Package 'MetaVolcanoR'

November 13, 2023

Type Package

Title Gene Expression Meta-analysis Visualization Tool

Version 1.17.0

Description MetaVolcanoR combines differential gene expression results. It implements three strategies to summarize differential gene expression from different studies. i) Random Effects Model (REM) approach, ii) a p-value combining-approach, and iii) a vote-counting approach. In all cases, MetaVolcano exploits the Volcano plot reasoning to visualize the gene expression meta-analysis results.

Depends R (>= 4.1.1)

Imports methods, data.table, dplyr, tidyr, plotly, ggplot2, cowplot, parallel, metafor, metap, rlang, topconfects, grDevices, graphics, stats, htmlwidgets

Suggests knitr, markdown, rmarkdown, testthat

License GPL-3

Encoding UTF-8

biocViews GeneExpression, DifferentialExpression, Transcriptomics, mRNAMicroarray, RNASeq

LazyData true

roxygennote 6.1.1.9000

VignetteBuilder knitr

RoxygenNote 6.1.1

git_url https://git.bioconductor.org/packages/MetaVolcanoR

git_branch devel

git_last_commit ff2314c

git_last_commit_date 2023-10-24

Date/Publication 2023-11-13

Author Cesar Prada [aut, cre], Diogenes Lima [aut], Helder Nakaya [aut, ths]

Maintainer Cesar Prada <cesar.prada.medina@gmail.com>

Table of contents:

`calc_vi` *A function to calculate variance from confidence interval limits*

Description

This function takes the limits of a confidence interval (95 a calculate a variance

Usage

```
calc_vi(diffexp, llcol, rlcol)
```

Arguments

`diffexp`	data.frame/data.table containing differential expression results
`llcol`	column name of the fold change coinfidence interval left limit name <string>
`rlcol`	column name of the fold change coinfidence interval left limit name <string>

Value

`data.table/data.frame` with a new `vi` variable

Examples

```
data(diffexplist)
diffexp <- calc_vi(diffexplist[[1]], "CI.L", "CI.R")
head(diffexp, 3)
```

collapse_deg *A function to filter out geneIDs standing for the same gene name*

Description

This function to remove redundant geneIDs standing for the same gene name

Usage

```
collapse_deg(diffexp, genenamecol, pcriteria)
```

Arguments

diffexp	data.frame/data.table output of the deg.def() function
genenamecol	the column name of the gene name variable <string>
pcriteria	the column name of the pvalue criteria to consider <string>

Value

data.table differential expression results with unique gene names

Examples

```
data(diffexplist)
diffexp <- collapse_deg(diffexplist[[1]], "Symbol", "pvalue")
head(diffexp, 3)
```

combining_mv *A function to draw the 'Combining meta-analysis' MetaVolcano*

Description

This function draws the 'Combining meta-analysis' MetaVolcano

Usage

```
combining_mv(diffexp = list(), pcriteria = "pvalue",
  foldchangecol = "Log2FC", genenamecol = "Symbol", geneidcol = NULL,
  metafc = "Mean", metathr = 0.01, collaps = "FALSE",
  jobname = "MetaVolcano", outputfolder = ".", draw = "HTML")
```

Arguments

diffexp	list of data.frame/data.table (s) with DE results where lines are genes
pcriteria	the column name of the Pval criteria to consider c("adj.P.Val", "P.Value") <string>
foldchangecol	the column name of the foldchange variable <string>
genenamecol	the column name of the gene name variable <string>
geneidcol	the column name of the gene ID/probe/oligo/transcript variable <string>
metafc	method for summarizing gene fold-changes across studies c("Mean", "Median") <string>
metathr	top percentage of perturbed genes to be highlighted <double>
collaps	if probes should be collapsed based on the DE direction <logical>
jobname	name of the running job <string>
outputfolder	/path where to write the results/
draw	wheather or not to draw the .pdf or .html visualization <c(NULL, "PDF", "HTML")>

Value

MetaVolcano object

Examples

```
data(diffexplist)
mv <- combining_mv(diffexplist)
str(mv)
```

---

cum_freq_data *A data formating function for inverse-cummulative DEG distribution*

