Package: netZooR (via r-universe)

May 30, 2026

Type Package

Title A menagerie of methods for the inference and analysis of gene regulatory networks

Version 1.16.0

Date 2025-07-31

Description netZooR unifies the implementations of several Network Zoo methods (netzoo, netzoo.github.io) into a single package by creating interfaces between network inference and network analysis methods. Currently, the package has 3 methods for network inference including PANDA and its optimized implementation OTTER (network reconstruction using mutliple lines of biological evidence), LIONESS (single-sample network inference), and EGRET (genotype-specific networks). Network analysis methods include CONDOR (community detection), ALPACA (differential community detection), CRANE (significance estimation of differential modules), MONSTER (estimation of network transition states). In addition, YARN allows to process gene expresssion data for tissue-specific analyses and SAMBAR infers missing mutation data based on pathway information.

Depends R (>= 4.2.0), igraph, reticulate, pandaR, Biobase,

Remotes stan-dev/cmdstanr, jnpaulson/pandaR, cytoscape/RCy3

biocViews GeneExpression, GeneRegulation, GraphAndNetwork, Microarray, Network, NetworkInference, Transcription

Imports cmdstanr, AnnotationDbi, assertthat, biomaRt, cmdstanr, corpcor, data.table, doParallel, downloader, dplyr, edgeR, foreach, GeneNet, ggdendro, ggplot2, GO.db, GOstats, gplots, graphics, grid, limma, loo, MASS, Matrix, matrixcalc, matrixStats, matrixTests, methods, nnet, org.Hs.eg.db, parallel, penalized, preprocessCore, quantro, rARPACK, RColorBrewer, RCy3, readr, reshape, reshape2, stats, STRINGdb, tidyr, utils, vegan, viridisLite,

License GPL-3

Encoding UTF-8

LazyData false

Suggests dorothea, knitr, pkgdown, rmarkdown, testthat (>= 2.1.0),

Additional_repositories https://stan-dev.r-universe.dev

VignetteEngine knitr

VignetteBuilder knitr

RoxygenNote 7.3.2

BugReports https://github.com/netZoo/netZooR/issues

URL https://github.com/netZoo/netZooR, https://netzoo.github.io/

PackageStatus Deprecated

Config/pak/sysreqs cmake libglpk-dev make libbz2-dev libicu-dev liblzma-dev libpng-dev libuv1-dev libxml2-dev libssl-dev python3 libx11-dev xz-utils libzmq3-dev zlib1g-dev

Repository https://bioc-release.r-universe.dev

Date/Publication 2026-04-28 13:06:30 UTC

RemoteUrl https://github.com/bioc/netZooR

RemoteRef RELEASE_3_23

RemoteSha b26220b66c9754b2cc16294b90106e065cbafbad

Contents

adj2el *Convert a bipartite adjacency matrix to an edgelist*

Description

Convert a bipartite adjacency matrix to an edgelist

Usage

```
adj2el(adj)
```

Arguments

adj	An adjacency matrix, with rows as TFs and columns as genes.

Value

An edge list dataframe with three columns. First column is TF name, second column is gene name, and third column is edge weight.

adj2regulon *Convert bipartite adjacency to regulon*

Description

Convert bipartite adjacency to regulon

Usage

```
adj2regulon(adj)
```

Arguments

adj	An adjacency matrix, with rows as TFs and columns as genes.

Value

A VIPER required regulon object.

adjMatToElist *converts adjacency matrix to edge list*

Description

converts adjacency matrix to edge list

Usage

```
adjMatToElist(adj_mat)
```

Arguments

adj_mat	adjacency matrix

Value

edge list

alpaca *Main ALPACA function*

Description

This function compares two networks and finds the sets of nodes that best characterize the change in modular structure

Usage

```
alpaca(net.table, file.stem, verbose = FALSE)
```

Arguments

net.table	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.
file.stem	The folder location and title under which all results will be stored.
verbose	Indicates whether the full differential modularity matrix should also be written to a file. Defaults to FALSE. modularity

Value

List where first element is the membership vector and second element is the contribution score of each node to its module's total differential modularity

Examples

```
example_path <- system.file("extdata", "Example_2comm.txt",
package = "netZooR", mustWork = TRUE)
simp.mat <- read.table(example_path,header=TRUE)
simp.alp <- alpaca(simp.mat,NULL,verbose=FALSE)
```

---

alpacaCommunityStructureRotation

Comparing node community membership between two networks

Description

This function uses the pseudo-inverse to find the optimal linear transformation mapping the community structures of two networks, then ranks nodes in the network by how much they deviate from the linear mapping.

Usage

```
alpacaCommunityStructureRotation(net1.memb, net2.memb)
```

Arguments

net1.memb	The community membership for Network 1.
net2.memb	The community membership for Network 2.

Value

A ranked list of nodes.

Examples

```
a <- 1 #place holder
```

---

alpacaComputeDifferentialScoreFromDWBM

Compute Differential modularity score from differential modularity matrix

---

Description

This functions takes the precomputed differential modularity matrix and the genLouvain membership to compute the differential modularity score.

Usage

```
alpacaComputeDifferentialScoreFromDWBM(dwbm, louv.memb)
```

Arguments

dwbm	differential modularity matrix
louv.memb	louvain community membership

Value

Vector of differntial modularity score

---

alpacaComputeDWBMmatmScale

Differential modularity matrix

Description

This function computes the differential modularity matrix for weighted bipartite networks. The community structure of the healthy network is rescaled by the ratio of m (the total edge weight) of each network.

Usage

```
alpacaComputeDWBMmatmScale(edge.mat, ctrl.memb)
```

Arguments

edge.mat	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.
ctrl.memb	The community membership for the control (healthy) network.

Value

The differential modularity matrix, with rows representing "from" nodes and columns representing "to" nodes.

Examples

```
a <- 1 # place holder
```

alpacaComputeWBMmat *Compute modularity matrix for weighted bipartite network*

Description

This function computes the modularity matrix for a weighted bipartite network.

Usage

```
alpacaComputeWBMmat(edge.mat)
```

Arguments

edge.mat	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the network of interest.

Value

Modularity matrix with rows representing TFs ("from" nodes) and columns repesenting targets ("to" nodes)

Examples

```
a <- 1 # example place holder
```

alpacaCrane *Find the robust nodes in ALPACA community using CRANE*

Description

Find the robust nodes in ALPACA community using CRANE

Usage

```
alpacaCrane(input, alp, alpha = 0.1, beta = 0, iteration = 30, isParallel = F)
```

Arguments

input	same input for alpaca: first column TF, second column Genes, third column edge weights from baseline condition, fourth column edge weights from disease condition.
alp	alpca object in list format (output from alpaca package)
alpha	alpha paramter perturbs each edge weights

beta	beta parameter perturbs the strength of each node. Set this to 0 if you want nodes to have node strength identical to the orignal network.
iteration	Number of CRANE distributions to create. Higher value leads to better ranking but longer runtime.
isParallel	TRUE = use Multithread / FALSE = do not use Multithread

Value

list of data frames

Examples

```
## Not run:

input=cbind(nonAng,ang[,3])
alp=alpaca(input,NULL,verbose = F)
alpListObject=alpacaCrane(input, alp, isParallel = T)


## End(Not run)
```

---

alpacaDeltaZAnalysis *Edge subtraction method (CONDOR optimizaton)*

Description

Takes two networks, subtracts edges individually, and then clusters the subtracted network using CONDOR.

Usage

```
alpacaDeltaZAnalysis(net.table, file.stem)
```

Arguments

net.table	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.
file.stem	The folder location and title under which all results will be stored.

Value

List where first element is the membership vector and second element is the contribution score of each node to its community's modularity in the final edge-subtracted network

Examples

```
a <- 1 # example place holder
```

---

alpacaDeltaZAnalysisLouvain *Edge subtraction method (Louvain optimizaton)*

Description

Takes two networks, subtracts edges individually, and then clusters the subtracted network using Louvain method.

Usage

```
alpacaDeltaZAnalysisLouvain(net.table, file.stem)
```

Arguments

net.table	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.
file.stem	The folder location and title under which all results will be stored.

Value

List where first element is the membership vector and second element is the contribution score of each node to its community's modularity in the final edge-subtracted network

Examples

```
a <- 1 # example place holder
```

---

alpacaExtractTopGenes *Extract core target genes in differential modules*

Description

This function outputs the top target genes in each module, ranked by their contribution to the differential modularity of the particular module in which they belong.

Usage

```
alpacaExtractTopGenes(module.result, set.lengths)
```

Arguments

module.result	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.
set.lengths	The desired lengths of the top gene lists.

Value

List with two elements. First element is a list of the top target genes in each cluster. Second element is a vector with the names of the gene sets. The names are in the format "number_length", where number is the module number label and length is the length of the gene set.

Examples

```
example_path <- system.file("extdata", "Example_2comm.txt",
package = "netZooR", mustWork = TRUE)
simp.mat <- read.table(example_path,header=TRUE)
simp.alp <- alpaca(simp.mat,NULL,verbose=FALSE)
alpacaExtractTopGenes(simp.alp, set.lengths=c(2,2))
```

---

alpacaGenLouvain *Generalized Louvain optimization*

Description

This function implements the Louvain optimization scheme on a general symmetric matrix. First, nodes are all placed in separate communities, and merged iteratively according to which merge moves result in the greatest increase in the modularity sum. Note that nodes are iterated in the order of the input matrix (not randomly) so that all results are reproducible. Second, the final community membership is used to form a alpacaMetaNetwork whose nodes represent communities from the prevous step, and which are connected by effective edge weights. The merging process is then repeated on the alpacaMetaNetwork. These two steps are repeated until the modularity sum does not increase more than a very small tolerance factor. New

Usage

```
alpacaGenLouvain(B)
```

Arguments

B	Symmetric modularity matrix

Value

The community membership vector

Examples

```
a <- 1 # example place holder
```

`alpacaGetMember` *get the member vector from alpaca object*

Description

get the member vector from alpaca object

Usage

```
alpacaGetMember(alp, target = "all")
```

Arguments

`alp`	alpaca object
`target`	tf, gene, or all

Value

member vector

`alpacaGOtabtogenes` *The top GO term associated genes in each module*

Description

Get all the genes in the top-scoring lists which are annotated with the enriched GO terms. Only GO terms with at least 3 genes in the overlap are included.

