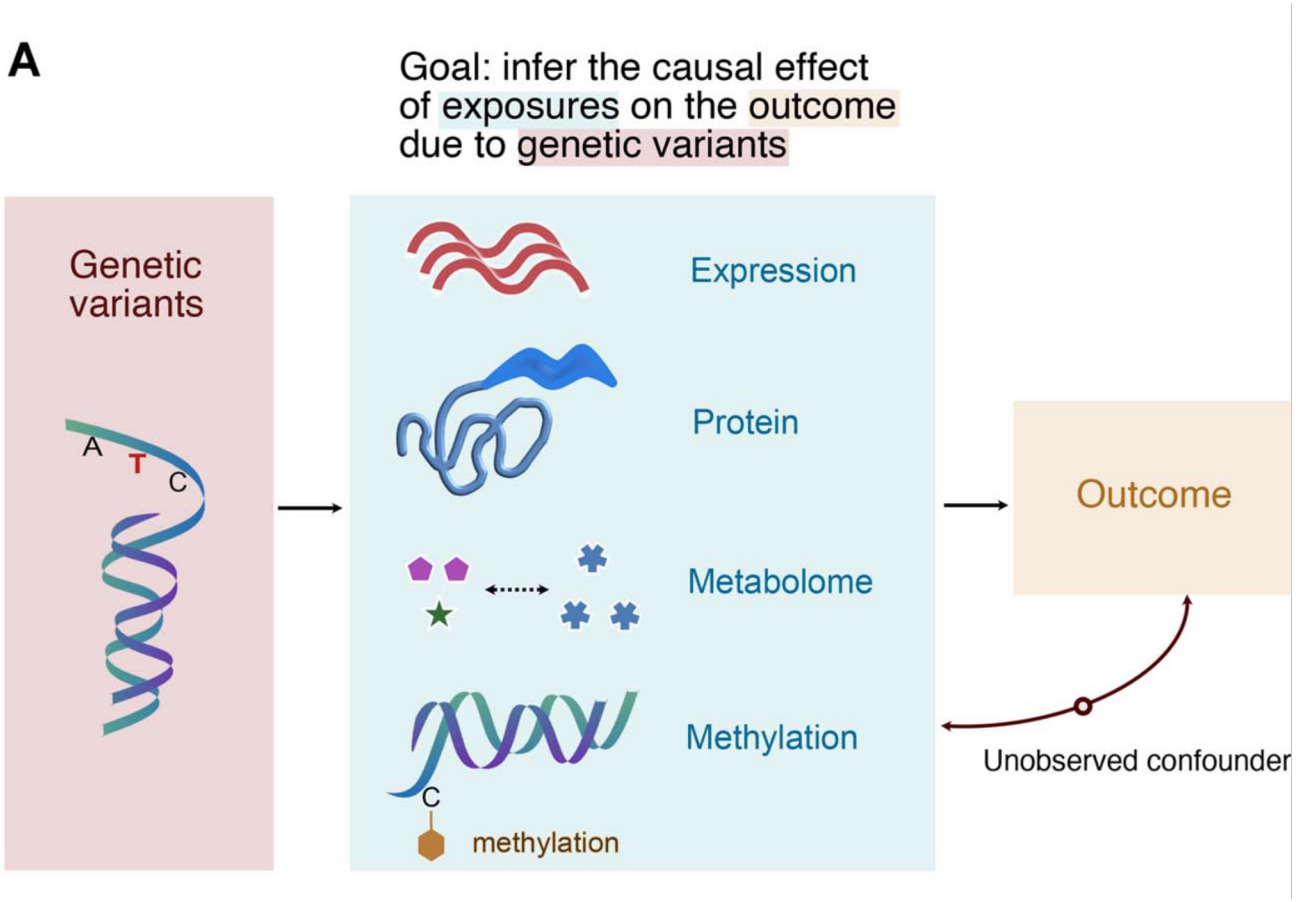
Outline

1. Mendelian randomization (MR)
2. MR methods for multi-omics summary statistics
3. Methods in *p*-value combinations
4. Simulations
5. Analyzing an Alzheimer's disease dataset

Alzheimer's disease

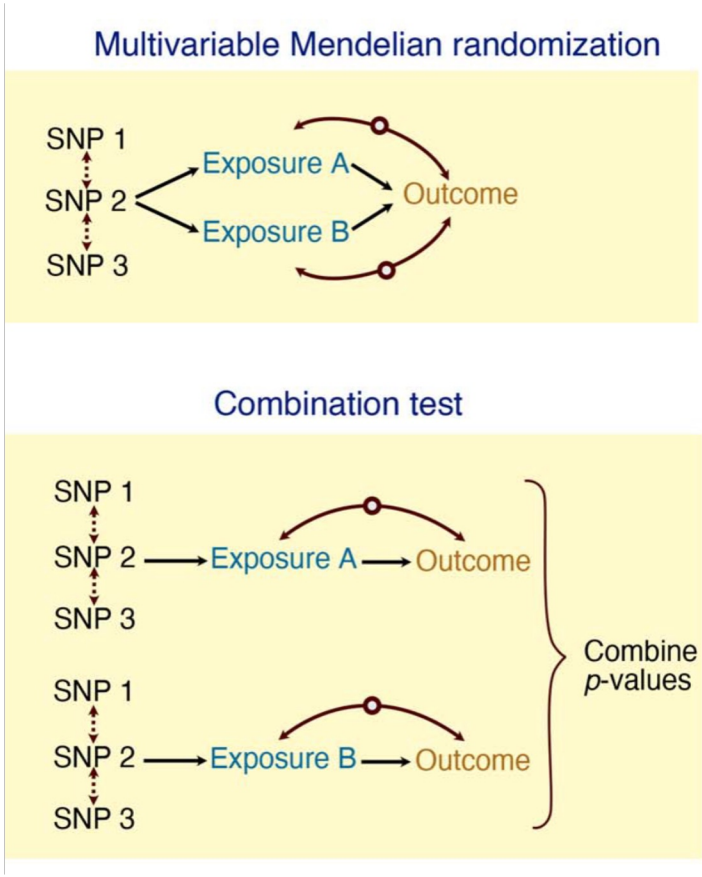
- In 2019, the estimated total worldwide cost of dementia was US$ 1.3 trillion (WHO report)
- Agora is compiling a "Nominated Target List" as investigators nominate genes that are potentially good targets from human genomic, proteomic and/or metabolomic data

Multi-omics Mendelian randomization



Multi-omics Mendelian randomization

In the project, we aim to evaluate which is the better strategy:



Inverse variance-weighted average method

When the genetic variants are uncorrelated, the IVW estimate of the causal effect size of the outcome on the exposure k is

$$\hat{\theta}_{\mathrm{IVW},k} = \frac{\sum_{j=1}^{p} \widehat{\beta}_{Xkj} \widehat{\beta}_{Yj} \widehat{\sigma}_{Yj}^{-2}}{\sum_{j=1}^{p} \widehat{\beta}_{Yj}^{2} \widehat{\sigma}_{Yj}^{-2}},$$

where the standard error of the estimate is

$$\mathrm{se}(\hat{\theta}_{\mathrm{IVW},k}) = \sqrt{\frac{1}{\sum_{j=1}^{p} \widehat{\beta}_{Yj}^{2} \widehat{\sigma}_{Yj}^{-2}}}.$$

