Package 'DEprot'

May 26, 2024

Type Package

Title An R-package for proteomics differential analyses

Version 0.1.0

Author Sebastian Gregoricchio

Maintainer Sebastian Gregoricchio <sebastian.gregoricchio@gmail.com>

Description An R-package for proteomics differential analyses using LFQ values.

License GNU GENERAL PUBLIC LICENSE version 3

Depends R (>= 4.0.3)

Imports BiocManager, cowplot, data.table, doParallel, dplyr, doRNG, ggplot2 (>= 3.3.3), ggdendro, ggh4x, ggtext, ggrepel, matrixStats, MBQN, patchwork, plyr, pryr, prettydoc, S4Vectors, stringr, tidyr, purrr, randomForest, methods, missForest, reshape2, R.utils, devtools, knitr, rmarkdown, patchwork, viridis

biocViews

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

VignetteBuilder knitr

URL https://sebastian-gregoricchio.github.io/DEprot/
https://github.com/sebastian-gregoricchio/DEprot/
https://sebastian-gregoricchio.github.io/

BugReports https://github.com/sebastian-gregoricchio/DEprot/issues

R topics documented:

`DEprot` *DEprot class*

Description

DEprot class

`DEprot.analyses` *DEprot.analyses class*

Description

DEprot.analyses class

`DEprot.correlation` *DEprot.correlation class*

Description

DEprot.correlation class

`DEprot.PCA` *DEprot.PCA class*

Description

DEprot.PCA class

`diff.analyses` *diff.analyses*

Description

Allows for the computation of differential analyses. Includes means, Fold Changes, and pvalues.

Usage

```
## S3 method for class 'analyses'
diff(
  DEprot.object,
  contrast.list,
  linear.FC.th = 2,
  linear.FC.unresp.range = c(1/1.1, 1.1),
  padj.th = 0.05,
  padj.method = "BH",
  stat.test = "t.test",
  up.color = "indianred",
  down.color = "steelblue",
  unresponsive.color = "purple",
  null.color = "gray",
  which.data = "imputed",
  overwrite.analyses = FALSE
)
```

Arguments

`DEprot.object`	An object of class `DEprot`.
`contrast.list`	List of 3-elements vectors indicating (in order): metadata_column, variable_1, variable_2.
`linear.FC.th`	Number indicating the (absolute) fold change threshold (linear scale) to use to define differential proteins. Default: `2`.
`linear.FC.unresp.range`	A numeric 2-elements vector indicating the range (linear scale) used to define the unresponsive fold changes. Default: `c(1/1.1,1.1)`.

padj.th — Numeric value indicating the p.adjusted threshold to apply to the differential analyses. Default: 0.05.

padj.method — String indicating the method to use to correct the p-values. One among: "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none". Default: BH.

stat.test — String indicating the type of statistic test to use. One among: "t-test" and "wilcoxon". Default: "t.test".

up.color — String indicating the color to use for up-regulated proteins in the plots. Default: "indianred".

down.color — String indicating the color to use for up-regulated proteins in the plots. Default: "steelblue".

unresponsive.color — String indicating the color to use for unresponsive proteins in the plots. Default: "purple".

null.color — String indicating the color to use for null proteins in the plots. Default: "gray".

which.data — String indicating which type of counts should be used. One among: 'raw', 'normalized', 'norm', 'imputed', 'imp'. Default: "imputed".

overwrite.analyses — Logical value to indicate whether overwrite analyses already generated. Default: FALSE.

`get.metadata` *get.metadata*

Description

Function to extract the metadata from a DEprot object.

Usage

```
get.metadata(DEprot.object)
```

Arguments

DEprot.object — Any object of class DEprot.

Value

Data.frame corresponding to the metadata of the provided object.

get.results *get.results*

Description

Simplifies the access to the differential expression results table

Usage

```
get.results(DEprot.analyses.object, contrast = 1)
```

Arguments

DEprot.analyses.object	An object of class DEprot.analyses.
contrast	Number indicating the position of the contrast to use for the plotting.

Value

A data.frame.

impute.counts *impute.counts*

Description

Function that allows for the imputation of missing values using the missForest algorithm.

