Package 'tmod'

March 10, 2018

Type Package

Title Feature Set Enrichment Analysis for Metabolomics and Transcriptomics

Version 0.36

Date 2018-03-09

Author January Weiner

Maintainer January Weiner <january.weiner@gmail.com>

Description Methods and feature set definitions for feature or gene set enrichment analysis in transcriptional and metabolic profiling data. Package includes tests for enrichment based on ranked lists of features, functions for visualisation and multivariate functional analysis.

License GPL (>= 2.0)

Depends R (>= 2.10)

LazyData false

VignetteBuilder knitr

URL http://bioinfo.mpiib-berlin.mpg.de/tmod/

Imports beeswarm,tagcloud,XML,methods,plotwidgets

Suggests testthat,knitr,rmarkdown,pca3d,limma

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-03-10 16:16:10 UTC

R topics documented:

`tmod-package` *Transcriptional Module Analysis*

Description

Transcriptional Module Analysis

Details

The primary role of this package is to provide published module assignments between genes and transcriptional modules, as well as tools to analyse and visualize the modules.

See Also

`tmodHGtest`, `tmodUtest`

Egambia *Gene expression in TB patients and Healthy controls*

Description

Gene expression in TB patients and Healthy controls

Details

This data set has been constructed from the gene expression data set accessible in the Gene Expression Omnibus (GEO) at the accession number GSE28623. Ten healthy donors (NID, non-infected donors) and 10 tubercolosis patients (TB) have been randomly selected from the full data set, and top 25 genes with the highest IQR have been selected for further analysis. Genes without an Entrez gene (EG) identifier have likewise been omitted.

The Egambia object is a data frame. The first three columns are the gene symbol, gene name and Entrez gene (EG) identifier. The remaining columns correspond to the gene expression data.

Examples

```
## Not run:
# The data set has been generated as follows:
# get the data set from GEO
library( GEOquery )
gambia <- getGEO( "GSE28623" )[[1]]

# Convert to limma and normalize
library( limma )
e <- new( "EListRaw", list( E= exprs( gambia ), genes= fData( gambia ), targets= pData( gambia ) ) )
e.bg <- backgroundCorrect( e, method= "normexp" )
en <- normalizeBetweenArrays( e.bg, method= "q" )
en <- avereps( en, ID= en$genes$NAME )
en <- en[ en$genes$CONTROL_TYPE == "FALSE", ]
en$targets$group <- factor( gsub( " whole blood RNA *", "", en$targets$description ) )

# Fill in Entrez Gene IDs
library( org.Hs.eg.db )
en$genes$EG <- ""
sel <- en$genes$REFSEQ %in% ls( org.Hs.egREFSEQ2EG )
en$genes$EG[sel] <- mget( as.character( en$genes$REFSEQ[sel] ), org.Hs.egREFSEQ2EG )

# Filter by IQR and missing EG's
iqrs <- apply( en$E, 1, IQR )
en2 <- en[ iqrs > quantile( iqrs, 0.75 ) & en$genes$EG != "", ]

# Select 10 random samples from NID and TB groups
en2 <- en2[ , c( sample( which( en2$targets$group == "NID" ), 10 ),
                 sample( which( en2$targets$group == "TB" ), 10 ) ) ]
colnames( en2$E ) <- en2$targets$group
Egambia <- cbind( en2$genes[ , c( "GENE_SYMBOL", "GENE_NAME", "EG" ) ], en2$E )
```

```
## End(Not run)
```

`evidencePlot` *Create an evidence plot for a module*

Description

Create an evidence plot for a module

Usage

```
evidencePlot(l, m, mset = "all", scaled = TRUE, rug = TRUE, roc = TRUE,
  filter = FALSE, unique = TRUE, add = FALSE, col = "black",
  col.rug = "#eeeeee", gene.labels = NULL, gl.cex = 1, style = "roc",
  lwd = 1, lty = 1, rug.size = 0.2, legend = NULL, ...)
```

Arguments

`l`	sorted list of HGNC gene identifiers
`m`	character vector of modules for which the plot should be created
`mset`	Which module set to use (see tmodUtest for details)
`scaled`	if TRUE, the cumulative sums will be divided by the total sum (default)
`rug`	if TRUE, draw a rug-plot beneath the ROC curve
`roc`	if TRUE, draw a ROC curve above the rug-plot
`filter`	if TRUE, genes not defined in the module set will be removed
`unique`	if TRUE, duplicates will be removed
`add`	if TRUE, the plot will be added to the existing plot
`col`	a character vector color to be used
`col.rug`	a character value specifying the color of the rug
`gene.labels`	if TRUE, gene names are shown; alternatively, a named character vector with gene labels to be shown, or NULL (default) for no labels (option evaluated only if rug is plotted)
`gl.cex`	Text cex (magnification) for gene labels
`style`	"roc" for receiver-operator characteristic curve (default), and "gsea" for GSEA-style (Kaplan-Meier like plot)
`lwd`	line width (see par())
`lty`	line type (see par())
`rug.size`	fraction of the plot that should show the rug. If rug.size is 0, rug is not drawn. If rug.size is 1, ROC curve is not drawn.
`legend`	position of the legend. If NULL, no legend will be drawn
`...`	Further parameters passed to the plotting function

Details

This function creates an evidence plot for a module, based on an ordered list of genes. The plot shows the receiving operator characteristic (ROC) curve and a rug below, which indicates the distribution of the module genes in the sorted list.

