

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

Requirements for MR

- Independent SNPs (IVs)
- Summary statistics for both exposure and outcome
- Harmonization potential between datasets
- Attention to: palindromic SNPs, genome build differences

Where to Find GWAS Data?

-
- GWAS Catalog
(ebi.ac.uk/gwas)

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

GWAS Catalog

- URL:
<https://www.ebi.ac.uk/gwas/>

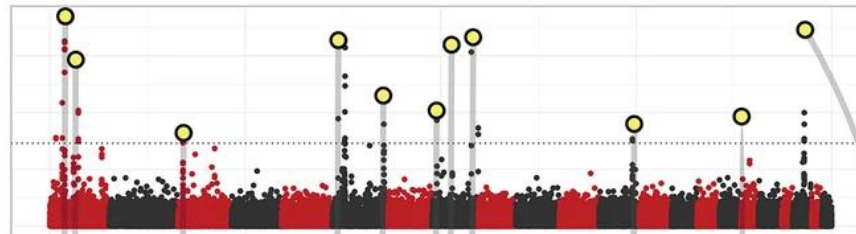
- Curated repository of published GWAS findings

- Limitation: may not provide full summary stats

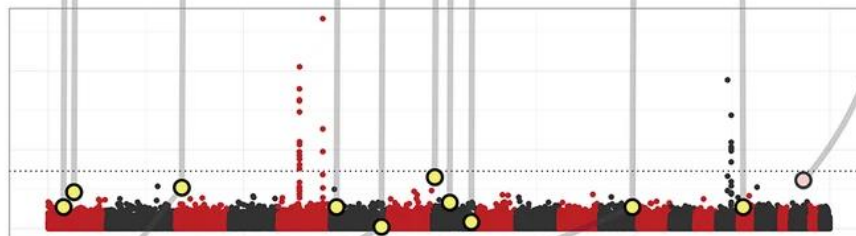
MR-Base & IEU OpenGWAS

- 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 exposure GWAS



Extract SNP effects from outcome GWAS



LD Proxies

If an exposure instrument is not available in the outcome GWAS then look for LD proxies in 1000 genomes



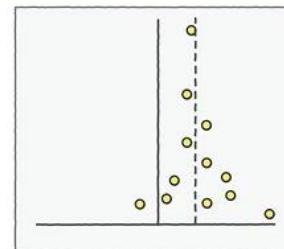
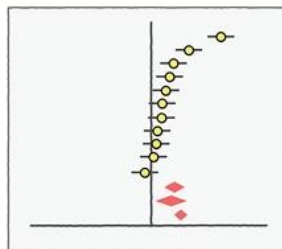
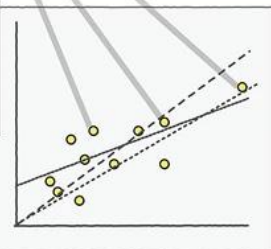
Harmonise exposure and outcome effects

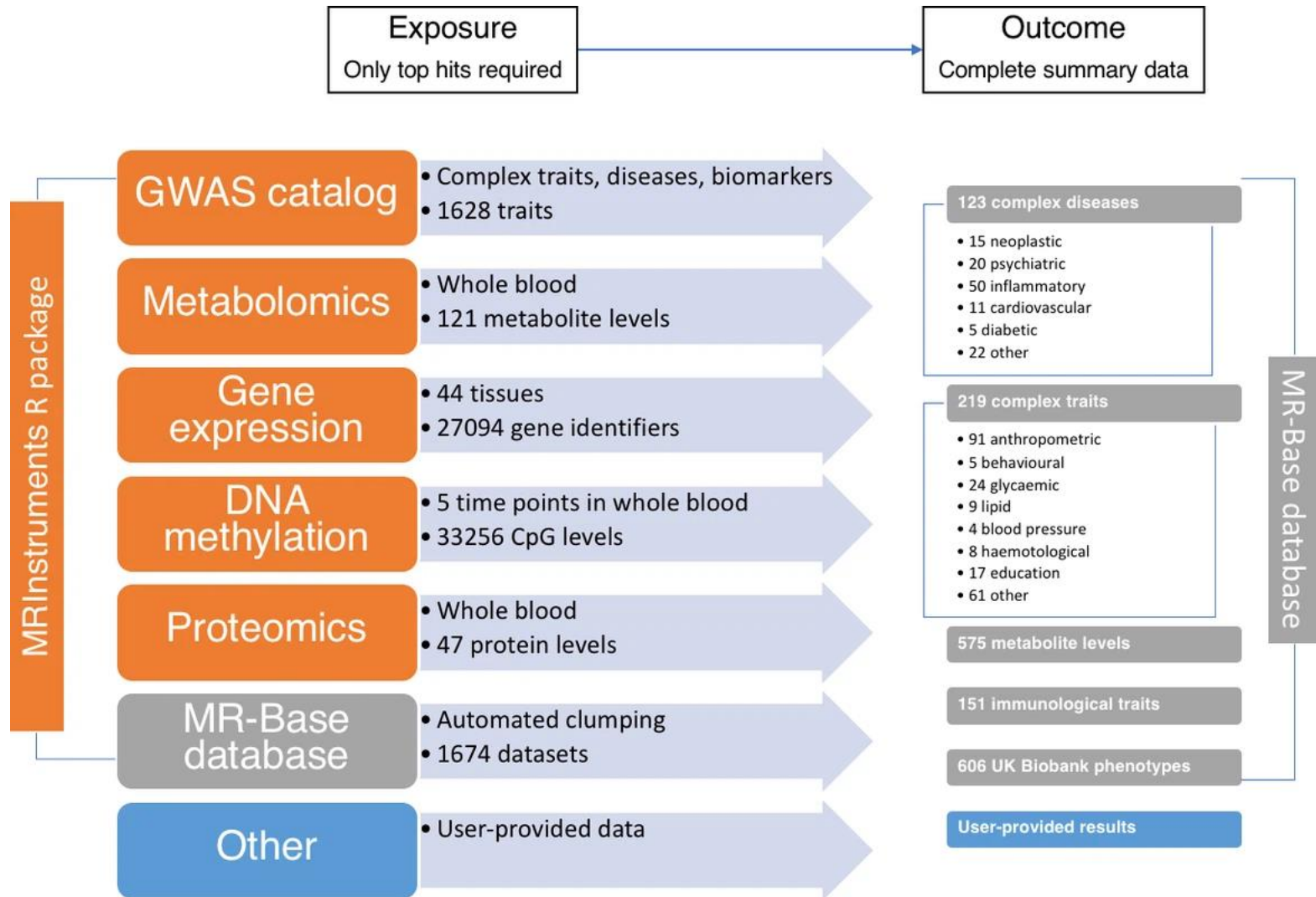
SNP	Exposure GWAS				Outcome GWAS			
	Effect	Effect allele	Other allele	Effect allele frequency	Effect	Effect allele	Other allele	Effect allele frequency
rs12345	0.132	A	G	0.28	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	0.056	T	G	0.61
rs34567	0.203	G	C	0.11	-0.046	G	C	0.88



