Package 'MApckg'

September 11, 2015

Title R package of MetaboAnalyst functions

Description Function from MetaboAnalyst adapted to use in R.

Version 0.1

Suggests knitr

VignetteBuilder knitr

URL http://github.com/flajole/MApckg

BugReports http://github.com/flajole/MApckg/issues

Imports
png,xtable,caret,ellipse,RColorBrewer,Cairo,randomForest,e1071,grDevices,grid,som,pheatmap,pls,siggenes,genefilte

License GPL-3

LazyData true

R topics documented:

ANOVA.Anal   *One-way ANOVA*

Description

Performs one-way ANOVA. Creates a file with results: either "anova_nonparametric.csv" or "anova_posthoc.csv" depending on nonpar argument.

Usage

```
ANOVA.Anal(dataSet, analSet, nonpar = FALSE, thresh = 0.05,
  post.hoc = "fisher")
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
nonpar	If FALSE - use classical ANOVA; if TRUE - Kruskal Wallis Test
thresh	Threshold of significance.
post.hoc	Post-hoc statistics: "tukey" or "fisher"

Value

Native `analSet` with one added `$aov` element consisting of:

- `$aov.nm` - method of analysis
- `$raw.thresh` - value of `thresh` argument
- `$thresh` - `-log10(thresh)`
- `$p.value` - p-values
- `$p.log` - `-log10(p.value)`
- `$inx.imp` - logical vector of features with significant difference
- `$post.hoc` - post-hoc statistics, value of `post.hoc` argument
- `$sig.mat` - data frame of significant features with corresponding statistics

---

ANOVA2.Anal   *Two-way ANOVA*

Description

Performs two-way ANOVA. Creates a file with results: either `"anova_within_sbj.csv"` or `"anova_between_sbj.csv"` depending on `type` argument.

Usage

```
ANOVA2.Anal(dataSet, analSet, thresh = 0.05, p.cor = "fdr", type = "b")
```

Arguments

dataSet	List, data set object generated by `MS_to_MA` function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
thresh	Threshold of significance.
p.cor	Method of p-value correction, one of `p.adjust` methods
type	Type of comparison. If `"b"` - between subjects; if `"w"` - within subjects

Value

Native `analSet` with one added `$aov2` element consisting of:

- `$type` - value of `type` argument
- `$sig.nm` - name of file with the resulting ANOVA table
- `$thresh` - value of `thresh` argument
- `$multi.c` - value of `p.cor` argument
- `$sig.mat` - data frame of significant features
- `$vennC` - Venn counts

EBAM.A0.Anal   *EBAM Fudge Factor*

Description

`EBAM.A0.Anal` - finds EBAM Fudge Factor. Uses `find.a0` function.

`EBAM.Cmpd.Anal` - performs Empirical Bayesian Analysis of Microarray (and Metabolites). Uses `ebam` function.

Usage

```
EBAM.A0.Anal(dataSet, analSet, paired = FALSE, var.equal = TRUE)

EBAM.Cmpd.Anal(dataSet, analSet, method = "z.ebam", A0 = 0,
  paired = FALSE, var.equal = TRUE)
```

Arguments

dataSet	List, data set object generated by `MS_to_MA` function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
paired	Are values in data set paired or not.
var.equal	Are variances assumed equal or not.
method	A character string or name specifying the method or function that should be used in the computation of the expression score z. If "z.ebam", a modified t- or F-statistic, respectively, will be computed as proposed by Efron et al. (2001). If "wilc.ebam", a (standardized) Wilcoxon sum / signed rank statistic will be used as expression score. Note: if method is "wilc.ebam", then A0 and var.equal parameters will be ignored
A0	Fudge Factor. `EBAM.A0.Anal`

Value

`EBAM.A0.Anal` - native `analSet` with one added $ebam.a0 element containing standard `find.a0` function output

`EBAM.Cmpd.Anal` - native `analSet` with one added $ebam element containing standard `ebam` function output

Functions

- `EBAM.A0.Anal`:

FC.Anal *Fold change*

Description

Fold change (FC) analysis is to compare the absolute value change between two group means. Since column-wise normalization (i.e. log transformation, mean-centering) will significantly change the absolute values, FC is calculated as the ratio between two group means using data before column-wise normalization was applied.

Usage

```
FC.Anal(dataSet, analSet, fcthresh = 2, percent.thresh = 0.75,
  cmpType = 0)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
fcthresh	Fold-change threshold.
percent.thresh	Sig. count threshold (for paired data only).
cmpType	Comparison type. If equal 0 then group 1 is compared against group 2. Otherwise, vice versa.