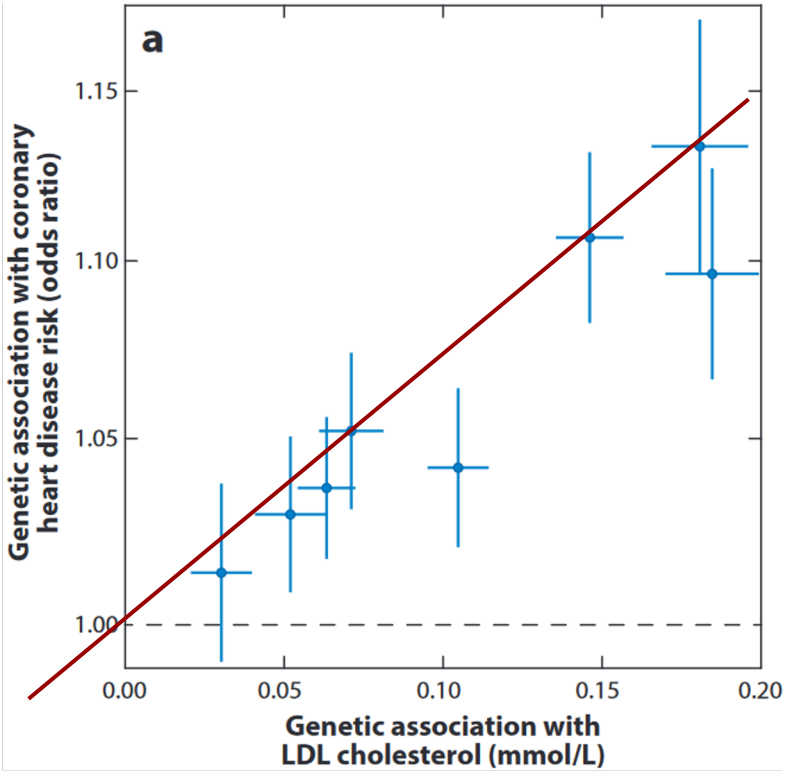


Figure from: Burgess, S., Foley, C. N. & Zuber, V. Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. *Annual Review of Genomics and Human Genetics* **19**, 303–327 (2018).

Generalized least squares method

The generalized least squares method (GLS) builds upon IVW and accounts for remaining LD within SNPs after LD clumping

$$\widehat{\theta}_{\mathrm{GLS},k} = (\widehat{\boldsymbol{\beta}}_{Xk}^{T}\Omega^{-1}\widehat{\boldsymbol{\beta}}_{Xk})^{-1}(\widehat{\boldsymbol{\beta}}_{Xk}^{T}\Omega^{-1}\widehat{\boldsymbol{\beta}}_{Y}),$$

$$\mathbf{var}\left(\widehat{\theta}_{\mathrm{GLS},k}\right) = (\widehat{\boldsymbol{\beta}}_{Xk}^{T}\Omega^{-1}\widehat{\boldsymbol{\beta}}_{Xk})^{-1},$$

where the covariance matrix Ω is constructed from the unsquared LD matrix R

$$\Omega = (\widehat{\sigma}_{Yi}\widehat{\sigma}_{Yj}r_{ij}) = \begin{bmatrix} \widehat{\sigma}_{Y1} & & \\ & \ddots & \\ & & \widehat{\sigma}_{Yp} \end{bmatrix} R \begin{bmatrix} \widehat{\sigma}_{Y1} & & \\ & \ddots & \\ & & \widehat{\sigma}_{Yp} \end{bmatrix},$$

and r_{ij}'s are the entries in the unsquared LD matrix R.

Summary data-based MR

Applicable when we have only one SNP as instrumental variable

Under the assumption that the GWAS z-value z_Y and QTL z-value z_{Xk} are estimated using independent samples, we have an approximate χ^2 test statistic with one degree of freedom

$$T_{\mathrm{SMR},k} = \frac{z_Y^2 z_{Xk}^2}{z_Y^2 + z_{Xk}^2}.$$

Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nature Genetics 48, 481–487 (2016).

Generalized summary data-based MR

$$\widehat{\boldsymbol{\theta}}_k = (\widehat{\theta}_{k1}, \widehat{\theta}_{k2}, \ldots, \widehat{\theta}_{km})^{\mathrm{T}} = \left(\frac{\widehat{\beta}_{\mathrm{Y}1}}{\widehat{\beta}_{\mathrm{X}k1}}, \frac{\widehat{\beta}_{\mathrm{Y}2}}{\widehat{\beta}_{\mathrm{X}k2}}, \ldots, \frac{\widehat{\beta}_{\mathrm{Y}p}}{\widehat{\beta}_{\mathrm{X}kp}} \right)^{\mathrm{T}},$$

$$\mathrm{var}\left(\widehat{\theta}_{kj}\right) = \frac{\widehat{\beta}_{\mathrm{Y}j}^2}{\widehat{\beta}_{\mathrm{X}kj}^2} \left[\frac{\mathrm{var}(\widehat{\beta}_{\mathrm{X}kj})}{\widehat{\beta}_{\mathrm{X}kj}^2} + \frac{\mathrm{var}(\widehat{\beta}_{\mathrm{Y}j})}{\widehat{\beta}_{\mathrm{Y}j}^2} - \frac{\mathrm{var}(\widehat{\beta}_{\mathrm{X}kj})^2}{\widehat{\beta}_{\mathrm{X}kj}^4} \right],$$

$$\mathrm{cov}\left(\widehat{\theta}_{ki}, \widehat{\theta}_{kj}\right) = \frac{r_{ij}\sqrt{\mathrm{var}(\widehat{\beta}_{\mathrm{Y}i})\mathrm{var}(\widehat{\beta}_{\mathrm{Y}j})}}{\widehat{\beta}_{\mathrm{X}ki}\widehat{\beta}_{\mathrm{X}kj}} + \frac{\widehat{\beta}_{\mathrm{Y}i}\widehat{\beta}_{\mathrm{Y}j}}{\widehat{\beta}_{\mathrm{X}ki}\widehat{\beta}_{\mathrm{X}kj}}$$

$$\left[\frac{r_{ij}\sqrt{\mathrm{var}(\widehat{\beta}_{\mathrm{X}ki})\mathrm{var}(\widehat{\beta}_{\mathrm{X}kj})}}{\widehat{\beta}_{\mathrm{X}ki}\widehat{\beta}_{\mathrm{X}kj}} - \frac{\mathrm{var}(\widehat{\beta}_{\mathrm{X}ki})\mathrm{var}(\widehat{\beta}_{\mathrm{X}kj})}{\widehat{\beta}_{\mathrm{X}ki}^2\widehat{\beta}_{\mathrm{X}kj}^2} \right],$$

$$\mathrm{V}_k = \left(\mathrm{cov}\left(\widehat{\theta}_{ki}, \widehat{\theta}_{kj}\right)\right)_{p\times p}.$$

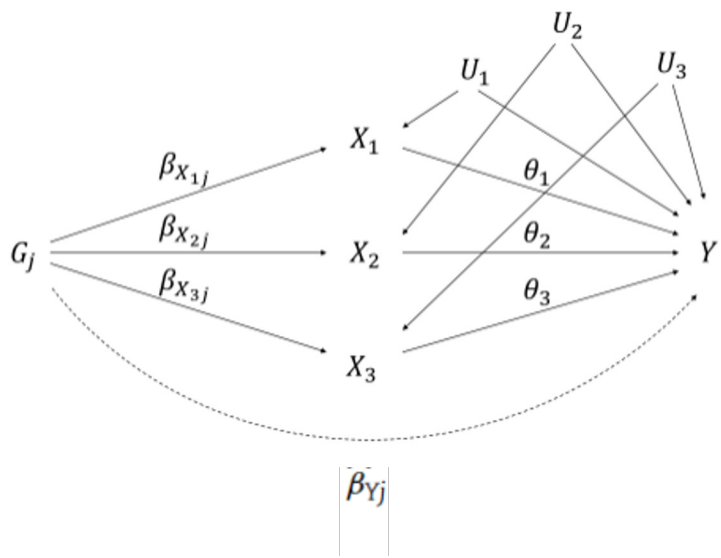
$$\widehat{\theta}_{\mathrm{GSMR},k} = (\mathbf{1}^{\mathrm{T}}\mathrm{V}_k^{-1}\mathbf{1})^{-1}\mathbf{1}^{\mathrm{T}}\mathrm{V}_k^{-1}\widehat{\boldsymbol{\theta}}_{\mathbf{k}},$$

$$\mathrm{var}\left(\widehat{\theta}_{\mathrm{GSMR},k}\right) = (\mathbf{1}^{\mathrm{T}}\mathrm{V}_k^{-1}\mathbf{1})^{-1},$$

$$T_{\mathrm{GSMR},k} = \widehat{\theta}_{\mathrm{GSMR},k}^2 / \mathrm{var}\left(\widehat{\theta}_{\mathrm{GSMR},k}\right).$$

