

## Supplementary source code

This file provides source code so that, combined with the provided input data, readers should be able to reproduce the results from Mendelian randomization and MVMR analyses as described in ‘Gallbladder Cancer Risk and Indigenous South American Mapuche Ancestry: Instrumental Variable Analysis Using Ancestry-Informative Markers’. Note: The path needs to be set accordingly.

```
#####  
#                                                                 #  
# Title: Perform mendelian randomization analyses for           #  
#   - Mapuche ancestry (MAP) -> GBC                            #  
#   - Mapuche ancestry (MAP) -> gallstone disease (GST)        #  
#   - Mapuche ancestry (MAP) -> BMI                             #  
#                                                                 #  
# Author: Linda Zollner                                         #  
# Version: 2.0                                                   #  
# Date: 11.08.2023                                              #  
#                                                                 #  
# Available at:                                                 #  
#   www.biometrie.uni-heidelberg.de/StatisticalGenetics/Software_and_Data #  
#                                                                 #  
#####  
  
rm(list=ls())  
  
#####  
### required packages  
library('TwoSampleMR') # version 0.5.2  
# followed the detailed guide provided by the authors  
# https://mrcieu.github.io/TwoSampleMR/articles/index.html  
library('RadialMR') # version 0.4  
library('ggplot2') # version 3.3.5  
  
### required files:  
# sum_data_MAP.txt, sum_data_GBC.txt, sum_data_GST.txt, sum_data_BMI.txt  
  
# text file with header line; each line describes one IV:  
  
# SNP   - name of IV  
# association statistics from the association  
# MAP proportion / GBC (y/n) / GST (y/n) / BMI ~ SNP:  
#   beta - beta/log(OR) estimate  
#   se - standard error  
#   pval - p-value  
# allele information in the underlying population:  
#   ea - effect allele  
#   oa - other allele  
#   eaf - effect allele frequency
```

```

# (min_pval - threshold for p-value [IV -> exposure])
#####

### to do: set path to data
path <- ''

### choose outcome
# choose between 'GBC', 'GST' or 'BMI'
outcome <- 'GBC'

### read summary data (IV-exposure[MAP]; IV-outcome)
exp_dat <- read_exposure_data(
  paste0(path, 'sum_data_MAP.txt'),
  sep = '\t',
  snp_col = 'SNP',
  beta_col = 'beta',
  se_col = 'se',
  eaf_col = 'eaf',
  effect_allele_col = 'ea',
  other_allele_col = 'oa',
  pval_col = 'pval'
)
exp_dat$exposure <- 'Mapuche ancestry'

out_dat <- read_outcome_data(
  snps = exp_dat$SNP,
  filename = paste0(path, 'sum_data_', outcome, '.txt'),
  sep = '\t',
  snp_col = 'SNP',
  beta_col = 'beta',
  se_col = 'se',
  effect_allele_col = 'ea',
  other_allele_col = 'oa',
  eaf_col = 'eaf',
  pval_col = 'pval'
)
out_dat$outcome <- outcome

### radial MR
# find and extract outlying IVs using IVW and Egger method
for_rad <- format_radial(
  BXG = exp_dat$beta.exposure,
  BYG = out_dat$beta.outcome,
  seBXG = exp_dat$se.exposure,
  seBYG = out_dat$se.outcome,
  RSID = exp_dat$SNP
)
ivw <- ivw_radial(for_rad)
egger <- egger_radial(for_rad)

outlier <- unique(
  c(as.character(ivw$data$SNP[ivw$data$Outliers == 'Outlier']),
    as.character(egger$data$SNP[egger$data$Outliers == 'Outlier']))
)