Usage

```
impute.counts(
  DEprot.object,
  max.iterations = 100,
  variable.wise.OOBerror = TRUE,
  use.normalized.data = TRUE,
  overwrite.imputation = FALSE,
  cores = 1,
  parallel.mode = "variables"
)
```

Arguments

DEprot.object	A DEprot object, as generated by load.counts.
max.iterations	Max number of iterations for the missForest algorithm. Default: 100.
variable.wise.OOBerror	Logical value to define whether the OOB error is returned for each variable separately. Default: TRUE.
use.normalized.data	Logical value indicating whether the imputation should be performed based on the rationalized data. Default: TRUE.
overwrite.imputation	Logical value to indicate whether, in the case already available, the table of imputed counts should be overwritten. Default: FALSE.
cores	Number of cores used to run the missForest algorithm. If cores is 1 (or lower), the imputation will be run in parallel. Two modes are possible and can be defined by the parameter parallel.mode. Default: 1.
parallel.mode	Define the mode to use for the parallelization, ignored when cores is more than 1. One among: 'variables', 'forests'. Default: "variables". See also the documentation of the missForest function.

Value

A DEprot object. The boxplot showing the distribution of the protein intensity is remade and added to the slot (boxplot.imputed). A list with parameters and other info about the imputation is added as well in the imputation slot.

See Also

missForest package.

---

load.counts *load.counts*

Description

Function used to generate a DEprot object starting from counts and metadata.

Usage

```
load.counts(
  counts,
  metadata,
  log.base = NA,
  imputation = NA,
  normalization.method = NA,
  column.id = "column.id"
)
```

Arguments

counts	A data.frame or a matrix in which the rownames are the proteins and the columns the samples.
metadata	A data.frame containing at least one column called column.id which corresponds to the colnames of counts. Any other column can be added and will correspond to a "feature"of each sample.
log.base	Number indicating the base of the log used to transform the counts. If none transformation is applied, indicate the default value NA.
imputation	A string indicating the imputation method used. If none, use the default value NA.
normalization.method	String or list indicating the normalization method used. If none, use the default value NA.
column.id	String indicating the name of the column to use as "column.id" from the metadata data.frame. This column must contain all the colnames of counts.

Value

A DEprot object (S4 vector).

---

normalize.counts *normalize.counts*

Description

Function that allows for the tail-robust quantile normalization of the counts using the MBQN package.

Usage

```
normalize.counts(
  DEprot.object,
  balancing.function = "median",
  NRI.RI.ratio.threshold = 0.5,
  overwrite.normalization = FALSE
)
```

Arguments

DEprot.object	A DEprot object, as generated by load.counts.
balancing.function	A string indicating the function to use to balance the quantile normalization. Default: "median".
NRI.RI.ratio.threshold	Value indicating the NRI/RI (nearly rank invariant (NRI)/rank invariant (RI)) ratio threshold. Default: 0.5.

overwrite.normalization
: Logical value to indicate whether, in the case already available, the table of normalized counts should be overwritten. Default: FALSE.

Value

A DEprot object. A boxplot showing the distribution of the protein intensity after normalization for each sample is added to the slot (boxplot.norm).

See Also

load.counts, MBQN package.

---

perform.PCA *perform.PCA*

Description

This function performs principal component analyses (PCA).

Usage

```
perform.PCA(DEprot.object, sample.subset = NULL, which.data = "imputed")
```

Arguments

DEprot.object	An object of class DEprot.
sample.subset	String vector indicating the column names (samples) to keep in the counts table (the 'column.id' in the metadata table). Default: NULL (no subsetting).
which.data	String indicating which type of counts should be used. One among: 'raw', 'normalized', 'norm', 'imputed', 'imp'. Default: "imputed".

Value

A DEprot.PCA, containing the PC values (PCs) and the importance summary (importance).

plot.correlation.heatmap *plot.correlation.heatmap*

Description

Function to generate a correlation heatmap. Includes dendrogram and clustering of the data.

Usage

```
## S3 method for class 'correlation.heatmap'
plot(
  DEprot.object,
  correlation.method = "pearson",
  sample.subset = NULL,
  which.data = "imputed",
  palette = viridis::mako(100, direction = -1),
  correlation.scale.limits = c(0, 1),
  exclude.diagonal = FALSE,
  dendrogram.position = "left",
  dendrogram.color = "black",
  dendrogram.linewidth = 0.5,
  display.values = TRUE,
  values.color = "white",
  values.decimals = 2,
  values.font.size = 2,
  values.transparency = 1,
 plot.title = paste0("**", stringr::str_to_title(correlation.method), " correlation**"),
  plot.subtitle = NULL,
  clustering.method = "complete"
)
```

Arguments

DEprot.object	An object of class DEprot.
correlation.method	String indicating the clustering method to use to generate the correlation matrix. Possible options: 'pearson', 'spearman', 'kendall'. Default: "pearson".
sample.subset	Vector indicating the name of the columns (column.id in the metadata table) to use/subset for the correlation. Default: NULL (no subsetting).
which.data	String indicating which type of counts should be used. One among: 'raw', 'normalized', 'norm', 'imputed', 'imp'. Default: "imputed".
palette	Vector of colors corresponding to the palette to use for the heatmap color scale. Default: viridis::mako(100,direction = -1).