Usage

```
alpacaGOtabtogenes(go.result, dm.top)
```

Arguments

`go.result`	The result of the GO term analysis (alpacaListToGo)
`dm.top`	The result of extracting the top genes of the differential modules (dm.top)

Value

A vector with strings representing gene lists, each element of the vector has the genes in that GO term and community pasted together with spaces in between.

Examples

```
a <- 1 # example place holder
```

alpacaGoToGenes *Map GO terms to gene symbols*

Description

This function extracts all the gene symbols associated with a GO term and its descendants. (v1)

Usage

```
alpacaGoToGenes(go.term)
```

Arguments

go.term	The GO Biological Process ID (string).

Value

A vector of all gene symbols associated with the GO term.

Examples

```
a <- 1 # example place holder
```

alpacaListToGo *GO term enrichment for a list of gene sets*

Description

GO term enrichment is run using the GOstats package, and corrected for multiple testing using the Benjamini-Hochberg method.

Usage

```
alpacaListToGo(gene.list, univ.vec, comm.nums)
```

Arguments

gene.list	A list consisting of vectors of genes; genes must be identified by their official gene symbols.
univ.vec	A vector of all gene symbols that were present in the original network. This set is used as the universe for running the hypergeometric test in GOstats.
comm.nums	A vector of names for the gene sets in the input parameter "gene.list". These are used to create the table of final results.

Value

A table whose rows represent enriched GO terms (p_adj<0.05) and columns describe useful properties, like the name of the GO term, the label of the gene set which is enriched in that GO term, the adjusted p-value and Odds Ratio.

Examples

```
a <- 1 # example place holder
```

`alpacaMetaNetwork` *Create alpacaMetaNetwork for Louvain optimization*

Description

Computes the "effective" adjacency matrix of a alpacaMetaNetwork whose nodes represent communities in the larger input matrix.

Usage

```
alpacaMetaNetwork(J, S)
```

Arguments

`J`	The modularity matrix
`S`	The community membership vector from the previous round of agglomeration.

Value

The differential modularity matrix, with rows representing "from" nodes and columns representing "to" nodes.

Examples

```
a <- 1 # example place holder
```

alpacaNodeToGene *Remove tags from gene names*

Description

In gene regulatory networks, transcription factors can act as both "from" nodes (regulators) and "to" nodes (target genes), so the network analysis functions automatically tag the two columns to differentiate them. This function removes those tags from the gene identifiers.

Usage

```
alpacaNodeToGene(x)
```

Arguments

x	Tagged node identifier

Value

Untagged node name

Examples

```
a <- 1 # example place holder
```

alpacaObjectToDfList *Converts alpaca output into list of data frames*

Description

Converts alpaca output into list of data frames

Usage

```
alpacaObjectToDfList(alp)
```

Arguments

alp	alpaca object

Value

list of data frames

alpacaRotationAnalysis

Community comparison method (CONDOR optimizaton)

Description

Takes two networks, finds community structure of each one individually using CONDOR, and then ranks the nodes that show the biggest difference in their community membership.

Usage

```
alpacaRotationAnalysis(net.table)
```

Arguments

net.table	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.

Value

Vector of nodes ordered by how much they change their community membership between the two networks.

Examples

```
a <- 1 # example place holder
```

alpacaRotationAnalysisLouvain

Community comparison method (CONDOR optimizaton)

Description

Takes two networks, finds community structure of each one individually using a generalization of the Louvain method, and then ranks the nodes that show the biggest difference in their community membership.

Usage

```
alpacaRotationAnalysisLouvain(net.table)
```

Arguments

net.table	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the control or healthy network, and the fourth column contains the edge weights for the disease network or network of interest.

Value

Vector of nodes ordered by how much they change their community membership between the two networks.

Examples

```
a <- 1 # example place holder
```

---

alpacaSimulateNetwork *Simulated networks*

Description

This function creates a pair of networks given user-defined parameters for the modular structure of the first (healthy) network and the type of added module in the second (disease) network.

Usage

```
alpacaSimulateNetwork(
  comm.sizes,
  edge.mat,
  num.module,
  size.module,
  dens.module
)
```

Arguments

comm.sizes	A two-column matrix indicating the number of "from" nodes (left column) and number of "to" nodes (right column) in each community (row).
edge.mat	A matrix indicating the number of edges from the TFs in community i (rows) to target genes in community j (columns).
num.module	The number of modules that will be added to simulate the disease network.
size.module	A two-column matrix indicating the number of "from" and "to" nodes in each new module (row) that will be added to simulate the disease network.
dens.module	A vector of length num.module, indicating the edge density of each added module.

Value

A list with two elements. The first element is a four-column edge table of the same form that is input into the differential modularity function. The second element is a list of all the new nodes in the modules that were added to create the disease network.

Examples

```
a <- 1 # example place holder
```

`alpacaTestNodeRank` *Enrichment in ranked list*

Description

This function computes the enrichment of selected nodes in a ranked list, using Wilcoxon, Kolmogorov-Smirnov, and Fisher exact tests.

Usage

```
alpacaTestNodeRank(node.ordered, true.pos)
```

Arguments

`node.ordered`	An ordered list of nodes (high-scoring to low-scoring).
`true.pos`	The selected set of nodes being tested for enrichment among the ranked list.

Value

A vector of 4 values. 1) Wilcoxon p-value, 2) KS p-value, 3) Fisher p-value, 4) Fisher odds ratio.

Examples

```
a <- 1 # example place holder
```

alpacaTidyConfig *Renumbering community membership vector*

Description

This is a helper function alpacaGenLouvain. It re-numbers the communities so that they run from 1 to N increasing through the vector.

Usage

```
alpacaTidyConfig(S)
```

Arguments

S	The community membership vector derived from the previous round of agglomeration.

Value

The renumbered membership vector.

Examples

```
a <- 1 # example place holder
```

alpacaTopEnsembltoTopSym *Translating gene identifiers to gene symbols*

Description

Takes a list of gene sets named using gene identifiers and converts them to a list of symbols given a user-defined annotation table.

Usage

```
alpacaTopEnsembltoTopSym(mod.top, annot.vec)
```

Arguments

mod.top	A list of gene sets (gene identifiers)
annot.vec	A vector of gene symbols with gene identifiers as the names of the vector, that defines the translation between annotations.

Value

A list of sets of gene symbols.

Examples

```
a <- 1 # example place holder
```

`alpacaWBMlouvain` *Generalized Louvain method for bipartite networks*

Description

This function implements a generalized form of the Louvain method for weighted bipartite networks.

Usage

```
alpacaWBMlouvain(net.frame)
```

Arguments

`net.frame`	A table of edges, with the first column representing the TFs ("from" nodes) and the second column representing the targets ("to" nodes). The third column contains the edge weights corresponding to the network of interest.

Value

List where first element is the community membership vector and second element is the contribution score of each node to its community's portion of the bipartite modularity.

Examples

```
a <- 1 # example place holder
```

`annotateFromBiomart` *Annotate your Expression Set with biomaRt*

Description

Annotate your Expression Set with biomaRt

Usage

```
annotateFromBiomart(
  obj,
  genes = featureNames(obj),
  filters = "ensembl_gene_id",
 attributes = c("ensembl_gene_id", "hgnc_symbol", "chromosome_name", "start_position",
    "end_position"),
  biomart = "ensembl",
  dataset = "hsapiens_gene_ensembl",
  ...
)
```

Arguments

obj	ExpressionSet object.
genes	Genes or rownames of the ExpressionSet.
filters	getBM filter value, see getBM help file.
attributes	getBM attributes value, see getBM help file.
biomart	BioMart database name you want to connect to. Possible database names can be retrieved with teh function listMarts.
dataset	Dataset you want to use. To see the different datasets available within a biomaRt you can e.g. do: mart = useMart('ensembl'), followed by listDatasets(mart).
...	Values for useMart, see useMart help file.

Value

ExpressionSet object with a fuller featureData.

Examples

```
data(skin)
# subsetting and changing column name just for a silly example
skin <- skin[1:10,]
colnames(fData(skin)) = paste("names",1:6)
biomart<-"ENSEMBL_MART_ENSEMBL";
genes <- sapply(strsplit(rownames(skin),split="\\."),function(i)i[1])
newskin <-annotateFromBiomart(skin,genes=genes,biomart=biomart)
head(fData(newskin)[,7:11])
```

---

bladder *Bladder RNA-seq data from the GTEx consortium*

Description

Bladder RNA-seq data from the GTEx consortium. V6 release.

Usage

```
data(bladder)
```

Format

An object of class "ExpressionSet", see ExpressionSet.

Value

ExpressionSet object

Source

GTEx Portal

References

GTEx Consortium, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235), pp.648-660. (PubMed)

Examples

```
data(bladder);
checkMisAnnotation(bladder, "GENDER");
```

BuildSubnetwork *Find the subnetwork of significant edges connecting the genes.*

Description

Find the subnetwork of significant edges connecting the genes.

Usage

```
BuildSubnetwork(
  geneSet,
  networks,
  alpha,
  hopConstraint,
  nullDistribution,
  verbose = FALSE,
  topX = NULL,
  doFDRAdjustment = TRUE,
  pValueChunks = 100,
  loadPValues = FALSE,
  pValueFile = "pvalues.RDS"
)
```

Arguments

geneSet	A character vector of genes comprising the targets of interest.
networks	A list of bipartite (PANDA-like) networks, where each network is a data frame with the following format: tf,gene,score
alpha	The significance cutoff for the statistical test.
hopConstraint	The maximum number of hops to be considered between gene pairs. Must be an even number.
nullDistribution	The null distribution, specified as a vector of values.

verbose	Whether or not to print detailed information about the run.
topX	Select the X lowest significant p-values for each gene. NULL by default.
doFDRAdjustment	Whether or not to perform FDR adjustment.
pValueChunks	The number of chunks to split when calculating the p-value. This parameter allows the edges to be split into chunks to prevent memory errors.
loadPValues	Whether p-values should be loaded from pValueFile or re-generated. Default is FALSE.
pValueFile	The file where the p-values should be saved. If NULL, they are not saved and need to be recalculated.