Description

This function counts how many genes consistly appears as DE along the input studies

Usage

```
cum_freq_data(meta_diffexp, nstud)
```

Arguments

meta_diffexp	data.frame/data.table containing all the input studies
nstud	the number of inputed GEO2R outputs <integer>

Value

data.frame inverse cummulative distribution

Examples

```
library(dplyr)
data(diffexplist)
diffexp <- lapply(diffexplist, function(...) deg_def(..., "pvalue",
           "Log2FC", 0.05, 0))
diffexp <- rename_col(diffexp, "Symbol")
meta_diffexp <- Reduce(function(...) merge(..., by = "Symbol", all = TRUE),
           diffexp)
meta_diffexp %>%
dplyr::select(dplyr::matches("deg_")) %>%
    data.matrix -> n_deg
meta_diffexp[['ndeg']] <- rowSums(n_deg^2, na.rm = TRUE)
cfd <- cum_freq_data(meta_diffexp, length(diffexplist))
head(cfd, 3)
```

---

deg_def *A DEG definition function*

Description

This function creates a new variable indicating DEGs as -1, 0, 1 based on the user-defined fold-change and p-value criteria

Usage

```
deg_def(diffexp, pcriteria, foldchangecol, pv, fc)
```

Arguments

diffexp	data.frame/data.table with differential expression results
pcriteria	column name of the pvalue variable <strings>
foldchangecol	column name of the foldchange variable <string>
pv	pvalue threshold <double>
fc	foldchange threshold <double>

Value

data.table/data.frame with a new deg variable

Examples

```
data(diffexplist)
diffexp <- deg_def(diffexplist[[1]], "pvalue", "Log2FC", 0.05, 0)
table(diffexp[['deg']])
```

diffexplist *Differential expression results from five studies*

Description

A named list with five differential expresion results.

Usage

```
diffexplist
```

Format

A named list with 5 data frames with ~20k genes and 5 variables:

GSE12050 differential expression result, disease vs healthy

GSE24883 differential expression result, disease vs healthy ...

Source

https://www.ncbi.nlm.nih.gov/geo/

draw_cum_freq *A function to visualize the inverse-cummulative DEG distribution*

Description

This function create a ggplot object with the inverse-cummulative DEG distribution

Usage

```
draw_cum_freq(meta_diffexp, nstud)
```

Arguments

meta_diffexp	data.frame/data.table containing all the input studies
nstud	the number of inputed GEO2R outputs <integer>

Value

ggplot2 object

Examples

```
library(dplyr)
data(diffexplist)
diffexp <- lapply(diffexplist, function(...) deg_def(..., "pvalue",
           "Log2FC", 0.05, 0))
diffexp <- rename_col(diffexp, "Symbol")
meta_diffexp <- Reduce(function(...) merge(..., by = "Symbol", all = TRUE),
           diffexp)
meta_diffexp %>%
dplyr::select(dplyr::matches("deg_")) %>%
    data.matrix -> n_deg
meta_diffexp[['ndeg']] <- rowSums(n_deg^2, na.rm = TRUE)
gg <- draw_cum_freq(meta_diffexp, length(diffexplist))
plot(gg)
```

---

draw_degbar *A function for DEG barplot visualization*

Description

This function visualize as barplots the number of DEGs across the input studies

Usage

```
draw_degbar(degbar_data)
```

Arguments

degbar_data	output of the `set_degbar_data()` function <data.fram/data.table>

Value

`ggplot2` object

Examples

```
data(diffexplist)
diffexp <- lapply(diffexplist, function(...) deg_def(..., "pvalue",
           "Log2FC", 0.05, 0))
bardat <- set_degbar_data(diffexp)
gg <- draw_degbar(bardat)
plot(gg)
```

`draw_forest` *A function to draw a forest plot from the REM MetaVolcano result*

Description

This function draws a forest plot for a given gene based on the REM MetaVolcano result

Usage

```
draw_forest(remres, gene = "MMP9", genecol = "Symbol",
  foldchangecol = "Log2FC", llcol = "CI.L", rlcol = "CI.R",
  jobname = "MetaVolcano", outputfolder = ".", draw = "PDF")
```