See Also

tmod-package, hgEnrichmentPlot

Examples

```
# artificially enriched list of genes
set.seed(123)
data(tmod)
bg <- tmod$GENES$ID
fg <- sample(c(tmod$MODULES2GENES[["LI.M127"]], bg[1:1000]))
l <- unique(c(fg, bg))
evidencePlot(l, "LI.M127")
evidencePlot(l, filter=TRUE, "LI.M127")
```

---

getGenes *Get genes belonging to a module*

Description

Get genes belonging to a module

Usage

```
getGenes(modules, genelist = NULL, fg = NULL, mset = "LI")
```

Arguments

modules	module IDs
genelist	list of genes
fg	genes which are in the foreground set
mset	module set to use

Details

Create a data frame mapping each module to a comma separated list of genes. If genelist is provided, then only genes in that list will be shown. An optional column, "fg" informs which genes are in the "foreground" data set.

Value

data frame containing module to gene mapping

getModuleMembers *Return the contents of a module*

Description

Return the contents of a module

Usage

```
getModuleMembers(x, mset = "all")
```

Arguments

x	a character vector of module names
mset	optional, a module set

Details

This function returns the selected modules from a module set.

Value

A list of modules

Examples

```
# show the interferon related modules
getModuleMembers(c("LI.M127", "LI.M158.0", "LI.M158.0"))
```

hgEnrichmentPlot *Create a visualisation of enrichment*

Description

Create a visualisation of enrichment

Usage

```
hgEnrichmentPlot(fg, bg, m, mset = "all", ...)
```

Arguments

fg	the foreground set of genes
bg	the background set of genes (gene universe)
m	module for which the plot should be created
mset	Which module set to use (see tmodUtest for details)
...	additional parameters to be passed to the plotting function

Details

This functions creates a barplot visualizing the enrichment of a module in the foreground (fg) set as compared to the background (bg) set. It is the counterpart

See Also

tmod-package, evidencePlot

Examples

```
set.seed(123)
data(tmod)
bg <- tmod$GENES$ID
fg <- sample(c(tmod$MODULES2GENES[["LI.M127"]], bg[1:1000]))
hgEnrichmentPlot(fg, bg, "LI.M127")
```

---

makeTmod *S4 class for tmod*

Description

S4 class for tmod

Usage

```
makeTmod(modules, modules2genes, genes2modules = NULL, genes = NULL)
```

Arguments

modules	A data frame with at least columns ID and Title
modules2genes	A list with module IDs as names. Each member of the list is a character vector with IDs of genes contained in that module
genes2modules	A list with gene IDs as names. Each member of the list is a character vector with IDs of modules in which that gene is contained. This object will be created automatically if the provided parameter is NULL
genes	A data frame with meta-information on genes. Must contain the column ID. If NULL, then a data frame with only one column (ID) will be created automatically.

Details

An object of class tmod contains the module annotations (tmod$MODULES), gene to module (tmod$GENES2MODULES) and module to gene (tmod$MODULES2GENES) mappings and a gene list (tmod$GENES).

tmod$MODULES and tmod$GENES are data frames, while tmod$MODULES2GENES and tmod$GENES2MODULES are lists with, respectively, module and gene identifiers as names. The data frames MODULES and

GENES must contain the column "ID", and the data frame MODULES must contain additionally the column "Title".

Objects of class tmod should be constructed by calling the function makeTmod(). This function check the validity and consistency of the provided objects, and, if necessary, automatically creates the members GENES and GENES2MODULES.

See the package vignette for more on constructing custom module sets.

See Also

tmod-data

Examples

```
# A minimal example
m <- data.frame(ID=letters[1:3], Title=LETTERS[1:3])
m2g <- list(a=c("g1", "g2"), b=c("g3", "g4"), c=c("g1", "g2", "g4"))
mymset <- makeTmod(modules=m, modules2genes=m2g)
```

---

modmetabo *Modules for metabolic profiling*

Description

Feature and data sets for metabolic profiling

Details

The module set "modmetabo" can be used with tmod to analyse metabolic profiling data. The clusters defined in this set are based on hierarchical clustering of metabolic compounds from human serum and have been published in a paper on metabolic profiling in tuberculosis by Weiner et al. (2012).

For an example analysis, "tbmprof" is a data set containing metabolic profiles of serum isolated from tuberculosis (TB) patients and healthy individuals. The tbmprof is a data frame containing observations in rows and metabolite id's (corresponding to the id's in the modmetabo object). See examples below.

References

Weiner et al. "Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients." PloS one 7.7 (2012): e40221.

See Also

tmod-data

Examples

```
data(modmetabo)  # module definitions
data(tbmprof)    # example data set
ids <- rownames(tbmprof)
tb  <- factor(gsub("\\..*", "", ids))

## use Wilcoxon test to calculate significant differences
wcx <- apply(tbmprof, 2, function(x) wilcox.test(x ~ tb)$p.value)
wcx <- sort(wcx)
tmodCERNOtest(names(wcx), mset=modmetabo)
```

`mtx2mset` *Convert between matrix representation of modules and tmod objects*

Description

Converts a matrix where columns correspond to modules and rows to individual variables into an tmod object ("mset") represenation and back

Usage

```
mtx2mset(x)

mset2mtx(x)
```

Arguments

`x`	for mtx2mset, a numeric matrix with named rows and columns; for mset2mtx, an object of the class tmod

Details

A matrix mapping genes onto modules and vice versa can be converted with mtx2mset into a tmod object. The numeric matrix either only contains '0' and '1' values indicating presence or absence of a variable (gene) in a module (gene set), or any other numbers which are treated as weights. In the latter case, '0' and only '0' is interpreted as absence of a variable in a module; any other value is interpreted as a weight and stored in the WEIGHTS slot of the tmod object.