Value

A bipartite subnetwork in the same format as the original networks.

`CalculatePValues` *Calculate p-values for all edges in the network using a Wilcoxon two-sample test for each edge.*

Description

Calculate p-values for all edges in the network using a Wilcoxon two-sample test for each edge.

Usage

```
CalculatePValues(
  network,
  nullDistribution,
  pValueChunks = 100,
  doFDRAdjustment = TRUE,
  pValueFile = "pvalues.RDS",
  verbose = FALSE
)
```

Arguments

network	A combination of PANDA-like networks, with the following format (e.g., 3 networks), provided as a data frame: tf,gene,score1,score2,score3
nullDistribution	The null distribution, specified as a vector of values.
pValueChunks	The number of chunks to split when calculating the p-value. This parameter allows the edges to be split into chunks to prevent memory errors.
doFDRAdjustment	Whether or not to perform FDR adjustment.
pValueFile	The file where the p-values should be saved. If NULL, they are not saved and need to be recalculated.
verbose	Whether or not to print detailed information about the run.

Value

A vector of p-values, one for each edge.

`checkMisAnnotation` *Check for wrong annotation of a sample using classical MDS and control genes.*

Description

Check for wrong annotation of a sample using classical MDS and control genes.

Usage

```
checkMisAnnotation(
  obj,
  phenotype,
  controlGenes = "all",
  columnID = "chromosome_name",
  plotFlag = TRUE,
  legendPosition = NULL,
  ...
)
```

Arguments

`obj`	ExpressionSet object.
`phenotype`	phenotype column name in the phenoData slot to check.
`controlGenes`	Name of controlGenes, ie. 'Y' chromosome. Can specify 'all'.
`columnID`	Column name where controlGenes is defined in the featureData slot if other than 'all'.
`plotFlag`	TRUE/FALSE Whether to plot or not
`legendPosition`	Location for the legend.
`...`	Extra parameters for `plotCMDS` function.

Value

Plots a classical multi-dimensional scaling of the 'controlGenes'. Optionally returns co-ordinates.

Examples

```
data(bladder)
checkMisAnnotation(bladder,'GENDER',controlGenes='Y',legendPosition='topleft')
```

checkTissuesToMerge *Check tissues to merge based on gene expression profile*

Description

Check tissues to merge based on gene expression profile

Usage

```
checkTissuesToMerge(
  obj,
  majorGroups,
  minorGroups,
  filterFun = NULL,
  plotFlag = TRUE,
  ...
)
```

Arguments

obj	ExpressionSet object.
majorGroups	Column name in the phenoData slot that describes the general body region or site of the sample.
minorGroups	Column name in the phenoData slot that describes the specific body region or site of the sample.
filterFun	Filter group specific genes that might disrupt PCoA analysis.
plotFlag	TRUE/FALSE whether to plot or not
...	Parameters that can go to checkMisAnnotation

Value

CMDS Plots of the majorGroupss colored by the minorGroupss. Optional matrix of CMDS loadings for each comparison.

See Also

checkTissuesToMerge

Examples

```
data(skin)
checkTissuesToMerge(skin,'SMTS','SMTSD')
```

cobra *Run COBRA in R*

Description

Description: COBRA decomposes a (partial) gene co-expression matrix as a linear combination of covariate-specific components. It can be applied for batch correction, differential co-expression analysis controlling for variables, and to understand the impact of variables of interest to the observed co-expression.

Usage

```
cobra(X, expressionData, method = "pearson")
```

Arguments

X	: design matrix of size (n, q), n = number of samples, q = number of covariates
expressionData	: gene expression as a matrix of size (g, n), g = number of genes
method	: if pearson, the decomposition of the co-expression matrix is compouted. If pcorsh, COBRA decomposes the regularized partial co-expression
	Outputs:

Details

Inputs:

Value

psi : impact of each covariate on the eigenvalues as a matrix of size (q, n)

Q : eigenvectors corresponding to non-zero eigenvalues as a matrix of size (g, n)

D : non-zero eigenvalues as a list of length n

G : (standardized) gene expression as a matrix of size (g, n)

Examples

```
g <- 100 # number of genes
n <- 10 # number of samples
q <- 2 # number of covariates
X <- X <- cbind(rep(1, n), rbinom(n, 1, 0.5))
expressionData=matrix(rnorm(g*n, 1, 1), ncol = n, nrow = g)

# Run COBRA algorithm
cobra_output <- cobra(X, expressionData)
```

condorCluster *Main clustering function for condor.*

Description

This function performs community structure clustering using the bipartite modularity described in condorModularityMax. This function uses a standard (non-bipartite) community structure clustering of the uni-partite, weighted projection of the original bipartite graph as an initial guess for the bipartite modularity.

Usage

```
condorCluster(
  condor.object,
  cs.method = "LCS",
  project = TRUE,
  low.memory = FALSE,
  deltaQmin = "default"
)
```

Arguments

condor.object	Output of make.condor.object. This function uses condor.object$edges
cs.method	is a string to specify which unipartite community structure algorithm should be used for the seed clustering. Options are LCS (cluster_louvain), LEC (leading.eigenvector.community), FG (fastgreedy.community).
project	Provides options for initial seeding of the bipartite modularity maximization. If TRUE, the nodes in the first column of condor.object$edges are projected and clustered using cs.method. If FALSE, the complete bipartite network is clustered using the unipartite clustering methods listed in cs.method.
low.memory	If TRUE, uses condorModularityMax instead of condorMatrixModularity. This is a slower implementation of the modularity maximization, which does not store any matrices in memory. Useful on a machine with low RAM. However, runtimes are (much) longer.
deltaQmin	convergence parameter determining the minimum required increase in the modularity for each iteration. Default is min(10^(-4),1/(number of edges)), with number of edges determined by nrow(condor.object$edges). User can set this parameter by passing a numeric value to deltaQmin.

Value

condor.object with condorModularityMax output included.

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
condor.object <- condorCluster(condor.object)
```

`condorCoreEnrich` *Compare qscore distribution of a subset of nodes to all other nodes.*

Description

Compute one-sided KS and wilcox tests to determine if a subset of nodes has a stochastically larger qscore distribution.

Usage

```
condorCoreEnrich(test_nodes, q, perm = FALSE, plot.hist = FALSE, nsamp = 1000)
```

Arguments

`test_nodes`	is a list containing the subset of nodes (of one node class –blue or red–only) to be tested
`q`	is a two column data frame containing the node names in the first column and the q-scores in the second column.
`perm`	if TRUE, run permutation tests. Else, run `ks.test` and `wilcox.test` only.
`plot.hist`	if TRUE, produces two histograms of test statistics from permutation tests, one for KS and one for wilcoxon and a red dot for true labeling. Only works if perm=TRUE.
`nsamp`	Number of permutation tests to run

Value

if `perm=FALSE`, the analytical p-values from `ks.test` and `wilcox.test`

if `perm=TRUE`, the permutation p-values are provided in addition to the analytical values.

Note

`ks.test` and `wilcox.test` will throw warnings due to the presence of ties, so the p-values will be approximate. See those functions for further details.

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
condor.object <- condorCluster(condor.object)
condor.object <- condorQscore(condor.object)
q_in <- condor.object$qscores$red.qscore
out <- condorCoreEnrich(c("Alice","Mary"),q=q_in,perm=TRUE,plot.hist=TRUE)
```

`condorCreateObject` *creates condor object*

Description

creates condor object

Usage

```
condorCreateObject(elist)
```

Arguments

`elist`	edge list

Value

condor object

`condorMatrixModularity`
Iteratively maximize bipartite modularity.

Description

This function is based on the bipartite modularity as defined in "Modularity and community detection in bipartite networks" by Michael J. Barber, Phys. Rev. E 76, 066102 (2007) This function uses a slightly different implementation from the paper. It does not use the "adaptive BRIM" method for identifying the number of modules. Rather, it simply continues to iterate until the difference in modularity between iterations is less that 10^(-4). Starting from a random initial condition, this could take some time. Use `condorCluster` for quicker runtimes and likely better clustering, it initializes the blue node memberships by projecting the blue nodes into a unipartite "blue" network and then identify communities in that network using a standard unipartite community detection algorithm run on the projected network. See `condorCluster` for more details on that. This function loads the entire adjacency matrix in memory, so if your network has more than ~50,000 nodes, you may want to use `condorModularityMax`, which is slower, but does not store the matrices in memory. Or, of course, you could move to a larger machine.

Usage

```
condorMatrixModularity(
  condor.object,
  T0 = cbind(seq_len(q), rep(1, q)),
  weights = 1,
  deltaQmin = "default"
)
```

Arguments

condor.object	is a list created by createCondorObject. condor.object$edges must contain the edges in the giant connected component of a bipartite network
T0	is a two column data.frame with the initial community assignment for each "blue" node, assuming there are more reds than blues, though this is not strictly necessary. The first column contains the node name, the second column the community assignment.
weights	edgeweights for each edge in edgelist.
deltaQmin	convergence parameter determining the minimum required increase in the modularity for each iteration. Default is min(10^(-4),1/(number of edges)), with number of edges determined by nrow(condor.object$edges). User can set this parameter by passing a numeric value to deltaQmin.

