Package 'DEGreport'

October 17, 2017

Version 1.12.0

Date 2017-04-10

Type Package

Title Report of DEG analysis

Description Creation of a HTML report of differential expression analyses of count data. It integrates some of the code mentioned in DESeq2 and edgeR vignettes, and report a ranked list of genes according to the fold changes mean and variability for each selected gene.

biocViews DifferentialExpression, Visualization, RNASeq, ReportWriting, GeneExpression

Suggests biomaRt, RUnit, BiocStyle, BiocGenerics, org.Hs.eg.db, DESeq2, AnnotationDbi, BiocParallel

Depends R (>= 3.2.0), quantreg

Imports utils, methods, ggplot2, ggrepel, Nozzle.R1, coda, edgeR, cluster, logging, dplyr, tidyr, reshape, pheatmap, grid, gridExtra, knitr, grDevices, stats

Maintainer Lorena Pantano <lorena.pantano@gmail.com>

License MIT + file LICENSE

VignetteBuilder knitr

RoxygenNote 5.0.1

NeedsCompilation no

Author Lorena Pantano [aut, cre], John Hutchinson [ctb], Victor Barrera [ctb], Mary Piper [ctb]

R topics documented:

`createReport` *Create report of RNAseq DEG anlaysis*

Description

This function get the count matrix, pvalues, and FC of a DEG analysis and create a report to help to detect possible problems with the data.

Usage

```
createReport(g1, g2, counts, tags, pvalues, fc, path, colors = "",
  pop = 400, name = "DEGreport", ncores = NULL)
```

Arguments

`g1`	group 1
`g2`	group 2
`counts`	matrix with counts for each samples and each gene. Should be same length than pvalues vector.
`tags`	genes of DEG analysis
`pvalues`	pvalues of DEG analysis
`fc`	FC for each gene
`path`	path to save the figure
`colors`	data frame with colors for each gene
`pop`	random genes for background
`name`	name of the html file
`ncores`	num cores to be used to create report

Value

create a html file with all figures and tables

`degBI` *Get the estimates of the fold change (FC) mean from a FC distribution using bayesian inference*

Description

Get the estimates of the fold change (FC) mean from a FC distribution using bayesian inference

Usage

```
degBI(fc, iter = 1000, ncores = NULL)
```

Arguments

`fc`	list of FC
`iter`	number of iteration in the mcmc model
`ncores`	number of cores to use

Value

matrix with values from degBIcmd

`degBIcmd` *Apply bayesian inference to estimate the average fold change (FC) of a distribution*

Description

code based on http://www.johnmyleswhite.com/notebook/2010/08/20/ using-jags-in-r-with-the-rjags-package/ http://public.wsu.edu/~jesse.brunner/classes/bio572/Lab7_Bayesian.html

Usage

```
degBIcmd(x, iter = 1000)
```

Arguments

`x`	list of values
`iter`	number of iteration in the mcmc model

Value

vector with mu and its confidence intervales (2.5

degCheckFactors *Distribution of gene ratios used to calculate Size Factors.*

Description

Distribution of gene ratios used to calculate Size Factors.

Usage

```
degCheckFactors(counts)
```

Arguments

counts	matrix with counts for each samples and each gene. row number should be the same length than pvalues vector.

Details

This function will plot the gene ratios for each sample. To calculate the ratios, it follows the simliar logic than DESeq2/edgeR uses, where the expression of each gene is divided by the mean expression of that gene. The distribution of the ratios should approximate to a normal shape and the factors should be similar to the median of distributions. If some samples show different distribution, the factor may be bias due to some biological or technical factor.

Value

ggplot2 object

Examples

```
data(DEGreportSet)
degCheckFactors(DEGreportSet$counts[,1:10])
```

degComb *Get random combinations of two groups*

Description

Get random combinations of two groups

Usage

```
degComb(g1, g2, pop)
```

Arguments

g1	list of samples in group 1
g2	list of samples in group 2
pop	number of combinations to be return

Value

matrix with different combinatios of two vector

`degFC` *get the FC for each gene between two groups*

Description

get the FC for each gene between two groups

Usage

```
degFC(g1, g2, counts, popsize)
```

Arguments

`g1`	list of samples in group 1
`g2`	list of samples in group 2
`counts`	count matrix of deregulated genes
`popsize`	number of combinations to generate

Value

FC for different combinations of samples in each group for each gene

---

`degMB` *Distribution of expression of DE genes compared to the background*

Description

Distribution of expression of DE genes compared to the background

Usage

```
degMB(tags, g1, g2, counts, pop = 400)
```

Arguments

`tags`	list of genes that are DE
`g1`	list of samples in group 1
`g2`	list of samples in group 2
`counts`	matrix with counts for each samples and each gene Should be same length than pvalues vector
`pop`	number of random samples taken for background comparison

