

GROUP4_MR

Sandra

2023-07-03

```
#{.tabset} ##Load packages required for the MR
```

Loading in the exposure dataset

```
# Specify the file path
file_path <- "/home/sandra/Documents/EDA_DATA/Mendelian Randomization/mrc_uganda_hackathon_202306_group

#Reading in the exposure data using function read_exposure
bmi_exp_data <- read_exposure_data(
  filename = file_path,
  sep = "\t",
  snp_col = "SNP",
  beta_col = "BETA",
  se_col = "SE",
  effect_allele_col = "Tested_Allele",
  other_allele_col = "Other_Allele",
  eaf_col = "Freq_Tested_Allele_in_HRS",
  pval_col = "P",
  samplesize_col = "N"
)
```

No phenotype name specified, defaulting to 'exposure'.

```
#How many rows and columns are in the dataset
```

```
dim(bmi_exp_data)
```

```
[1] 656 13
```

```
head(bmi_exp_data)
```

	SNP	effect_allele.exposure	other_allele.exposure	eaf.exposure
1	rs12044597	A	G	0.4971
2	rs7535528	A	G	0.3741
3	rs6577584	T	G	0.6537
4	rs7556169	A	G	0.3857
5	rs10779751	A	G	0.2744
6	rs761423	T	C	0.5500

	beta.exposure	se.exposure	pval.exposure	samplesize.exposure	exposure
1	-0.0143	0.0016	1.7e-18	789125	exposure
2	-0.0152	0.0018	1.4e-16	632868	exposure
3	-0.0131	0.0018	3.7e-13	691824	exposure
4	-0.0103	0.0018	5.4e-09	692527	exposure
5	0.0139	0.0018	2.5e-14	794997	exposure

	0.0113	0.0017	5.5e-11	692545	exposure
	mr_keep.exposure	pval_orig.in.exposure	id.exposure	data_source.exposure	
1	TRUE		reported	Xt6nIW	textfile
2	TRUE		reported	Xt6nIW	textfile
3	TRUE		reported	Xt6nIW	textfile
4	TRUE		reported	Xt6nIW	textfile
5	TRUE		reported	Xt6nIW	textfile
6	TRUE		reported	Xt6nIW	textfile

Clumping the exposure data

```
Bmi_exp_data <- clump_data(bmi_exp_data)
```

API: public: <http://gwas-api.mrcieu.ac.uk/>

Please look at vignettes for options on running this locally if you need to run many instances of this

Clumping Xt6nIW, 656 variants, using EUR population reference

Removing 364 of 656 variants due to LD with other variants or absence from LD reference panel

#364 variants have been removed due to LD with other variants or absence from the LD reference panel

```
dim(Bmi_exp_data)
```

```
[1] 292 13
```

#Renaming the column for Exposure to BMI

```
Bmi_exp_data$exposure = "BMI"
```

Loading in the outcome dataset

```
file_path = "/home/sandra/Documents/EDA_DATA/Mendelian Randomization/mrc_uganda_hackathon_202306_group4"
outcome_data <- read_outcome_data(
  snps = Bmi_exp_data$SNP,
  filename = file_path,
  sep = "\t",
  snp_col = "rsid",
  beta_col = "beta_AFR",
  se_col = "se_AFR",
  effect_allele_col = "alt",
  other_allele_col = "ref",
  eaf_col = "af_AFR",
  pval_col = "neglog10_pval_AFR",
)
```

No phenotype name specified, defaulting to 'outcome'.

```
dim(outcome_data)
```

```
[1] 292 14
```

```
outcome_data$outcome = "BMI"
```

Harmonizing the data

```
H_data <- harmonise_data(
  exposure_dat = Bmi_exp_data,
  outcome_dat = outcome_data
)
```

Harmonising BMI (Xt6nIW) and BMI (3oKmf1)

Removing the following SNPs for being palindromic with intermediate allele frequencies:

rs10887578, rs12575252, rs1324110, rs1454687, rs2163188, rs2504674, rs2875762, rs6011457, rs6595205, rs

```
dim(H_data)
```

```
[1] 292 32
```

```
head(H_data)
```