Details

For paired analysis, the program first counts the number of pairs with consistent change above the given FC threshold. If this number exceeds a given count threshold, the variable will be reported as significant. Writes an output file "fold_change.csv".

Value

Native analSet with one added $fc element consisting of:

- $paired - are data paired or not
- $raw.thresh - value of fcthresh argument
- $max.thresh - upper log-threshold of fold change
- $min.thresh - lower log-threshold of fold change
- $fc.all - fold changes of all features
- $fc.mat - matrix of fold changes (for paired data only)
- $fc.log - log2 of fold changes (for unpaired data only)
- $inx.up - logical vector of increasing features (for paired data only)
- $inx.down - logical vector of decreasing features (for paired data only)
- $inx.imp - logical vector of features with significant difference (for unpaired data only)
- $sig.mat - data frame of significant features

Kmeans.Anal   *K-means analysis*

Description

Perform K-means analysis. Uses [kmeans](#) function.

Usage

```
Kmeans.Anal(dataSet, analSet, clust.num)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
clust.num	The cluster number.

Value

Native `analSet` with one added `$kmeans` element containing standard [kmeans](#) output.

PatternHunter   *Pattern hunter*

Description

Correlation analysis.

`FeatureCorrelation` performs analysis against a given feature. Writes an output file ″correlation_feature.csv″

`Match.Pattern` performs analysis against a given pattern. Writes an output file ″correlation_pattern.csv″ respectively.

Usage

```
FeatureCorrelation(dataSet, analSet, dist.name = ″pearson″, varName)

Match.Pattern(dataSet, analSet, dist.name = ″pearson″, pattern = NULL)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
dist.name	Method of correlation calculating, one of ″pearson″, ″kendall″, ″spearman″
varName	The name of the feature to perform correlation analysis against.
pattern	A character vector, the pattern to search. The pattern is specified as a series of numbers separated by "-". Each number corresponds to the expected expression pattern in the corresponding group. For example, a 1-2-3-4 pattern is used to search for features that increase linearly with time in a time-series data with four time points (or four groups). The order of the groups is given as the first item in the predefined patterns.

Value

Native `analSet` with one added `$corr` element consisting of

- `$corr$sig.nm` - ???
- `$corr$cor.mat` - correlation matrix
- `$corr$pattern` - name of the feature of comparison or the used pattern.

PCA.Anal   *Perform PCA*

Description

Uses [prcomp](#) function. Writes two output files: "pca_score.csv" and "pca_loadings.csv".
Adds `analSet$pca` element with [prcomp](#) function output and some basic statistics.

Usage

```
PCA.Anal(dataSet, analSet)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).

Value

Native `analSet` with one added `$pca` element consisting of:

- all the elements of the standard [prcomp](#) output
- `$std` - standard deviation
- `$variance` - variance explained by each PC
- `$cum.var` - cummulated variance explained

PCA.Loadings   *PCA loadings*

Description

PCA loadings. Adds loading matrix to `analSet$pca` element.

Usage

```
PCA.Loadings(dataSet, analSet, inx1 = 1, inx2 = 2)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
inx1,inx2	The order numbers of PCs.

Value

Native `analSet` with two added elements:

- `$pca$load.x.uniq` - ???
- `$pca$imp.loads` - loading matrix

`PlotANOVA`   *Plot one-way ANOVA results*

Description

Plot one-way ANOVA results

Usage

```
PlotANOVA(dataSet, analSet, imgName = "aov_", format = "png", dpi = 72,
  width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](MS_to_MA) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

`PlotANOVA2`   *Plot two-way ANOVA results*

Description

Plot two-way ANOVA results

Usage

```
PlotANOVA2(dataSet, analSet, imgName = "aov2_", format = "png", dpi = 72,
  width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](MS_to_MA) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotCorr   *Plot pattern correlations*

Description

Plot the top 25 features correlated to a feature or a pattern.

