Workshop: Mendelian Randomization

Summary of the packages used in this workshop

MendelianRandomization

Package developed to carry out various Mendelian randomisation analyses on genetic data in R. The package uses several methods to assess whether a risk factor (or exposure) has a causal effect on an outcome.

metafor

Provides functions for conducting meta-analyses in R. Includes functions for fitting fixed and random effects models. Allows the inclusion of moderator variables in the models.

TwoSampleMR

Two sample Mendelian randomisation (2SMR) is a method to estimate the causal effect of an exposure on an outcome using only summary statistics from genome wide association studies (GWAS).

MRPRESSO

Mendelian Randomization Pleiotropy RESidual Sum and Outlier is a method that allows for the evaluation of horizontal pleiotropy in multi-instrument Mendelian Randomization utilizing genome-wide summary association statistics.

Data sets

For this workshop we use the data set from the study: "A Genome-wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic Community Health Study/Study of Latinos", which is a GWAS on the human metabolome from which we obtain the uric acid exposure data in the Latino population containing 1,169 observations. The article can be accessed through this **link**.

The database used to obtain the coronary artery disease (CAD) outcome is obtained from the article: "A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease". Which has 933 CAD variants. This can be accessed via this **link**

Data cleaning and merge

Part of the analysis process involves cleaning the data sets and homogenizing the information. After downloading the databases and managing the information (See 01 - MR Analysis.R, section 1, 2 and 3) we must obtain a database with 1,023 rows with the following structure:

```r
head(exp_out)
```

```
##            cpaid beta.outcome se.outcome effect_allele.outcome
## 1 10:1012315:A:G    -0.013016  0.0109204                     A
## 2 10:1012372:A:G    -0.013041  0.0109193                     A
## 3 10:1013535:A:C    -0.012598  0.0109019                     A
## 4 10:1014603:A:G    -0.013341  0.0109008                     A
## 5 10:1014603:A:G    -0.013341  0.0109008                     A
## 6 10:1019138:A:T    -0.012900  0.0109106                     T
##   other_allele.outcome eaf.outcome pval.outcome        SNP beta.exposure
## 1                    G    0.725704    0.2332993 rs10795302     0.1325959
## 2                    G    0.725832    0.2323564 rs10795303     0.1324959
## 3                    C    0.725339    0.2478531  rs6560863     0.1307491
## 4                    G    0.725719    0.2210069  rs1904671     0.1318013
## 5                    G    0.725719    0.2210069  rs1904671     0.1315361
## 6                    A    0.723641    0.2370712 rs10795306     0.1289839
##   se.exposure effect_allele.exposure other_allele.exposure eaf.exposure
## 1  0.02808845                      G                     A     0.211479
## 2  0.02808579                      G                     A     0.211491
## 3  0.02804180                      C                     A     0.211538
## 4  0.02804180                      G                     A     0.211861
## 5  0.02804489                      G                     A     0.211825
## 6  0.02804754                      A                     T     0.212326
##   pval.exposure
## 1   2.35084e-06
## 2   2.38735e-06
## 3   3.12169e-06
## 4   2.59926e-06
## 5   2.72924e-06
## 6   4.25009e-06
```

Independent SNPs

We proceed to find all independent SNPs by means of the Linkage Disequilibrium (LD) Clumplig process. This process reports the most significant genetic associations in a region in terms of a smaller number of genetically linked SNP "clumps". This can help assess how many independent loci are associated with a given trait.

```r
all <- TwoSampleMR::clump_data(exp_out,                # base de datos
                               clump_r2 = 0.05,
                               pop = "AMR"             # Population reference panel - America
                               )
```

```
## pval.exposure and pval.outcome columns present. Using pval.exposure for clumping.

## 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 th

## Clumping zQyFlj, 1023 variants, using AMR population reference
```

```
## Removing 1000 of 1023 variants due to LD with other variants or absence from LD reference panel
```

```
head(all)
```

```
##              cpaid beta.outcome se.outcome effect_allele.outcome
## 1   10:1012315:A:G    -0.013016  0.0109204                     A
## 7  11:35544832:C:T    -0.001765  0.0147908                     T
## 9  15:54797917:C:T    -0.009535  0.0186090                     T
## 10 16:50646908:A:G    -0.027037  0.0413166                     G
## 11 16:80891881:C:T    -0.063719  0.0320083                     C
## 12 16:88146661:C:G     0.029501  0.0172718                     C
##    other_allele.outcome eaf.outcome pval.outcome        SNP beta.exposure
## 1                     G    0.725704    0.2332993 rs10795302     0.1325959
## 7                     C    0.890695    0.9050130 rs12800242     0.1960886
## 9                     C    0.933329    0.6083812 rs72738425    -0.1967064
## 10                    A    0.974936    0.5128630 rs74460821    -0.5409063
## 11                    T    0.974680    0.0465134 rs76233293    -0.2558357
## 12                    G    0.135256    0.0876285 rs71392336    -0.1541312
##    se.exposure effect_allele.exposure other_allele.exposure eaf.exposure
## 1   0.02808845                      G                     A     0.211479
## 7   0.04110638                      T                     C     0.918110
## 9   0.04171195                      T                     C     0.918888
## 10  0.11143716                      G                     A     0.986327
## 11  0.05314069                      C                     T     0.951605
## 12  0.03355297                      G                     C     0.837493
##    pval.exposure id.exposure
## 1    2.35084e-06      zQyFlj
## 7    1.83979e-06      zQyFlj
## 9    2.40722e-06      zQyFlj
## 10   1.21050e-06      zQyFlj
## 11   1.47713e-06      zQyFlj
## 12   4.35529e-06      zQyFlj
```