`correlation.scale.limits`
: Two-elements vector to indicate lower and higher limits, respectively, to apply to the correlation coefficient color scale. Default: `c(0,1)`.

`dendrogram.position`
: String indicating the position of the dendrogram. One among: "top", "bottom", "left", "right". Default: `"left"`.

`dendrogram.color`
: String indicating the color of the dendrogram lines. Default: `"black"`.

`dendrogram.linewidth`
: Numeric value indicating the line.width of the dendrogram. Default: `"black"`.

`display.values` Logical value indicating whether the correlation coefficient should be displayed for each cell. Default: `TRUE`.

`values.color` String indicating the color to use for the correlation coefficient labels. Default: `"white"`.

`values.decimals`
: Numeric value indicating the number of decimals at which round the correlation coefficient labels. Default: `2`.

`values.font.size`
: Numeric value indicating the font size of the correlation coefficient labels. Default: `2`.

`values.transparency`
: Numeric value between 0-1 indicating the transparency (alpha) of the correlation coefficient labels. Default: `1`, full color.

`plot.title` String indicating the main title of the plot. Default: `paste(stringr::str_to_title(correlation.met`

`plot.subtitle` String indicating the subtitle of the plot. Default: `NULL`.

`clustering.method`
: String indicating the clustering method to use. The value should be (an unambiguous abbreviation of) one among: 'ward.D', 'ward.D2', 'single', 'complete', 'average' (= UPGMA), 'mcquitty' (= WPGMA), 'median' (= WPGMC) or 'centroid' (= UPGMC).

Value

A `DEprot.correlation` with the correlation heatmap in ggplot format.

`plot.MA` *plot.MA*

Description

Plots a MA plot log2(basemean) x log2(FoldChange) of differential expression results

Usage

```
## S3 method for class 'MA'
plot(
  DEprot.analyses.object,
  contrast = 1,
  up.color = "indianred",
  down.color = "steelblue",
 density.colors = colorRampPalette(colors = RColorBrewer::brewer.pal(9, "Blues"))(101),
  point.size = 2,
  point.alpha = 0.5,
  title = NULL,
  use.uncorrected.pvalue = FALSE,
  symmetric.y = TRUE
)
```

Arguments

DEprot.analyses.object	An object of class DEprot.analyses.
contrast	Number indicating the position of the contrast to use for the plotting.
up.color	String indicating the color to use for up-regulated proteins in the plots. Default: "indianred".
down.color	String indicating the color to use for up-regulated proteins in the plots. Default: "steelblue".
density.colors	List of colors, passed to scale_fill_gradientn, to use for the density gradient. Default: "colorRampPalette(colors = RColorBrewer::brewer.pal(9,"Blues"))(101)".
point.size	Numeric value indicating the size of the dots. Default: 2.
point.alpha	Numeric value between 0 and 1 to indicate the transparency (alpha) of the dots. Default: 0.5.
title	String indicating the title to use. Default: NULL (automatic title).
use.uncorrected.pvalue	Logical value indicating whether it should be used the normal p-value instead of the adjusted one (differential proteins numbers are recomputed). Default: FALSE, padj is used.
symmetric.x	Logical values indicating whether the x-axis scale should be symmetric or not. Default: TRUE.

Value

A ggplot object.

plot.PC.cumulative *plot.PC.cumulative*

Description

Function to plot the cumulative variance of all the principal components of a PCA.

Usage

```
## S3 method for class 'PC.cumulative'
plot(
  DEprot.PCA.object,
  bar.color = "steelblue",
  line.color = "navyblue",
  title = NULL
)
```

Arguments

DEprot.PCA.object	An object of class DEprot.PCA.
bar.color	String indicating the color to use for the bar fill. Default: "steelblue".
line.color	String indicating the color to use for the line and the dots of the cumulative curve. Default: "navyblue".
title	String indicating the title of the plot (markdown annotation supported).

Value

A barplot in ggplot format.

plot.PC.scatter *plot.PC.scatter*

Description

Plot a scatter of the PCs from an object generated by perform.PCA.