Usage

```
PlotCorr(dataSet, analSet, imgName = "ptn_", format = "png", dpi = 72,
  width = NA)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotCorrHeatMap   *Plot correlations heatmap*

Description

Plot correlations heatmap and write an output file "correlation_table.csv". Note, the heatmap will only show correlations for a maximum of 1000 features. For larger datasets, only top 1000 features will be selected based on their interquantile range (IQR). When color distribution is fixed, you can potentially compare the correlation patterns among different data sets. In this case, you can choose "do not perform clustering" for all data set, or only to perform clustering on a single reference data set, then manually re-arranged other data sets according to the clustering pattern of the reference data set.

Usage

```
PlotCorrHeatMap(dataSet, analSet, imgName = "corr_heat_", format = "png",
  dpi = 72, width = NA, cor.method = "pearson", colors = "default",
  viewOpt = "overview", fix.col = F, no.clst = FALSE, top = F,
  top.num = 999)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
cor.method	Method of correlation calculation, one of "pearson", "kendall", "spearman"
colors	The color contrast. One of "default", "gbr" (red/green), "heat", "topo", "gray"
viewOpt	View mode, "overview" or "detailed"
fix.col	If TRUE, color distribution is fixed.
no.clst	If TRUE, no clustering lines are plotted.
top	????????????????????
top.num	The number of the features with the best contrast to be plotted.

PlotEBAM	*Plot EBAM*

Description

Functions for plotting the results of EBAM.

PlotEBAM.A0 - plot EBAM A0 plot.

PlotEDAM.Cmpd - plot significant features.

Usage

```
PlotEBAM.A0(dataSet, analSet, imgName = "ebam_view_", format = "png",
  dpi = 72, width = NA)

PlotEBAM.Cmpd(dataSet, analSet, imgName = "ebam_imp_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotFC *Plot fold change*

Description

Plot fold change analysis results.

Usage

```
PlotFC(dataSet, analSet, imgName = "fc_", format = "png", dpi = 72,
  width = NA)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotHCTree *Plot dendrogram*

Description

Plot hierarchical clustering tree.

Usage

```
PlotHCTree(dataSet, analSet, imgName = "tree_", format = "png", dpi = 72,
  width = NA, smplDist = "euclidean", clstDist = "ward.D")
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
smplDist	The distance measure. One of "euclidean", "spearman"
clstDist	The agglomeration method to be used, one of "ward.D", "ward.D2", "single", "complete", "average". For details: hclust

Description

Uses [pheatmap](#) function. This method displays data in the form of colored cells. It provides direct visualization of the relative levels of individual samples or variables. Each colored cell on the map corresponds to a concentration value in your data table, with samples in rows and features/compounds in columns. You can use heatmap to identify samples/features that are unusually high/low.

Usage

```
PlotHeatMap2(dataSet, analSet, imgName = "heatmap2_", format = "png",
  dpi = 72, width = NA, smplDist = "pearson", clstDist = "average",
  colors = "bwm", viewOpt = "overview", hiRes = FALSE, sortInx = 2,
  var.inx = 1:ncol(dataSet$norm))
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
smplDist	The distance measure, one of "euclidean", "pearson", "minkowski"
clstDist	The agglomeration method to be used, one of "ward.D", "ward.D2", "single", "complete", "average". For details: [pheatmap](#)
colors	The color contrast. One of "default", "gbr" (red/green), "heat", "topo", "gray"
viewOpt	View mode, "overview" or "detailed"
hiRes	If TRUE, then produces hi-resolution plot.
sortInx	If "A", samples are arranged by the first factor; if "B" - by the second one.
var.inx	Vector of the numbers of plotted features.

Value

Native analSet with one added $htmap2 element containing:

- $htmap2$dist.par - equal smplDist argument value

- $htmap2$clust.par - equal clstDist argument value

PlotKmeans   *Plot K-means*

Description

Plot K-means map. Please note: only the cluster members will be calculated if the specified cluster number > 20. The blue lines represent the median intensities of each cluster.