This gives us the final data set with 23 SNPs that we will be working on.

F-statistics

F-statistics describes the statistically expected level of heterozygosity in a population; more specifically the expected degree of (usually) a reduction in heterozygosity when compared to Hardy–Weinberg expectation.

```
all$f1 <-(all$beta.exposure * all$beta.exposure) / (all$se.exposure * all$se.exposure)
all$f1
```

```
##  [1]  22.28460  22.75549  22.23903  23.56047  23.17756  21.10180  22.85019
##  [8]  21.49357  21.51265  24.23331  21.10087  23.39318  21.27883  46.73277
## [15]  25.57593  50.97732  23.28218 164.93471  21.50016  20.98004  20.93580
## [22]  24.72514  23.65520
```

```
mean(all$f1)
```

```
## [1] 31.05569
```

Align the SNPs on the same effect allele for exposure and outcome

We align SNPs with the same effect allele for exposure and event by changing the exposure beta symbol if the allele effects are not similar.

```
head(all)
```

```
##              cpaid beta.outcome se.outcome effect_allele.outcome
## 1   10:1012315:A:G    -0.013016  0.0109204                     A
## 7  11:35544832:C:T    -0.001765  0.0147908                     T
## 9  15:54797917:C:T    -0.009535  0.0186090                     T
## 10 16:50646908:A:G    -0.027037  0.0413166                     G
## 11 16:80891881:C:T    -0.063719  0.0320083                     C
## 12 16:88146661:C:G     0.029501  0.0172718                     C
##    other_allele.outcome eaf.outcome pval.outcome        SNP beta.exposure
## 1                     G    0.725704    0.2332993 rs10795302    -0.1325959
## 7                     C    0.890695    0.9050130 rs12800242     0.1960886
## 9                     C    0.933329    0.6083812 rs72738425    -0.1967064
## 10                    A    0.974936    0.5128630 rs74460821    -0.5409063
## 11                    T    0.974680    0.0465134 rs76233293    -0.2558357
## 12                    G    0.135256    0.0876285 rs71392336     0.1541312
##    se.exposure effect_allele.exposure other_allele.exposure eaf.exposure
## 1   0.02808845                      G                     A     0.211479
## 7   0.04110638                      T                     C     0.918110
## 9   0.04171195                      T                     C     0.918888
## 10  0.11143716                      G                     A     0.986327
## 11  0.05314069                      C                     T     0.951605
## 12  0.03355297                      G                     C     0.837493
##    pval.exposure id.exposure       f1
## 1    2.35084e-06      zQyFlj 22.28460
## 7    1.83979e-06      zQyFlj 22.75549
## 9    2.40722e-06      zQyFlj 22.23903
## 10   1.21050e-06      zQyFlj 23.56047
## 11   1.47713e-06      zQyFlj 23.17756
## 12   4.35529e-06      zQyFlj 21.10180
```