Multivariable MR

$$\widehat{\beta}_{Yj} = \sum_{k=1}^{m} \theta_k \widehat{\beta}_{Xkj} + \epsilon, \quad \textbf{weights} = \widehat{\sigma}_{Yj}^{-2},$$



Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Statistics in Medicine* **36**, 4705–4718 (2017).

Outline

1. Mendelian randomization (MR)
2. MR methods for multi-omics summary statistics
3. Methods in *p*-value combinations
4. Simulations
5. Analyzing an Alzheimer's disease dataset

Methods in p-value combinations

Cauchy combination test:

Cauchy distribution

Harmonic mean p-value:

Stable distribution (1,1)

$$T_{\text{Cauchy}} = \sum_{k=1}^{m} w_k \tan\{(0.5 - p_k)\pi\}$$

$$T_{\text{HMP}} = \frac{\sum_{k=1}^{m} w_k}{\sum_{k=1}^{m} w_k / p_k}$$

The p-values do not need to be independent in Cauchy combination test or HMP.

Methods in p-value combinations

Cauchy combination test:

Cauchy distribution

Harmonic mean p-value:

Stable distribution (1,1)

Minimum p-value:

Uniform distribution

Fisher combination test:

(Fisher_chisq) chi-squared distribution

(Fisher_gamma) gamma distribution to approximate the null distribution can *account for the correlation between p-values*

$$T_{\text{Cauchy}} = \sum_{k=1}^{m} w_k \tan\{(0.5 - p_k)\pi\}$$

$$T_{\text{HMP}} = \frac{\sum_{k=1}^{m} w_k}{\sum_{k=1}^{m} w_k / p_k}$$

$$T_{\text{MinP}} = \min_{k \in \{1 \ldots m\}} p_k$$

$$T_{\text{Fisher}} = -2 \sum_{k=1}^{m} \ln(p_k)$$

Methods in p-value combinations

Cauchy combination test:

Cauchy distribution

$$T_{\text{Cauchy}} = \sum_{k=1}^{m} w_k \tan\{(0.5 - p_k)\pi\}$$

Harmonic mean p-value:

Stable distribution (1,1)

$$T_{\text{HMP}} = \frac{\sum_{k=1}^{m} w_k}{\sum_{k=1}^{m} w_k / p_k}$$

Minimum p-value:

Uniform distribution

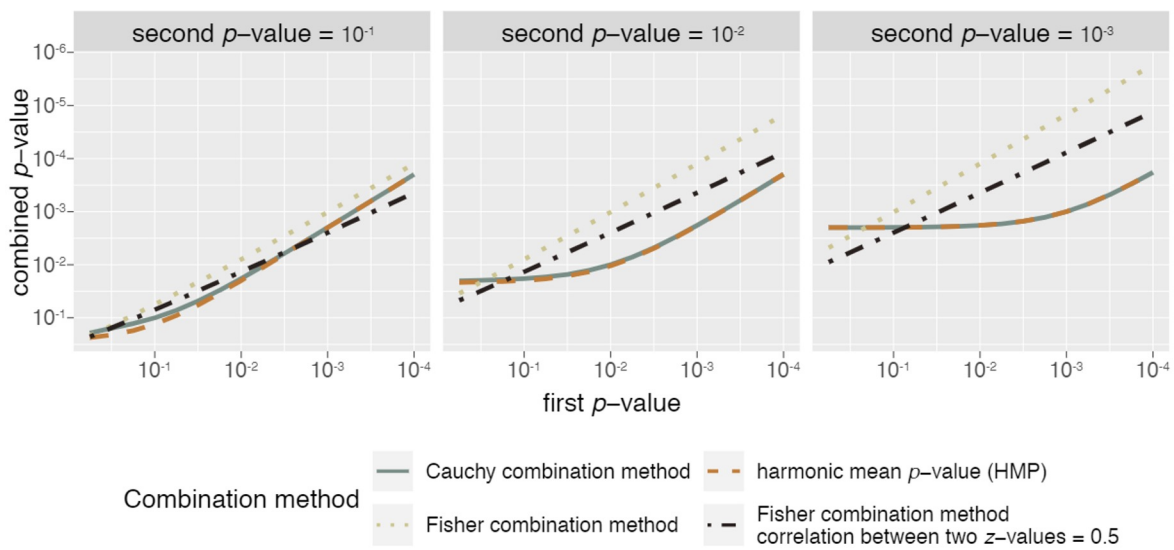
$$T_{\text{MinP}} = \min_{k \in \{1 \ldots m\}} p_k$$

Fisher combination test:

chi-squared distribution

gamma distribution

$$T_{\text{Fisher}} = -2 \sum_{k=1}^{m} \ln(p_k)$$



Fisher combination test: Fisher_gamma

- Plugging in $\boldsymbol{\gamma}_{Xk}$ and $\mathbf{z}_Y$ to the GLS estimator, we have

$$\widehat{\theta}_{\mathrm{GLS},k} = (\widehat{\boldsymbol{\gamma}}_{Xk}^T \mathrm{R}^{-1} \widehat{\boldsymbol{\gamma}}_{Xk})^{-1} (\widehat{\boldsymbol{\gamma}}_{Xk}^T \mathrm{R}^{-1} \mathbf{z}_Y),$$

$$\mathrm{var}\left(\widehat{\theta}_{\mathrm{GLS},k}\right) = (\widehat{\boldsymbol{\gamma}}_{Xk}^T \mathrm{R}^{-1} \widehat{\boldsymbol{\gamma}}_{Xk})^{-1},$$

$$z_{\mathrm{GLS},k} = \frac{\widehat{\theta}_{\mathrm{GLS},k}}{\sqrt{\mathrm{var}\left(\widehat{\theta}_{\mathrm{GLS},k}\right)}},$$