Usage

```
PlotKmeans(dataSet, analSet, imgName = "kmeans_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotPCA   *Plot PCA*

Description

Set of functions for plotting the results of PCA.

PlotPCAPairSummary - plot summary.

PlotPCAScree - plot scree plot of varience explained by PC.

PlotPCA2DScore - plot 2D score plot.

PlotPCABiplot - plot the biplot.

PlotPCALoadings - plot loadings.

Usage

```
PlotPCAPairSummary(dataSet, analSet, imgName = "pca_pair_", format = "png",
  dpi = 72, width = NA, pc.num = 2)

PlotPCAScree(dataSet, analSet, imgName = "pca_scree_", format = "png",
  dpi = 72, width = NA, scree.num)

PlotPCA2DScore(dataSet, analSet, imgName = "pca_score2d_", format = "png",
  dpi = 72, width = NA, inx1 = 1, inx2 = 2, reg = 0.95, show = TRUE,
  grey.scale = FALSE)
```

```
PlotPCABiplot(dataSet, analSet, imgName = "pca_biplot_", format = "png",
  dpi = 72, width = NA, inx1 = 1, inx2 = 2)

PlotPCALoadings(dataSet, analSet, imgName = "pca_loading_", format = "png",
  dpi = 72, width = NA, plotType = "scatter", show = TRUE)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
pc.num	The number of plotted principal components.
scree.num	The total number of plotted PCs.
inx1,inx2	The order numbers of PCs.
reg	Set the confidence level for plotting confidence region ellipse.
show	If TRUE then points at the plot are labeled.
grey.scale	If TRUE then plot is colored in 50 shades of gray.
plotType	"scatter" for scatter plot or "boxplot" for box plot

---

PlotPLS   *Plot PLS*

Description

Set of functions for plotting the results of PLS.

PlotPLSPairSummary - plot summary.

PlotPLS2DScore - plot 2D score plot.

PlotPLSLoadings - plot loadings.

PlotPLS.Classification - plot PLSDA classification performance using different components

PlotPLS.Permutation - plot PLSDA permutation plot.

Usage

```
PlotPLSPairSummary(dataSet, analSet, imgName = "pls_pair_", format = "png",
  dpi = 72, width = NA, pc.num = 2)

PlotPLS2DScore(dataSet, analSet, imgName = "pls_score2d_", format = "png",
  dpi = 72, width = NA, inx1 = 1, inx2 = 2, reg = 0.95, show = TRUE,
  grey.scale = F)

PlotPLSLoading(dataSet, analSet, imgName = "pls_loading_", format = "png",
  dpi = 72, width = NA, plotType = "scatter", show = T)
```

```
PlotPLS.Classification(dataSet, analSet, imgName = "plsda_", format = "png",
  dpi = 72, width = NA)

PlotPLS.Permutation(dataSet, analSet, imgName = "pls_perm_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
pc.num	The number of plotted principal components.
inx1,inx2	The order numbers of PCs.
reg	Set the confidence level for plotting confidence region ellipse.
show	If TRUE then points at the plot are labeled.
grey.scale	If TRUE then plot is colored in 50 shades of gray.
plotType	"scatter" for scatter plot or "boxplot" for box plot

---

PlotPLS.Imp   *Plot PLS important features*

Description

There are two importance measures in PLS-DA : one is variable importance in projection (VIP) and the other is weighted sum of absolute regression coefficients (coef.). The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study.

Usage

```
PlotPLS.Imp(dataSet, analSet, imgName = "pls_imp_", format = "png",
  dpi = 72, width = NA, type = "vip", feat.nm, feat.num = 15,
  color.BW = FALSE)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
type	Importance measure. "vip" - for variable importance in projection (VIP); "coef" - for weighted sum of absolute regression coefficients.
feat.nm	?????????????????????????????????????
feat.num	The number of top features to plot.
color.BW	If TRUE, grey scale is used.

PlotRF   *Plot Random Forest*

Description

Set of functions for plotting the results of Random Forest analysis.

`PlotRF.Classification` - plot feature classification.

`PlotRF.VIP` - plot variable importance of features ranked by their contributions to classification accuracy (MeanDecreaseAccuracy).

`PlotRF.Outlier` - plot outlying measures, only top 5 potential outliers are labeled.

Usage

```
PlotRF.Classification(dataSet, analSet, imgName = "rf_cls_", format = "png",
  dpi = 72, width = NA)

PlotRF.VIP(dataSet, analSet, imgName = "rf_imp_", format = "png",
  dpi = 72, width = NA)

PlotRF.Outlier(dataSet, analSet, imgName = "rf_outiler_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by `MS_to_MA` function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotRSVM   *Plot RSVM*

Description

Set of functions for plotting the results of recursive Support Vector Machine analysis.

`PlotRSVM.Classification` - plot feature classification.

`PlotRSVM.VIP` - please note : features are ranked by their frequencies being selected in the best classifiers (only top 15 will be shown)

Usage

```
PlotRSVM.Classification(dataSet, analSet, imgName = "svm_cls_",
  format = "png", dpi = 72, width = NA)

PlotRSVM.Cmpd(dataSet, analSet, imgName = "svm_imp_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotSAM   *Plot SAM*

Description

Functions for plotting the results of SAM.