See Also

tmod-class

pcaplot *Plot a PCA object returned by prcomp*

Description

Plot a PCA object returned by prcomp

Usage

```
pcaplot(pca, components = 1:2, group = NULL, col = "black", pch = 19,
  ...)
```

Arguments

pca	PCA object returned by prcomp
components	a vector of length two indicating the components to plot
group	a factor determining shapes of the points to show (unless overriden by pch=...)
col	Color for plotting (default: grey)
pch	Type of character to plot (default: 19)
...	any further parameters will be passed to the plot() function (e.g. col, cex, ...)

Value

If group is NULL, then NULL; else a data frame containing colors and shapes matching each group

pvalEffectPlot *Create an effect size / p-value plot*

Description

Create a heatmap-like plot showing information about both effect size and p-values.

Usage

```
pvalEffectPlot(e, p, pval.thr = 0.01, pval.cutoff = 1e-06,
  row.labels = NULL, col.labels = NULL, plot.func = NULL, grid = "at",
  grid.color = "#33333333", plot.cex = 1, text.cex = 1,
  col.labels.style = "top", legend.style = "auto")
```

Arguments

e	matrix with effect sizes
p	matrix with probabilities
pval.thr	The p-value must be this or lower in order for a test result to be visualized
pval.cutoff	On visual scale, all p-values below pval.cutoff will be replaced by pval.cutoff
row.labels	Labels for the modules. This must be a named vector, with module IDs as vector names. If NULL, module titles from the analyses results will be used.
col.labels	Labels for the columns. If NULL, names of the elements of the list x will be used.
plot.func	Optionally, a function to be used to draw the dots. See "Details"
grid	Style of a light-grey grid to be plotted; can be "none", "at" and "between"
grid.color	Color of the grid to be plotted (default: light grey)
plot.cex	a numerical value giving the amount by which the plot symbols will be magnified
text.cex	a numerical value giving the amount by which the plot text will be magnified, or a vector containing three cex values for row labels, column labels and legend, respectively
col.labels.style	Style of column names: "top" (default), "bottom", "both", "none"
legend.style	Style of the legend: "auto" – automatic; "broad": pval legend side by side with effect size legend; "tall": effect size legend above pval legend; "none" – no legend.

Details

pvalEffectPlot shows a heatmap-like plot. Each row corresponds to one series of tests (e.g. one module), and each column corresponds to the time points or conditions for which a given analysis was run. Each significant result is shown as a red dot. Size of the dot corresponds to the effect size (or any arbitrary value), and intensity of the color corresponds to the log10 of p-value.

Just like a heatmap corresponds to a single numeric matrix, the pvalue / effect plot corresponds to two matrices: one with the effect size, and another one with the p-values. Each cell in the matrix corresponds to the results of a single statistical test.

For example, a number of genes or transcriptional modules might be tested for differential expression or enrichment, respectively, in several conditions.

By default, each test outcome is represented by a dot of varying size and color. Alternatively, a function may be specified with the parameter 'plot.func'. It will be called for each test result to be drawn. The plot.func function must take the following arguments:

- row, coleither row / column number or the id of the row / column to plot; NULL if drawing legend
- x, yuser coordinates of the result to visualize
- w, hwidth and height of the item to plot

- eEnrichment – a relative value between 0 and 1, where 0 is the minimum and 1 is the maximum enrichment found
- pP-value – an absolute value between 0 and 1

For the purposes of drawing a legend, the function must accept NULL p-value or a NULL enrichment parameter.

Value

Invisibly returns a NULL value.

---

show *Shows the tmod object*

Description

Shows the tmod object

Extracts parts of a tmod object

Usage

```
## S4 method for signature 'tmod'
show(object)

## S4 method for signature 'tmod'
x[i]
```

Arguments

object	a tmod object
x	a tmod object
i	indices specifying elements to extract or replace

---

showGene *A combined beeswarm / boxplot*

Description

A combined beeswarm / boxplot

Usage

```
showGene(data, group, main = "", pch = 19, xlab = "",
  ylab = "log2 expression", las = 2, pwcol = NULL, ...)
```

Arguments

data	a vector of numeric values to be plotted
group	factor describing the groups
main	title of the plot
pch	character to plot the points
xlab, ylab	x and y axis labels
las	see par()
pwcol	colors of the points (see beeswarm)
...	any additional parameters to be passed to the beeswarm command

Details

This is just a simple wrapper around the beeswarm() and boxplot() commands.

Examples

```
data(Egambia)
E <- as.matrix(Egambia[,-c(1:3)])
showGene(E["20799",], rep(c("CTRL", "TB"), each=15))
```

---

showModule *Select genes belonging to a module from a data frame*

Description

Select genes belonging to a module from a data frame or vector

Usage

```
showModule(x, genes, module, mset = "all", extra = TRUE)
```

Arguments

x	a data frame or a vector
genes	a character vector with gene IDs
module	a single character value, ID of the module to be shown
mset	Module set to use; see "tmodUtest" for details
extra	If TRUE, additional information about the features will be shown

Details

showModule filters a data frame or a vector such that only genes from a module are shown. Use it, for example, to show a subset of topTable result from limma in order to see which genes from a module are significantly regulated. In essence, this is just a wrapper around "subset()".

Value

Either a filtered vector or (if extra==TRUE) a data frame.

`simplePie` *Simple Pie Chart*

Description

The simplePie function draws a simple pie chart at specified coordinates with specified width, height and color. The simpleRug function draws a corresponding rug plot, while simpleBoxpie creates a "rectangular pie chart" that is considered to be better legible than the regular pie.

