

# Package ‘ROTS’

April 15, 2017

**Type** Package

**Title** Reproducibility-Optimized Test Statistic

**Version** 1.2.0

**Author** Fatemeh Seyednasrollah, Tomi Suomi, Laura L. Elo

**Maintainer** Fatemeh Seyednasrollah <fatsey@utu.fi>

**Description** Calculates the Reproducibility-Optimized Test Statistic (ROTS) for differential testing in omics data.

**biocViews** Software, GeneExpression, DifferentialExpression, Microarray, RNASeq, Proteomics

**License** GPL (>= 2)

**LazyData** yes

**LinkingTo** Rcpp

**Imports** Rcpp, stats, Biobase, methods

**Depends** R (>= 3.3)

**NeedsCompilation** yes

## R topics documented:

affySpikeIn . . . . .	1
plot.ROTS . . . . .	2
ROTS . . . . .	3
summary.ROTS . . . . .	5
upsSpikeIn . . . . .	6
<b>Index</b>	<b>7</b>

---

affySpikeIn

*Gene expression data from the Affymetrix spike-in experiment*

---

**Description**

The Affymetrix HG-U95Av2 spike-in data contains two groups of 12 samples. In this carefully controlled experiment, it is known that 14 genes are differentially expressed between the two sample groups, while the rest of the genes are equally expressed. In order to reduce the running time of the vignette example, this package uses a subset of 1000 genes of the original dataset including all the spiked genes with 5 samples from each group. The preprocessed RMA-normalized data were obtained from the Bioconductor 2.2 package DESeq. For more details about the spike-in experiment, see <http://www.affymetrix.com>.

**Usage**

```
affySpikeIn
```

**Format**

`affySpikeIn` a numeric matrix of gene expression levels with 12626 rows (genes) and 24 columns (samples).

`affySpikeIn.L` a vector indicating the sample groups.

`affySpikeIn.gnames` a character vector containing the names of the genes.

`spikedgene` a numeric vector indicating the locations of the 14 spiked genes.

---

plot.ROTS

*Plotting of a ROTs object*

---

**Description**

Plots the ROTs objects created with the ROTs package.

**Usage**

```
## S3 method for class 'ROTS'
plot(x, fdr=0.05, type=NULL, labels=FALSE, ...)
```

**Arguments**

<code>x</code>	A ROTs object created from differential expression testing run by ROTs.
<code>fdr</code>	Selected cutoff for FDR value.
<code>type</code>	Type of plot to be created. Options are: volcano, heatmap, ma, reproducibility, pvalue, pca.
<code>labels</code>	Option to print labels for differentially expressed features.
<code>...</code>	Other arguments passed to the plot function.

**Details**

This function plots the results from a ROTs object using given false discovery rate threshold.

**Value**

Plots the results from ROTs object.

**Author(s)**

Fatemeh Seyednasrollah, Tomi Suomi, Laura L. Elo Maintainer: Fatemeh Seyednasrollah <fatsey@utu.fi>

**See Also**

[affySpikeIn](#)

**Examples**

```
## ROTS-statistic for the Affymetrix spike-in data.
rots.out <- ROTS(data = affySpikeIn, groups = c(rep(0,5), rep(1,5)),
  B = 100, K = 500 , seed = 1234)
## Plotting of the ROTS results.
plot(rots.out, type="volcano")
```

---

ROTS

*Reproducibility-Optimized Test Statistic (ROTS)*

---

**Description**

Calculates the reproducibility-optimized test statistic (ROTS) for ranking genes in order of evidence for differential expression in two-group comparisons.

**Usage**

```
ROTS(data, groups, B, K, paired = FALSE, seed = NULL, a1 = NULL, a2 = NULL, log = TRUE, progress = F
```

**Arguments**

data	A numeric data matrix or an ExpressionSet instance, in which rows correspond to genes and columns correspond to samples.
groups	A vector indicating the sample groups.
B	An integer specifying the number of bootstrap and permutation resamplings.
K	An integer indicating the largest top list size considered.
paired	A logical indicating whether a paired test is performed. The samples are expected to be in the same order in both groups.
seed	An integer seed for the random number generator.
a1, a2	Non-negative parameters. See details section for further information.
log	A logical (default TRUE) indicating whether input data is log2 scaled. This information is only used to calculate log fold change.
progress	A logical indicating if additional progress bars are shown.

## Details

The reproducibility-optimization procedure ROTS enables the selection of a suitable gene ranking statistic directly from the given dataset. The statistic is optimized among a family of t-type statistics  $d = m/(a_1+a_2*s)$ , where  $m$  is the absolute difference between the group averages,  $s$  is the pooled standard error, and  $a_1$  and  $a_2$  are the non-negative parameters to be optimized. Two special cases of this family are the ordinary t-statistic ( $a_1=0$ ,  $a_2=1$ ) and the signal log-ratio ( $a_1=1$ ,  $a_2=0$ ). The optimality is defined in terms of maximal overlap of top-ranked genes in group-preserving bootstrap datasets. Importantly, besides the group labels, no a priori information about the properties of the data is required and no fixed cutoff for the gene rankings needs to be specified. For more details about the reproducibility-optimization procedure, see Elo et al. (2008).

The user is given the option to adjust the largest top list size considered in the reproducibility calculations, since lowering this size can markedly reduce the computation time. In large data matrices with thousands of rows, we generally recommend using a size of several thousands. In smaller data matrices, and especially if there are many rows with only a few non-missing entries, the size of  $K$  should be decreased accordingly.