Value

Qcoms data.frame with modularity of each community.

modularity modularity value after each iteration.

red.memb community membership of the red nodes

blue.memb community membership of the blue.nodes

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
#randomly assign blues to their own community
T0 <- data.frame(nodes=blues,coms=seq_len(4))
condor.object <- condorMatrixModularity(condor.object,T0=T0)
```

`condorModularityMax` *Iteratively maximize bipartite modularity.*

Description

This function is based on the bipartite modularity as defined in "Modularity and community detection in bipartite networks" by Michael J. Barber, Phys. Rev. E 76, 066102 (2007) This function uses a slightly different implementation from the paper. It does not use the "adaptive BRIM" method for identifying the number of modules. Rather, it simply continues to iterate until the difference in modularity between iterations is less that 10^(-4). Starting from a random initial condition, this could take some time. Use `condorCluster` for quicker runtimes and likely better clustering, it initializes the blue node memberships by projecting the blue nodes into a unipartite "blue" network and then identify communities in that network using a standard unipartite community detection algorithm run on the projected network. See `condorCluster` for more details that.

Usage

```
condorModularityMax(
  condor.object,
  T0 = cbind(seq_len(q), rep(1, q)),
  weights = 1,
  deltaQmin = "default"
)
```

Arguments

`condor.object`	is a list created by `createCondorObject`. `condor.object$edges` must contain the edges in the giant connected component of a bipartite network
`T0`	is a two column data.frame with the initial community assignment for each "blue" node, assuming there are more reds than blues, though this is not strictly necessary. The first column contains the node name, the second column the community assignment.
`weights`	edgeweights for each edge in `edgelist`.
`deltaQmin`	convergence parameter determining the minimum required increase in the modularity for each iteration. Default is min(10^(-4),1/(number of edges)), with number of edges determined by `nrow(condor.object$edges)`. User can set this parameter by passing a numeric value to deltaQmin.

Value

Qcoms data.frame with modularity of each community.

modularity modularity value after each iteration.

red.memb community membership of the red nodes

blue.memb community membership of the blue.nodes

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
#randomly assign blues to their own community
T0 <- data.frame(nodes=blues,coms=1)
condor.object <- condorModularityMax(condor.object,T0=T0)
```

`condorPlotCommunities` *Plot adjacency matrix with links grouped and colored by community*

Description

This function will generate the network link 'heatmap' with colored dots representing within-community links and black dots between-community links

Usage

```
condorPlotCommunities(
  condor.object,
  color_list,
  point.size = 0.01,
  xlab = "SNP",
  ylab = "Gene"
)
```

Arguments

`condor.object`	output of either `condorCluster` or `condorModularityMax`
`color_list`	vector of colors accepted by `col` inside the `plot` function. There must be as many colors as communities.
`point.size`	passed to `cex` in the `plot`
`xlab`	x axis label
`ylab`	y axis label

Value

produces a `plot` output.

Note

For the condor paper `http://arxiv.org/abs/1509.02816`, I used 35 colors from the "Tarnish" palette with "hard" clustering

References

http://tools.medialab.sciences-po.fr/iwanthue/ for a nice color generator at

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
condor.object <- condorCluster(condor.object)
condorPlotCommunities(condor.object,
color_list=c("darkgreen","darkorange"),point.size=2,
xlab="Women",ylab="Men")
```

---

condorPlotHeatmap *Plot weighted adjacency matrix with links grouped by community*

Description

This function will generate the network link 'heatmap' for a weighted network

Usage

```
condorPlotHeatmap(condor.object, main = "", xlab = "blues", ylab = "reds")
```

Arguments

condor.object	output of either condorCluster or condorModularityMax
main	plot title
xlab	x axis label
ylab	y axis label

Value

produces a plot output.

Examples

```
data(small1976)
condor.object <- createCondorObject(small1976)
condor.object <- condorCluster(condor.object, project=FALSE)
condorPlotHeatmap(condor.object)
```

condorQscore *Calculate Qscore for all nodes*

Description

Qscore is designed to calculate the fraction of the modularity contributed by each node to its community's modularity

Usage

```
condorQscore(condor.object)
```

Arguments

condor.object	output of condorCluster or condorModularityMax

Value

condor.object list has condor.object$qscores added to it. this includes two data.frames, blue.qscore and red.qscore which have the qscore for each red and blue node.

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
condor.object <- condorCluster(condor.object)
condor.object <- condorQscore(condor.object)
```

condorRun *Run CONDOR clustering*

Description

Run CONDOR clustering

Usage

```
condorRun(elist, qscore = F)
```

Arguments

elist	edge list
qscore	TRUE = output qscore / FALSE = do not output qscore

Value

condor object

---

craneBipartite *Pertrubs the bipartite network with fixed node strength*

Description

Pertrubs the bipartite network with fixed node strength

Usage

```
craneBipartite(df, alpha = 0.1, beta = 0, getAdj = F, randomStart = F)
```

Arguments

df	Adjacency Matrix or Edge list
alpha	alpha paramter perturbs each edge weights
beta	beta parameter perturbs the strength of each node. Set this to 0 if you want nodes to have node strength identical to the orignal network.
getAdj	TRUE = this will return adjacency matrix instead of edge list
randomStart	FALSE = initialize the first row with completely random edge weights.

Value

edge list

Examples

```
## Not run:

# Using Edge list as input
elist=craneBipartite(nonAng)
elist=craneBipartite(nonAng,alpha=0.3)

# Using Edge list as input and Adjcency Matrix as output
adjMatrix=craneBipartite(nonAng,alpha=0.1,getAdj=T)

# Using Edge list as input and Adjcency Matrix as output
A=elistToAdjMat(nonAng)
elist=craneBipartite(A)


## End(Not run)
```

craneUnipartite *Pertrubs the unipartite network with fixed node strength from adjacency matrix*

Description

Pertrubs the unipartite network with fixed node strength from adjacency matrix

Usage

```
craneUnipartite(A, alpha = 0.1, isSelfLoop = F)
```

Arguments

A	Adjacency Matrix
alpha	alpha paramter perturbs each edge weights
isSelfLoop	TRUE/FALSE if self loop exists. co-expression matrix will have a self-loop of 1. Thus TRUE

Value

adjacency matrix

createCondorObject *Create list amenable to analysis using* condor *package.*

Description

Converts an edge list into a list which is then an input for other functions in the condor package.

Usage

```
createCondorObject(edgelist, return.gcc = TRUE)
```

Arguments

edgelist	a data.frame with 'red' nodes in the first column and 'blue' nodes in the second column, representing links from the node in the first column to the node in the second column. There must be more unique 'red' nodes than 'blue' nodes. Optionally, a third column may be provided to create a weighted network.
return.gcc	if TRUE, returns the giant connected component

Value

G is an igraph graph object with a 'color' attribute based on the colnames of edgelist. This can be accessed via V(g)$color, which returns a vector indicating red/blue. Use V(g)$name with V(g)$color to identify red/blue node names

edges corresponding to graph G. If return.gcc=TRUE, includes only those edges in the giant connected component.

Qcoms output from `condorCluster` or `condorModularityMax`

modularity `NULL` output from `condorCluster` or `condorModularityMax`

red.memb `NULL` output from `condorCluster` or `condorModularityMax`

blue.memb `NULL` output from `condorCluster` or `condorModularityMax`

qscores `NULL` output from `condorQscore`

Examples

```
r = c(1,1,1,2,2,2,3,3,3,4,4);
b = c(1,2,3,1,2,4,2,3,4,3,4);
reds <- c("Alice","Sue","Janine","Mary")
blues <- c("Bob","John","Ed","Hank")
elist <- data.frame(red=reds[r],blue=blues[b])
condor.object <- createCondorObject(elist)
```

---

createPandaStyle *Create a Cytoscape visual style for PANDA network*

Description

This function is able to create a Cytoscape visual style for any PANDA network output.

Usage

```
createPandaStyle(style_name = "PandaStyle")
```

Arguments

`style_name`	Character string indicating the style name. Defaults to "PandaStyle"

Value

A visual style in Cytoscape Control Panel under "Style" button.

degreeAdjust — *Function to adjust the degree so that the hub nodes are not penalized in z-score transformation*

Description

Function to adjust the degree so that the hub nodes are not penalized in z-score transformation

Usage

```
degreeAdjust(A)
```

Arguments

A	Input adjacency matrix

domonster — *MONSTER quick-start with pre-made regulatory networks*

Description

This function is a wrapper to simplify usage of the monster function in the case where the pair of regulatory networks to be contrasted have already been estimated, either as panda objects, or represented as an adjacency matrix with regulators in rows and genes in columns.

Usage

```
domonster(exp_graph, control_graph, nullPerms = 1000, numMaxCores = 3, ...)
```

Arguments

exp_graph	matrix or PANDA object (generated by netzooR::panda); if matrix, should be the adjacency matrix for the network with regulators in rows and genes in columns, both named. This should be the network for the experimental (case) group.
control_graph	matrix or PANDA object (generated by netzooR::panda); if matrix, should be the adjacency matrix for the network with regulators in rows and genes in columns, both named. This should be the network for the control (reference) group.
nullPerms	numeric; defaults to 1000. Number of null permutations to perform. See monster for more details.
numMaxCores	numeric; defaults to 3. Maximum number of cores to use; will be the minimum of this number and the actual available cores. See monster for more details.
...	other arguments for monster may be passed.

Value

monster object

Examples

```
# Generating PANDA networks for demonstration:
# For the purposes of this example, first partition the pandaToyData samples, then perform panda:
pandaResult_exp <- panda(pandaToyData$motif, pandaToyData$expression[,1:25], pandaToyData$ppi)
pandaResult_control <- panda(pandaToyData$motif, pandaToyData$expression[,26:50], pandaToyData$ppi)

# function takes both panda objects and matrices, or a mixture
monster_res1 <- domonster(pandaResult_exp, pandaResult_control, numMaxCores = 1)
monster_res2 <- domonster(pandaResult_exp@regNet, pandaResult_control@regNet, numMaxCores = 1)
monster_res3 <- domonster(pandaResult_exp@regNet, pandaResult_control, numMaxCores = 1)
```

---

downloadGTEx *Download GTEx files and turn them into ExpressionSet object*

Description

Downloads the V6 GTEx release and turns it into an ExpressionSet object.

Usage

```
downloadGTEx(type = "genes", file = NULL, ...)
```

Arguments

type	Type of counts to download - default genes.
file	File path and name to automatically save the downloaded GTEx expression set. Saves as a RDS file.
...	Does nothing currently.

Value

Organized ExpressionSet set.