Value

ggplot2 object

Examples

```
data(DEGreportSet)
detag <- row.names(DEGreportSet$deg[1:10,])
degMB(detag,DEGreportSet$g1,DEGreportSet$g2,DEGreportSet$counts)
```

---

degMean *Distribution of pvalues by expression range*

Description

Distribution of pvalues by expression range

Usage

```
degMean(pvalues, counts)
```

Arguments

pvalues	pvalues of DEG analysis
counts	matrix with counts for each samples and each gene. row number should be the same length than pvalues vector.

Value

ggplot2 object

Examples

```
data(DEGreportSet)
degMean(DEGreportSet$deg[,4],DEGreportSet$counts)
```

---

degMerge *Integrate data comming from degPattern into one data object*

Description

The simplest case is if you want to convine the pattern profile for gene expression data and proteomic data. It will use the first element as the base for the integration. Then, it will loop through clusters and run degPatterns in the second data set to detect patterns that match this one.

Usage

```
degMerge(matrix_list, cluster_list, metadata_list, summarize = "group",
  time = "time", col = "condition", scale = TRUE, mapping = NULL)
```

Arguments

matrix_list	list expression data for each element
cluster_list	list df item from degPattern output
metadata_list	list data.frames from each element with design experiment. Normally colData output
summarize	character column to use to group samples
time	character column to use as x-axes in figures
col	character column to color samples in figures
scale	boolean scale by row expression matrix
mapping	data.frame mapping table in case elements use different ID in the row.names of expression matrix. For instance, when integrating miRNA/mRNA.

Value

A data.frame with information on what genes are in each cluster in all data set, and the correlation value for each pair cluster comparison.

---

degMV *Correlation of the standard desviation and the mean of the abundance of a set of genes.*

Description

Correlation of the standard desviation and the mean of the abundance of a set of genes.

Usage

```
degMV(group, pvalues, counts, sign = 0.01)
```

Arguments

group	character vector with group name for each sample in the same order than counts column names.
pvalues	pvalues of DEG analysis
counts	matrix with counts for each samples and each gene.
sign	defining the cutoff to label significant features. row number should be the same length than pvalues vector.

Value

ggplot2 object

Examples

```
data(DEGreportSet)
degMV(c(rep("M", length(DEGreportSet$g1)), rep("F", length(DEGreportSet$g2))),
      DEGreportSet$deg[,4],
      DEGreportSet$counts)
```

degNcomb — *Get number of potential combinations of two vectors*

Description

Get number of potential combinations of two vectors

Usage

```
degNcomb(g1, g2)
```

Arguments

g1	list of samples in group 1
g2	list of samples in group 2

Value

maximum number of combinations of two vectors

degObj — *Create a deg object that can be used to plot expression values at shiny server:runGist(9930881)*

Description

Create a deg object that can be used to plot expression values at shiny server:runGist(9930881)

Usage

```
degObj(counts, design, outfile)
```

Arguments

counts	output from get_rank function
design	colour used for each gene
outfile	file that will contain the object

Value

R object to be load into vizExp

Examples

```
data(DEGreportSet)
de = data.frame(row.names=colnames(DEGreportSet$counts),
sex = c(rep("M", length(DEGreportSet$g1)),
        rep("F", length(DEGreportSet$g2))))
degObj(DEGreportSet$counts, de, NULL)
```

degPatterns *Make groups of genes using expression profile*

Description

Make groups of genes using expression profile

Usage

```
degPatterns(ma, metadata, minc = 15, summarize = "group", time = "time",
  col = "condition", reduce = FALSE, cutoff = 0.7, scale = TRUE,
  plot = TRUE, fixy = NULL)
```

Arguments

ma	log2 normalized count matrix
metadata	data frame with sample information. Rownames should match ma column names row number should be the same length than p-values vector.
minc	integer minimum number of genes in a group that will be return
summarize	character column name in metadata that will be used to gorup replicates. For instance, a merge between summarize and time parameters: control_point0 ... etc
time	character column name in metadata that will be used as variable that changes, normally a time variable.
col	character column name in metadata to separate samples. Normally control/mutant
reduce	boolean reduce number of clusters using correlation values between them.
cutoff	integer threshold for correlation expression to merge clusters (0 - 1)
scale	boolean scale the ma values by row
plot	boolean plot the clusters found
fixy	vector integers used as ylim in plot

Details

It would be used diana function to detect a value to cut the expression based clustering at certain height. It can work with one or more groups with 2 or more several time points. The different patterns can be merged to get similar ones into only one pattern. The expression correlation of the patterns will be used to decide whether some need to be merged or not.

