Package 'ToPASeq'

March 30, 2021

Version 1.24.0

Date 2019-03-19

Title Topology-based pathway analysis of RNA-seq data

Author Ivana Ihnatova, Eva Budinska, Ludwig Geistlinger

Maintainer Ivana Ihnatova <ihnatova@iba.muni.cz>

Description Implementation of methods for topology-based pathway analysis of RNA-seq data. This includes Topological Analysis of Pathway Phenotype Association (TAPPA; Gao and Wang, 2007), PathWay Enrichment Analysis (PWEA; Hung et al., 2010), and the Pathway Regulation Score (PRS; Ibrahim et al., 2012).

Depends R(>= 3.5.0), graphite

Imports Rcpp, graph, methods, Biobase, RBGL, SummarizedExperiment, gRbase, limma, corpcor

Suggests BiocStyle, airway, knitr, rmarkdown, DESeq2, DESeq, edgeR, plotrix, breastCancerVDX, EnrichmentBrowser

LinkingTo Rcpp

LazyData yes

License AGPL-3

biocViews ImmunoOncology, GeneExpression, RNASeq, DifferentialExpression, GraphAndNetwork, Pathways, NetworkEnrichment, Visualization

VignetteBuilder knitr

RoxygenNote 6.1.1

Encoding UTF-8

git_url https://git.bioconductor.org/packages/ToPASeq

git_branch RELEASE_3_12

git_last_commit c5c4a49

git_last_commit_date 2020-10-27

Date/Publication 2021-03-29

R topics documented:

`CePa` *Centrlity-based Pathway entrichment (CePa)*

Description

The function runs CePa method on microarray or RNA-Seq data. The implementation includes the identification of differentially expressed genes and transformation of pathways' topologies to an appropriate form. Only the ORA version of the CePa method is implemented and covers centralities: equal-weight, in-degree, out-degree, in-reach, out-reach and betweenness.

Usage

```
CePa(x, group, pathways, type, which = "proteins", edgeType = NULL,
  preparePaths = TRUE, norm.method = NULL, test.method = NULL,
  p.th = 0.05, logFC.th = 2, nperm = 1000, both.directions = TRUE,
  maxNodes = 150, minEdges = 0, commonTh = 2, filterSPIA = FALSE,
  convertTo = "none", convertBy = NULL)
```

Arguments

`x`	An `ExpressionSet` object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
`group`	Name or number of the phenoData column or a character vector or factor that contains required class assigments
`pathways`	A list of pathways in a form from `graphite` package or created by `preparePathways()`
`type`	Type of the input data, `"MA"` for microarray and `"RNASeq"` for RNA-Seq
`which`	Character, which type of nodes is preserved in a pathway. Possible values are `"proteins"`,`"metabolites"`,`"mixed"`
`edgeType`	Character, which type of edges is preserved in a pathway. If `NULL`, all edges are kept.
`preparePaths`	Logical, by default the pathways are transformed with `preparePathways()`. Use `FALSE`, if you have done this transformation separately
`norm.method`	Character, the method to normalize RNAseq data. If `NULL` then vst-normalization is performed. Possible values are: `"edgeR"`,`"vst"`,`"rLog"`,`"none"`

test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR
p.th	Numeric, threshold for p-values of tests for differential expression of genes. Use 1 if you don't want any threshold to be applied
logFC.th	Numeric, threshold for log fold-change of a gene to identify the gene as differentially expressed. Use negative if you don't want any threshold to be applied
nperm	Numeric, number of permutations
both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy	Arguments for the preparePathways()

Value

A list:

res	A matrix, each row refers to one pathway, each column to one centrality and the value is a p-value.
topo.sig	A list of weights for genes (nodes) in individual pathways
degtest	A numeric vector of gene-level differential expression statistics of all genes in the dataset

Author(s)

Ivana Ihnatova

References

Gu Z., Liu J., Cao K., Zhang J., Wang J.: Centrality-based pathway enrichment: a systematic approach for finding significant pathways dominated by key genes. BMC Systems Biology 2012, 6:56

Examples

```
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

CePa(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL")
}
```

---

clipper *clipper*

Description

clipper is a method for topological gene set analysis. It implements a two-step empirical approach based on the exploitation of graph decomposition into a junction tree to reconstruct the most relevant signal path. In the first step clipper selects significant pathways according to statistical tests on the means and the concentration matrices of the graphs derived from pathway topologies. Then, it "clips" the whole pathway identifying the signal paths having the greatest association with a specific phenotype.