Examples

```
# obj <- downloadGTEx(type='genes',file='~/Desktop/gtex.rds')
```

dragon *Run DRAGON in R.*

Description

Description: Estimates a multi-omic Gaussian graphical model for two input layers of paired omic data.

Usage

```
dragon(
  layer1,
  layer2,
  pval = FALSE,
  gradient = "finite_difference",
  verbose = FALSE
)
```

Arguments

layer1	: first layer of omics data; rows: samples (order must match layer2), columns: variables
layer2	: second layer of omics data; rows: samples (order must match layer1), columns: variables.
pval	: calculate p-values for network edges. Not yet implemented in R; available in netZooPy.
gradient	: method for estimating parameters of p-value distribution, applies only if p-val == TRUE. default = "finite_difference"; other option = "exact"
verbose	: verbosity level (TRUE/FALSE)

Value

A list of model results. cov : the shrunken covariance matrix

- cov: the shrunken covariance matrix
- prec: the shrunken precision matrix
- ggm: the shrunken Gaussian graphical model; matrix of partial correlations. Self-edges (diagonal elements) are set to zero.
- lambdas: Vector of omics-specific tuning parameters (lambda1, lambda2) for layer1 and layer2
- gammas: Reparameterized tuning parameters; gamma = 1 - lambda^2
- risk_grid: Risk grid, for assessing optimization. Grid boundaries are in terms of gamma.

el2adj *Convert bipartite edge list to adjacency mat*

Description

Convert bipartite edge list to adjacency mat

Usage

```
el2adj(el)
```

Arguments

el	An edge list dataframe with three columns. First column is TF name, second column is gene name, and third column is edge weight.

Value

An adjacency matrix with rows as TFs and columns as genes.

el2regulon *Convert a bipartite edgelist to regulon*

Description

Convert a bipartite edgelist to regulon

Usage

```
el2regulon(el)
```

Arguments

el	An edge list dataframe with three columns. First column is TF name, second column is gene name, and third column is edge weight.

Value

A VIPER required regulon object

elistAddTags *Adds "_A" to first column and "_B" to second column*

Description

Adds "_A" to first column and "_B" to second column

Usage

```
elistAddTags(elist)
```

Arguments

elist	edge list

Value

edge list

elistIsEdgeOrderEqual *check if first two columns are identical*

Description

check if first two columns are identical

Usage

```
elistIsEdgeOrderEqual(elist1, elist2)
```

Arguments

elist1	edge list
elist2	edge list

Value

boolean

elistRemoveTags *undo elistAddTags*

Description

undo elistAddTags

Usage

```
elistRemoveTags(elist)
```

Arguments

elist	edge list

Value

edge list

elistSort *Sorts the edge list based on first two columns in alphabetical order*

Description

Sorts the edge list based on first two columns in alphabetical order

Usage

```
elistSort(elist)
```

Arguments

elist	edge list

Value

edge list

elistToAdjMat *Converts edge list to adjacency matrix*

Description

Converts edge list to adjacency matrix

Usage

```
elistToAdjMat(elist, isBipartite = F)
```

Arguments

elist	edge list
isBipartite	TRUE = for bipartite / FALSE = for unipartite

Value

Adjcency Matrix

exon.size *Gene length*

Description

A vector of gene lengths. This will be used to normalize the gene mutation scores by the gene's length. This example is based on hg19 gene symbols. The gene length is based on the number of non-overlapping exons. Data were downloaded and pre-processed as described in Kuijjer et al.

This data is a toy example data for SAMBAR, it contains length of Exons.

This data is a toy example data for SAMBAR, it contains gene annotations

Usage

```
data(exon.size)

data(exon.size)

data(genes)
```

Format

A integer vector of size 23459, with gene symbols as names

A list containing Exon sizes for 23459 genes

A vector containing names of 23459 genes

Value

A list of length 1

A vector of length 23459

References

Kuijjer, Marieke Lydia, et al. "Cancer subtype identification using somatic mutation data." British journal of cancer 118.11 (2018): 1492-1501.

Kuijjer, Marieke Lydia, et al. "Cancer subtype identification using somatic mutation data." British journal of cancer 118.11 (2018): 1492-1501.

`extractMatrix` *Extract the appropriate matrix*

Description

This returns the raw counts, log2-transformed raw counts, or normalized expression. If normalized = TRUE then the log paramater is ignored.

Usage

```
extractMatrix(obj, normalized = FALSE, log = TRUE)
```

Arguments

`obj`	ExpressionSet object or objrix.
`normalized`	TRUE / FALSE, use the normalized matrix or raw counts
`log`	TRUE/FALSE log2-transform.

Value

matrix

Examples

```
data(skin)
head(netZooR:::extractMatrix(skin,normalized=FALSE,log=TRUE))
head(netZooR:::extractMatrix(skin,normalized=FALSE,log=FALSE))
```

filterGenes *Filter specific genes*

Description

The main use case for this function is the removal of sex-chromosome genes. Alternatively, filter genes that are not protein-coding.

Usage

```
filterGenes(
  obj,
  labels = c("X", "Y", "MT"),
  featureName = "chromosome_name",
  keepOnly = FALSE
)
```

Arguments

obj	ExpressionSet object.
labels	Labels of genes to filter or keep, eg. X, Y, and MT
featureName	FeatureData column name, eg. chr
keepOnly	Filter or keep only the genes with those labels

Value

Filtered ExpressionSet object

Examples

```
data(skin)
filterGenes(skin,labels = c('X','Y','MT'),featureName='chromosome_name')
filterGenes(skin,labels = 'protein_coding',featureName='gene_biotype',keepOnly=TRUE)
```

filterLowGenes *Filter genes that have less than a minimum threshold CPM for a given group/tissue*

Description

Filter genes that have less than a minimum threshold CPM for a given group/tissue

Usage

```
filterLowGenes(obj, groups, threshold = 1, minSamples = NULL, ...)
```

Arguments

obj	ExpressionSet object.
groups	Vector of labels for each sample or a column name of the phenoData slot. for the ids to filter. Default is the column names.
threshold	The minimum threshold for calling presence of a gene in a sample.
minSamples	Minimum number of samples - defaults to half the minimum group size.
...	Options for cpm.

Value

Filtered ExpressionSet object

See Also

cpm function defined in the edgeR package.

Examples

```
data(skin)
filterLowGenes(skin,'SMTSD')
```

---

filterMissingGenes *Filter genes not expressed in any sample*

Description

The main use case for this function is the removal of missing genes.

Usage

```
filterMissingGenes(obj, threshold = 0)
```

Arguments

obj	ExpressionSet object.
threshold	Minimum sum of gene counts across samples – defaults to zero.

Value

Filtered ExpressionSet object

Examples

```
data(skin)
filterMissingGenes(skin)
```

`filterSamples` *Filter samples*

Description

Filter samples

Usage

```
filterSamples(obj, ids, groups = colnames(obj), keepOnly = FALSE)
```

Arguments

`obj`	ExpressionSet object.
`ids`	Names found within the groups labels corresponding to samples to be removed
`groups`	Vector of labels for each sample or a column name of the phenoData slot for the ids to filter. Default is the column names.
`keepOnly`	Filter or keep only the samples with those labels.

Value

Filtered ExpressionSet object

Examples

```
data(skin)
filterSamples(skin,ids = "Skin - Not Sun Exposed (Suprapubic)",groups="SMTSD")
filterSamples(skin,ids=c("GTEX-OHPL-0008-SM-4E3I9","GTEX-145MN-1526-SM-5SI9T"))
```

`FindConnectionsForAllHopCounts`
For all hop counts up to the maximum, find subnetworks connecting each pair of genes by exactly that number of hops. For instance, find each

Description

For all hop counts up to the maximum, find subnetworks connecting each pair of genes by exactly that number of hops. For instance, find each

Usage

```
FindConnectionsForAllHopCounts(subnetworks, verbose = FALSE)
```

Arguments

subnetworks	A list of bipartite (PANDA-like) subnetworks for each gene, containing only the significant edges meeting the hop count criteria and where each network is a data frame with the following format: tf,gene
verbose	Whether or not to print detailed information about the run.

Value

A bipartite subnetwork in the same format as the original networks.

FindSignificantEdgesForHop

Find the subnetwork of significant edges n / 2 hops away from each gene.

Description

Find the subnetwork of significant edges n / 2 hops away from each gene.

Usage

```
FindSignificantEdgesForHop(
  geneSet,
  combinedNetwork,
  hopConstraint,
  pValues,
  verbose = FALSE,
  topX = NULL
)
```

Arguments

geneSet	A character vector of genes comprising the targets of interest.
combinedNetwork	A concatenation of n PANDA-like networks with the following format: tf,gene,score_net1, score_net2, ... , score_netn
hopConstraint	The maximum number of hops to be considered for a gene.
pValues	The p-values for all edges.
verbose	Whether or not to print detailed information about the run.
topX	Select the X lowest significant p-values for each gene. NULL by default.

Value

A bipartite subnetwork in the same format as the original networks.

GenerateNullPANDADistribution

Generate a null distribution of edge scores for PANDA-like networks; that is, the set of edges where (1) the TF does not have a binding motif in the gene region, (2) the TF does not form a complex with any other TF that has a binding motif in the gene region, and (3) the genes regulated by the TF are not coexpressed with the gene in question. We obtain this by inputting an empty prior and an identity coexpression matrix.

Description

Generate a null distribution of edge scores for PANDA-like networks; that is, the set of edges where (1) the TF does not have a binding motif in the gene region, (2) the TF does not form a complex with any other TF that has a binding motif in the gene region, and (3) the genes regulated by the TF are not coexpressed with the gene in question. We obtain this by inputting an empty prior and an identity coexpression matrix.

Usage

```
GenerateNullPANDADistribution(
  ppiFile,
  motifFile,
  sampSize = 20,
  numberOfPandas = 10
)
```

Arguments

ppiFile	The location of the protein-protein interaction network between transcription factors. This should be a TSV file where the first two columns are the transcription factors and the third is whether there is a PPI between them.
motifFile	The location of the motif prior from genes to transcription factors. This should be a TSV file where the first column is the transcription factors, the second is the genes, and the third is whether the transcription factor's binding motif is in the gene promoter region.
sampSize	Number of samples to simulate
numberOfPandas	Number of null PANDA networks to generate

genes *Example of a gene list*

Description

List of cancer-associated genes to subset the mutation data to, as described in Kuijjer et al.