Usage

```
simplePie(x, y, w, h, v, col, res = 100, border = NA)

simpleRug(x, y, w, h, v, col, border = NULL)

simpleBoxpie(x, y, w, h, v, col, border = NA, grid = 3)
```

Arguments

`x, y`	coordinates at which to draw the plot
`w, h`	width and height of the plot
`v`	sizes of the slices
`col`	colors of the slices
`res`	resolution (number of polygon edges in a full circle)
`border`	color of the border. Use NA (default) or NULL for no border
`grid`	boxpie only: the grid over which the areas are distributed. Should be roughly equal to the number of areas shown.

Details

simplePie() draws a pie chart with width w and height h at coordinates (x,y). The size of the slices is taken from the numeric vector v, and their color from the character vector col.

Examples

```
# demonstration of the three widgets
plot.new()
par(usr=c(0,3,0,3))
x <- c(7, 5, 11)
col <- tmodPal()
b <- "black"
simpleRug(0.5, 1.5, 0.8, 0.8, v=x, col=col, border=b)
simplePie(1.5, 1.5, 0.8, 0.8, v=x, col=col, border=b)
```

```
simpleBoxpie(2.5, 1.5, 0.8, 0.8, v=x, col=col, border=b)

# using pie as plotting symbol
plot(NULL, xlim=1:2, ylim=1:2, xlab="", ylab="")
col <- c("#cc000099", "#0000cc99")
for(i in 1:125) {
  x <- runif(1) + 1
  y <- runif(1) + 1
  simplePie( x, y, 0.05, 0.05, c(x,y), col)
}

# square filled with box pies
n <- 10
w <- h <- 1/(n+1)
plot.new()
for(i in 1:n) for(j in 1:n)
 simpleBoxpie(1/n*(i-1/2), 1/n*(j-1/2), w, h,
 v=runif(3), col=tmodPal())
```

`tmod-data` *Default gene expression module data*

Description

Gene expression module data from Chaussabel et al. (2008) and Li et al. (2014)

Details

The tmod package includes one data set of class tmod which can be loaded with data(tmod). This data set is derived from two studies (see package vignette for details). By default, enrichment analysis with tmod uses this data set; however, it is not loaded into user workspace by default.

References

Chaussabel, Damien, Charles Quinn, Jing Shen, Pinakeen Patel, Casey Glaser, Nicole Baldwin, Dorothee Stichweh, et al. 2008. "A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus." Immunity 29(1):150-64.

Li, Shuzhao, Nadine Rouphael, Sai Duraisingham, Sandra Romero-Steiner, Scott Presnell, Carl Davis, Daniel S Schmidt, et al. 2014. "Molecular Signatures of Antibody Responses Derived from a Systems Biology Study of Five Human Vaccines." Nature Immunology 15(2):195-204.

See Also

tmod-class, modmetabo

Examples

```
# list of first 10 modules
data(tmod)
tmod
tmod$MODULES[1:10, ]
tmod[1:10]
```

---

tmodAUC *Calculate AUC*

Description

Calculate AUC

Usage

```
tmodAUC(l, ranks, modules = NULL, stat = "AUC", remove.dups = TRUE,
  recalculate.ranks = TRUE, filter = FALSE, mset = "LI")
```

Arguments

l	List of gene names corresponding to rows from the ranks matrix
ranks	a matrix with ranks, where columns correspond to samples and rows to genes from the l list
modules	optional list of modules for which to make the test
stat	Which statistics to generate. Default: AUC
remove.dups	Remove duplicate gene names in l and corresponding rows from ranks
recalculate.ranks	Filtering and removing duplicates will also remove ranks, so that they should be recalculated. Use FALSE if you don't want this behavior. If unsure, stay with TRUE
filter	Remove gene names which have no module assignments
mset	Which module set to use. "LI", "DC" or "all" (default: LI)

Details

tmodAUC calculates the AUC and U statistics. The main purpose of this function is the use in randomization tests. While tmodCERNOtest and tmodUtest both calculate, for each module, the enrichment in a single sorted list of genes, tmodAUC takes any number of such sorted lists. Or, actually, sortings – vectors with ranks of the genes in each replicate. Note that the input for this function is different from tmodUtest and related functions: the ordering of l and the matrix ranks does not matter, as long as the matrix ranks contains the actual rankings. Each column in the ranks matrix is treated as a separate sample.

Value

A matrix with the same number of columns as "ranks" and as many rows as there were modules to be tested.

See Also

tmod-package

Examples

```
data(tmod)
l <- tmod$GENES$ID
ranks <- 1:length(l)
res <- tmodAUC(l, ranks)
head(res)
```

---

tmodDecideTests *Count the Up- or Down-regulated genes per module*

Description

For each module in a set, calculate the number of genes which are in that module and which are significantly up- or down-regulated.

Usage

```
tmodDecideTests(g, lfc = NULL, pval = NULL, lfc.thr = 0.5,
  pval.thr = 0.05, labels = NULL, filter.unknown = FALSE, mset = "all")
```

Arguments

g	a character vector with gene symbols
lfc	a numeric vector or a matrix with log fold changes
pval	a numeric vector or a matrix with p-values. Must have the same dimensions as lfc
lfc.thr	log fold change threshold
pval.thr	p-value threshold
labels	Names of the comparisons. Either NULL or a character vector of length equal to the number of columns in lfc and pval.
filter.unknown	If TRUE, modules with no annotation will be omitted
mset	Which module set to use. Either a character vector ("LI", "DC" or "all", default: LI) or a list (see "Custom module definitions" below)

Details

This function can be used to decide whether a module, as a whole, is up- or down regulated. For each module, it calculates the number of genes which are up-, down- or not regulated. A gene is considered to be up- regulated if the associated p-value is smaller than pval.thr and the associated log fold change is greater than lfc.thr. A gene is considered to be down- regulated if the associated p-value is smaller than pval.thr and the associated log fold change is smaller than lfc.thr.