Usage

```
clipper(x, group, pathways, type, which = "proteins", edgeType = NULL,
  preparePaths = TRUE, norm.method = NULL, test.method = NULL,
  method = "mean", testCliques = FALSE, nperm = 1000,
  alphaV = 0.05, both.directions = TRUE, maxNodes = 150,
  minEdges = 0, commonTh = 2, filterSPIA = FALSE,
  convertTo = "none", convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"
test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR This analysis is needed only for the visualization.
method	Character, "mean" or "var", the kind of test to perform on the cliques
testCliques	Logical, if TRUE then the test is applied also on the cliques of the each pathway. It is a very time consuming calculation, especially for many or big pathways
nperm	Number of permutations, if 0 then asymptotic distribution is used. May not be valid when shrinked estimator is used.
alphaV	Numeric, the threshold for variance test. The calculation of mean test depends on the result of variance test.
both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy	Arguments for the preparePathways()

Value

A list:

res	A list. First slot is a data frame containing p-values and q-values of mean and variance tests on pathways. The second slot is a list containing data.frames of the most affected paths in each pathway. The columns of the data frames contain: 1 - Index of the starting clique 2 - Index of the ending clique 3 - Index of the clique where the maximum value is reached 4 - length of the path 5 - maximum score of the path 6 - average score along the path 7 - percentage of path activation 8 - impact of the path on the entire pathway 9 - clique involved and significant 10 - clique forming the path 11 - genes forming the significant cliques 12 - genes forming the path

topo.sig	if testCliques=TRUE, a list where each slot contains the pvalues and a list of cliques in one pathway. NULL otherwise
degtest	A data.frame of gene-level differential expression statistics

Author(s)

Ivana Ihnatova

References

Martini P, Sales G, Massa MS, Chiogna M, Romualdi C. Along signal paths: an empirical gene set approach exploiting pathway topology. Nucleic Acids Res. 2013 Jan 7;41(1):e19. doi: 10.1093/nar/gks866. Epub 2012 Sep 21. PubMed PMID: 23002139; PubMed Central PMCID: PMC3592432.

Examples

```
## Not run:
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

clipper(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL", nperm=10)
}

## End(Not run)
```

---

convertIdentifiersByVector

Convert pathway identifiers

Description

This function converts identifiers of nodes in a pathway. It uses the user specified named vector for the conversion.

Usage

```
convertIdentifiersByVector(pathway, conv.table)
```

Arguments

pathway	An object of class Pathway
conv.table	A data.frame, in which the first column contains the type and the identifiers present in the pathway separeated by : and the second column contains the new identifiers and the third columns contains the types of the new identifiers

Value

A Pathway with new identifiers of the nodes

Author(s)

Ivana Ihnatova

See Also

Pathway-class

Examples

```
g<-pathways("hsapiens","kegg")
ng<-sapply(g, function(x) length(nodes(x,"mixed")))
g<-g[[which.min(ng)]]
conv<-data.frame(orig=nodes(g,"mixed"), new=LETTERS[seq_len(min(ng))],newtype=rep("LETTERS",min(ng)))
gc<-convertIdentifiersByVector(g, conv.table = conv)@protEdges
```

DEGraph *Differential Expression of Graph (DEGraph)*

Description

DEGraph implements recent hypothesis testing methods which directly assess whether a particular gene network is differentially expressed between two conditions. In employs Graph Laplacian, Fourier transformation and multivariate T2-statistic

Usage