Usage

```
data(genes)
```

Format

A character vector of length 2352

isElist *Check if data frame is an edge list*

Description

Check if data frame is an edge list

Usage

```
isElist(df)
```

Arguments

df	some data frame

Value

Boolean

jutterDegree *CRANE Beta perturbation function. This function will add noice to the node strength sequence.*

Description

CRANE Beta perturbation function. This function will add noice to the node strength sequence.

Usage

```
jutterDegree(nodeD, beta, beta_slope = T)
```

Arguments

nodeD	Vector of node strength
beta	beta
beta_slope	TRUE=use predetermined slope to add noise / FALSE = use constant value for noise

Value

vector with new strength distribution

lioness *Compute LIONESS (Linear Interpolation to Obtain Network Estimates for Single Samples)*

Description

Compute LIONESS (Linear Interpolation to Obtain Network Estimates for Single Samples)

Usage

```
lioness(
  expr,
  motif = NULL,
  ppi = NULL,
  network.inference.method = "panda",
  ncores = 1,
  mode = "union",
  ...
)
```

Arguments

expr	A mandatory expression dataset, as a genes (rows) by samples (columns) data.frame
motif	A motif dataset, a data.frame, matrix or exprSet containing 3 columns. Each row describes an motif associated with a transcription factor (column 1) a gene (column 2) and a score (column 3) for the motif.
ppi	A Protein-Protein interaction dataset, a data.frame containing 3 columns. Each row describes a protein-protein interaction between transcription factor 1(column 1), transcription factor 2 (column 2) and a score (column 3) for the interaction.
network.inference.method	String specifying choice of network inference method. Default is "panda". Options include "pearson".
ncores	int specifying the number of cores to be used. Default is 1. (Note: constructing panda networks can be memory-intensive, and the number of cores should take into consideration available memory.)
mode	Aggregation mode between three input networks: Union (default), intersection, legacy (maps on motif network).
...	additional arguments for panda analysis

Value

A list of length N, containing objects of class "panda" corresponding to each of the N samples in the expression data set.
"regNet" is the regulatory network
"coregNet" is the coregulatory network
"coopNet" is the cooperative network

References

Kuijjer, M.L., Tung, M., Yuan, G., Quackenbush, J. and Glass, K., 2015. Estimating sample-specific regulatory networks. arXiv preprint arXiv:1505.06440. Kuijjer, M.L., Hsieh, PH., Quackenbush, J. et al. lionessR: single sample network inference in R. BMC Cancer 19, 1003 (2019). https://doi.org/10.1186/s12885-019-6235-7

Examples

```
data(pandaToyData)
lionessRes <- lioness(expr = pandaToyData$expression[,1:3], motif = pandaToyData$motif,
  ppi = pandaToyData$ppi,hamming=1,progress=FALSE)
```

lionessPy *Run python implementation of LIONESS*

Description

LIONESS(Linear Interpolation to Obtain Network Estimates for Single Samples) is a method to estimate sample-specific regulatory networks. [(LIONESS publication)]).

Usage

```
lionessPy(
  expr_file,
  motif_file = NULL,
  ppi_file = NULL,
  computing = "cpu",
  precision = "double",
  save_tmp = TRUE,
  modeProcess = "union",
  remove_missing = FALSE,
  start_sample = 1,
  end_sample = "None",
  save_single_network = FALSE,
  save_dir = "lioness_output",
  save_fmt = "npy"
)
```

Arguments

expr_file	Character string indicating the file path of expression values file, with each gene(in rows) across samples(in columns).
motif_file	An optional character string indicating the file path of a prior transcription factor binding motifs dataset. When this argument is not provided, analysis will continue with Pearson correlation matrix.
ppi_file	An optional character string indicating the file path of protein-protein interaction edge dataset. Also, this can be generated with a list of proteins of interest by sourcePPI.
computing	'cpu' uses Central Processing Unit (CPU) to run PANDA; 'gpu' use the Graphical Processing Unit (GPU) to run PANDA. The default value is "cpu".
precision	'double' computes the regulatory network in double precision (15 decimal digits); 'single' computes the regulatory network in single precision (7 decimal digits) which is fastaer, requires half the memory but less accurate. The default value is 'double'.
save_tmp	'TRUE' saves middle data like expression matrix and normalized networks; 'FALSE' deletes the middle data. The default value is 'TURE'.

modeProcess 'legacy' refers to the processing mode in netZooPy<=0.5, 'union': takes the union of all TFs and genes across priors and fills the missing genes in the priors with zeros; 'intersection': intersects the input genes and TFs across priors and removes the missing TFs/genes. Default values is 'union'.

remove_missing Only when modeProcess='legacy': remove_missing='TRUE' removes all unmatched TF and genes; remove_missing='FALSE' keeps all tf and genes. The default value is FALSE.

start_sample Numeric indicating the start sample number, The default value is 1.

end_sample Numeric indicating the end sample number. The default value is 'None' meaning no end sample, i.e. print out all samples.

save_single_network Boolean vector, "TRUE" wirtes out the single network in npy/txt/mat formats, directory and format are specifics by params "save_dir" and "save_fmt". The default value is 'FALSE'

save_dir Character string indicating the folder name of output lioness networks for each sample by defined. The default is a folder named "lioness_output" under current working directory. This paramter is valid only when save_single_network = TRUE.

save_fmt Character string indicating the format of lioness network of each sample. The dafault is "npy". The option is txt, npy, or mat. This paramter is valid only when save_single_network = TRUE.

Value

A data frame with columns representing each sample, rows representing the regulator-target pair in PANDA network generated by pandaPy. Each cell filled with the related score, representing the estimated contribution of a sample to the aggregate network.

Examples

```
# refer to the input datasets files of control in inst/extdat as example
control_expression_file_path <- system.file("extdata", "expr10_reduced.txt", package = "netZooR",
    mustWork = TRUE)
motif_file_path <- system.file("extdata", "chip_reduced.txt", package = "netZooR", mustWork = TRUE)
   ppi_file_path <- system.file("extdata", "ppi_reduced.txt", package = "netZooR", mustWork = TRUE)

# Run LIONESS algorithm

control_lioness_result <- lionessPy(expr_file = control_expression_file_path,
    motif_file = motif_file_path, ppi_file = ppi_file_path,
    modeProcess="union",start_sample=1, end_sample=1, precision="single")

unlink("lioness_output", recursive=TRUE)
```

monster *MOdeling Network State Transitions from Expression and Regulatory data (MONSTER)*

Description

This function runs the MONSTER algorithm. Biological states are characterized by distinct patterns of gene expression that reflect each phenotype's active cellular processes. Driving these phenotypes are gene regulatory networks in which transcriptions factors control when and to what degree individual genes are expressed. Phenotypic transitions, such as those that occur when disease arises from healthy tissue, are associated with changes in these networks. MONSTER is an approach to understanding these transitions. MONSTER models phenotypic-specific regulatory networks and then estimates a "transition matrix" that converts one state to another. By examining the properties of the transition matrix, we can gain insight into regulatory changes associated with phenotypic state transition. Important note: the direct regulatory network observed from gene expression is currently implemented as a regular correlation as opposed to the partial correlation described in the paper. Citation: Schlauch, Daniel, et al. "Estimating drivers of cell state transitions using gene regulatory network models." BMC systems biology 11.1 (2017): 139. https://doi.org/10.1186/s12918-017-0517-y

Usage

```
monster(
  expr,
  design,
  motif,
  nullPerms = 100,
  ni_method = "BERE",
  ni.coefficient.cutoff = NA,
  numMaxCores = 1,
  outputDir = NA,
  alphaw = 0.5,
  mode = "buildNet"
)
```

Arguments

expr	Gene Expression dataset, can be matrix or data.frame of expression values or ExpressionSet.
design	Binary vector indicating case control partition. 1 for case and 0 for control.
motif	Regulatory data.frame consisting of three columns. For each row, a transcription factor (column 1) regulates a gene (column 2) with a defined strength (column 3), usually taken to be 0 or 1
nullPerms	number of random permutations to run (default 100). Set to 0 to only calculate observed transition matrix. When mode is is 'buildNet' it randomly permutes the case and control expression samples, if mode is 'regNet' it will randomly permute the case and control networks.

ni_method — String to indicate algorithm method. Must be one of "bere","pearson","cd","lda", or "wcd". Default is "bere"

ni.coefficient.cutoff — numeric to specify a p-value cutoff at the network inference step. Default is NA, indicating inclusion of all coefficients.

numMaxCores — requires doParallel, foreach. Runs MONSTER in parallel computing environment. Set to 1 to avoid parallelization, NA will take the default parallel pool in the computer.

outputDir — character vector specifying a directory or path in which which to save MONSTER results, default is NA and results are not saved.

alphaw — A weight parameter between 0 and 1 specifying proportion of weight to give to indirect compared to direct evidence. The default is 0.5 to give an equal weight to direct and indirect evidence.

mode — A parameter telling whether to build the regulatory networks ('buildNet') or to use provided regulatory networks ('regNet'). If set to 'regNet', then the parameters motif, ni_method, ni.coefficient.cutoff, and alphaw will be set to NA. Gene regulatory networks are supplied in the 'expr' variable as a TF-by-Gene matrix, by concatenating the TF-by-Gene matrices of case and control, expr has size nTFs x 2nGenes.