Usage

```
## S3 method for class 'PC.scatter'
plot(
  DEprot.PCA.object,
  PC.x = 2,
  PC.y = 1,
  color.column = "column.id",
```

```
  shape.column = NULL,
  label.column = NULL,
  plot.zero.lines = TRUE
)
```

Arguments

DEprot.PCA.object	An object of class DEprot.PCA.object, as generated by perform.PCA.
PC.x	Number indicating which Principal Component (PC) display on the x-axis. Default: 2 (PC2).
PC.y	Number indicating which Principal Component (PC) display on the y-axis. Default: 1 (PC1).
color.column	String indicating the name of the column in the metadata to use as factor for the dot colors. Default: "column.id" (each sample a color).
shape.column	String indicating the name of the column in the metadata to use as factor for the dot shapes. Default: NULL (all dots).
label.column	String indicating the name of the column in the metadata to use as label of the dots. Default: NULL (no labeling).
plot.zero.lines	Logical value to indicate whether to plot two gray dashed lines in correspondence of y=0 and x=0. Default: TRUE.

Value

A ggplot object.

See Also

perform.PCA

---

plot.volcano *plot.volcano*

Description

Plots a volcano plot log2(FoldChange) x -log10(p.adjusted) of differential expression results

Usage

```
## S3 method for class 'volcano'
plot(
  DEprot.analyses.object,
  contrast = 1,
  up.color = "indianred",
  down.color = "steelblue",
```

```
  unresponsive.color = "purple",
  null.color = "gray",
  point.size = 2,
  point.alpha = 0.5,
  title = NULL,
  use.uncorrected.pvalue = FALSE,
  symmetric.x = TRUE
)
```

Arguments

DEprot.analyses.object	An object of class DEprot.analyses.
contrast	Number indicating the position of the contrast to use for the plotting.
up.color	String indicating the color to use for up-regulated proteins in the plots. Default: "indianred".
down.color	String indicating the color to use for up-regulated proteins in the plots. Default: "steelblue".
unresponsive.color	String indicating the color to use for unresponsive proteins in the plots. Default: "purple".
null.color	String indicating the color to use for null proteins in the plots. Default: "gray".
point.size	Numeric value indicating the size of the dots. Default: 2.
point.alpha	Numeric value between 0 and 1 to indicate the transparency (alpha) of the dots. Default: 0.5.
title	String indicating the title to use. Default: NULL (automatic title).
use.uncorrected.pvalue	Logical value indicating whether it should be used the normal p-value instead of the adjusted one (differential proteins numbers are recomputed). Default: FALSE, padj is used.
symmetric.x	Logical values indicating whether the x-axis scale should be symmetric or not. Default: TRUE.

Value

A ggplot object.

---

rename.samples *rename.samples*

Description

Function that allows for the renaming of the columns of the counts table in a DEprot object.

Usage

```
## S3 method for class 'samples'
rename(DEprot.object, metadata.column = "column.id")
```

Arguments

DEprot.object	An object of class DEprot.
metadata.column	A string indicating any column from the metadata table to use to rename the counts table columns. Default "column.id" (no renaming).

Value

An object of class DEprot (S4 vector). A column called old.column.id will be added to the metadata in order to keep track of the original names.

---

sample.config *Example of metadata table*

Description

Dummy table for metadata corresponding to the sample configuratiuon of unimputed.counts.

Usage

```
sample.config
```

Format

A data frame with 12 rows and 6 columns:

column.id IDs of columns in the 'unimputed.counts'

sample.id Actual IDs of the samples

cell cell line ID

condition Culture and treatment conditions

combined.id ID combining cell and condition columns

replicate biological replicate ID

Source

Simulated data

show,DEprot-method *DEprot show-method*

Description

DEprot show-method

Usage

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

show,DEprot.analyses-method
DEprot.analyses show-method

Description

DEprot.analyses show-method

Usage

```
## S4 method for signature 'DEprot.analyses'
show(object)
```

show,DEprot.correlation-method
DEprot.correlation show-method

Description

DEprot.correlation show-method

Usage

```
## S4 method for signature 'DEprot.correlation'
show(object)
```

show,DEprot.PCA-method *DEprot.PCA show-method*

Description

DEprot.PCA show-method

Usage

```
## S4 method for signature 'DEprot.PCA'
show(object)
```

summary,DEprot.analyses-method *DEprot.analyses summary-method*

Description

DEprot.analyses summary-method

Usage

```
## S4 method for signature 'DEprot.analyses'
summary(object)
```

unimputed.counts *Unimputed proteomics (LFQ) data*

Description

Dummy example of full proteomics data (LFQ values in log2). Not imputed.

Usage

```
unimputed.counts
```

Format

A data frame with 13,239 rows and 12 columns:

`rows` proteins

`columns` samples

Source

Simulated data

Index