	SNP	effect_allele.exposure	other_allele.exposure	effect_allele.outcome			
1	rs1003081	T	C	T			
2	rs10083803	T	C	T			
3	rs10118701	A	G	A			
4	rs10132280	A	C	A			
5	rs10182181	A	G	A			
6	rs10211055	T	C	T			
	other_allele.outcome	beta.exposure	beta.outcome	eaf.exposure	eaf.outcome		
1	C	0.0121	0.017770	0.4497	0.1901		
2	C	-0.0127	0.002660	0.2605	0.2767		
3	G	-0.0163	0.023250	0.6773	0.3986		
4	C	-0.0223	0.005292	0.3017	0.5511		
5	G	-0.0325	-0.077310	0.5247	0.0957		
6	C	-0.0157	-0.003052	0.6544	0.7692		
	remove	palindromic	ambiguous	id.outcome	chr.outcome	pos.outcome	se.outcome
1	FALSE	FALSE	FALSE	3oKmf1	11	118913993	0.02224
2	FALSE	FALSE	FALSE	3oKmf1	16	6701400	0.01960
3	FALSE	FALSE	FALSE	3oKmf1	9	103061366	0.01756
4	FALSE	FALSE	FALSE	3oKmf1	14	25928179	0.01756
5	FALSE	FALSE	FALSE	3oKmf1	2	25150296	0.03049
6	FALSE	FALSE	FALSE	3oKmf1	2	229016917	0.02064
	pval.outcome	outcome	mr_keep.outcome	pval_origin.outcome	data_source.outcome		
1	0.372500000	BMI	TRUE	reported	textfile		
2	0.049610000	BMI	TRUE	reported	textfile		
3	0.731800000	BMI	TRUE	reported	textfile		
4	0.117400000	BMI	TRUE	reported	textfile		
5	0.005612977	BMI	TRUE	inferred	textfile		
6	0.054300000	BMI	TRUE	reported	textfile		
	se.exposure	pval.exposure	samplesize.exposure	exposure	mr_keep.exposure		
1	0.0016	1.3e-13	788712	BMI	TRUE		
2	0.0019	5.9e-11	690845	BMI	TRUE		
3	0.0018	1.1e-18	692315	BMI	TRUE		
4	0.0018	5.6e-35	786578	BMI	TRUE		
5	0.0016	6.7e-90	792111	BMI	TRUE		
6	0.0018	7.2e-18	686577	BMI	TRUE		
	pval_origin.exposure	id.exposure	data_source.exposure	action	mr_keep		
1	reported	Xt6nIW	textfile	2	TRUE		
2	reported	Xt6nIW	textfile	2	TRUE		

3	reported	Xt6nIW	textfile	2	TRUE
4	reported	Xt6nIW	textfile	2	TRUE
5	reported	Xt6nIW	textfile	2	TRUE
6	reported	Xt6nIW	textfile	2	TRUE

	samplesize.outcome
1	NA
2	NA
3	NA
4	NA
5	NA
6	NA

Checking for palindromic SNPs

```
palindromic_at <- subset(H_data, effect_allele.outcome %in% "A" & other_allele.outcome %in% "T")
palindromic_ta <- subset(H_data, effect_allele.outcome %in% "T" & other_allele.outcome %in% "A")
palindromic_gc <- subset(H_data, effect_allele.outcome %in% "G" & other_allele.outcome %in% "C")
palindromic_cg <- subset(H_data, effect_allele.outcome %in% "C" & other_allele.outcome %in% "G")
dim(palindromic_at)
```

```
[1] 11 32
```

```
dim(palindromic_ta)
```

```
[1] 0 32
```

```
dim(palindromic_gc)
```

```
[1] 0 32
```

```
dim(palindromic_cg)
```

```
[1] 23 32
```

```
#Scatter plot for the beta estimate of the exposure and outcome to access for directionality {.tabset}
```

```
# Create the scatter plot
```

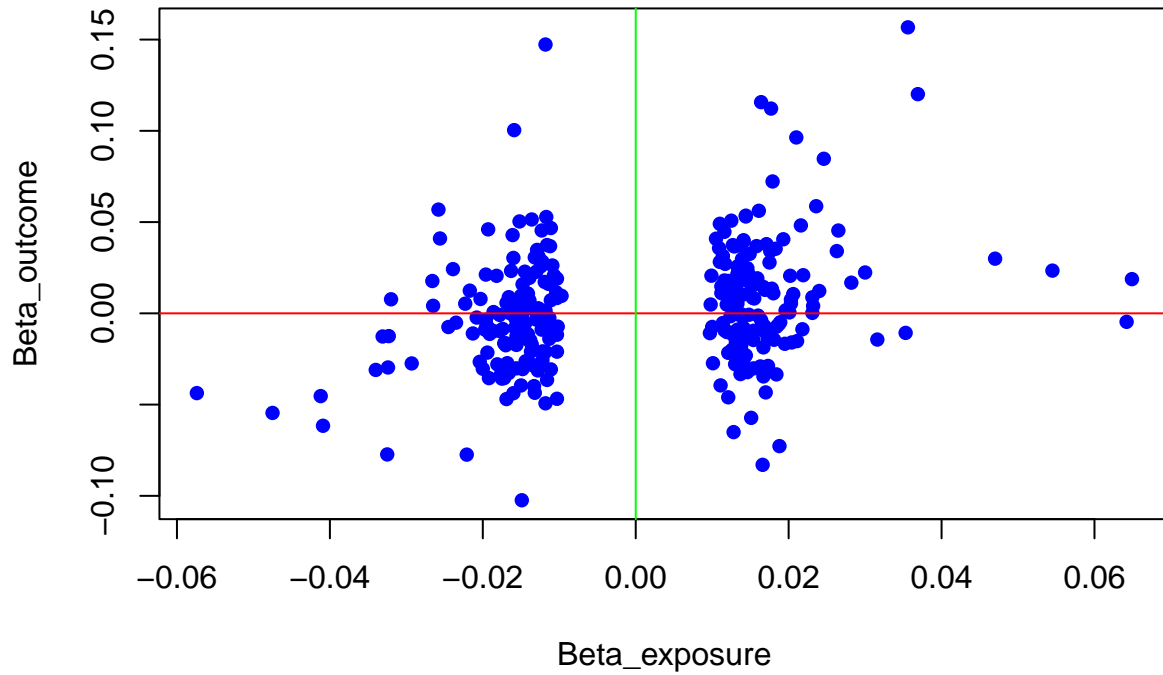
```
plot(H_data$beta.exposure, H_data$beta.outcome, pch = 16, col = "blue", xlab = "Beta_exposure", ylab = "Beta_outcome")
```

```
# Add a horizontal line
```

```
abline(h = 0, col = "red")
```

```
# Add a vertical line
```

```
abline(v = 0, col = "green")
```