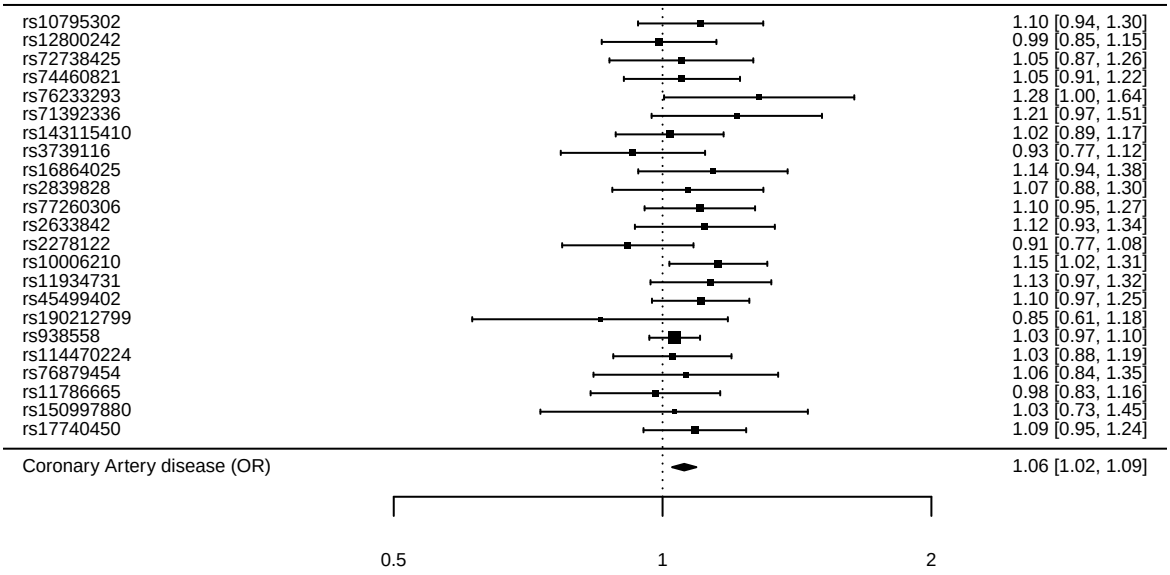
Wald test & Forest plot

We run the Wald test with fixed effects and generate a forest plot with the fixed effects method Mendelian randomization.

```
dmres
```

```
##
## Fixed-Effects Model (k = 23)
##
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  0.80
##
## Test for Heterogeneity:
## Q(df = 22) = 17.5035, p-val = 0.7350
```

```
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.0558  0.0160  3.4886  0.0005  0.0245  0.0872  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



MR estimates

We obtain the estimates for CAD using the Mendelian randomization package with the fixed effects model by the inverse-variance weighted method.

```
##
## Inverse-variance weighted method
## (variants uncorrelated, fixed-effect model)
##
## Number of Variants : 23
##
## ------------------------------------------------------------------
##  Method Estimate Std Error 95% CI       p-value
##     IVW    0.056     0.016 0.024, 0.087   0.000
## ------------------------------------------------------------------
```

```
## Residual standard error =  0.892
## Residual standard error is set to 1 in calculation of confidence interval by fixed-effect assumption
## Residual standard error is set to 1 in calculation of confidence interval when its estimate is less t
## Heterogeneity test statistic (Cochran's Q) = 17.5035 on 22 degrees of freedom, (p-value = 0.7350). I
## F statistic = 31.1.
```

We obtain estimates for CAD using the Mendelian randomization package with the random effects model.

```
##
## Inverse-variance weighted method
## (variants uncorrelated, random-effect model)
##
## Number of Variants : 23
##
## ------------------------------------------------------------------
##  Method Estimate Std Error 95% CI       p-value
##     IVW    0.056     0.016 0.024, 0.087   0.000
## ------------------------------------------------------------------
## Residual standard error =  0.892
## Residual standard error is set to 1 in calculation of confidence interval when its estimate is less t
## Heterogeneity test statistic (Cochran's Q) = 17.5035 on 22 degrees of freedom, (p-value = 0.7350). I
## F statistic = 31.1.
```

We can also obtain weighted median and MR-Egger estimates.

```
##
##  Weighted median method
##
## Number of Variants : 23
## ------------------------------------------------------------------
##                   Method Estimate Std Error 95% CI       p-value
##  Weighted median method    0.049     0.023 0.003, 0.094   0.037
## ------------------------------------------------------------------
```