and

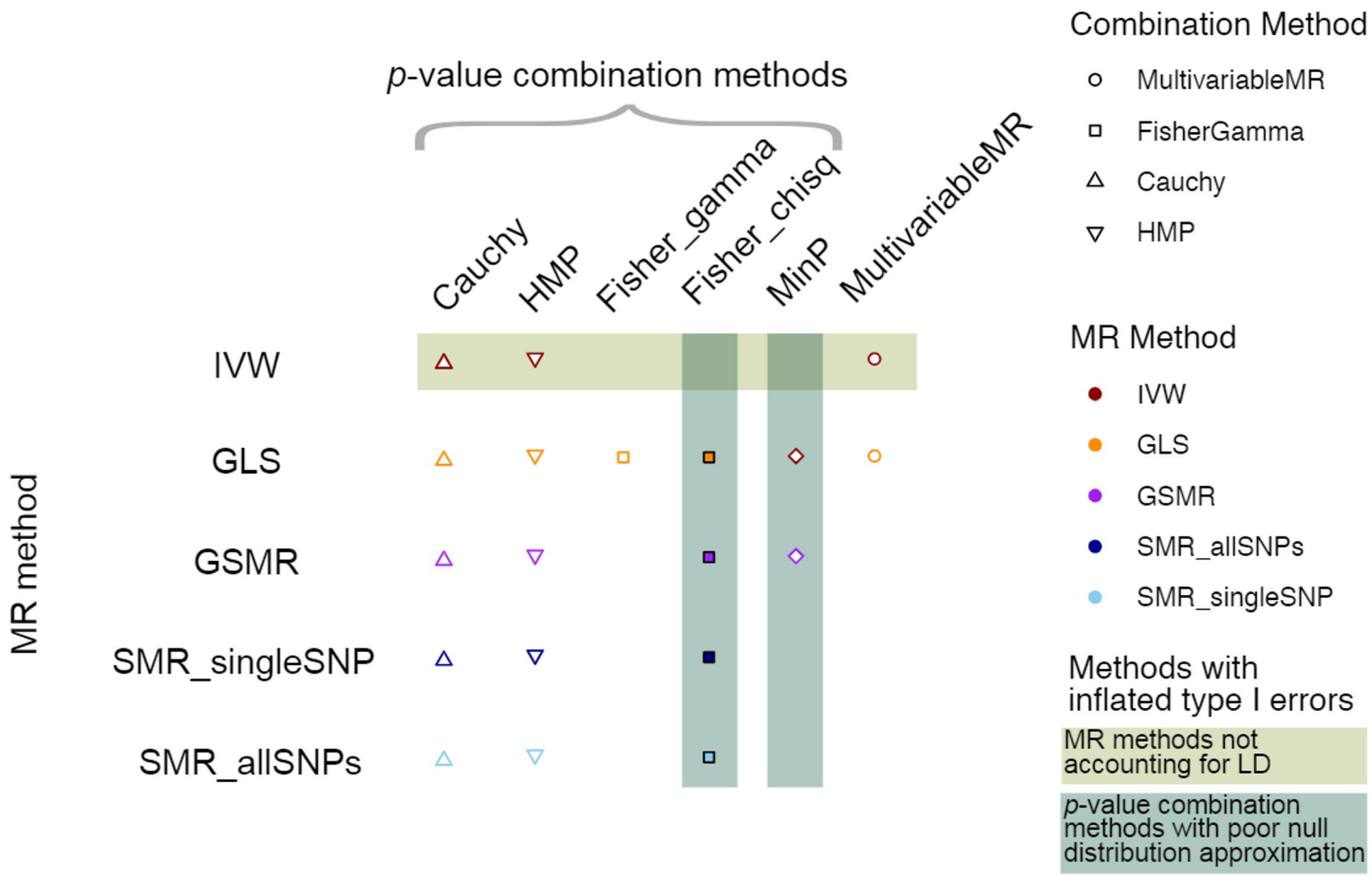
$$\begin{aligned}
&\mathrm{cov}\left(z_{\mathrm{GLS},k_1}, z_{\mathrm{GLS},k_2}\right) \\
=&\mathrm{cov}\left(\frac{\widehat{\boldsymbol{\beta}}_{Xk_1}^T \mathrm{R}^{-1} \mathbf{z}_Y}{\sqrt{\widehat{\boldsymbol{\beta}}_{Xk_1}^T \mathrm{R}^{-1} \widehat{\boldsymbol{\beta}}_{Xk_1}}}, \frac{\widehat{\boldsymbol{\beta}}_{Xk_2}^T \mathrm{R}^{-1} \mathbf{z}_Y}{\sqrt{\widehat{\boldsymbol{\beta}}_{Xk_2}^T \mathrm{R}^{-1} \widehat{\boldsymbol{\beta}}_{Xk_2}}}\right) \\
=&\left(\frac{\widehat{\boldsymbol{\beta}}_{Xk_1}^T \mathrm{R}^{-1}}{\sqrt{\widehat{\boldsymbol{\beta}}_{Xk_1}^T \mathrm{R}^{-1} \widehat{\boldsymbol{\beta}}_{Xk_1}}}\right) \mathrm{R} \left(\frac{\widehat{\boldsymbol{\beta}}_{Xk_2}^T \mathrm{R}^{-1}}{\sqrt{\widehat{\boldsymbol{\beta}}_{Xk_2}^T \mathrm{R}^{-1} \widehat{\boldsymbol{\beta}}_{Xk_2}}}\right)^T.
\end{aligned}$$

Then we can use a Gamma distribution approximation in Yang *et al.* 2016 when the Fisher combination statistic is the sum of dependent chi-squared statistics:

$$T_{\mathrm{Fisher}} = -2 \sum_{k=1}^{m} \ln(p_k)$$

Mendelian randomization (MR) methods and combination methods

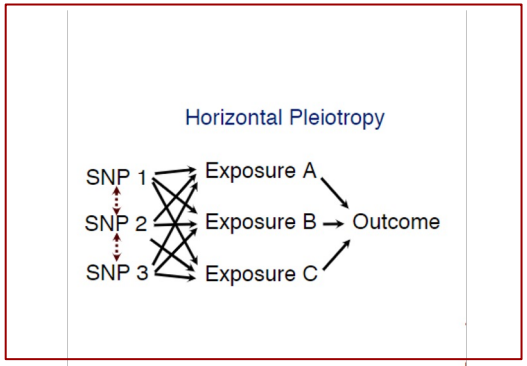
The methods we investigate in the simulation are marked by dots:

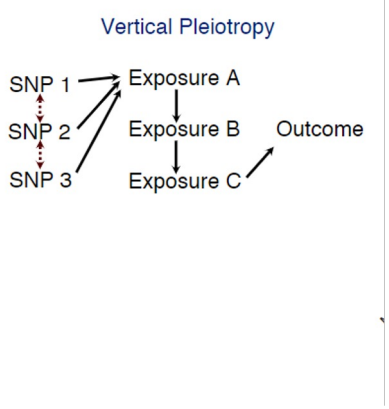


Outline

1. Mendelian randomization (MR)
2. MR methods for multi-omics summary statistics
3. Methods in *p*-value combinations
4. Simulations
5. Analyzing an Alzheimer's disease dataset

Simulation settings: **Horizontal** vs. Vertical Pleiotropy





$$Z_j \sim \text{Binom}(2, 0.3) \text{ for } j = 1, \ldots, p,$$

$$r_{ij}^2 = 0, 0.01, \text{or } 0.2 \text{ for } i \neq j,$$

$$X_1 = \frac{\alpha}{p} \sum_{j=1}^{p} Z_j + U + \epsilon_{X_1},$$

$$X_2 = 2X_1 - U + \epsilon_{X_2},$$

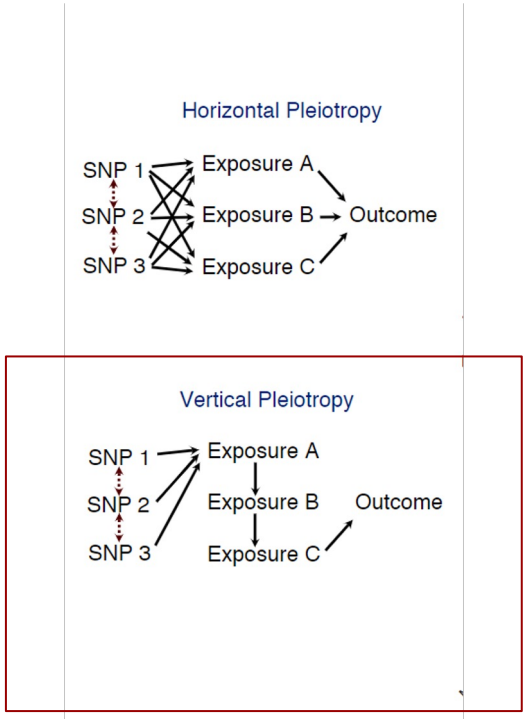
$$X_3 = -0.5X_2 + U + \epsilon_{X_3},$$

$$Y_B \sim \text{Binom}(1, \tau), \text{ where } \text{logit}(\tau) = -2 + \beta_X X_3 - U,$$

$$Y_C = \beta_X X_3 - U + \epsilon_Y,$$

$$U \sim \mathcal{N}(0, 1), \ \epsilon_{X_k} \sim \mathcal{N}(0, 1), \ \epsilon_Y \sim \mathcal{N}(0, 1) \text{ independently}.$$

