

Package ‘MultiMeta’

February 19, 2015

Type Package

Title Meta-analysis of Multivariate Genome Wide Association Studies

Version 0.1

Date 2014-08-21

Author Dragana Vuckovic

Maintainer Dragana Vuckovic <dragana.vuckovic@burlo.trieste.it>

Description Allows running a meta-analysis of multivariate Genome Wide Association Studies (GWAS) and easily visualizing results through custom plotting functions. The multivariate setting implies that results for each single nucleotide polymorphism (SNP) include several effect sizes (also known as “beta coefficients”, one for each trait), as well as related variance values, but also covariance between the betas. The main goal of the package is to provide combined beta coefficients across different cohorts, together with the combined variance/covariance matrix. The method is inverse-variance based, thus each beta is weighted by the inverse of its variance-covariance matrix, before taking the average across all betas. The default options of the main function `multi_meta` will work with files obtained from GEMMA multivariate option for GWAS (Zhou & Stephens, 2014). It will work with any other output, as soon as columns are formatted to have the according names. The package also provides several plotting functions for QQ-plots, Manhattan Plots and custom summary plots.

License GPL (>= 2)

Imports mvtnorm, expm, ggplot2, reshape2

Depends gtable, grid

Collate 'betas_plot.R' 'mhplot.R' 'multi_meta.R' 'qqplotter.R'
'Example_file_1.R' 'Example_file_2.R' 'Example_output_file.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2015-01-15 11:44:10

R topics documented:

betas_plot	2
Example_file_1	3
Example_file_2	3
Example_output_file	4
mhplot	4
multi_meta	5
qqplotter	7

Index	8
--------------	----------

betas_plot	<i>Summary plot of the effect sizes and their correlations for a SNP of interest</i>
------------	--

Description

betas_plot returns a pdf file containing information about the meta-analysis combined effect sizes (beta coefficients) for a chosen SNP of interest.

Usage

```
betas_plot(SNP, res.file = NULL, data = NULL,
           trait.names = NULL, out.file = "Betas.pdf", col = NULL,
           orig.files = NULL, cohort.names = NULL)
```

Arguments

SNP	A SNP of interest, e.g. a SNP that is significantly associated to the analysed traits.
res.file	File name of the multi_meta results containing effect sizes for the above SNP. This is an alternative to data.
data	Object containing results from the multi_meta function and the above SNP. This is an alternative to file.
trait.names	Vector of analysed trait names to appear on the plot.
out.file	File name for the output plot.
col	Colours for the heat-map plot. Combinations of several colours are possible. If left blank a heat-map in scales of grey is produced.
orig.files	List of files with multivariate results for each cohort.
cohort.names	Names of the cohorts included in the meta-analysis.

Details

The plot produced can be a useful tool for visualizing effect sizes across all traits, and the correlation between them. The plot will have two panels:

1. effect sizes with 95% confidence interval for each trait;
2. a heat-map showing correlations between betas.

Two different types of input are allowed:

- a data.frame, containing the output from the `multi_meta` function for at least the specified SNP. This is the best choice if results have already been loaded into the R workspace.
- a file name containing the output from `multi_meta` for at least the specified SNP. The separator field is set to white space. This file can be a subset of the complete results file (suggested), e.g. a file containing only significant SNPs.

Value

The output is a pdf file containing the described plot.

Example_file_1

Example data file 1

Description

Example data file 1

Author(s)

Dragana Vuckovic <dragana.vuckovic@burlo.trieste.it>

Example_file_2

Example data file 2

Description

Example data file 2

Author(s)

Dragana Vuckovic <dragana.vuckovic@burlo.trieste.it>

Example_output_file *Example output file*

Description

This file is an example of the output obtained with `multi_meta` function. The input files were `Example_file_1.txt` and `Example_file_2.txt`, the command line was the one in the example (see `multi_meta help`)

The columns are:

- **chr** Chromosome
- **SNP** SNP name
- **Position** Position
- **allele1** Effect allele
- **allele0** Non-effect allele
- **tot_af** Effect-allele frequency computed on all the cohorts analysed
- **n_pops** Number of cohorts (populations) analysed
- **pops** String of 1 or 0 values indicating which populations were analysed: e.g. if the first number is 1, the first cohort was included in the analysis etc.
- **beta_1, beta_2, ..., beta_p** Effect sizes for each of the p traits; in this example $p=6$
- **Vbeta_1_1, Vbeta_1_2, ..., Vbeta_1_p, Vbeta_2_2, ..., Vbeta_2_p, ..., Vbeta_p_p** Variance-covariance matrix entries (diagonal and upper triangle values only, since this matrix is symmetric)
- **p_value** P-value

Author(s)

Dragana Vuckovic <dragana.vuckovic@burlo.trieste.it>

mhplot

Manhattan Plot for meta-analysis results

Description

`mhplot` returns a pdf file containing the Manhattan plot of meta-analysis results.