Note that unlike decideTests from limma, tmodDecideTests does not correct the p-values for multiple testing – therefore, the p-values should already be corrected.

Value

A list with as many elements as there were comparisons (columns in lfc and pval). Each element of the list is a data frame with the columns "Down", "Zero" and "Up" giving the number of the down-, not- and up-regulated genes respectively. Rows of the data frame correspond to module IDs.

See Also

tmodSummary, tmodPanelPlot, tmodDecideTestsLimma

---

tmodImportMSigDB *Import data from MSigDB*

Description

Import data from an MSigDB file in either XML or GMT format

Usage

```
tmodImportMSigDB(file = NULL, format = "xml", organism = "Homo sapiens",
  fields = c("STANDARD_NAME", "CATEGORY_CODE", "SUBCATEGORY_CODE"))
```

Arguments

file	The name of the file to parse
format	Format (either "xml" or "gmt")
organism	Select the organism to use. Use "all" for all organisms in the file (only for "xml" format; default: "Homo sapiens")
fields	Which fields to import to the MODULES data frame (only for "xml" format)

Details

This command parses a file from MSigDB. Both XML and the MSigDB-specific "GMT" format are supported (however, the latter is discouraged, as it contains less information).

Examples

```
## Not run:
## First, download the file "msigdb_v5.0.xml" from http://www.broadinstitute.org/gsea/downloads.jsp
msig <- tmodImportMSigDB( "msigdb_v5.0.xml" )

## End(Not run)
```

---

tmodLimmaDecideTests *Up- and down-regulated genes in modules based on limma object*

Description

For each module in mset and for each coefficient in f$coefficients, this function calculates the numbers of significantly up- and down-regulated genes.

Usage

```
tmodLimmaDecideTests(f, genes, lfc.thr = 0.5, pval.thr = 0.05,
  filter.unknown = FALSE, adjust.method = "BH", mset = "all")
```

Arguments

f	result of linear model fit produced by limma functions lmFit and eBayes
genes	Either the name of the column in f$genes which contains the gene symbols corresponding to the gene set collection used, or a character vector with gene symbols
lfc.thr	log fold change threshold
pval.thr	p-value threshold
filter.unknown	If TRUE, modules with no annotation will be omitted
adjust.method	method used to adjust the p-values for multiple testing. See p.adjust(). Default:BH.
mset	Which module set to use (see tmodUtest for details)

Details

For an f object returned by eBayes(), tmodLimmaDecideTests considers every coefficient in this model (every column of f$coefficients). For each such coefficient, tmodLimmaDecideTests calculates, for each module, the number of genes which are up- or down-regulated.

In short, tmodLimmaDecideTests is the equivalent of tmodDecideTests, but for limma objects returned by eBayes().

Value

A list with as many elements as there were coefficients in f. Each element of the list is a data frame with the columns "Down", "Zero" and "Up" giving the number of the down-, not- and up-regulated genes respectively. Rows of the data frame correspond to module IDs. The object can directly be used in tmodPanelPlot as the pie parameter.

See Also

tmodDecideTests, tmodLimmaTest, tmodPanelPlot

Examples

```
data(Egambia)
design <- cbind(Intercept=rep(1, 30), TB=rep(c(0,1), each= 15))
if(require(limma)) {
  fit <- eBayes( lmFit(Egambia[,-c(1:3)], design))
  ret <- tmodLimmaTest(fit, Egambia$GENE_SYMBOL)
  pie <- tmodLimmaDecideTests(fit, Egambia$GENE_SYMBOL)
  tmodPanelPlot(ret, pie=pie)
}
```

---

tmodLimmaTest *Run tmod enrichment tests directly on a limma object*

Description

Order the genes according to each of the coefficient found in a limma object and run an enrichment test on the ordered list.

Usage

```
tmodLimmaTest(f, genes, sort.by = "msd", tmodFunc = tmodCERNOtest,
  coef = NULL, ...)
```

Arguments

f	result of linear model fit produced by limma functions lmFit and eBayes
genes	Either the name of the column in f$genes which contains the gene symbols corresponding to the gene set collection used, or a character vector with gene symbols
sort.by	How the gene names should be ordered: "msd" (default), "pval" or "lfc"
tmodFunc	The function to run the enrichment tests. Either tmodCERNOtest or tmodUtest
coef	If not NULL, only run tmod on these coefficients
...	Further parameters passed to the tmod test function

Details

For each coefficient in the fit returned by the eBayes / lmFit functions from the limma package, tmodLimmaTest will order the genes run an enrichment test and return the results.

The ordering of the genes according to a certain metric is the fundament for gene enrichment analysis. tmodLimmaTest allows three orderings: p-values, "MSD" and log fold changes. The default MSD ("minimal significant difference") is the lower boundary of the 95 confidence interval for positive log fold changes, and 0 minus the upper boundary of the 95 better than ordering by p-value or by log fold change. See discussion in the package vignette.

Value

A list with length equal to the number of coeffients. Each element is the value returned by tmod test function. The list can be directly passed to the functions tmodSummary and tmodPanelPlot.