```
##
## MR-Egger method
## (variants uncorrelated, random-effect model)
##
## Number of Variants =  23
##
## ------------------------------------------------------------------
##       Method Estimate Std Error  95% CI       p-value
##     MR-Egger    0.062     0.040 -0.015, 0.140   0.116
##  (intercept)   -0.001     0.008 -0.017, 0.014   0.855
## ------------------------------------------------------------------
## Residual Standard Error :  0.912
## Residual standard error is set to 1 in calculation of confidence interval when its estimate is less t
## Heterogeneity test statistic = 17.4701 on 21 degrees of freedom, (p-value = 0.6822)
## I^2_GX statistic: 82.4%
```

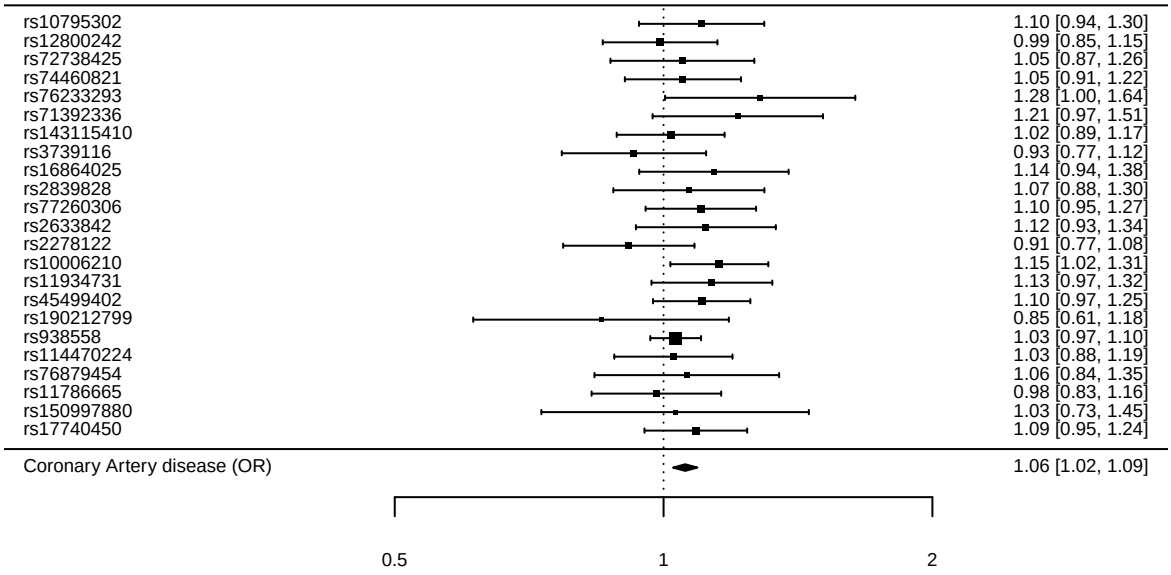
Mendelian Randomization Pleiotropy

With our data we ran the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-Presso) test, this method helps to identify SNPs that compose the CAD by checking which of these are outliers.

```
## Warning in mr_presso(BetaOutcome = "beta.outcome", BetaExposure =
## "beta.exposure", : No outlier were identified, therefore the results for the
## outlier-corrected MR are set to NA
```

```
## $`Main MR results`
##        Exposure       MR Analysis Causal Estimate         Sd  T-stat
## 1 beta.exposure               Raw      0.05581901 0.01427209 3.91106
## 2 beta.exposure Outlier-corrected              NA         NA      NA
##        P-value
## 1 0.0007490214
## 2           NA
##
## $`MR-PRESSO results`
## $`MR-PRESSO results`$`Global Test`
## $`MR-PRESSO results`$`Global Test`$RSSobs
## [1] 19.04688
##
## $`MR-PRESSO results`$`Global Test`$Pvalue
## [1] 0.766
```

```
##
## Random-Effects Model (k = 23; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0 (SE = 0.0014)
## tau (square root of estimated tau^2 value):      0
## I^2 (total heterogeneity / total variability):   0.00%
## H^2 (total variability / sampling variability):  1.00
##
## Test for Heterogeneity:
## Q(df = 22) = 17.5035, p-val = 0.7350
##
## Model Results:
##
## estimate      se    zval    pval   ci.lb   ci.ub
##   0.0558  0.0160  3.4886  0.0005  0.0245  0.0872  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


rs10795302 1.10 [0.94, 1.30]
rs12800242 0.99 [0.85, 1.15]
rs72738425 1.05 [0.87, 1.26]
rs74460821 1.05 [0.91, 1.22]
rs76233293 1.28 [1.00, 1.64]
rs71392336 1.21 [0.97, 1.51]
rs143115410 1.02 [0.89, 1.17]
rs3739116 0.93 [0.77, 1.12]
rs16864025 1.14 [0.94, 1.38]
rs2839828 1.07 [0.88, 1.30]
rs77260306 1.10 [0.95, 1.27]
rs2633842 1.12 [0.93, 1.34]
rs2278122 0.91 [0.77, 1.08]
rs10006210 1.15 [1.02, 1.31]
rs11934731 1.13 [0.97, 1.32]
rs45499402 1.10 [0.97, 1.25]
rs190212799 0.85 [0.61, 1.18]
rs938558 1.03 [0.97, 1.10]
rs114470224 1.03 [0.88, 1.19]
rs76879454 1.06 [0.84, 1.35]
rs11786665 0.98 [0.83, 1.16]
rs150997880 1.03 [0.73, 1.45]
rs17740450 1.09 [0.95, 1.24]
Coronary Artery disease (OR) 1.06 [1.02, 1.09]
0.5
1
2