```
DEGraph(x, group, pathways, type, which = "proteins", edgeType = NULL,
  preparePaths = TRUE, norm.method = NULL, test.method = NULL,
  overall = "biggest", useInteractionSigns = TRUE, EdgeAttrs = NULL,
  both.directions = TRUE, maxNodes = 150, minEdges = 0,
  commonTh = 2, filterSPIA = FALSE, convertTo = "none",
  convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"

`test.method` Character, the method for differentiall expression analysis of RNAseq data. If `NULL` then `"voomlimma"` is used. Possible values are: `"DESeq2"`,`"voomlimma"`,`"vstlimma"`,`"edgeR`
This analysis is needed only for the visualization.

`overall` Character, how should the overall p-value for a pathway be calculated. The possible values are: "mean", "min", "biggest". "biggest" returns the p-value of the biggest connected component.

`useInteractionSigns`
Logical, should types of interaction be included in the analysis?

`EdgeAttrs` A list containing two data.frames. See `edgeData` for the details. The interactions are assigned signs according to the `beta` column of the second data.frame.

`both.directions`, `maxNodes`, `minEdges`, `commonTh`, `filterSPIA`, `convertTo`, `convertBy`
Arguments for the `preparePathways()`

Value

A list:

`res` Results from analysis of individual pathways. The first column refers to the overall p-value for a pathway. Then groups of four columns follows. One group refers to one connected component and contains a pair of p-values (without and with Fourier transformation), graph and number of Fourier componets used in the test. The number of groups is equal to the highest number of components in analysed pathways. Components are sorted in the decreasing order of their nodes number.

`topo.sig` `NULL`, present for the compatibility with outputs from other methods

`degtest` A data.frame of gene-level statistics of all genes in the dataset

A list:

Author(s)

Ivana Ihnatova

References

L. Jacob, P. Neuvial, and S. Dudoit. Gains in power from structured two-sample tests of means on graphs. Technical Report arXiv:q-bio/1009.5173v1, arXiv, 2010.

Examples

```
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

DEGraph(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL")
}
```

prs *Pathway regulation score (PRS)*

Description

This function implements the PRS method to analyze pathway enrichment of gene expression data. For PRS, a gene weight correspond to the number of downstream differentially expressed genes.

Usage

```
prs(de, all, pwys, nperm = 1000)

prsWeights(pwy, de, all)
```

Arguments

de	A named numeric vector containing log2 fold-changes of the differentially expressed genes. Recommended names are Entrez gene IDs.
all	A character vector with the gene IDs in the reference set. If the data was obtained from a gene expression experiment, this set will contain all genes measured in the experiment. This vector should contain *all* names of the de argument.
pwys	A linkS4class{PathwayList} containing the pathways that should be analyzed for enrichment.
nperm	Integer. Number of permutations.
pwy	A linkS4class{Pathway} for which the weights should be computed.

Value

A data.frame with normalized score and p-value for each pathway analyzed.

Author(s)

Ivana Ihnatova

References

Ibrahim et al. (2012) A topology-based score for pathway enrichment. J Comput Biol, 19(5):563-73.

See Also

pathways

Examples

```
# pathways
library(graphite)
pwys <- pathways("hsapiens","kegg")[1:10]

# expression data
```

```
all <- nodes(pwys[[1]])
nds <- sample(all, 30)
de <- setNames(rnorm(30), nds)

# executing PRS
prsWeights(pwys[[1]], de, all)
prs(de, all, pwys, nperm=100)
```

PRS_wrapper *Pathway Regulation Score (PRS)*

Description

A function runs PRS method on a gene expression data matrix or count matrix and vector dividing samples into two groups and a set of pathways from graphite package. The PRS method (please see Reference for the details) was adapted to graphite's graphs where each node is represented only by one gene.