Usage

```
mhplot(res.file = NULL, data = NULL,
       col = c("darkgrey", "lightgrey"),
       out.file = "mhplot.pdf", main = "MHplot",
       plab = "p_value", CHRlab = "chr", POSlab = "Position",
       sig.p = 5e-08, sug.p = 5e-07, sep = " ")
```

Arguments

res.file	File name for the file(s) containing meta-analysis results. For multiple files (e.g. one file for each chromosome) the special character "*" is allowed. This is an alternative to data.
data	Dataset containing meta-analysis results. This is an alternative to file, if results are already loaded in the R workspace.
col	Choice of two colours for the plot. They will be alternating for different chromosomes.
out.file	File name for the output.
main	Title to appear on plot.
plab	p-values label i.e. column name for the p-values column.
CHRlab	Chromosome label i.e. column name for the chromosome numbers column.
POSlab	Position label i.e. column name for the position numbers column.
sig.p	Significant p-values level.
sug.p	Suggestive p-values level.
sep	Separator used in file. Default is white space.

Details

The function is useful for a quick overview of meta-analysis results. Two different types of input are allowed:

- a data frame containing chromosome, position and p-values to be plotted.
- a file name for retrieving information from file. Note that separator is set to white space, can be changed accordingly if needed. Special character "*" is allowed for selecting more than one file (e.g. in case there is one file for each chromosome).

Value

The output is a pdf file containing the Manhattan Plot.

multi_meta

Meta-analysis of multivariate GWAS results

Description

multi_meta returns the meta-analysis results for multivariate GWAS across different cohorts.

Usage

```
multi_meta(files = c(), N = c(),  
           output.file = "Meta_Results.txt", size.chunks = 5e+06,  
           min.pop = 2, sep = "\t")
```

Arguments

files	A vector containing the names of the results files to meta-analyse. These can be outputs from GEMMA multivariate analysis or similar (see Details). Furthermore they can be single-chromosome or genome-wide results.
N	A vector containing sample sizes for each of the above files. This parameter is optional and is only required for computing the overall allele frequency.
output.file	The name of the output file.
size.chunks	Size of each chunk to be read and processed. Default is 5,000,000 (5 Mb). This size will require very low memory usage. Increase this parameter if more memory is allocated or if the number of cohorts is limited. Read more about the chunks in Details .
min.pop	Minimum number of populations required per SNP to compute meta-analysis. Default is 2, it can be any number up to the total number of cohorts analysed.
sep	Separator for reading input files.

Details

This function applies an inverse-variance based method to meta-analyse multivariate GWAS results. In particular, given n different cohorts, for which p phenotypes have been tested for genome-wide association, the results for each cohort will have p different effect size coefficients i.e. beta values (one per each phenotype) and a variance/covariance $p \times p$ matrix representing beta's variances and covariances. In particular, the function is built to consider the output from the GEMMA software multivariate association testing. If your output is not produced with GEMMA, the function works on any results file containing the following column names:

- **chr** Chromosome
- **ps** Position
- **rs** SNP name
- **allele1** Effect allele
- **allele0** Non-effect allele
- **af** Effect-allele frequency
- **beta_1, beta_2, ..., beta_p** Effect sizes for each of the p traits
- **Vbeta_1_1, Vbeta_1_2, ..., Vbeta_1_p, Vbeta_2_2, ..., Vbeta_2_p, ..., Vbeta_p_p** variance-covariance matrix entries (diagonal and upper triangle values only, since this matrix is symmetric)

The function divides input files into chunks based on position. Only one chunk at a time is read and analysed; thus a limited amount of data is loaded in the workspace at one given time. Default chunk dimension is 5 Mb for which low memory is required (<250 MB for 2 cohorts). If you have larger RAM availability, sparse markers or a limited number of cohorts, change chunks' dimension from the command line.

Examples

```
file1=system.file("extdata", "Example_file_1.txt", package="MultiMeta")
file2=system.file("extdata", "Example_file_2.txt", package="MultiMeta")
multi_meta(files=c(file1,file2), N=c(1200,600), sep=" ",
output.file="Output_from_running_example.txt")
```

qqplotter

Quantile-quantile plot for observed vs. expected p-values

Description

qqplotter returns a pdf file containing the QQ-plot for a given set of p-values

Usage

```
qqplotter(p_values = NULL, res.file = NULL,
p.num.col = NULL, sep = " ", out.file = "QQplot.pdf",
title = "")
```

Arguments

p_values	A vector or a column in a data-set containing p-values to plot. This is an alternative to file.
res.file	File name of the file containing p-values to retrieve. This alternative does not require loading the results and is an alternative to the p_values parameter.
p.num.col	Number indicating which column of the file contains p-values. Specify only if you are using the file option.
sep	File separator to be used when reading file. Default is white space.
title	Title to appear on the plot.
out.file	File name for the output plot.

Details

The function is useful for comparing expected and observed distributions of statistics arising from genome-wide testing. Two different types of input are allowed:

- an object such as vector or data-frame column, containing p-values;
- a file name and the column number for retrieving p-values from file. Note that separator is set to white space, can be changed accordingly if needed. The special character "*" is allowed for selecting more than one file (e.g. in case there is one file for each chromosome) but the p-values column position must be the same for all files.

Value

The output will be a pdf file containing the QQplot.

Index

*Topic **datasets**

Example_file_1, 3

Example_file_2, 3

Example_output_file, 4

*Topic **data**

Example_file_1, 3

Example_file_2, 3

Example_output_file, 4

betas_plot, 2

Example_file_1, 3

Example_file_2, 3

Example_output_file, 4

mhplot, 4

multi_meta, 5

qqplotter, 7