Simulation settings: Horizontal vs. **Vertical** Pleiotropy



$$Z_j \sim \text{Binom}(2,\ 0.3) \text{ for } j = 1, \ldots, p,$$

$$r_{ij}^2 = 0, 0.01, \text{or } 0.2 \text{ for } i \neq j,$$

$$X_1 = \frac{\alpha}{p} \sum_{j=1}^{p} Z_j + U + \epsilon_{X_1},$$

$$X_2 = -\frac{\alpha}{p} \sum_{j=1}^{\lfloor p/2 \rfloor} Z_j - U + \epsilon_{X_2},$$

$$X_3 = \frac{\alpha}{p} \sum_{j=\lfloor p/2 \rfloor + 1}^{p} Z_j - U + \epsilon_{X_3},$$

$$Y_B \sim \text{Binom}(1, \tau), \text{ where } \text{logit}(\tau) = -2 + \beta_X \sum_{k=1}^{3} X_k + U,$$

$$Y_C = \beta_X \sum_{k=1}^{3} X_k + u_i + \epsilon_Y,$$

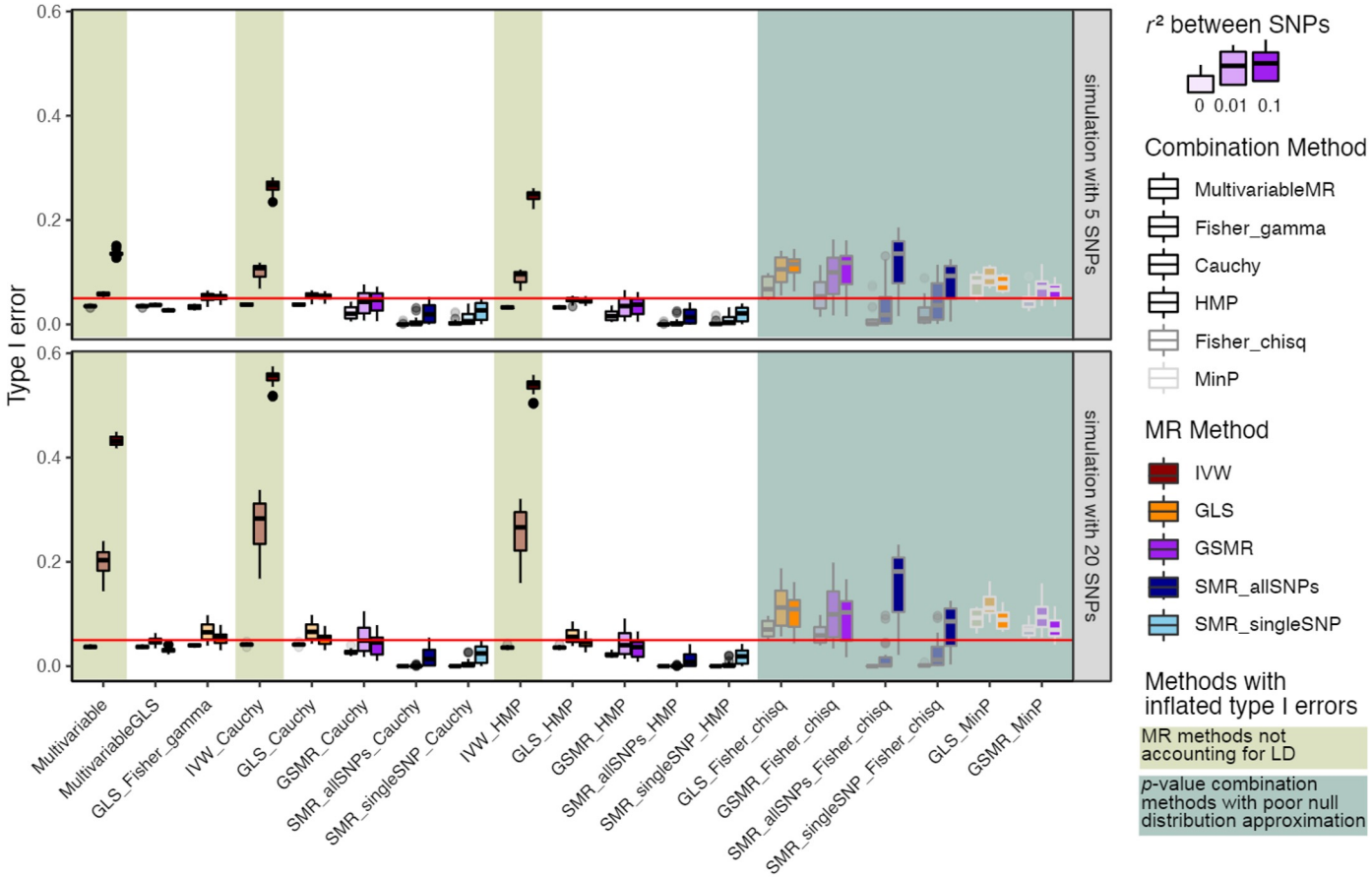
$$U \sim \mathcal{N}(0,1),\ \epsilon_{X_k} \sim \mathcal{N}(0,1),\ \epsilon_Y \sim \mathcal{N}(0,1) \text{ independently}.$$

Simulation settings

- (i) Pleiotropy: horizontal or vertical;
- (ii) Number of SNPs used as IVs: 5 or 20;
- (iii) Overlap between samples involved in the GWAS studies and the QTL studies: 0 (two-sample), 0.5 (half of the samples overlap) or 1 (one-sample);
- (iv) Use same samples or independent samples to calculate QTLs belonging to multiple omics biomarkers under a two-sample setting;
- (v) Outcome: continuous (Y_C) or binary (Y_B). When the outcome is binary, we also need to decide whether we will only use control samples to estimate the QTLs;
- (vi) LD between different SNPs r_2: 0, 0.01 or 0.2;
- (vii) Strength of IV α, which is proportional to the association between IVs and exposures: 0.5, 1 or 2;
- (viii) Effect size of the pleiotropic association between exposure and outcome β_X: 0 or 0.1.