Value

list wiht two items. df is a data.frame with two columns. The first one with genes, the second with the clusters they belong. pass_to_plot is a vector of the clusters that pass the minc cutoff.

Examples

```
data(humanSexDEedgeR)
ma <- humanSexDEedgeR$counts[1:100,]
des <- data.frame(row.names=colnames(ma),
group=as.factor(humanSexDEedgeR$samples$group))
res <- degPatterns(ma, des, time="group", col=NULL)
```

`degPlot` *Plot top genes allowing more variables to color and shape points*

Description

Plot top genes allowing more variables to color and shape points

Usage

```
degPlot(dds, res, n = 9, xs = "time", group = "condition", batch = NULL)
```

Arguments

`dds`	DESeqDataSet object
`res`	DESeqResults object
`n`	integer number of genes to plot.
`xs`	character, colname in colData that will be used as X-axes
`group`	character, colname in colData to color points and add different lines for each level
`batch`	character, colname in colData to shape points, normally used by batch effect visualization

Value

ggplot showing the expresison of the genes

`degPlotWide` *Plot selected genes on a wide format*

Description

Plot selected genes on a wide format

Usage

```
degPlotWide(dds, genes, group = "condition", batch = NULL)
```

Arguments

`dds`	DESeqDataSet object
`genes`	character genes to plot.
`group`	character, colname in colData to color points and add different lines for each level
`batch`	character, colname in colData to shape points, normally used by batch effect visualization

Value

ggplot showing the expresison of the genes on the x axis

Examples

```
data(humanSexDEedgeR)
library(DESeq2)
idx <- c(1:10, 75:85)
dse <- DESeqDataSetFromMatrix(humanSexDEedgeR$counts[1:1000, idx],
humanSexDEedgeR$samples[idx,], design=~group)
dse <- DESeq(dse)
degPlotWide(dse, rownames(dse)[1:10], group="group")
```

---

degPR *plot the correlation between the rank according estimator and the rank according FC*

Description

plot the correlation between the rank according estimator and the rank according FC

Usage

```
degPR(rank, colors = "")
```

Arguments

rank	output from degRank function
colors	colour used for each gene

Value

ggplot2 object

Examples

```
data(DEGreportSet)
degPR(DEGreportSet$rank)
```

degQC — *Plot main figures showing p-values distribution and mean-variance correlation*

Description

This function joins the output of degMean, degVar and degMV in a single plot. See these functions for further information.

Usage

```
degQC(pvalue, counts, groups)
```

Arguments

pvalue	pvalues of DEG analysis
counts	matrix with counts for each samples and each gene.
groups	character vector with group name for each sample in the same order than counts column names.

Value

ggplot2 object

Examples

```
library(DESeq2)
data(humanSexDEedgeR)
idx <- c(1:10, 75:85)
dse <- DESeqDataSetFromMatrix(humanSexDEedgeR$counts[1:1000, idx],
humanSexDEedgeR$samples[idx,], design=~group)
dse <- DESeq(dse)
res <- results(dse)
degQC(res$pvalue, counts(dse, normalized=TRUE),colData(dse)$group)
```

degRank — *Get rank data frame with best score on the top*

Description

Get rank data frame with best score on the top

Usage

```
degRank(g1, g2, counts, fc, popsize = 400, iter = 1000, ncores = NULL)
```

Arguments

g1	list of samples in group 1
g2	list of samples in group 2
counts	count matrix for each gene and each sample that is deregulated
fc	list of FC of deregulated genes. Should be same length than counts row.names
popsize	number of combinations to generate
iter	number of iteration in the mcmc model
ncores	number of cores to use

Value

data frame with the output of degBIcmd for each gene

Examples

```
data(DEGreportSet)
#library(rjags)
#degRank(DEGreportSet$g1,DEGreportSet$g2,
#    DEGreportSet$counts[DEGreportSet$detag[1:5],],
#    DEGreportSet$deg[DEGreportSet$detag[1:5],1],400,500)
```

DEGreportSet *List of process geuvadis data to test the package*

Description

It contains gene counts matrix, group1 of samples, group2 of samples, differential expression analysis table, set of genes, output from degRank function.