Arguments

`remres`	MetaVolcano object. Output of the rem_mv() function <MetaVolcano>
`gene`	query gene to plot
`genecol`	name of the variable with genes <string>
`foldchangecol`	the column name of the foldchange variable <string>
`llcol`	left limit of the fold change coinfidence interval variable name <string>
`rlcol`	right limit of the fold change coinfidence interval variable name <string>
`jobname`	name of the running job <string>
`outputfolder`	/path where to write the results/ <string>
`draw`	either 'PDF' or 'HTML' to save metaolcano as .pdf or .html respectively <string>

Value

`ggplot2` object

Examples

```
data(diffexplist)
diffexplist <- lapply(diffexplist, function(del) {
    dplyr::filter(del, grepl("MP", Symbol))
})
mv <- rem_mv(diffexplist, metathr = 0.1)
gg <- draw_forest(mv, gene="MMP9")
plot(gg)
```

MetaVolcano-class *An S4 class to represent MetaVolcanoR results*

Description

An S4 class to represent MetaVolcanoR results

Slots

`input` merged diiferential expression inputs `data.frame`

`inputnames` names of the differential expression inputs `character`

`metaresult` meta-analysis results `data.frame`

`MetaVolcano` plot with meta-analysis results

`degfreq` supplementary figure of the vote-counting MetaVolcano

plot_mv *A MetaVolcano ploting function*

Description

This function plots either the combining- or the vote-counting- MetaVolcanos

Usage

```
plot_mv(meta_diffexp, nstud, genecol, comb, metafc)
```

Arguments

`meta_diffexp`	data.frame/data.table containing the differential expression inputs
`nstud`	the number of differential expression inputs <integer>
`genecol`	column name of the variable to label genes in the .html file <string>
`comb`	wheather or not the drawing is for the combining-metavolcano <logical>
`metafc`	method for summarizing gene fold-changes across studies c("Mean", "Median") <string>

Value

`ggplot2` object

Examples

```
data(diffexplist)
mv <- votecount_mv(diffexplist)
gg <- plot_mv(mv@metaresult, length(diffexplist), "Symbol", FALSE, "Mean")
plot(gg)
```

plot_rem — *A function to plot the Random Effect Model (REM) MetaVolcano*

Description

This function plots the REM MetaVolcano using ggplot2

Usage

```
plot_rem(meta_diffexp, jobname, outputfolder, genecol, metathr)
```

Arguments

meta_diffexp	data.frame/data.table containing the REM results from rem_mv() <data.table/data.frame>
jobname	name of the running job <string>
outputfolder	/path where to write the results/ <string>
genecol	column name of the variable to label genes in the .html file <string>
metathr	top percentage of perturbed genes to be highlighted <double>

Value

ggplot2 object

Examples

```
data(diffexplist)
diffexplist <- lapply(diffexplist, function(del) {
    dplyr::filter(del, grepl("MP", Symbol))
})
mv <- rem_mv(diffexplist, metathr = 0.1)
gg <- plot_rem(mv@metaresult, "MV", ".", "Symbol", 0.01)
plot(gg)
```

remodel — *A function to model foldchange variance along several studies This function calculate the REM-summary fold-change*

Description

A function to model foldchange variance along several studies This function calculate the REM-summary fold-change

Usage

```
remodel(gene, foldchangecol, vcol)
```

Arguments

gene	named vector with foldchanges and variances <vector<
foldchangecol	the column name of the foldchange variable <string>
vcol	name of the fold change variance variable <string>

Value

data.frame with REM results for a gene

Examples

```
g <- data.frame('Symbol'="XGENE", 'Log2FC_1'=1.2, 'Log2FC'=0.8,
                'vi_1'=0.01, 'vi_2'=0.1)
remodel(g, 'Log2FC', 'vi')
```