```

```

)
)

for (i in 1:length(outlier)){
  exp_dat <- exp_dat[-which(exp_dat$SNP == outlier[i]),]
  out_dat <- out_dat[-which(out_dat$SNP == outlier[i]),]
}

### perform mendelian randomization
# harmonize data
dat <- harmonise_data(
  exposure_dat = exp_dat,
  outcome_dat = out_dat
)

### MR
res_mr <- mr(
  dat, method_list=c('mr_ivw', 'mr_egger_regression', 'mr_weighted_median')
)
mr_het <- mr_heterogeneity(dat)
mr_pleio <- mr_pleiotropy_test(dat)
res_odds <- generate_odds_ratios(res_mr)

# results
if (outcome == 'GBC' | outcome == 'GST'){
  result <- data.frame(
    outcome = outcome,
    IV = dim(dat)[1],
    beta_OR = res_odds$or[which(res_odds$method == 'Inverse variance weighted')],
    OR_l_CI = res_odds$or_lci95[which(res_odds$method == 'Inverse variance weighted')],
    OR_u_CI = res_odds$or_uci95[which(res_odds$method == 'Inverse variance weighted')],
    p_val = res_odds$pval[which(res_odds$method == 'Inverse variance weighted')],
    Q_p_val = mr_het$Q_pval[which(mr_het$method == 'Inverse variance weighted')],
    intercept_p_val = mr_pleio$pval
  )
} else if (outcome == 'BMI'){
  result <- data.frame(
    outcome = 'BMI',
    IV = dim(dat)[1],
    beta_OR = res_odds$b[which(res_odds$method == 'Inverse variance weighted')],
    OR_l_CI = res_odds$lo_ci[which(res_odds$method == 'Inverse variance weighted')],
    OR_u_CI = res_odds$up_ci[which(res_odds$method == 'Inverse variance weighted')],
    p_val = res_odds$pval[which(res_odds$method == 'Inverse variance weighted')],
    Q_p_val = mr_het$Q_pval[which(mr_het$method == 'Inverse variance weighted')],
    intercept_p_val = mr_pleio$pval
  )
}
}

