Package 'lhcMR'

October 25, 2021

Title Latent Heritable Confounder - Mendelian Randomisation

Version 0.0.0.9000

Description lhcMR esimates a causal effect between two traits while accounting for a possible latent heritable confounder acting on them, as well as sample overlap.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Imports data.table,
dplyr,
GenomicSEM,
magrittr,
MASS,
mixtools,
parallel,
rslurm,
stats,
stringr,
TwoSampleMR

URL https://github.com/LizaDarrous/lhcMR

BugReports https://github.com/LizaDarrous/lhcMR/issues

Suggests rmarkdown,
knitr

VignetteBuilder knitr

R topics documented:

`calculate_SP` *Calculate starting points to be used in the likelihood function optimisation*

Description

Calculate starting points to be used in the likelihood function optimisation

Usage

```
calculate_SP(
  input.df,
  trait.names,
  run_ldsc = TRUE,
  run_MR = TRUE,
  saveRFiles = TRUE,
  hm3 = NA,
  ld = NA,
  nStep = 2,
  SP_single = 3,
  SP_pair = 50,
  SNP_filter = 10,
  SNP_filter_ldsc = NA,
  nCores = NA,
  M = 1e+07
)
```

Arguments

`input.df`	The resulting data frame from merge_sumstats(), where the effect size, SE, RSID and other columns are present, in addition to columns representing LD scores, weights and local LD structure
`trait.names`	Vector containing the trait names in the order they were used in merge_sumstats(): Exposure, Outcome
`run_ldsc`	Boolean. Whether GenomicSEM::ldsc should be run to obtain the cross trait-intercept (i_XY). If FALSE, a random value will be generated. Default value = TRUE
`run_MR`	Boolean. Whether TwoSampleMR::mr should be run to obtain the bidirectional causal effects (axy_MR, ayx_MR). If FALSE, random values will be generated. Default value = TRUE
`saveRFiles`	Boolean, whether to write the results of GenomicSEM::ldsc,TwoSampleMR::mr, and the single trait analysis of LHC-MR (returns trait intercept and polygenicity) Default value = TRUE
`hm3`	Path to the input file (HAPMAP3 SNPs) required by GenomicSEM::ldsc
`ld`	Path to the input file (LD scores) required by GenomicSEM::ldsc
`nStep`	Can take two numerical values: 1 or 2. Represents the number of steps the lhcMR analysis will undertake. One single step estimates all 9 parameters simultaneously while fixing only the traits' intercepts iX and iY, while two steps

estimates 7 parameters after having estimated traits' intercepts and polygenicity (iX, piX, iY, piY) from the single trait analysis and fixed their values in the likelihood optimisation and parameter estimation

`SP_single` Numerical value indicating how many starting points should the single trait analysis use in the likelihood optimisation. Best to range between 3-5, default value = 3

`SP_pair` Numerical value indicating how many starting points should the pair trait analysis use in the likelihood optimisation. Best to range between 50-100, default value = 50

`SNP_filter` Numerical value indicating the filtering of every nth SNP to reduce large datasets and speed up analysis. Default value = 10

`SNP_filter_ldsc` Numerical value indicating the filtering of every nth SNP to reduce large datasets and speed up the LDSC analysis. Set to 1 if no filtering is needed, otherwise default = 10

`nCores` Numerical value indicating number of cores to be used in 'mclapply' to parallelise the analysis. If not set (default value = NA), then it will be calculated as 2/3 of the available cores

`M` Numerical value indicating the number of SNPs used to calculate the LD reported in the LD file (for genotyped SNPs). Default value = 1e7

Value

Returns a list containing the filtered dataset (by every `SNP_filter`th SNP), the starting points to be used in the pair trait optimisation, the traits' intercepts, the traits' polygenicity if nStep = 2, as well as some extra parameters like the cross-trait intercept and bidirectional causal effect estimated by IVW

`lhc_mr` *Main trait pair analysis using LHC-MR*

Description

Main trait pair analysis using LHC-MR

Usage

```
lhc_mr(
  SP_list,
  trait.names,
  partition = NA,
  paral_method = "rslurm",
  nCores = NA,
  nBlock = 200,
  M = 1e+07
)
```

Arguments

SP_list	List resulting from calculate_SP. Contains the filtered dataset (by every 'SNP_filter'th SNP), the starting points to be used in the pair trait optimisation, the traits' intercepts, the traits' polygenicity if nStep = 2, as well as some extra parameters like the cross-trait intercept and bidirectional causal effect estimated by IVW.
trait.names	Vector containing the trait names in the order they were used in merge_sumstats(): Exposure, Outcome
partition	String indicating the partition name to be used for the "rslurm" parallelisation
paral_method	String indicating which method to parallelise the optimisation over the number of sets of starting points. "rslurm" will submit the calculation to a SLurm cluster using a 'Slurm' workload manager, "lapply" will parallelise the optimisation using 'mclapply' over a set number of cores but will go sequentially over the sets of starting points and thus take more time.
nCores	Numerical value indicating number of cores to be used in 'mclapply' to parallelise the analysis. If not set (default value = NA), then it will be calculated as 2/3 of the available cores
nBlock	Numerical value indicating the number of blocks to create from the block jackknife analysis, where at each iteration one block is left out and the optimisation is ran again for a single starting point to obtain eventually 'nBlock' estimates and calculate the SE of the parameter estimates
M	Numerical value indicating the number of SNPs used to calculate the LD reported in the LD file (for genotyped SNPs). Default value = 1e7

Value

Prints out a summary of the results

---

merge_sumstats *Merge summary statistics into a single input data frame*

Description

Merge summary statistics into a single input data frame

Usage

```
merge_sumstats(
  input.files,
  trait.names,
  LD.filepath,
  rho.filepath,
  mafT = 0.005,
  infoT = 0.99
)
```

Arguments

`input.files`	• list of data frames, where each data frame contains the summary statistics of a trait to use in the order of Exposure - Outcome
`trait.names`	• Vector containing the trait names in the order they're found in 'input files'
`mafT`	• Minor allele frequency threshold of selection, to be used if a MAF column is found in the summary statistics file. Default value = 0.005
`infoT`	• SNP imputation quality threshold, to be used if an INFO column is found in the summary statistics file. Default value = 0.99
`LD.file`	• LD scores file, either obtained from Alkes group (1000G) or the one provided in the github (UK10K)
`rho.file`	• Genotyped SNP-specific (local) LD scores

Value

Returns a data frame where the summary statistics file, the LD file, and the SNP-specific LD file are merged

Index