Usage

```
DEGreportSet
```

Format

List

Author(s)

Lorena Pantano, 2014-05-31

Source

gEUvadis

degResults *Complete report from DESeq2 analysis*

Description

Complete report from DESeq2 analysis

Usage

```
degResults(res = NULL, dds, rlogMat = NULL, name, org = NULL,
  FDR = 0.05, do_go = TRUE, FC = 0.1, group = "condition",
  xs = "time", path_results = ".", contrast = NULL)
```

Arguments

res	output from results function.
dds	DESeqDataSet object.
rlogMat	matrix from rlog function.
name	string to identify results
org	an organism annotation object, like org.Mm.eg.db. NULL if you want to skip this step.
FDR	int cutoff for false discovery rate.
do_go	boolean if GO enrichment is done.
FC	int cutoff for log2 fold change.
group	string column name in colData(dds) that separates samples in meaninful groups.
xs	string column name in colData(dss) that will be used as X axes in plots (i.e time)
path_results	character path where files are stored. NULL if you don't want to save any file.
contrast	list with character vector indicating the fold change values from different comparisons to add to the output table.

Value

ggplot2 object

Examples

```
data(humanSexDEedgeR)
library(DESeq2)
idx <- c(1:10, 75:85)
dse <- DESeqDataSetFromMatrix(humanSexDEedgeR$counts[1:1000, idx],
humanSexDEedgeR$samples[idx,], design=~group)
dse <- DESeq(dse)
res <- degResults(dds=dse, name="test", org=NULL,
do_go=FALSE, group="group", xs="group", path_results = NULL)
```

degVar — *Distribution of pvalues by standard desviation range*

Description

Distribution of pvalues by standard desviation range

Usage

```
degVar(pvalues, counts)
```

Arguments

pvalues	pvalues of DEG analysis
counts	matrix with counts for each samples and each gene. row number should be the same length than pvalues vector.

Value

ggplot2 object

Examples

```
data(DEGreportSet)
degVar(DEGreportSet$deg[,4],DEGreportSet$counts)
```

degVB — *Distribution of the standard desviation of DE genes compared to the background*

Description

Distribution of the standard desviation of DE genes compared to the background

Usage

```
degVB(tags, g1, g2, counts, pop = 400)
```

Arguments

tags	list of genes that are DE
g1	list of samples in group 1
g2	list of samples in group 2
counts	matrix with counts for each samples and each gene. Should be same length than pvalues vector.
pop	number of random samples taken for background comparison

Value

ggplot2 object

Examples

```
data(DEGreportSet)
detag <- row.names(DEGreportSet$deg[1:10,])
degVB(detag,DEGreportSet$g1,DEGreportSet$g2,DEGreportSet$counts)
```

`degVolcano` *Create volcano plot from log2FC and adjusted pvalues data frame*

Description

Create volcano plot from log2FC and adjusted pvalues data frame

Usage

```
degVolcano(stats, side = "both",
  title = "Volcano Plot with Marginal Distributions", pval.cutoff = 0.05,
  lfc.cutoff = 1, shade.colour = "orange", shade.alpha = 0.25,
  point.colour = "gray", point.alpha = 0.75,
  point.outline.colour = "darkgray", line.colour = "gray",
  plot_text = NULL)
```

Arguments

`stats`	data.frame with two columns: logFC and Adjusted.Pvalue
`side`	plot UP, DOWN or BOTH de-regulated points
`title`	title for the figure
`pval.cutoff`	cutoff for the adjusted pvalue. Default 0.05
`lfc.cutoff`	cutoff for the log2FC. Default 1
`shade.colour`	background color. Default orange.
`shade.alpha`	transparency value. Default 0.25
`point.colour`	colours for points. Default gray
`point.alpha`	transparency for points. Default 0.75
`point.outline.colour`	Default darkgray
`line.colour`	Defaul gray
`plot_text`	data.frame with three columns: logFC, Pvalue, Gene name

Details

This function was mainly developed by @jnhutchinson.

Value

The function will plot volcano plot together with density of the fold change and p-values on the top and the right side of the volcano plot.

Examples

```
data(DEGreportSet)
stats = DEGreportSet$deg[,c("logFC", "PValue")]
degVolcano(stats)
```

---

geneInfo *data.frame with chromose information for each gene*

Description

data.frame with chromose information for each gene

Usage

```
colors
```

Format

data.frame

Author(s)

Lorena Pantano, 2014-08-14

Source

biomart

---

humanSexDEedgeR *DGEList object for DE genes betwen Male and Females*

Description

DGEList object for DE genes betwen Male and Females

Usage

```
humanSexDEedgeR
```

Format

DGEList

Author(s)

Lorena Pantano, 2014-05-31

Source

gEUvadis

Index