See Also

tmodCERNOtest, tmodUtest, tmodPlotPanel, tmodSummary

Examples

```
data(Egambia)
design <- cbind(Intercept=rep(1, 30), TB=rep(c(0,1), each= 15))
if(require(limma)) {
  fit <- eBayes( lmFit(Egambia[,-c(1:3)], design))
  ret <- tmodLimmaTest(fit, genes=Egambia$GENE_SYMBOL)
  tmodSummary(ret)
  tmodPanelPlot(ret)
}
```

---

tmodLimmaTopTable *tmod's replacement for the limma topTable function*

Description

Produce a data frame for all or for selected coefficients of a linear fit object, including log fold changes, q-values, confidence intervals and MSD.

Usage

```
tmodLimmaTopTable(f, genelist = NULL, coef = NULL, adjust.method = "BH",
  confint = 0.95)
```

Arguments

f	result of linear model fit produced by limma functions lmFit and eBayes
genelist	A data frame or a character vector with additional information on the genes / probes
coef	Which coefficients to extract
adjust.method	Which method of p-value adjustment; see "p.adjust()"
confint	Confidence interval to be calculated

Details

Produce a data frame for all or for selected coefficients of a linear fit object, including log fold changes, q-values, confidence intervals and MSD. For each coefficient, these four columns will be created in the output file, with the name consisting of a prefix indicating the type of the column ("msd", "logFC", "qval", "SE", "ci.L", "ci.R") and the name of the coefficient.

Value

A data frame with all genes.

See Also

tmodLimmaTest

`tmodPal` *A selection of color palettes*

Description

Return a preset selection of colors, adjusted by alpha

Usage

```
tmodPal(n = NULL, set = "friendly", alpha = 0.7, func = FALSE)
```

Arguments

`n`	Number of colors to return (default: all for "friendly", 3 for everything else)
`set`	Which palette set (see Details).
`alpha`	0 for maximum transparency, 1 for no transparency.
`func`	if TRUE, the returned object will be a function rather than a character vector

Details

A few palettes have been predefined in tmod, and this function can be used to extract them. The following palettes have been defined: * friendly – a set of distinct, colorblind-friendly colors * bwr, rwb, ckp, pkc – gradients (b-blue, r-red, w-white, c-cyan, k-blacK, p-purple) By default, either all colors are returned, or, if it is a gradient palette, only three.

Yes, I wrote this function to save myself the slightest amount of typing.

Value

Either a character vector, or, when the func parameter is TRUE, a function that takes only one argument (a single number)

`tmodPanelPlot` *Plot a summary of multiple tmod analyses*

Description

Plot a summary of multiple tmod analyses

Usage

```
tmodPanelPlot(x, pie = NULL, clust = "qval", filter.empty.cols = FALSE,
  filter.empty.rows = TRUE, filter.unknown = TRUE,
  filter.rows.pval = 0.05, filter.rows.auc = 0.5, filter.by.id = NULL,
  col.labels = NULL, col.labels.style = "top", row.labels = NULL,
  row.labels.auto = "both", pval.thr = 10^-2, pval.thr.lower = 10^-6,
  plot.func = NULL, grid = "at", pie.colors = c("#0000FF", "#cccccc",
  "#FF0000"), plot.cex = 1, text.cex = 1, pie.style = "pie",
  legend.style = "auto", ...)
```

Arguments

`x`	list, in which each element has been generated with a tmod test function
`pie`	a list of data frames with information for drawing a pie chart
`clust`	whether, in the resulting data frame, the modules should be ordered by clustering them with either q-values ("qval") or the effect size ("effect"). If NULL, the modules are sorted alphabetically by their ID.
`filter.empty.cols`	If TRUE, all elements (columns) with no enrichment below pval.thr in any row will be removed
`filter.empty.rows`	If TRUE, all modules (rows) with no enrichment below pval.thr in any column will be removed
`filter.unknown`	If TRUE, modules with no annotation will be omitted
`filter.rows.pval`	Rows in which no p value is below this threshold will be omitted
`filter.rows.auc`	Rows in which no AUC value is above this threshold will be omitted
`filter.by.id`	if provided, show only modules with IDs in this character vector
`col.labels`	Labels for the columns. If NULL, names of the elements of the list x will be used.
`col.labels.style`	Style of column names: "top" (default), "bottom", "both", "none"
`row.labels`	Labels for the modules. This must be a named vector, with module IDs as vector names. If NULL, module titles from the analyses results will be used.

row.labels.auto	Automatic generation of row labels from module data: "both" (default, ID and title), "id" (only ID), "title" (only title), "none" (no row label)
pval.thr	Results with p-value above pval.thr will not be shown
pval.thr.lower	Results with p-value below pval.thr.lower will look identical on the plot
plot.func	Optionally, a function to be used to draw the dots. See "pvalEffectPlot"
grid	Style of a light-grey grid to be plotted; can be "none", "at" and "between"
pie.colors	character vector of length equal to the cardinality of the third dimension of the pie argument. By default: blue, grey and red.
plot.cex	a numerical value giving the amount by which the plot symbols will be magnified
text.cex	a numerical value giving the amount by which the plot text will be magnified, or a vector containing three cex values for row labels, column labels and legend, respectively
pie.style	Can be "pie", "boxpie" or "rug"
legend.style	Style of the legend: "auto" – automatic; "broad": pval legend side by side with effect size legend; "tall": effect size legend above pval legend
...	Any further arguments will be passed to the pvalEffectPlot function (for example, grid.color)

Details

This function is useful if you run an analysis for several conditions or time points and would like to summarize the information on a plot. You can use lapply() to generate a list with tmod results and use tmodPanelPlot to visualize it.

tmodPanelPlot shows a heatmap-like plot. Each row corresponds to one module, and columns correspond to the time points or conditions for which the tmod analyses were run. Each significantly enriched module is shown as a red dot. Size of the dot corresponds to the effect size (for example, AUC in the CERNO test), and intensity of the color corresponds to the q-value.