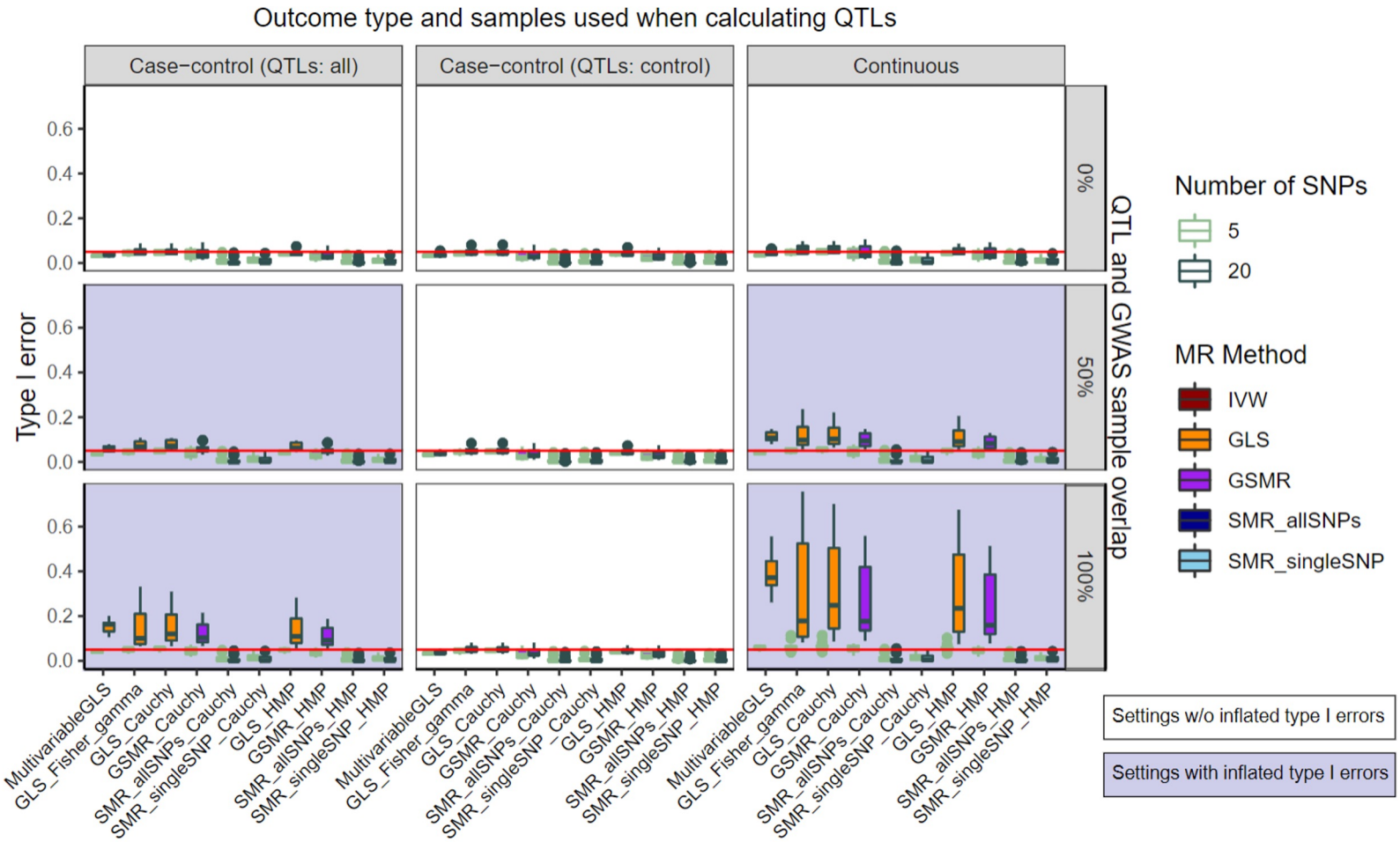
A comparison of the type I error across methods

The unshaded methods can control type I error for a large number of simulation settings

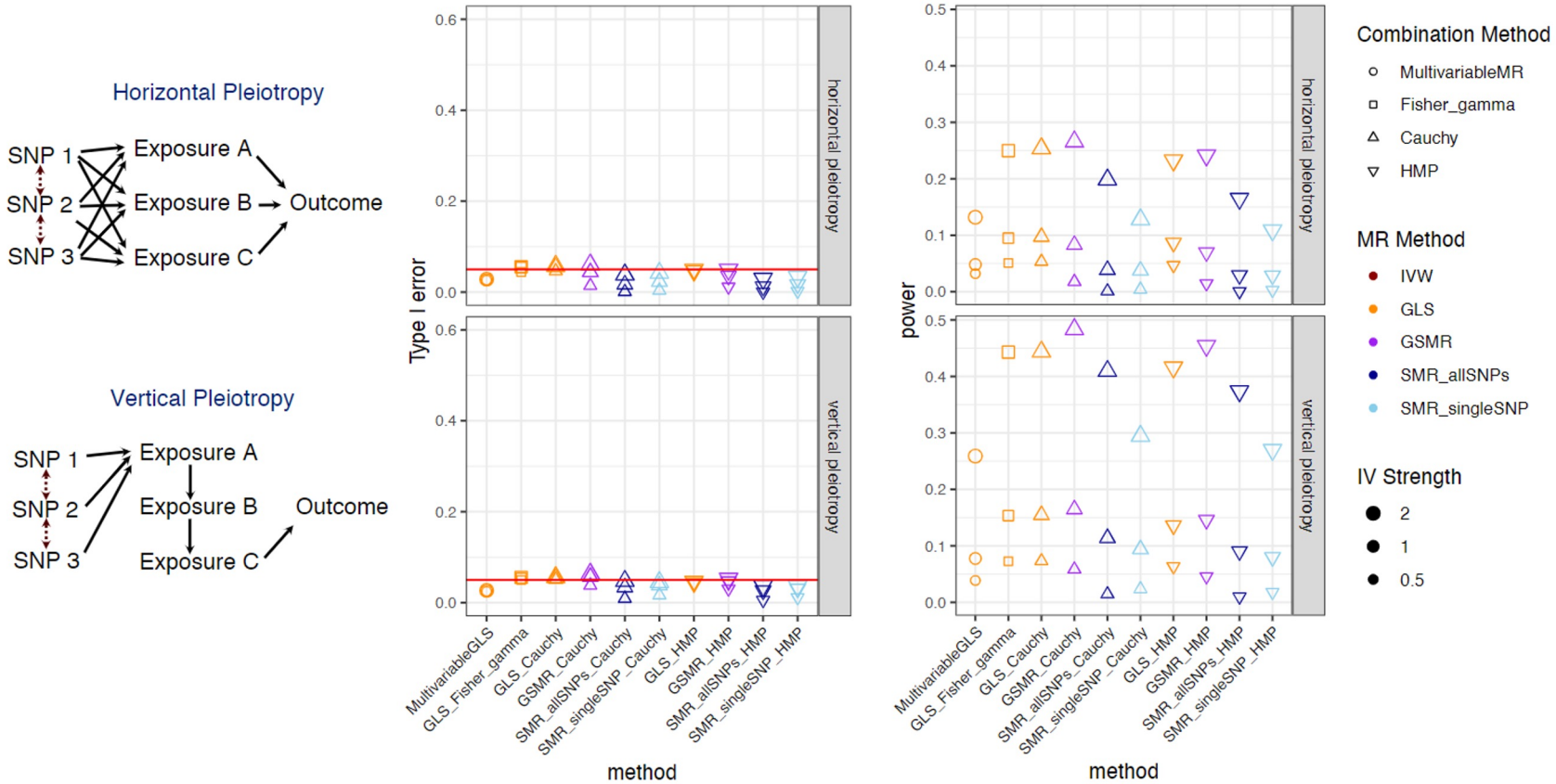


Type I error when the QTL and the GWAS datasets are simulated with various degrees of overlap