SNP	Exposure GWAS				Outcome GWAS			
	Effect	Effect allele	Other allele	Effect allele frequency	Effect	Effect allele	Other allele	Effect allele frequency
rs12345	0.132	A	G	0.28	0.022	A	G	0.26
rs23456	-0.485	G	T	0.41	-0.056	G	T	0.39
rs34567	0.203	G	C	0.11	0.046	G	C	0.12

MR estimates and sensitivity analyses





Choosing GWAS for MR

- Large sample size
- Clear phenotype definition
- Ancestry match with outcome
- No sample overlap (2-sample MR)
- Full summary statistics availability

Required Packages

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

Next: Hands- On Preview

- Querying GWAS from different sources

- 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

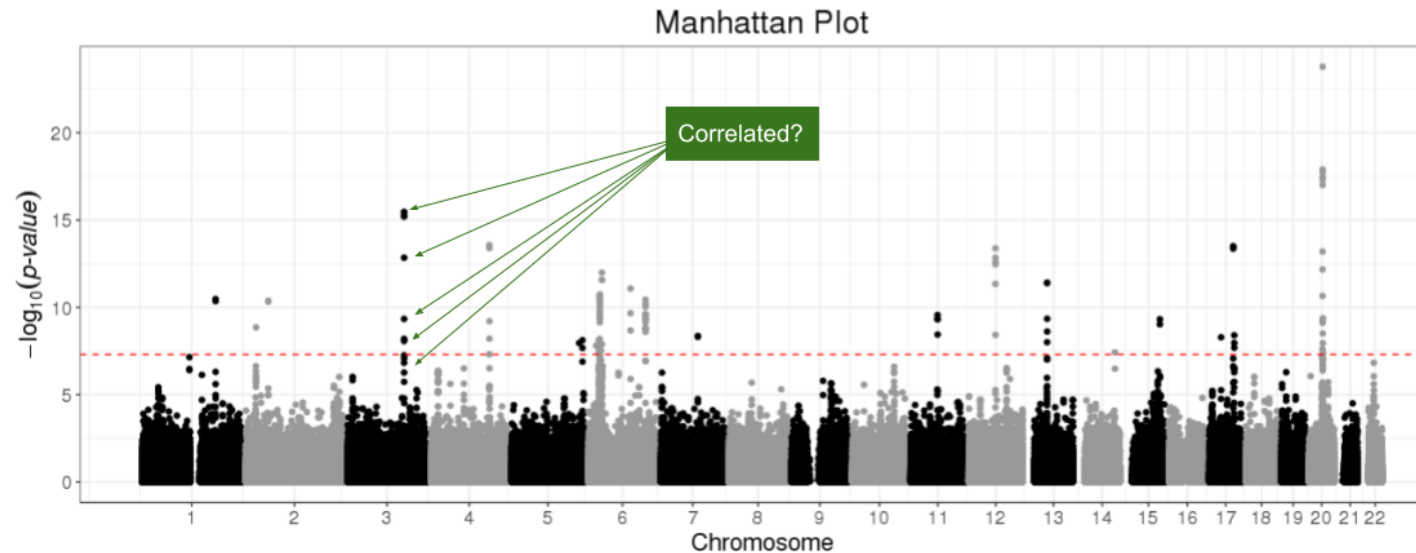
exposure effect = 0.5
effect allele = A
other allele = T
effect allele frequency = 0.11
outcome effect = -0.05
effect allele = A
other allele = T
effect allele frequency = 0.91

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 non-random association of alleles at different loci on a chromosome.



Next: Hands- On Preview

-
- Performing MR

-
- 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!