By default, tmodPanelPlot visualizes each the results of a single statistical test by a red dot. However, it is often interesting to know how many of the genes in a module are significantly up- or down regulated. tmodPanelPlot can draw a pie chart based on the optional argument "pie". The argument must be a list of length equal to the length of x. Note also that the names of the pie list must be equal to the names of x. Objects returned by the function tmodDecideTests can be directly used here. The rownames of either the data frame or the array must be the module IDs.

Value

a data frame with a line for each module encountered anywhere in the list x, two columns describing the module (ID and module title), and two columns(effect size and q value) for each element of list x.

See Also

tmodDecideTests, tmodSummary, pvalEffectPlot, simplePie

Examples

```
data(Egambia)
E <- Egambia[,-c(1:3)]
pca <- prcomp(t(E), scale.=TRUE)

# Calculate enrichment for first 5 PCs
gs   <- Egambia$GENE_SYMBOL
gn.f <- function(r) {
    o <- order(abs(r), decreasing=TRUE)
    tmodCERNOtest(gs[o],
                qval=0.01)
}
x <- apply(pca$rotation[,3:4], 2, gn.f)
tmodPanelPlot(x, text.cex=0.7)
```

---

tmodPCA *PCA plot annotated with tmod*

Description

Generate a PCA plot on which each dimension is annotated by a tag cloud based on tmod enrichment test.

Usage

```
tmodPCA(pca, loadings = NULL, genes, tmodfunc = "tmodCERNOtest",
  plotfunc = pcaplot, mode = "simple", components = c(1, 2),
  plot.params = NULL, filter = TRUE, simplify = TRUE, legend = FALSE,
  ...)
```

Arguments

pca	Object returned by prcomp or a matrix of PCA coordinates. In the latter case, a loading matrix must be provided separately.
loadings	A matrix with loadings
genes	A character vector with gene identifiers
tmodfunc	Name of the tmod enrichment test function to use. Either
plotfunc	Function for plotting the PCA plot. See Details
mode	Type of the plot to generate; see Details. tmodCERNOtest or tmodUtest (tmodHGtest is not suitable)
components	integer vector of length two: components which components to show on the plot. Must be smaller than the number of columns in pca.
plot.params	A list of parameters to be passed to the plotting function. See Details
filter	Whether "uninteresting" modules (with no annotation) should be removed from the tag cloud

simplify	Whether the names of the modules should be simplified
legend	whether a legend should be shown
...	Any further parameters passed to the tmod test function

Details

There are three types of plots that can be generated (parameter "mode"): simple, leftbottom and cross. In the "simple" mode, two enrichments are run, on on each component, sorted by absolute loadings of the PCA components. Both "leftbottom" and "cross" run two enrichment analyses on each component, one on the loadings sorted from lowest to largest, and one on the loadings sorted from largetst to lowest. Thus, two tag clouds are displayed per component. In the "leftbottom" mode, the tag clouds are displayed to the left and below the PCA plot. In the "cross" mode, the tag clouds are displayed on each of the four sides of the plot.

By default, the plotting function is pca2d from the pca3d package. Any additional parametrs for pca2d can be passed on using the plot.params parameter. You can define your own function instead of pca2d, however, mind that in any case, there will be two parameters passed to it on the first two positions: pca and components, named "pca" and "components" respectively.

Value

A list containing the calculated enrichments as well as the return value from the plotting function

Examples

```
data(Egambia)
E <- as.matrix(Egambia[,-c(1:3)])
pca <- prcomp(t(E), scale.=TRUE)
group <- rep(c("CTRL", "TB"), each=15)
tmodPCA(pca,
  genes=Egambia$GENE_SYMBOL,
  components=4:3,
  plot.params=list(group=group))
```

---

tmodSummary *Create a summary of multiple tmod analyses*

Description

Create a summary of multiple tmod analyses

Usage

```
tmodSummary(x, clust = NULL, filter.empty = FALSE, filter.unknown = TRUE)
```

Arguments

x	list, in which each element has been generated with a tmod test function
clust	whether, in the resulting data frame, the modules should be ordered by clustering them with either q-values ("qval") or the effect size ("effect"). If NULL, the modules are sorted alphabetically by their ID.
filter.empty	If TRUE, all elements (columns) with no significant enrichment will be removed
filter.unknown	If TRUE, modules with no annotation will be omitted

Details

This function is useful if you run an analysis for several conditions or time points and would like to summarize the information in a single data frame. You can use lapply() to generate a list with tmod results and use tmodSummary to convert it to a data frame.

Value

a data frame with a line for each module encountered anywhere in the list x, two columns describing the module (ID and module title), and two columns(effect size and q value) for each element of list x.

See Also

tmodPanelPlot

Examples

```
data(Egambia)
E <- Egambia[,-c(1:3)]
pca <- prcomp(t(E), scale.=TRUE)

# Calculate enrichment for each component
gs   <- Egambia$GENE_SYMBOL
gn.f <- function(r) {
    tmodCERNOtest(gs[order(abs(r),
                decreasing=TRUE)],
                qval=0.01)
}
x <- apply(pca$rotation, 2, gn.f)
tmodSummary(x)
```

---

tmodTagcloud *Tag cloud based on tmod results*

Description

Plot a tag (word) cloud based on results from tmod enrichment.