Value

An object of class "monsterAnalysis" containing results

Examples

```
# Example with the network reconstruction step
data(yeast)
design <- c(rep(0,20),rep(NA,10),rep(1,20))
yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# Example with provided networks

pandaResult <- panda(pandaToyData$motif, pandaToyData$expression, pandaToyData$ppi)
case=pandaResult@regNet
nelemReg=dim(pandaResult@regNet)[1]*dim(pandaResult@regNet)[2]
nGenes=length(colnames(pandaResult@regNet))
control=matrix(rexp(nelemReg, rate=.1), ncol=nGenes)
colnames(control) = colnames(case)
rownames(control) = rownames(case)
expr = as.data.frame(cbind(control,case))
design=c(rep(0,nGenes),rep(1, nGenes))
monsterRes <- monster(expr, design, motif=NA, nullPerms=10, numMaxCores=1, mode='regNet')

# alternatively, if a gene regulatory network has already been estimated,
# see the domonster function for quick start
```

`monsterBereFull` *Bipartite Edge Reconstruction from Expression data (composite method with direct/indirect)*

Description

This function generates a complete bipartite network from gene expression data and sequence motif data. This NI method serves as a default method for inferring bipartite networks in MONSTER. Running monsterBereFull can generate these networks independently from the larger MONSTER method.

Usage

```
monsterBereFull(
  motif.data,
  expr.data,
  alpha = 0.5,
  lambda = 10,
  score = "motifincluded"
)
```

Arguments

`motif.data`	A motif dataset, a data.frame, matrix or exprSet containing 3 columns. Each row describes an motif associated with a transcription factor (column 1) a gene (column 2) and a score (column 3) for the motif.
`expr.data`	An expression dataset, as a genes (rows) by samples (columns) data.frame
`alpha`	A weight parameter specifying proportion of weight to give to indirect compared to direct evidence. See documentation.
`lambda`	if using penalized, the lambda parameter in the penalized logistic regression
`score`	String to indicate whether motif information will be readded upon completion of the algorithm

Value

An matrix or data.frame

Examples

```
data(yeast)
monsterRes <- monsterBereFull(yeast$motif, yeast$exp.cc, alpha=.5)
```

monsterCalculateTmPValues

Calculate p-values for a tranformation matrix

Description

This function calculates the significance of an observed transition matrix given a set of null transition matrices

Usage

```
monsterCalculateTmPValues(monsterObj, method = "z-score")
```

Arguments

monsterObj	monsterAnalysis Object
method	one of 'z-score' or 'non-parametric'

Value

vector of p-values for each transcription factor

Examples

```
# data(yeast)
# design <- c(rep(0,20),rep(NA,10),rep(1,20))
# yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# monsterRes <- monster(yeast$exp.cc, design, yeast$motif, nullPerms=100, numMaxCores=4)
data(monsterRes)
monsterCalculateTmPValues(monsterRes)
```

monsterCalculateTmStats

Calculate statistics for a tranformation matrix

Description

This function powers both the p-value and t-value calculations for a transformation matrix. It calculates the off-diagonal squared mass for each transition matrix, and then calculates the p-values and t-values for the observed transition matrix compared to the null transition matrices. It is used by the monsterCalculateTmPValues function and by the monsterdTFIPlot

Usage

```
monsterCalculateTmStats(monsterObj, method = "z-score")
```

Arguments

monsterObj	monsterAnalysis Object
method	one of 'z-score' or 'non-parametric'

Value

p-values, t-values, off-diagonal squared mass for the observed transition matrix, and the null off-diagonal squared mass matrix

Examples

```
# data(yeast)
# design <- c(rep(0,20),rep(NA,10),rep(1,20))
# yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# monsterRes <- monster(yeast$exp.cc, design, yeast$motif, nullPerms=100, numMaxCores=4)
data(monsterRes)
monsterCalculateTmStats(monsterRes)
```

---

monsterCheckDataType *Checks that data is something MONSTER can handle*

Description

Checks that data is something MONSTER can handle

Usage

```
monsterCheckDataType(expr)
```

Arguments

expr	Gene Expression dataset

Value

expr Gene Expression dataset in the proper form (may be the same as input)

Examples

```
expr.matrix <- matrix(rnorm(2000),ncol=20)
monsterCheckDataType(expr.matrix)
#TRUE
data(yeast)
class(yeast$exp.cc)
monsterCheckDataType(yeast$exp.cc)
#TRUE
```

`monsterdTFIPlot` *This function plots the Off diagonal mass of an observed Transition Matrix compared to a set of null TMs*

Description

This function plots the Off diagonal mass of an observed Transition Matrix compared to a set of null TMs

Usage

```
monsterdTFIPlot(
  monsterObj,
  rescale = "none",
  plot.title = NA,
  highlight.tfs = NA,
  nTFs = -1
)
```

Arguments

`monsterObj`	monsterAnalysis Object
`rescale`	string indicating whether to reorder transcription factors according to their statistical significance and to rescale the values observed to be standardized by the null distribution ('significance'), to reorder transcription factors according to the largest dTFIs ('magnitude') with the TF x axis labels proportional to their significance , or finally without ordering them ('none'). When rescale is set to 'significance', the results can be different between two MONSTER runs if the number of permutations is not large enough to sample the null, that is why it is the seed should be set prior to calling MONSTER to get reproducible results. If rescale is set to another value such as 'magnitude' or 'none', it will produce deterministic results between two identical MONSTER runs.
`plot.title`	String specifying the plot title
`highlight.tfs`	vector specifying a set of transcription factors to highlight in the plot
`nTFs`	number of TFs to plot in x axis. -1 takes all TFs.

Value

ggplot2 object for transition matrix comparing observed distribution to that estimated under the null

Examples

```
# data(yeast)
# yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# design <- c(rep(0,20),rep(NA,10),rep(1,20))
# monsterRes <- monster(yeast$exp.cc, design, yeast$motif, nullPerms=100, numMaxCores=4)#'
data(monsterRes)
monsterdTFIPlot(monsterRes)
```

monsterGetTm *monsterGetTm*

Description

acessor for the transition matrix in MONSTER object

Usage

```
monsterGetTm(x)
```

Arguments

x	an object of class "monsterAnalysis"

Value

Transition matrix

Examples

```
data(monsterRes)
tm <- monsterGetTm(monsterRes)
```

monsterHclHeatmapPlot *Transformation matrix plot*

Description

This function plots a hierachically clustered heatmap and corresponding dendrogram of a transaction matrix

Usage

```
monsterHclHeatmapPlot(monsterObj, method = "pearson")
```

Arguments

monsterObj	monsterAnalysis Object
method	distance metric for hierarchical clustering. Default is "Pearson correlation"

Value

ggplot2 object for transition matrix heatmap

Examples

```
# data(yeast)
# design <- c(rep(0,20),rep(NA,10),rep(1,20))
# yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# monsterRes <- monster(yeast$exp.cc, design, yeast$motif, nullPerms=10, numMaxCores=1)
data(monsterRes)
monsterHclHeatmapPlot(monsterRes)
```

`monsterMonsterNI` *Bipartite Edge Reconstruction from Expression data*

Description

This function generates a complete bipartite network from gene expression data and sequence motif data

Usage

```
monsterMonsterNI(
  motif.data,
  expr.data,
  verbose = FALSE,
  randomize = "none",
  method = "bere",
  ni.coefficient.cutoff = NA,
  alphaw = 1,
  regularization = "none",
  score = "motifincluded",
  cpp = FALSE
)
```

Arguments

`motif.data`	A motif dataset, a data.frame, matrix or exprSet containing 3 columns. Each row describes an motif associated with a transcription factor (column 1) a gene (column 2) and a score (column 3) for the motif.
`expr.data`	An expression dataset, as a genes (rows) by samples (columns)
`verbose`	logical to indicate printing of output for algorithm progress.
`randomize`	logical indicating randomization by genes, within genes or none
`method`	String to indicate algorithm method. Must be one of "bere","pearson","cd","lda", or "wcd". Default is "bere". Important note: the direct regulatory network observed from gene expression is currently implemented as a regular correlation as opposed to the partial correlation described in the paper (please see Schlauch et al., 2017, https://doi.org/10.1186/s12918-017-0517-y)

ni.coefficient.cutoff
: numeric to specify a p-value cutoff at the network inference step. Default is NA, indicating inclusion of all coefficients.

alphaw
: A weight parameter between 0 and 1 specifying proportion of weight to give to indirect compared to direct evidence. The default is 0.5 to give an equal weight to direct and indirect evidence.

regularization
: String parameter indicating one of "none", "L1", "L2"

score
: String to indicate whether motif information will be readded upon completion of the algorithm to give to indirect compared to direct evidence. See documentation.

cpp
: logical use C++ for maximum speed, set to false if unable to run.

Value

matrix for inferred network between TFs and genes

Examples

```
data(yeast)
cc.net <- monsterMonsterNI(yeast$motif,yeast$exp.cc)
```

---

monsterPlotMonsterAnalysis

monsterPlotMonsterAnalysis

Description

plots the sum of squares of off diagonal mass (differential TF Involvement)

Usage

```
monsterPlotMonsterAnalysis(x, ...)
```

Arguments

x
: an object of class "monsterAnalysis"

...
: further arguments passed to or from other methods.

Value

Plot of the dTFI for each TF against null distribution

Examples

```
data(yeast)
yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
design <- c(rep(1,25),rep(0,10),rep(NA,15))
#monsterRes <- monster(yeast$exp.cc, design,
#yeast$motif, nullPerms=10, numMaxCores=1)
#monsterPlotMonsterAnalysis(monsterRes)
```

`monsterPrintMonsterAnalysis` *monsterPrintMonsterAnalysis*

Description

summarizes the results of a MONSTER analysis

Usage

```
monsterPrintMonsterAnalysis(x, ...)
```

Arguments

x	an object of class "monster"
...	further arguments passed to or from other methods.

Value

Description of transition matrices in object

Examples

```
data(yeast)
yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
design <- c(rep(1,25),rep(0,10),rep(NA,15))
#monster(yeast$exp.cc,design,yeast$motif, nullPerms=10, numMaxCores=1)
```

monsterRes *MONSTER results from example cell-cycle yeast transition*

Description

This data contains the MONSTER result from analysis of Yeast Cell cycle, included in data(yeast). This result arbitrarily takes the first 20 gene expression samples in yeast$cc to be the baseline condition, and the final 20 samples to be the final condition.