PlotSAM.FDR - plot FDR.

PlotSAM.Cmpd - plot significant features.

Usage

```
PlotSAM.FDR(dataSet, analSet, delta = NULL, imgName = "sam_view_",
  format = "png", dpi = 72, width = NA)

PlotSAM.Cmpd(dataSet, analSet, imgName = "sam_imp_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
delta	The delta to control FDR. If NULL, then it's suggested automaticly
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotSOM   *Plot SOM*

Description

Plot SOM map for less than 20 clusters. Please note : only cluster members will be calculated if
the total cluster number (xdim*ydim) > 20. The blue lines represent the median intensities of each
cluster.

Usage

```
PlotSOM(dataSet, analSet, imgName = "som_", format = "png", dpi = 72,
  width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotSubHeatMap   *Plot heatmap*

Description

Plot a sub heatmap based on results from t-tests/ANOVA, VIP or Random Forest analysis. Uses
[pheatmap](#) function. Heatmap provides intuitive visualization of the data table. Each colored cell
on the map corresponds to a concentration value in your data table, with samples in rows and features/compounds in columns. You can use heatmap to identify samples/features that are unusually
high/low.
Tip 1: Do not re-organize samples/rows to show the natural contrast among groups (with each group
a block).
Tip 2: Display top number of features ranked by t-tests to retain the most constrasting patterns.

Usage

```
PlotSubHeatMap(dataSet, analSet, imgName = "heatmap_", format = "png",
  dpi = 72, width = NA, smplDist = "euclidean", clstDist = "ward.D",
  colors = "default", method.nm = "tanova", top.num = 25,
  viewOpt = "overview", rowV = TRUE, colV = TRUE, border = T)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.
smplDist	The distance measure, one of "euclidean", "pearson", "minkowski"
clstDist	The agglomeration method to be used, one of "ward.D", "ward.D2", "single", "complete", "average". For details: [pheatmap](#)
colors	The color contrast. One of "default", "gbr" (red/green), "heat", "topo", "gray"
method.nm	Type of analysis applied to detect features with the best contrast. One of "tanova" - for t-test or ANOVA; "vip" - for PLS-DA VIP; "rf" - for Random Forest analysis
top.num	The number of the features with the best contrast to be plotted.
viewOpt	View mode, "overview" or "detailed"
rowV	If TRUE, samples are reorganized.
colV	If TRUE, features are reorganized.
border	If TRUE, show cell borders.

Value

Native analSet with one added $htmap element containing:

- $htmap$dist.par - equal smplDist argument value
- $htmap$clust.par - equal clstDist argument value

PlotTT	*Plot t-test results*

Description

Plot t-test results

Usage

```
PlotTT(dataSet, analSet, imgName = "tt_", format = "png", dpi = 72,
  width = NA)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PlotVolcano   *Plot Volcano results*

Description

The volcano plot is a combination of fold change and t-tests. Note, for unpaired samples, the x-axis is log (FC). For paired analysis, the x-axis is number of significant counts. Y-axis is -log10(p.value) for both cases.

Usage

```
PlotVolcano(dataSet, analSet, imgName = "volcano_", format = "png",
  dpi = 72, width = NA)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
imgName	Image file name prefix.
format	Image format, one of: "png", "tiff", "pdf", "ps", "svg"
dpi	Image resolution.
width	Image width.

PLS.Anal   *PLS analysis Uses* "oscorespls" *method of* plsr *function . Writes two output files:* "pls_score.csv" *and* "pls_loadings.csv". *Adds* analSet$pls *element with* plsr *function output.*

Description

PLS analysis Uses "oscorespls" method of plsr function . Writes two output files: "pls_score.csv" and "pls_loadings.csv". Adds analSet$pls element with plsr function output.

Usage

```
PLS.Anal(dataSet, analSet)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).