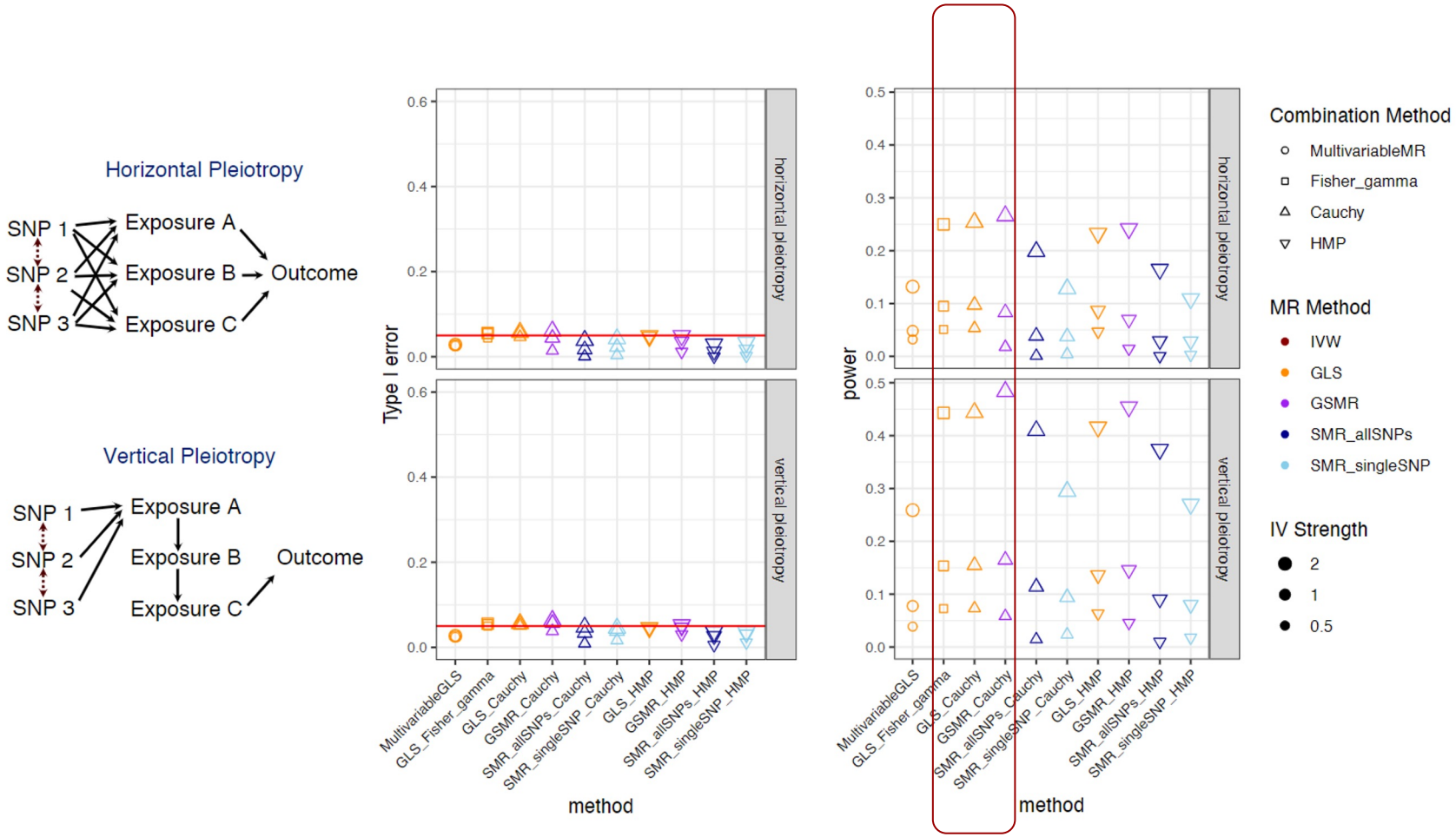
We can control type I error either when there is no overlap between samples *(row of grid)* or when outcome type is binary and only control samples are used to estimate QTLs *(column of grid).*



Type I error and power of selected combinations of MR methods and combination methods



Methods that have the highest powers



Outline

1. Mendelian randomization (MR)
2. MR methods for multi-omics summary statistics
3. Methods in *p*-value combinations
4. Simulations
5. Analyzing an Alzheimer's disease dataset

Prioritization of genes linked to Alzheimer's disease using expression, proteomics, and metabolomics biomarkers

(iv) Select a list of 584 genes with some evidence of association to Alzheimer's disease from Agora genes at https://agora.adknowledgeportal.org/genes.

(v) Within the Agora genes, select 41 genes where all three omics biomarkers (expression, proteomics and metabolomics) are present.

(vi) Select 27 genes that at least one of eQTL, pQTL and metQTL is significant (P-value $< \frac{0.05}{41}$).

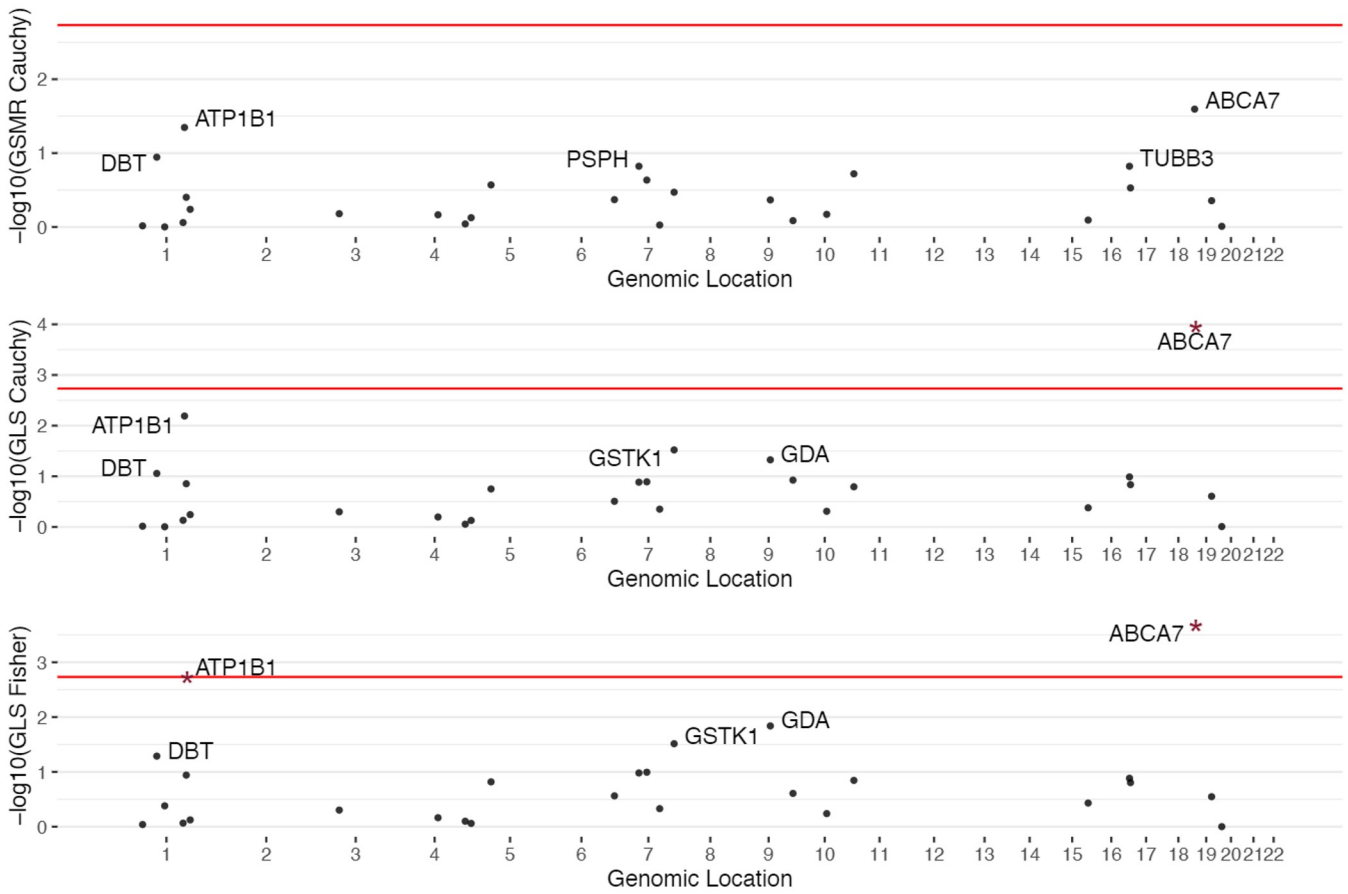
(vii) For each gene, select cis-SNPs (distance between gene and SNP less than 100 kb) and perform LD clumping in each omics biomarker (select SNP with the smallest QTL when possible, $r^2 < 0.2$, distance > 100 kb).