---

rem_mv *A function to perform the Random Effect Model (REM) MetaVolcano*

Description

This function runs the 'Random Effect Model' MetaVolcano section

Usage

```
rem_mv(diffexp = list(), pcriteria = "pvalue",
  foldchangecol = "Log2FC", genenamecol = "Symbol", geneidcol = NULL,
  collaps = FALSE, llcol = "CI.L", rlcol = "CI.R", vcol = NULL,
  cvar = TRUE, metathr = 0.01, jobname = "MetaVolcano",
  outputfolder = ".", draw = "HTML", ncores = 1)
```

Arguments

diffexp	list of data.frame/data.table (s) with DE results where lines are genes
pcriteria	the column name of the pvalue variable <string>
foldchangecol	the column name of the foldchange variable <string>
genenamecol	the column name of the gene name variable <string>
geneidcol	the column name of the gene ID/probe/oligo/transcript variable <string>
collaps	if probes should be collapsed based on the DE direction <logical>
llcol	left limit of the fold change coinfidence interval variable name <string>
rlcol	right limit of the fold change coinfidence interval variable name <string>
vcol	name of the fold change variance variable <string>
cvar	weather or not to calculate gene variance from confidence interval limits <logical>
metathr	top percentage of perturbed genes to be highlighted <double>

jobname	name of the running job <string>
outputfolder	/path where to write the results/
draw	wheather or not to draw the .html visualization <logical>
ncores	the number of processors the user wants to use <integer>

Value

MetaVolcano object

Examples

```
data(diffexplist)
diffexplist <- lapply(diffexplist, function(del) {
    dplyr::filter(del, grepl("MP", Symbol))
})
mv <- rem_mv(diffexplist, metathr = 0.1)
str(mv)
```

---

rename_col *A column renaming function merged inputs*

Description

This function rename the columns of the merged inputs

Usage

```
rename_col(diffexp, genecol)
```

Arguments

diffexp	list of data.frame/data.table (s) with DE results where lines are genes
genecol	the column name of the geneID or gene name variable <string>

Value

data.tabledata.frame with new colnames

Examples

```
data(diffexplist)
lapply(diffexplist, colnames)
diffexp <- rename_col(diffexplist, "Symbol")
lapply(diffexp, colnames)
```

set_degbar_data *A function setting data format for DEG barplot visualization*

Description

This function summarize the variable deg from the deg_def() function to visualize as barplots the number of DEGs per input study

Usage

```
set_degbar_data(diffexp)
```

Arguments

diffexp	list of data.frame/data.table (s) output of the deg_def() function <list>

Value

data.frame DEG by input

Examples

```
data(diffexplist)
diffexp <- lapply(diffexplist, function(...) deg_def(..., "pvalue",
           "Log2FC", 0.05, 0))
bardat <- set_degbar_data(diffexp)
head(bardat, 3)
```

votecount_mv *A function to draw the 'Vote-counting meta-analysis' MetaVolcano*

Description

This function draws the vote-counting meta-analysis MetaVolcano

Usage

```
votecount_mv(diffexp = list(), pcriteria = "pvalue",
  foldchangecol = "Log2FC", genenamecol = "Symbol", geneidcol = NULL,
  pvalue = 0.05, foldchange = 0, metathr = 0.01, collaps = FALSE,
  jobname = "MetaVolcano", outputfolder = ".", draw = "HTML")
```

Arguments

diffexp	list of data.frame/data.table (s) with DE results where lines are genes
pcriteria	the column name of the Pval criteria to consider <string>
foldchangecol	the column name of the foldchange variable <string>
genenamecol	the column name of the gene name variable <string>
geneidcol	the column name of the gene ID/probe/oligo/transcript variable <string>
pvalue	the Pval to use as threshold c(0:1) <double>
foldchange	the foldchange to use as DE threshold c(-Inf: Inf) <double>
metathr	the proportion of studies a gene has to be DEG to be considered cDEG <double>
collaps	if probes should be collapsed based on the DE direction <logical>
jobname	name of the running job <string>
outputfolder	/path where to write the results/
draw	wheather or not to draw a .pdf or .html visualization <c(NULL, 'PDF', 'HTML')>

Value

MetaVolcano object

Examples

```
data(diffexplist)
mv <- votecount_mv(diffexplist)
str(mv)
```

Index