Usage

```
PRS_wrapper(x, group, pathways, type, which = "proteins",
  edgeType = NULL, preparePaths = TRUE, norm.method = NULL,
  test.method = NULL, p.th = 0.05, logFC.th = 2, nperm = 1000,
  both.directions = TRUE, maxNodes = 150, minEdges = 0,
  commonTh = 2, filterSPIA = FALSE, convertTo = "none",
  convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"
test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR This analysis is needed only for the visualization.
p.th	Numeric, threshold for p-values of tests for differential expression of genes. Use 1 if you don't want any threshold to be applied

`logFC.th` Numeric, threshold for log fold-change of a gene to identify the gene as differentially expressed. Use negative if you don't want any threshold to be applied

`nperm` Numeric, number of permutations

`both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy` Arguments for the `preparePathways()`

Value

A list:

`res` A data frame with normalized score, p-value and FDR-adjusted p-value for each pathway

`topo.sig` A list with log fold-changes and number of downstream differentially expressed nodes for nodes of individual pathways

`degtest` A named vector of statistics from testing the differential expression of genes

Author(s)

Ivana Ihnatova

References

Maysson Al-Haj Ibrahim, Sabah Jassim, Michael Anthony Cawthorne, and Kenneth Langlands. A Topology-Based Score for Pathway Enrichment, Journal of Computational Biology. May 2012, 19(5): 563-573

Examples

```
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

PRS_wrapper(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL", logFC.th=-1, r
}
```

PWEA *PathWay Enrichment Analysis (PWEA)*

Description

The function runs PWEA method (please see References for the details) on gene expression data matrix, vector specifing to which group a sample belongs and a list of pathway graphs. Briefly, it is a weighted GSEA-like method. The weightes are based on the distance and Pearson's correlation between genes in a pathway.

Usage

```
PWEA(x, group, pathways, type, which = "proteins", edgeType = NULL,
  preparePaths = TRUE, norm.method = NULL, test.method = NULL,
  tif = NULL, alpha = 0.05, nperm = 1000, ncores = 1,
  both.directions = TRUE, maxNodes = 150, minEdges = 0,
  commonTh = 2, filterSPIA = FALSE, convertTo = "none",
  convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"
test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR This analysis is needed only for the visualization.
tif	A list of Topology Influence Factor's. One slot refers to one pathway. Use prepareTIF() to create it. It is required only if type=="DEtable"
alpha	Numeric, a theshold value used during TIF calculation
nperm	Numeric, number of permutations. Used only if x %in% c("MA","RNASeq")
ncores	Numeric, number of cores. Used only if x %in% c("MA","RNASeq"). The permutations are calculated in parallel way
both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy	Arguments for the preparePathways()

Value

A list:

res	A data frame, rows refer to pathways. It contains: Enrichment score for a pathway, p-value and p-value adjusted for multiple hypothesis testing by Benjamini-Hochberg's FDR method. NA's if less than 2 nodes are present in the data
topo.sig	A list, topology influence factors for the genes in individual pathways. NULL if less than 2 nodes are present in the data
degtest	A named vector of statistics from testing the differential expression

Author(s)

Ivana Ihnatova

References

Hung, JH., Whitfield, T. W., Yang, TH., Hu, Z., Weng, Z., DeLisi, Ch. (2010) Identification of functional modules that correlate with phenotypic difference: the influence of network topology, Genome Biology, 11:R23

Examples

```
## Not run:
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

PWEA(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL", nperm=10)
}

## End(Not run)
```

---

res *Functions to extract and display results*

Description

Functions to extract and display results

Usage

```
res(object)

## S3 method for class 'topResult'
res(object)

## S3 method for class 'topResult'
topo.sig(object)

## S3 method for class 'topResult'
degtable(object)
```

Arguments

object	an output from one of following function "SPIA","PRS","CePA","PWEA","TAPPA","TopologyGSA"
...	other arguments

Methods (by class)

- topResult: Extracts results of topology-based pathway analysis
- topResult: Extracts topological significance of genes
- topResult: Extracts results of differential expression analysis on genes

---

SPIA *Signaling Pathway Impact Analysis (SPIA)*

Description

The function runs SPIA method on microarray or RNA-Seq data. The implementatio includes the identification of differentially expressed genes and transformation of pathways' topologies to an appropriate form. The SPIA method combines two independent p-values. One p-value comes from overrepresentation analysis and the other is so called pertubation factor.