(viii) Perform MR analysis for each omics biomarker using the exposure datasets and the outcome datasets.

(ix) Integrate P-values using combination tests.

Prioritization of genes linked to Alzheimer's disease using expression, proteomics, and metabolomics biomarkers

GLS_Fisher can discover two significant genes *ABCA7* and *ATP1B1*



Discussion

- Prioritization of genes in a genome-wide analysis achieves an increase in power at the expense of deeper interrogations of relationship between biomarkers
- Hard to differentiate causality from pleiotropy and linkage
- Three-sample MR (Zhao *et al.* 2019) could be adopted in the future

Summary

- We propose combination tests that aggregate p-values related to a gene after Mendelian randomization (MR) analyses probing the causal effect (or pleiotropic effect) of omics biomarkers on the outcome.
- Both in simulations and a real example, the combination tests are more powerful in gene prioritization than the multivariable MR framework.
- We recommend using Fisher combination test with gamma distribution approximation. Its power is the highest among the compared methods especially when only weak IVs can be selected from the variants.

Thank you!

References

Burgess, Stephen, George Davey Smith, Neil M. Davies, Frank Dudbridge, Dipender Gill, M. Maria Glymour, Fernando P. Hartwig, et al. "Guidelines for Performing Mendelian Randomization Investigations." Wellcome Open Research 4 (April 28, 2020).

Rosenbaum, Paul R. "Identification of Causal Effects Using Instrumental Variables: Comment." Journal of the American Statistical Association 91, no. 434 (1996): 465–68. https://doi.org/10.2307/2291633.

Imbens, Guido W., and Paul R. Rosenbaum. "Robust, Accurate Confidence Intervals with a Weak Instrument: Quarter of Birth and Education." Journal of the Royal Statistical Society. Series A (Statistics in Society) 168, no. 1 (2005): 109–26.

Barfield, R. *et al.* Testing for the indirect effect under the null for genome-wide mediation analyses. *Genetic Epidemiology* **41**, 824–833 (2017).

Burgess, S., Dudbridge, F. & Thompson, S. G. Re: "Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects". *Am J Epidemiol* **181**, 290–291 (2015).

Burgess, S. & Thompson, S. G. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol* **181**, 251–260 (2015)..

Burgess, S., Timpson, N. J., Ebrahim, S. & Davey Smith, G. Mendelian randomization: where are we now and where are we going? *Int J Epidemiol* **44**, 379–388 (2015).

Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* k601 (2018).

Elliott, L. T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).

Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian Randomization. *JAMA* **318**, 1925 (2017).

Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* **47**, 1091–1098 (2015).

Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245–252 (2016).

Hannon, E., Weedon, M., Bray, N., O'Donovan, M. & Mill, J. Pleiotropic Effects of Trait-Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci. *The American Journal of Human Genetics* **100**, 954–959 (2017).

Huang, Y.-T., Liang, L., Moffatt, M. F., Cookson, W. O. C. M. & Lin, X. iGWAS: Integrative Genome-Wide Association Studies of Genetic and Genomic Data for Disease Susceptibility Using Mediation Analysis. *Genetic Epidemiology* **39**, 347–356 (2015).

Huang, Y.-T., VanderWeele, T. J. & Lin, X. Joint analysis of SNP and gene expression data in genetic association studies of complex diseases. *Ann. Appl. Stat.* **8**, 352–376 (2014).

Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45**, 1452–1458 (2013).

Leeuw, C. A. de, Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219 (2015).

Li, Q., Hu, J., Ding, J. & Zheng, G. Fisher's method of combining dependent statistics using generalizations of the gamma distribution with applications to genetic pleiotropic associations. *Biostatistics* **15**, 284–295 (2014).

Liu, D., Wang, Y., Jing, H., Meng, Q. & Yang, J. Mendelian randomization integrating GWAS and mQTL data identified novel pleiotropic DNA methylation loci for neuropathology of Alzheimer's disease. *Neurobiology of Aging* **97**, 18–27 (2021).

Liu, Y. et al. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet* **104**, 410–421 (2019).

Moskvina, V. *et al.* Evaluation of an approximation method for assessment of overall significance of multiple-dependent tests in a genomewide association study. *Genetic Epidemiology* **35**, 861–866 (2011).

Pingault, J.-B. *et al.* Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics* **19**, 566–580 (2018).

Poole, W., Gibbs, D. L., Shmulevich, I., Bernard, B. & Knijnenburg, T. A. Combining dependent P-values with an empirical adaptation of Brown's method. *Bioinformatics* **32**, i430–i436 (2016).

References

Qi, G. & Chatterjee, N. Mendelian randomization analysis using mixture models for robust and efficient estimation of causal effects. *Nature Communications* **10**, 1941 (2019).

Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Statistics in Medicine* **36**, 4705–4718 (2017).

Richardson, T. G., Hemani, G., Gaunt, T. R., Relton, C. L. & Davey Smith, G. A transcriptome-wide Mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. *Nature Communications* **11**, 185 (2020).

Richardson, T. G. *et al.* Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk. *The American Journal of Human Genetics* **101**, 590–602 (2017).

Shen, L. & Thompson, P. M. Brain Imaging Genomics: Integrated Analysis and Machine Learning. *Proceedings of the IEEE* **108**, 125–162 (2020).

Smith, G. D. & Ebrahim, S. *Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies*. *Biosocial Surveys* (National Academies Press (US), 2008).

Sun, R., Hui, S., Bader, G. D., Lin, X. & Kraft, P. Powerful gene set analysis in GWAS with the Generalized Berk-Jones statistic. *PLOS Genetics* **15**, e1007530 (2019).

Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* **50**, 693–698 (2018).

Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nature Genetics* **51**, 592–599 (2019).

Wu, Y. *et al.* Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nature Communications* **9**, 918 (2018).

Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369–375 (2012).

Yuan, Z. *et al.* Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nature Communications* **11**, 3861 (2020).

Yurko, R., Roeder, K., Devlin, B. & G'Sell, M. H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. *bioRxiv* 2020.08.20.260224 (2020).

Zhang, L. *et al.* Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nature Communications* **11**, 6114 (2020).

Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics* **50**, 1318–1326 (2018).

Zhang, Y., Xu, Z., Shen, X. & Pan, W. Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage* **96**, 309–325 (2014).

Zhao, B. *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nature Genetics* **51**, 1637–1644 (2019).

Zhao, B. *et al.* Transcriptome-wide association analysis of 211 neuroimaging traits identifies new genes for brain structures and yields insights into the gene-level pleiotropy with other complex traits. *bioRxiv* 842872 (2019).

Zhao, T., Hu, Y., Zang, T. & Wang, Y. Integrate GWAS, eQTL, and mQTL Data to Identify Alzheimer's Disease-Related Genes. *Front. Genet.* **10**, (2019).

Zheng, J. *et al.* Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep* **4**, 330–345 (2017).

Zhou, D. *et al.* A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nature Genetics* **52**, 1239–1246 (2020).

Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**, 481–487 (2016).