Value

Native analSet with one added $plsr element containing plsr function output

PLS.Loadings *PLS loadings*

Description

PLS loadings. Adds loading matrix to analSet$pls element.

Usage

```
PLS.Loadings(dataSet, analSet, inx1 = 1, inx2 = 2)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
inx1,inx2	The numbers of PC.

Value

Native analSet with two added elements:

- plsload.x.uniq - ???
- plsimp.loads - loading matrix

PLSDA.CV *PLSDA cross validation*

Description

Perform PLSDA classification and feature selection. Writes two output files: "plsda_vip.csv" and "plsda_coef.csv".

Usage

```
PLSDA.CV(dataSet, analSet, methodName = "CV",
  compNum = GetDefaultPLSCVComp(dataSet), choice = "Q2")
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
methodName	Cross validation method "LOOCV" or "CV"
compNum	The maximum number of components to search.
choice	The determining performance measure, one of "Q2", "R2", "Accuracy"

Value

Native `analSet` with one added `$plsda` element consisting of:

- `$best.num` - the detected best number of components

- `$choice` - value of `choice` argument

- `$coef.mat` - matrix of VIPs (variable importance in projection)

- `$vip.mat` - matrix of weighted sums of absolute regression coefficients

- `$fit.info` - all performance measures calculated for different number of components

---

`PLSDA.Permut`   *Permutation PLSDA test*

Description

Perform permutation, using training classification accuracy or separation distance as indicator, for two or multi-groups.

Usage

```
PLSDA.Permut(dataSet, analSet, num = 100, type = "accu")
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
num	The number of permutations.
type	The indicator. If "accu" - training classification accuracy; if "sep" - separation distance.

Value

Native `analSet` with some elements added:

- `$plsda$permut.p` -

- `$plsda$permut.inf` -

- `$plsda$permut.type` -

- `$plsda$permut` -

RF.Anal   *Random Forests*

Description

Perform Random Forest analysis. Uses randomForest function. Writes results to "randomforests_sigfeatures.csv" file.

Usage

```
RF.Anal(dataSet, analSet, treeNum = 500, tryNum = 10)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
treeNum	Number of trees to grow.
tryNum	Number of predictors to try for each node.

Value

Native analSet with two added elements:

- $rf - standard randomForest function output
- $rf.sigmat - matrix of features significance (sorted by MeanDecreaseAccuracy).

RSVM.Anal   *Support Vector Machine (SVM)*

Description

Perform recursive SVM for feature selection and classification. Uses svm function. Writes results to "svm_sigfeatures.csv" file. R-SVM uses SVM (with linear kernel) to perform classification recursively using different feature subsets. Features are selected based on their relative contribution in the classification using cross validation error rates. The least important features are eliminated in the subsequent steps. This process creates a series of SVM models (levels). The features used by the best model are plotted. LOOCV: leave one out cross-validation.

Usage

```
RSVM.Anal(dataSet, analSet, cvType = 10)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
cvType	Type of cross-validation of: integer - N-fold-CV is used; "LOO" - leave-one-out CV (LOOCV); "bootstrape" - bootstrape CV

Value

Native `analSet` with one added element $svm containing - standard randomForest function output

---

`SAM.Anal` *SAM analysis*

Description

Perform Significance Analysis of Microarray (and Metabolites). Uses [sam](sam) function.

Usage

```
SAM.Anal(dataSet, analSet, method = "d.stat", paired = FALSE,
  var.equal = TRUE)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](MS_to_MA) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
method	A character string or a name specifying the method/function that should be used in the computation of the expression scores d. If "d.stat", a modified t-statistic or F-statistic, respectively, will be computed as proposed by Tusher et al. (2001). If "wilc.stat", a Wilcoxon rank sum statistic or Wilcoxon signed rank statistic will be used as expression score.
paired	Are values in data set paired or not.
var.equal	Are variances assumed equal or not.

Value

Native `analSet` with one added $sam element containing standard [sam](sam) function output

---

`SetEBAMSigMat` *EBAM matrix of significance*

Description

EBAM matrix of significance. Writes results to "ebam_sigfeatures.csv" file.