Usage

```
SPIA(x, group, pathways, type, which = "proteins", edgeType = NULL,
  preparePaths = TRUE, norm.method = NULL, test.method = NULL,
  p.th = 0.05, logFC.th = 2, nperm = 1000, combine = "fisher",
  both.directions = TRUE, maxNodes = 150, minEdges = 0,
  commonTh = 2, filterSPIA = FALSE, convertTo = "none",
  convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"
test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR This analysis is needed only for the visualization.
p.th	Numeric, threshold for p-values of tests for differential expression of genes. Use 1 if you don't want any threshold to be applied
logFC.th	Numeric, threshold for log fold-change of a gene to identify the gene as differentially expressed. Use negative if you don't want any threshold to be applied

`nperm` Numeric, number of permutations

`combine` Character, the method to combine p-values. Defaults to `"fisher"` for Fisher's method. The other possible value is `"norminv"` for the normal inversion method.

`both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy` Arguments for the `preparePathways()`

Value

A list:

`res` A matrix with columns as descibed below: pSize - Pathway size, number of genes, NDE - Number of differentially expressed genes, pNDE - P-value of the overrepresentation part of the method, tA - The observed total preturbation accumulation in the pathway, pPERT - P-value of the pertubation part of the method, p - Combined p-value (overrepresentation and pertubation), pFdr - False discovery rate adjusted p, pFWER - FWER adjusted p, Status - If a pathway was identified as Acivated or Inhibited

`topo.sig` A list of accumulated pertubation factors and log fold-changes for genes in individual pathways

`degtest` A numeric vector of gene-level differential expression statistics of all genes in the dataset

Author(s)

Ivana Ihnatova

References

Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, Kim CJ, Kusanovic JP, Romero R. A novel signaling pathway impact analysis. Bioinformatics. 2009 Jan 1;25(1):75-82.

Adi L. Tarca, Sorin Draghici, Purvesh Khatri, et. al, A Signaling Pathway Impact Analysis for Microarray Experiments, 2008, Bioinformatics, 2009, 25(1):75-82.

Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R.: A systems biology approach for pathway level analysis. Genome Research, 17, 2007. Massa MS, Chiogna M, Romualdi C. Gene set analysis exploiting the topology of a pathway. BMC System Biol. 2010 Sep 1;4:121.

Examples

```
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

SPIA(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL", logFC.th=-1)
}
```

TAPPA *Topological Analysis of Pathway Phenotype Association (TAPPA)*

Description

The functions analyses the differential expression of pathways via TAPPA method. Expression is compared between two groups of samples by Mann-Whitney test. P-values are later adjusted for multiple hypothesis testing by Benjamini-Hochberg's FDR method.

Usage

```
TAPPA(x, group, pathways, type, which = "proteins", edgeType = NULL,
  preparePaths = TRUE, norm.method = NULL, test.method = NULL,
  test = t.test, normalize = TRUE, verbose = FALSE,
  both.directions = TRUE, maxNodes = 150, minEdges = 0,
  commonTh = 2, filterSPIA = FALSE, convertTo = "none",
  convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"
test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR This analysis is needed only for the visualization.
test	Function implementing a statistical test comparing PCI scores between groups. It is employed as test(PCI~group)$p.value, where PCI is a numeric vector of the same length as group
normalize	Logical, should data be normalized?
verbose	Logical, if TRUE names of the pathways are printed as they are analysed
both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy	Arguments for the preparePathways()

Value

A list,

res	A data frame, rows refer to pathways. Columns contain: number of valid PCI-scores, median, min and max of the PCI scores for each group of samples, p-value of the test (p.val) and adjusted p-value (p.adj). If less than two nodes are present in the data, the function puts NA's in all columns.
topo.sig	NULL, it is preserved for the compatibility with other methods implemented in this package
degtest	A numeric vector of gene-level differential expression statistics

Author(s)

Ivana Ihnatova

References

Gao, S. and Wang, X. (2007) TAPPA: topological analysis of pathway phenotype association. Bioinformatics, 23, pages 3100-3102

Examples

```
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:10]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

TAPPA(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL")
}
```

TopologyGSA *Gene set analysis exploiting the topology of a pathway (TopologyGSA)*

Description

TopologyGSA method uses graphical models to test the differential expression of a pathway. It also highlights pathway components involved in the deregulation.

