GWAS Data & Sources for Mendelian Randomization

What is GWAS?

- Genome-Wide Association Studies (GWAS) identify SNPs associated with traits.
- Output: SNP, effect size (beta), standard error, p-value, etc.
- Key for Mendelian
 Randomization: genetic
 instruments from GWAS.

GWAS Output for MR

Relevant fields in summary statistics:

- SNP ID, effect allele, other allele
- Beta (effect size), SE,
 p-value, EAF
- Sample size, trait description

Independent SNPs (IVs)

Requirements for MR

 Summary statistics for both exposure and outcome

 Harmonization potential between datasets

Attention to: palindromic
 SNPs, genome build differences

GWAS Catalog (ebi.ac.uk/gwas)

Where to Find GWAS Data?

 MR-Base / IEU OpenGWAS (gwas.mrcieu.ac.uk)

• URL:

https://www.ebi.ac.uk/gwas/

GWAS Catalog

 Curated repository of published GWAS findings

• Limitation: may not provide full summary stats

MR-Base & IEU OpenGWAS

URL: https://gwas.mrcieu.ac.uk/

Contains >50,000
 harmonized GWAS datasets

 API and R package support (ieugwasr)

 Integrated with TwoSampleMR for MR analysis

Obtain instruments from 00 exposure GWAS **LD Proxies** If an exposure instrument is not available in the outcome GWAS then look for LD proxies in 1000 genomes **Extract SNP effects from** -Target SNP outcome GWAS -Best LD proxy Harmonise exposure and **Exposure GWAS Outcome GWAS** outcome effects Effect Other Effect allele Effect Other Effect allele SNP Effect allele allele frequency Effect allele allele frequency O rs12345 0.132 A 0.28 0.022 A G 0.26 O rs23456 -0.485 G 0.41 0.056 T G 0.61 O rs34567 0.203 G 0.11 -0.046 G **Outcome GWAS Exposure GWAS** Effect Other Effect allele Effect Other Effect allele Effect allele allele frequency Effect allele allele frequency rs12345 0.132 A G 0.28 0.022 A G 0.26 rs23456 -0.485 G -0.056 G T 0.41 0.39 rs34567 0.203 G 0.046 G MR estimates and sensitivity analyses

 Complex traits, diseases, biomarkers **GWAS** catalog 1628 traits Whole blood MRInstruments R package Metabolomics • 121 metabolite levels Gene 44 tissues expression 27094 gene identifiers DNA 5 time points in whole blood methylation 33256 CpG levels Whole blood **Proteomics** 47 protein levels MR-Base Automated clumping database 1674 datasets User-provided data Other

123 complex diseases

- 15 neoplastic
- · 20 psychiatric
- 50 inflammatory
- 11 cardiovascular
- 5 diabetic
- 22 other

219 complex traits

• 91 anthropometric

MR-Base

data

- 5 behavioural
- · 24 glycaemic
- · 9 lipid
- · 4 blood pressure
- · 8 haemotological
- 17 education
- 61 other

575 metabolite levels

151 immunological traits

606 UK Biobank phenotypes

User-provided results

Choosing GWAS for MR

Large sample size

Clear phenotype definition

- Ancestry match with outcome
- No sample overlap (2sample MR)
- Full summary statistics availability

Required Packages

- library(TwoSampleMR)
- library(ieugwasr)
- library(VariantAnnotation)
- library(MRInstruments)
- library(gwasglue)
- library(ggplot2)
- library(dplyr)

Querying GWAS from different sources

Next:
HandsOn
Preview

Instrument selection and clumping

Harmonization and preparation for MR

Harmonize data

- Dealing with strand issues
- Palindromic SNP

Strand Issues

Correct, unambiguous

```
exposure effect = 0.5
effect allele = A
other allele = G
outcome effect = 0.05
effect allele = A
other allele = G
```

Ambiguous

```
exposure effect = 0.5
effect allele = A
other allele = G
outcome effect = 0.05
effect allele = A
other allele = C
```

Incorrect reference, unambiguous

```
exposure effect = 0.5
effect allele = A
other allele = G
outcome effect = -0.05
effect allele = C
other allele = T
```

Palindromic SNP

Inferrable

```
exposure effect = 0.5

effect allele = A

other allele = T

effect allele frequency = 0.11

outcome effect = -0.05

effect allele = A

other allele = T

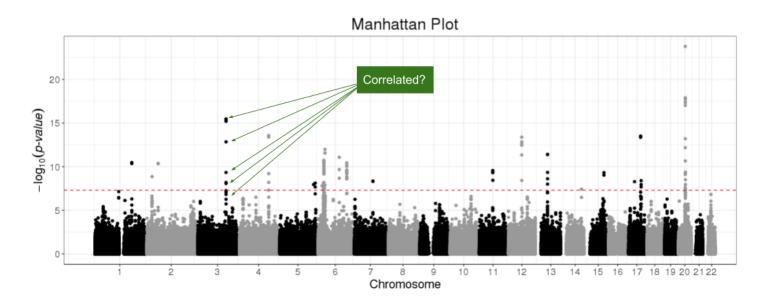
effect allele frequency = 0.91
```

Not Inferrable

```
exposure effect = 0.5
effect allele = A
other allele = T
effect allele frequency = 0.50
outcome effect = -0.05
effect allele = A
other allele = T
effect allele frequency = 0.50
```

Linkage Disequilibrium

Linkage disequilibrium (LD) refers to the nonrandom association of alleles at different loci on a chromosome.



Performing MR

Next:
HandsOn
Preview

Sensitivity Analysis

Plots and Interpretation

Summary Slide

Share your Feedback!

Scan the QR code to let us know your thoughts on this tutorial. Thank you!