Usage

```
SetEBAMSigMat(dataSet, analSet, delta = 0.9)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](MS_to_MA) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
delta	The delta to control FDR

Value

Native `analSet` with two added elements:

- `$ebam.cmpd` - matrix containing `z.value`, `posterior`, `local.fdr` columns
- `$ebam.delta` - value of `delta` argument

SetSAMSigMat   *SAM matrix of significance*

Description

SAM matrix of significance. Writes results to ″`sam_sigfeatures.csv`″ file.

Usage

```
SetSAMSigMat(dataSet, analSet, delta = 1.3)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
delta	The delta to control FDR

Value

Native `analSet` with two added elements:

- `$sam.cmpd` - matrix containing `d.value`, `stdev`, `rawp`, `q.value` columns
- `$sam.delta` - value of `delta` argument

SOM.Anal   *SOM analysis*

Description

Perform Self Organizing Map analysis. Uses [som](#) function. The parameters and outcome are identical.

Usage

```
SOM.Anal(dataSet, analSet, x.dim = 1, y.dim = 3, initMethod = ″linear″,
  neigb = ″gaussian″)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
x.dim	An integer specifying the x-dimension of the map.
y.dim	An integer specifying the y-dimension of the map.
initMethod	A character string specifying the initializing method. The following are permitted: "sample" uses a random sample from the data; "random" uses random draws from N(0,1); "linear" uses the linear grids upon the first two principle components directin.
neigb	A character string specifying the neighborhood function type. The following are permitted: "bubble", "gaussian"

Value

Native analSet with one added $som element containing standard [som](#) output.

Ttests.Anal   *T-test*

Description

Performs group comparison. Creates a file with results: either "Wilcoxon_rank.csv" or "t_test.csv" depending on nonpar argument.

Usage

```
Ttests.Anal(dataSet, analSet, nonpar = FALSE, thresh = 0.05,
  paired = FALSE, var.equal = TRUE)
```

Arguments

dataSet	List, data set object generated by [MS_to_MA](#) function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
nonpar	If FALSE - use classical t-test; if TRUE - Wilcoxon Rank Test
thresh	Threshold of significance.
paired	Are values in data set paired or not.
var.equal	Are variances assumed equal or not.

Value

Native analSet with one added $tt element consisting of:

- $tt.nm - method of analysis
- $raw.thresh - value of thresh argument
- $thresh - $-\log 10(\text{thresh})$
- $p.value - p value
- $p.log - $-\log 10(\text{p.value})$
- $inx.imp - logical vector of features with significant difference
- $sig.mat - data frame of significant features

Note

Note, for large data set (> 1000 variables), both the paired information and the group variance will be ignored, and the default parameters will be used for t-tests to save computational time. If you choose non-parametric tests (Wilcoxon rank-sum test), the group variance will be ignored.

Volcano.Anal   *Volcano*

Description

The volcano plot is a combination of fold change and t-tests. Note, for unpaired samples, the x-axis is log (FC). For paired analysis, the x-axis is number of significant counts. Y-axis is -log10(p.value) for both cases. Writes an output file "volcano.csv"

Usage

```
Volcano.Anal(dataSet, analSet, paired = FALSE, fcthresh = 2, cmpType = 0,
  percent.thresh = 0.75, nonpar = FALSE, thresh = 0.05,
  var.equal = TRUE)
```

Arguments

dataSet	List, data set object generated by MS_to_MA function.
analSet	List, containing the results of statistical analysis (can be just an empty list).
paired	Are values in data set paired or not.
fcthresh	Fold-change threshold.
cmpType	Comparison type. If equal 0 then group 1 is compared against group 2. Otherwise, vice versa.
percent.thresh	Sig. count threshold (for paired data).
nonpar	If FALSE - use classical t-test; if TRUE - Wilcoxon Rank Test
thresh	Threshold of significance.
var.equal	Are variances assumed equal or not.

Value

Native analSet with one added $volcano element consisting of:

- $raw.threshx - value of fcthresh argument
- $raw.threshy - value of thresh argument
- $paired - –log10(thresh)
- $max.xthresh - upper log-threshold of fold change
- $min.xthresh - lower log-threshold of fold change
- $thresh.y - –log10(thresh)
- $fc.all - fold changes of all features
- $fc.log - log of fold changes
- $fc.log.uniq - java-object

- `$inx.up` - logical vector of increasing features
- `$inx.down` - logical vector of decreasing features
- `$p.log` - `-log10(p.value)`
- `$inx.p` - logical vector of features with `p <= thresh`
- `$sig.mat` - data frame of significant features

Index