```

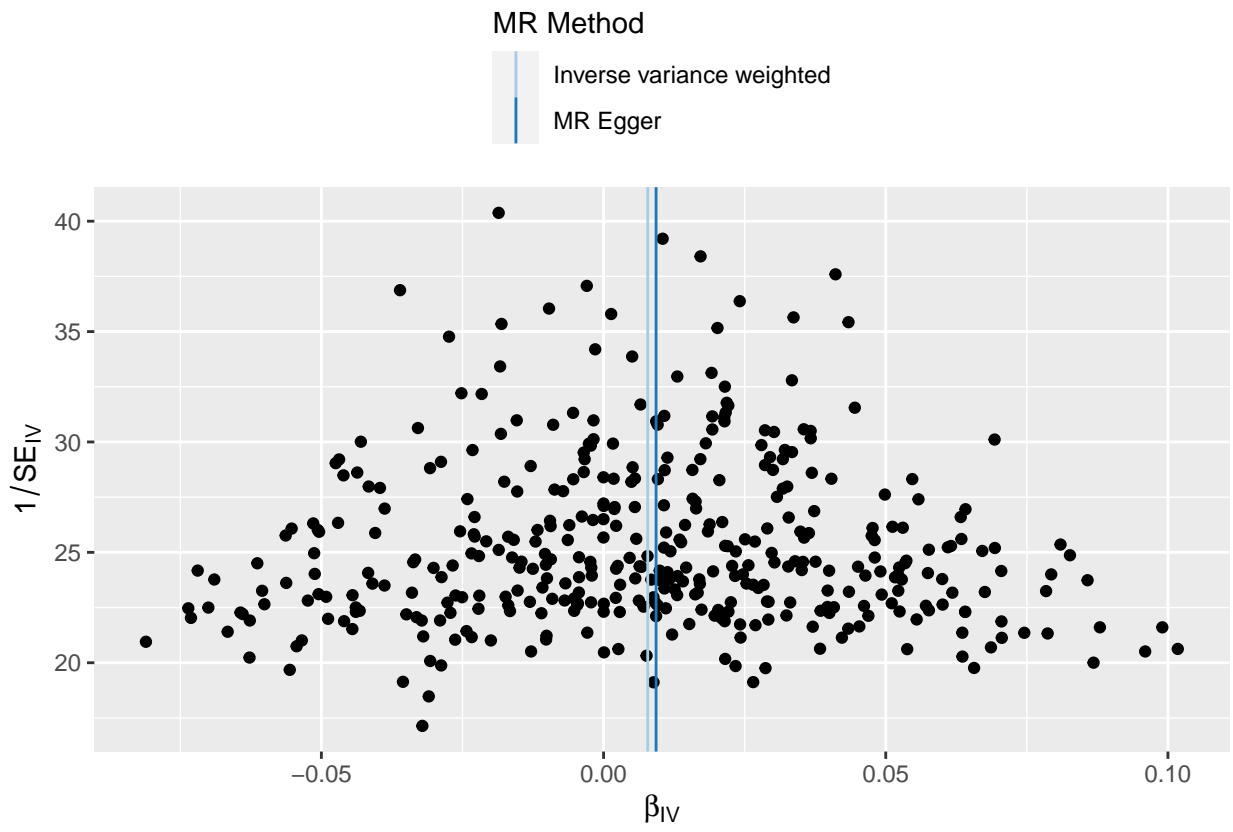
```
result
```

```
## outcome IV beta_OR OR_l_CI OR_u_CI p_val Q_p_val intercept_p_val
## 1 GBC 396 1.007854 1.003985 1.011738 6.701655e-05 0.9998928 0.8730317
```

```
### plots
# scatter plot
mr_scatter_plot(
  mr(dat, method_list=c('mr_ivw', 'mr_egger_regression', 'mr_weighted_median')),
  dat
)
```



```
# funnel plot
mr_funnel_plot(mr_singlesnp(dat))
```



```

#####
#
# Title: Perform MVMR analysis for
# - Mapuche ancestry (MAP) and gallstone disease (GST) -> GBC
#
# Author: Linda Zollner
# Version: 2.0
# Date: 28.07.2023
#
# Available at:
# www.biometrie.uni-heidelberg.de/StatisticalGenetics/Software_and_Data
#
#####

rm(list=ls())

#####
### required packages
library('MVMR') # version 0.3
# followed the guide provided by the authors
# vignette("MVMR")

### required data:
# sum_data_MAP_MVMR.txt, sum_data_GST_MVMR.txt, sum_data_GBC_MVMR.txt

# each in the form of:

# SNP - name of IV
# association statistics from the association
# MAP proportion/GBC (y/n)/GST (y/n) ~ SNP:
# beta - log(OR) estimate
# se - standard error

#####

### set path to data
path <- ''

### read summary data (IV-exposure1[MAP]; IV-exposure2[GST]; IV-outcome[GBC])
exp1_data <- read.table(
  paste0(path, 'sum_data_MAP_MVMR.txt'), header = T, stringsAsFactors = F
)
exp2_data <- read.table(
  paste0(path, 'sum_data_GST_MVMR.txt'), header = T, stringsAsFactors = F
)
out_data <- read.table(
  paste0(path, 'sum_data_GBC_MVMR.txt'), header = T, stringsAsFactors = F
)

### format data
data <- MVMR::format_mvnr(BXGs = cbind(exp1_data$beta, exp2_data$beta),
  BYG = out_data$beta,
  seBXGs = cbind(exp1_data$se, exp2_data$se),

```

```

        seBYG = out_data$se,
        RSID = exp1_data$SNP
    )

### assess instrument strength
sres <- MVMR::strength_mvmm(r_input = data, gencov = 0)

##
## Conditional F-statistics for instrument strength
##
##          exposure1 exposure2
## F-statistic 3.500939 1.377205

### assess horizontal pleiotropy
pres <- MVMR::pleiotropy_mvmm(r_input = data, gencov = 0)

## Q-Statistic for instrument validity:
## 424.9261 on 431 DF , p-value: 0.5734335

### estimate direct effects
res <- MVMR::ivw_mvmm(r_input = data)

##
## Multivariable MR
##
##          Estimate Std. Error   t value   Pr(>|t|)
## exposure1 0.0001544679 0.002395267 0.06448878 9.486109e-01
## exposure2 0.2333353888 0.038106842 6.12318874 2.062731e-09
##
## Residual standard error: 1.023 on 432 degrees of freedom

```