Usage

```
tmodTagcloud(results, filter = TRUE, simplify = TRUE, tag.col = "Title",
  weights.col = "auto", pval.col = "P.Value", ...)
```

Arguments

results	data frame produced by one of the tmod enrichment tests
filter	Whether redundant and not annotated modules should be removed
simplify	Whether module names should be simplified
tag.col	Which column from results should be used as tags on the plot
weights.col	Which column from results should be used as weights for the tag cloud
pval.col	Which column contains the P values which will be used to shade the tags
...	Any further parameters are passed to the tagcloud function

Details

The tags will be generated based on results from tmod or any other suitable data frame. The data frame must contain two numeric columns, specified with "weights.col" and "pval.col", which will be used to calculate the size and shade of the tags, respectively. Furthermore, it has to contain a column with tags (parameter "tag.col", by default "Title").

Any data frame can be used as long as it contains the specified columns.

Value

Either NULL or whatever tagcloud returns

Examples

```
data(tmod)
fg <- tmod$MODULES2GENES[["LI.M127"]]
bg <- tmod$GENES$ID
result <- tmodHGtest( fg, bg )
tmodTagcloud(result)
```

---

tmodUtest *Perform a statistical test of module expression*

Description

Perform a statistical test of module expression

Usage

```
tmodUtest(l, modules = NULL, qval = 0.05, order.by = "pval",
  filter = FALSE, mset = "LI", cols = "Title", useR = FALSE)

tmodCERNOtest(l, modules = NULL, qval = 0.05, order.by = "pval",
  filter = FALSE, mset = "LI", cols = "Title")

tmodZtest(l, modules = NULL, qval = 0.05, order.by = "pval",
  filter = FALSE, mset = "LI", cols = "Title")

tmodWZtest(l, modules = NULL, weights = NULL, qval = 0.05,
  order.by = "pval", filter = FALSE, mset = "LI", cols = "Title")

tmodHGtest(fg, bg, modules = NULL, qval = 0.05, order.by = "pval",
  filter = FALSE, mset = "LI", cols = "Title")
```

Arguments

`l`	sorted list of HGNC gene identifiers
`modules`	optional list of modules for which to make the test
`qval`	Threshold FDR value to report
`order.by`	Order by P value ("pval") or none ("none")
`filter`	Remove gene names which have no module assignments
`mset`	Which module set to use. Either a character vector ("LI", "DC" or "all", default: LI) or a list (see "Custom module definitions" below)
`cols`	Which columns from the MODULES data frame should be included in resulsts
`useR`	use the R `wilcox.test` function; slow, but with exact p-values for small samples
`weights`	for tmodWZtest
`fg`	foreground gene set for the HG test
`bg`	background gene set for the HG test

Details

Performs a test on either on an ordered list of genes (tmodUtest, tmodCERNOtest, tmodZtest) or on two groups of genes (tmodHGtest). tmodUtest is a U test on ranks of genes that are contained in a module.

tmodCERNOtest is also a nonparametric test working on gene ranks, but it originates from Fisher's combined probability test. This test weights genes with lower ranks more, the resulting p-values better correspond to the observed effect size. In effect, modules with small effect but many genes get higher p-values than in case of the U-test.

tmodZtest works very much like tmodCERNOtest, but instead of combining the rank-derived p-values using Fisher's method, it uses the Stouffer method (known also as the Z-transform test).

tmodWZtest is the weighted version of the tmodZtest. The weights can be provided as a named numeric vector (the "weights" parameter). In this case, the weights of any genes in the parameter

l which are not in the names of "weights" are set to 0. These weights will be constant for a given gene in all modules. Alternatively, weights associated with the "mset" parameter (an object of class tmod) in the "WEIGHTS" member of the object (if present).

For a discussion of the above three methods, read M. C. Whitlock, "Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach", J. Evol. Biol. 2005 (doi: 10.1111/j.1420-9101.2005.00917.x) for further details.

tmodHGtest is simply a hypergeometric test.

In tmod, two module sets can be used, "LI" (from Li et al. 2013), or "DC" (from Chaussabel et al. 2008). Using the parameter "mset", the module set can be selected, or, if mset is "all", both of sets are used.

Value

A data frame with module names, additional statistic (e.g. enrichment or AUC, depending on the test), P value and FDR q-value (P value corrected for multiple testing using the p.adjust function and Benjamini-Hochberg correction.

Custom module definitions

Custom and arbitrary module, gene set or pathway definitions can be also provided through the mset option, if the parameter is a list rather than a character vector. The list parameter to mset must contain the following members: "MODULES", "MODULES2GENES" and "GENES".

"MODULES" and "GENES" are data frames. It is required that MODULES contains the following columns: "ID", specifying a unique identifier of a module, and "Title", containing the description of the module. The data frame "GENES" must contain the column "ID".

The list MODULES2GENES is a mapping between modules and genes. The names of the list must correspond to the ID column of the MODULES data frame. The members of the list are character vectors, and the values of these vectors must correspond to the ID column of the GENES data frame.

See Also

tmod-package

Examples

```
data(tmod)
fg <- tmod$MODULES2GENES[["LI.M127"]]
bg <- tmod$GENES$ID
result <- tmodHGtest( fg, bg )

## A more sophisticated example
## Gene set enrichment in TB patients compared to
## healthy controls (Egambia data set)

data(Egambia)
  design <- cbind(Intercept=rep(1, 30), TB=rep(c(0,1), each= 15))
  if(require(limma)) {
  fit <- eBayes( lmFit(Egambia[,-c(1:3)], design))
  tt <- topTable(fit, coef=2, number=Inf, genelist=Egambia[,1:3] )
```

```
  tmodUtest(tt$GENE_SYMBOL)
}
```

Index