Usage

```
TopologyGSA(x, group, pathways, type, which = "proteins",
  edgeType = NULL, preparePaths = TRUE, norm.method = NULL,
  test.method = NULL, method = "mean", nperm = 1000, alpha = 0.05,
  testCliques = FALSE, both.directions = TRUE, maxNodes = 150,
  minEdges = 0, commonTh = 2, filterSPIA = FALSE,
  convertTo = "none", convertBy = NULL)
```

Arguments

x	An ExpressionSet object or a gene expression data matrix or count matrix, rows refer to genes, columns to samples
group	Name or number of the phenoData column or a character vector or factor that contains required class assigments
pathways	A list of pathways in a form from graphite package or created by preparePathways()
type	Type of the input data, "MA" for microarray and "RNASeq" for RNA-Seq
which	Character, which type of nodes is preserved in a pathway. Possible values are "proteins","metabolites","mixed"
edgeType	Character, which type of edges is preserved in a pathway. If NULL, all edges are kept.
preparePaths	Logical, by default the pathways are transformed with preparePathways(). Use FALSE, if you have done this transformation separately
norm.method	Character, the method to normalize RNAseq data. If NULL then vst-normalization is performed. Possible values are: "edgeR","vst","rLog","none"
test.method	Character, the method for differentiall expression analysis of RNAseq data. If NULL then "voomlimma" is used. Possible values are: "DESeq2","voomlimma","vstlimma","edgeR This analysis is needed only for the visualization.
method	Either "var" and "mean". Determine the type of test used by topologyGSA.
nperm	Numeric, number of permutations.
alpha	Numeric, threshold for statistical significance of variance test. It influences the method for the mean test
testCliques	Logical, if TRUE, then the test is also performed on individual cliques. It can be very computationally complex.
both.directions, maxNodes, minEdges, commonTh, filterSPIA, convertTo, convertBy	Arguments for the preparePathways()

Details

The method requires a Directed Acyclic Graph (DAG). Therefore if a pathway contain also undirected or bidirected edges and error is thrown.

The user can further specify for the mean test:

1. **perms** number of permutations of the test,
2. **paired** logical, if TRUE Hotelling test for paired samples is calculated and the test on the variances is not performed

Or for the variance test:

1. **variance** logical, if TRUE the estimates of the covariance matrices are included in the result.
2. **s1** First group covariance matrix estimation.
3. **s2** Second group covariance matrix estimation.

Value

A list

res	a list with one entry for each successfully analyzed pathway
topo.sig	if testCliques=TRUE, a list where each slot contains the pvalues and a list of cliques in one pathway. NULL otherwise
degtest	A numeric vector of gene-level differential expression statistics

Author(s)

Ivana Ihnatova

References

Massa MS, Chiogna M, Romualdi C. Gene set analysis exploiting the topology of a pathway. BMC System Biol. 2010 Sep 1;4:121.

Examples

```
## Not run:
if (require(breastCancerVDX)) {
data("vdx")
pathways<-pathways("hsapiens","biocarta")[1:3]
MAdata<-Biobase::exprs(vdx)[,1:10]
rownames(MAdata)<-Biobase::fData(vdx)[,"Gene.symbol"]
MAdata<-MAdata[!duplicated(rownames(MAdata)),]

TopologyGSA(MAdata, Biobase::pData(vdx)[,"er"][1:10], pathways, type="MA", convertTo="SYMBOL", nperm=10)
}

## End(Not run)
```

Index