Filtering for only SNPs with positive and negative beta estimates to obtain SNPs that are African associated

```
keep <- which(H_data$beta.exposure > 0 & H_data$beta.outcome > 0)
keep <- which(H_data$beta.exposure < 0 & H_data$beta.outcome < 0)
H_data = H_data[keep,]
head(H_data)
```

	SNP	effect_allele.exposure	other_allele.exposure				
5	rs10182181	A	G				
6	rs10211055	T	C				
10	rs1048932	A	C				
12	rs10744146	A	G				
13	rs1075901	T	C				
17	rs10876418	T	C				
		effect_allele.outcome	other_allele.outcome	beta.exposure	beta.outcome		
5		A	G	-0.0325	-0.077310		
6		T	C	-0.0157	-0.003052		
10		A	C	-0.0160	-0.043750		
12		A	G	-0.0122	-0.020850		
13		T	C	-0.0121	-0.022490		
17		T	C	-0.0123	-0.008909		
	eaf.exposure	eaf.outcome	remove	palindromic	ambiguous	id.outcome	chr.outcome
5	0.5247	0.0957	FALSE	FALSE	FALSE	3oKmf1	2
6	0.6544	0.7692	FALSE	FALSE	FALSE	3oKmf1	2

10	0.4162	0.3836	FALSE	FALSE	FALSE	3oKmf1	11
12	0.5146	0.3155	FALSE	FALSE	FALSE	3oKmf1	12
13	0.4361	0.2659	FALSE	FALSE	FALSE	3oKmf1	17
17	0.2325	0.1756	FALSE	FALSE	FALSE	3oKmf1	12
	pos.outcome	se.outcome	pval.outcome	outcome	mr_keep.outcome		
5	25150296	0.03049	0.005612977	BMI	TRUE		
6	229016917	0.02064	0.054300000	BMI	TRUE		
10	115044850	0.01756	0.006361003	BMI	TRUE		
12	17212881	0.01864	0.579700000	BMI	TRUE		
13	15943910	0.01952	0.603600000	BMI	TRUE		
17	39428802	0.02281	0.157300000	BMI	TRUE		
	pval_origin.outcome	data_source.outcome	se.exposure	pval.exposure			
5	inferred	textfile	0.0016	6.7e-90			
6	reported	textfile	0.0018	7.2e-18			
10	inferred	textfile	0.0017	3.8e-22			
12	reported	textfile	0.0017	1.4e-12			
13	reported	textfile	0.0016	1.2e-13			
17	reported	textfile	0.0019	2.8e-10			
	samplesize.exposure	exposure	mr_keep.exposure	pval_origin.exposure			
5	792111	BMI	TRUE	reported			
6	686577	BMI	TRUE	reported			
10	795167	BMI	TRUE	reported			
12	689141	BMI	TRUE	reported			
13	794789	BMI	TRUE	reported			
17	787067	BMI	TRUE	reported			
	id.exposure	data_source.exposure	action	mr_keep	samplesize.outcome		
5	Xt6nIW	textfile	2	TRUE	NA		
6	Xt6nIW	textfile	2	TRUE	NA		
10	Xt6nIW	textfile	2	TRUE	NA		
12	Xt6nIW	textfile	2	TRUE	NA		
13	Xt6nIW	textfile	2	TRUE	NA		
17	Xt6nIW	textfile	2	TRUE	NA		

```
dim(H_data)
```

```
[1] 83 32
```

```
View(H_data)
```