ROTS tolerates a moderate number of missing values in the data matrix by effectively ignoring their contribution during the operation of the procedure. However, each row of the data matrix must contain at least two values in both groups. The rows containing only a few non-missing values should be removed; or alternatively, the missing data entries can be imputed using, e.g., the K-nearest neighbour imputation, which is implemented in the Bioconductor package `impute`. ROTS assumes the input data matrix is log2 transformed (the default for log parameter is set to TRUE). Although, this only affects fold change values, we recommend setting log parameter to FALSE if the input matrix is not log transformed to avoid downstream confusions.

If the parameter values  $a_1$  and  $a_2$  are set by the user, then no optimization is performed but the statistic and FDR-values are calculated for the given parameters. The false discovery rate (FDR) for the optimized test statistic is calculated by permuting the sample labels. The results for all the genes can be obtained by setting the FDR cutoff to 1.

## Value

ROTS returns an object of class ROTS, which is a list containing the following components

<code>data</code>	the expression data matrix.
<code>B</code>	the number of bootstrap and permutation resamplings.
<code>d</code>	the value of the optimized ROTS-statistic for each gene.
<code>pvalue</code>	the corresponding pvalues.
<code>FDR</code>	the corresponding FDR-values.
<code>a1</code>	the optimized parameter $a_1$ .
<code>a2</code>	the optimized parameter $a_2$ .
<code>k</code>	the optimized top list size.
<code>R</code>	the optimized reproducibility value.
<code>Z</code>	the optimized reproducibility Z-score.

`print` prints the optimized parameters  $a_1$  and  $a_2$ , the optimized top list size and the corresponding reproducibility values.

`summary` summarizes the results of a ROTS analysis. If `fdr` and `num.genes` are not specified, then the optimized parameters  $a_1$  and  $a_2$ , the optimized top list size and the corresponding reproducibility values are shown. If `fdr` or `num.genes` is specified, then also the gene-specific information is shown for the genes at the specified FDR-level or top list size, respectively.

**Author(s)**

Fatemeh Seyednasrollah, Tomi Suomi, Laura L. Elo  
Maintainer: Fatemeh Seyednasrollah <fatsey@utu.fi>

**References**

L. L. Elo, S. Filen, R. Lahesmaa and T. Aittokallio: Reproducibility-optimized test statistic for ranking genes in microarray studies. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 5: 423–431, 2008.

**See Also**

[affySpikeIn](#)

**Examples**

```
## ROTS-statistic for the Affymetrix spike-in data.
rots.out <- ROTS(data = affySpikeIn, groups = c(rep(0,5), rep(1,5)),
  B = 100, K = 500 , seed = 1234)
## Summary of the ROTS results.
rots.summary <- summary(rots.out, fdr = 0.05)
```

---

summary.ROTS

*Summary of a ROTS object*

---

**Description**

Summarizes the differential testing results from ROTS package.

**Usage**

```
## S3 method for class 'ROTS'
summary(object, fdr=NULL, num.genes=NULL, verbose=TRUE, ...)
```

**Arguments**

object	a ROTS object created from differential expression testing run by ROTS.
fdr	selected cutoff for FDR value.
num.genes	selected cutoff number for number of differentially detected features.
verbose	If TRUE (default), summary function will print out 10 first detections which fulfill the cutoff criteria.
...	other arguments passed to the summary function.

**Details**

This function returns the summary information (including row number, test-statistic, pvalue and FDR value) for the selected features.

**Value**

Returns a matrix where the rows are the selected features and columns are the Row number, ROTS-statistic, pvalue and FDR.

**Author(s)**

Fatemeh Seyednasrollah, Tomi Suomi, Laura L. Elo  
Maintainer: Fatemeh Seyednasrollah <fatsey@utu.fi>

**References**

L. L. Elo, S. Filen, R. Lahesmaa and T. Aittokallio: Reproducibility-optimized test statistic for ranking genes in microarray studies. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 5: 423–431, 2008.

**See Also**

[affySpikeIn](#)

**Examples**

```
## ROTS-statistic for the Affymetrix spike-in data.  
rots.out <- ROTS(data = affySpikeIn, groups = c(rep(0,5), rep(1,5)),  
  B = 100, K = 500 , seed = 1234)  
## Summary of the ROTS results.  
rots.summary <- summary(rots.out, fdr = 0.05)
```

---

upsSpikeIn

*Protein expression data from the CPTAC Technology Assessment  
(Study 6)*

---

**Description**

The data contains preprocessed expression values from CPTAC Technology Assessment, where equimolar concentrations of 48 human proteins (Sigma UPS 1) are spiked in five different concentrations into yeast proteome background and then processed. The example data contains only two groups from the study 6.

**Usage**

```
upsSpikeIn
```

**Format**

upsSpikeIn A numeric matrix of protein expressions.

# Index

## \*Topic **datasets**

affySpikeIn, [1](#)

upsSpikeIn, [6](#)

## \*Topic **math**

plot.ROTS, [2](#)

ROTS, [3](#)

summary.ROTS, [5](#)

affySpikeIn, [1](#), [3](#), [5](#), [6](#)

plot.ROTS, [2](#)

print.ROTS (ROTS), [3](#)

ROTS, [3](#)

ROTS-package (ROTS), [3](#)

spikedgene (affySpikeIn), [1](#)

summary.ROTS, [5](#)

upsSpikeIn, [6](#)