Usage

```
data(monsterRes)
```

Format

MONSTER obj #' @references Schlauch, Daniel, et al. "Estimating drivers of cell state transitions using gene regulatory network models." BMC systems biology 11.1 (2017): 1-10.

monsterTransformationMatrix *Bi-partite network analysis tools*

Description

This function analyzes a bi-partite network.

Usage

```
monsterTransformationMatrix(
  network.1,
  network.2,
  by.tfs = TRUE,
  standardize = FALSE,
  remove.diagonal = TRUE,
  method = "ols"
)
```

Arguments

network.1	starting network, a genes by transcription factors data.frame with scores for the existence of edges between
network.2	final network, a genes by transcription factors data.frame with scores for the existence of edges between
by.tfs	logical indicating a transcription factor based transformation. If false, gives gene by gene transformation matrix

standardize	logical indicating whether to standardize the rows and columns
remove.diagonal	logical for returning a result containing 0s across the diagonal
method	character specifying which algorithm to use, default='ols'

Value

matrix object corresponding to transition matrix

Examples

```
data(yeast)
cc.net.1 <- monsterMonsterNI(yeast$motif,yeast$exp.cc[1:1000,1:20])
cc.net.2 <- monsterMonsterNI(yeast$motif,yeast$exp.cc[1:1000,31:50])
monsterTransformationMatrix(cc.net.1, cc.net.2)
```

---

monsterTransitionNetworkPlot

This function uses igraph to plot the transition matrix (directed graph) as a network. The edges in the network should be read as A 'positively/negatively contributes to' the targeting of B in the target state.

Description

This function uses igraph to plot the transition matrix (directed graph) as a network. The edges in the network should be read as A 'positively/negatively contributes to' the targeting of B in the target state.

Usage

```
monsterTransitionNetworkPlot(
  monsterObj,
  numEdges = 100,
  numTopTFs = 10,
  rescale = "significance"
)
```

Arguments

monsterObj	monsterAnalysis Object
numEdges	The number of edges to display
numTopTFs	The number of TFs to display, only when rescale='significance'
rescale	string to specify the order of edges. If set to 'significance', the TFs with the largest dTFI significance (smallest dTFI p-values) will be filtered first before plotting the edges with the largest magnitude in the transition matrix. Otherwise the filtering step will be skipped and the edges with the largest transitions will

be plotted. The plotted graph represents the top numEdges edges between the numTopTFs if rescale=='significance' and top numEdges edges otherwise. The edge weight represents the observed transition edges standardized by the null and the node size in the graph is proportional to the p-values of the dTFIs of each TF. When rescale is set to 'significance', the results can be different between two MONSTER runs if the number of permutations is not large enough to sample the null, that is why it is the seed should be set prior to calling MONSTER to get reproducible results. If rescale is set to another value such as 'none', it will produce deterministic results between two identical MONSTER runs.

Value

plot the transition matrix (directed graph) as a network.

Examples

```
# data(yeast)
# yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# design <- c(rep(0,20),rep(NA,10),rep(1,20))
# monsterRes <- monster(yeast$exp.cc, design, yeast$motif, nullPerms=100, numMaxCores=4)#'
data(monsterRes)
monsterTransitionNetworkPlot(monsterRes, rescale='significance')
monsterTransitionNetworkPlot(monsterRes, rescale='none')
```

---

monsterTransitionPCAPlot

Principal Components plot of transformation matrix

Description

This function plots the first two principal components for a transaction matrix

Usage

```
monsterTransitionPCAPlot(
  monsterObj,
  title = "PCA Plot of Transition",
  clusters = 1,
  alpha = 1
)
```

Arguments

monsterObj	a monsterAnalysis object resulting from a monster analysis
title	The title of the plot
clusters	A vector indicating the number of clusters to compute
alpha	A vector indicating the level of transparency to be plotted

Value

ggplot2 object for transition matrix PCA

Examples

```
# data(yeast)
# design <- c(rep(0,20),rep(NA,10),rep(1,20))
# yeast$exp.cc[is.na(yeast$exp.cc)] <- mean(as.matrix(yeast$exp.cc),na.rm=TRUE)
# monsterRes <- monster(yeast$exp.cc, design, yeast$motif, nullPerms=100, numMaxCores=4)#'
data(monsterRes)
# Color the nodes according to cluster membership
clusters <- kmeans(monsterGetTm(monsterRes),3)$cluster
monsterTransitionPCAPlot(monsterRes,
title="PCA Plot of Transition - Cell Cycle vs Stress Response",
clusters=clusters)
```

---

mut.ucec *Example of mutation data*

Description

Somatic mutations of Uterine Corpus Endometrial Carcinoma from The Cancer Genome Atlas. Data were downloaded and pre-processed as described in Kuijjer et al.

This data is a toy example data for SAMBAR, it contains gene annotations

Usage

```
data(mut.ucec)

data(mut.ucec)
```

Format

A table with 248 rows and 19754 columns

A binary dataframe where 1 indicates a mutation and 0 otherwise.

Value

A table of 19754 genes by 248 samples

References

Kuijjer, Marieke Lydia, et al. "Cancer subtype identification using somatic mutation data." British journal of cancer 118.11 (2018): 1492-1501.

`normalizeTissueAware` *Normalize in a tissue aware context*

Description

This function provides a wrapper to various normalization methods developed. Currently it only wraps qsmooth and quantile normalization returning a log-transformed normalized matrix. qsmooth is a normalization approach that normalizes samples in a condition aware manner.

Usage

```
normalizeTissueAware(
  obj,
  groups,
  normalizationMethod = c("qsmooth", "quantile"),
  ...
)
```

Arguments

`obj`	ExpressionSet object
`groups`	Vector of labels for each sample or a column name of the phenoData slot for the ids to filter. Default is the column names
`normalizationMethod`	Choice of 'qsmooth' or 'quantile'
`...`	Options for `limma::qsmooth` function or `limma::normalizeQuantiles`

Value

ExpressionSet object with an assayData called normalizedMatrix

Source

The function qsmooth comes from the qsmooth packages currently available on github under user 'kokrah'.

Examples

```
data(skin)
normalizeTissueAware(skin,"SMTSD")
```

`otter` *Run OTTER in R*

Description

Description: OTTER infers gene regulatory networks using TF DNA binding motif (W), TF PPI (P), and gene coexpression (C) through minimzing the following objective: min f(W) with f(W) = (1-lambda)*||WW' - P||^2 + lambda*||W'W - C||^2 + (gamma/2)*||W||^2

Usage

```
otter(W, P, C, lambda = 0.035, gamma = 0.335, Iter = 60, eta = 1e-05, bexp = 1)
```

Arguments

`W`	: TF-gene regulatory network based on TF motifs as a matrix of size (t,g), g=number of genes, t=number of TFs
`P`	: TF-TF protein interaction network as a matrix of size (t,t)
`C`	: gene coexpression as a matrix of size (g,g)
`lambda`	: tuning parameter in [0,1] (higher gives more weight to C)
`gamma`	: regularization parameter
`Iter`	: number of iterations of the algorithm
`eta`	: learning rate
`bexp`	: exponent influencing learning rate (higher means smaller)
	Outputs:

Details

Inputs:

Value

W : Predicted TF-gene complete regulatory network as an adjacency matrix of size (t,g).

Examples

```
W=matrix(rexp(100, rate=.1), ncol=10)
C=matrix(rexp(100, rate=.1), ncol=10)
P=matrix(rexp(100, rate=.1), ncol=10)

# Run OTTER algorithm
W <- otter(W, P, C)
```

pandaDiffEdges *Identify differential edges in two PANDA networks*

Description

To determine the probability that an edge is "different" between the networks, we first subtracted the z-score weight values estimated by PANDA for the two networks and then determined the value of the inverse cumulative distribution for this difference. The product of these two probabilities represents the probability than an edge is both "supported" and "different." We select edges for which this combined probability is greater than a threshold probability (default value is 0.8).

Usage

```
pandaDiffEdges(
  panda.net1,
  panda.net2,
  threshold = 0.8,
  condition_name = "cond.1"
)
```

Arguments

panda.net1	vector indicating the PANDA networks of one condition or phenotype.
panda.net2	vector indicating the PANDA networks of another compared condition orphenotype.
threshold	numerical vector indicating a threshold probability to select select edges.
condition_name	string vector indicating the condition name of net1

Value

a data.frame with five columns: tf, gene, motif, Score and defined condition name(the row with "T" in this column means this egde belongs to first condition or phenotype, "F" means this edge belongs to the second condition or phenotype)

Examples

```
# refer to four input datasets files in inst/extdat
treated_expression_file_path <- system.file("extdata", "expr4_matched.txt",
package = "netZooR", mustWork = TRUE)
control_expression_file_path <- system.file("extdata", "expr10_matched.txt",
package = "netZooR", mustWork = TRUE)
motif_file_path <- system.file("extdata", "chip_matched.txt", package = "netZooR",
  mustWork = TRUE)
ppi_file_path <- system.file("extdata", "ppi_matched.txt", package = "netZooR",
  mustWork = TRUE)
```

```
# Run PANDA for treated and control network
#treated_all_panda_result <- pandaPy(expr_file = treated_expression_file_path,
#motif_file= motif_file_path, ppi_file = ppi_file_path, modeProcess="legacy",
#  remove_missing = TRUE )
#control_all_panda_result <- pandaPy(expr_file = control_expression_file_path,
#motif_file= motif_file_path, ppi_file= ppi_file_path, modeProcess="legacy",
#  remove_missing = TRUE )

# access PANDA regulatory network
#treated_net <- treated_all_panda_result$panda
#control_net <- control_all_panda_result$panda

#merged.panda <- pandaDiffEdges(treated_net, control_net, condition_name="treated")
```

---

pandaPy *Run Python implementation PANDA in R*

Description

PANDA(Passing Attributes between Networks for Data Assimilation) is a message-passing model to reconstruct gene regulatory network, which integrates multiple sources of biological data-including protein-protein interaction data, gene expression data, and transcription factor binding motifs data to reconstruct genome-wide, condition-specific regulatory networks. [(Glass et al. 2013)]) This function is designed to run the a derived PANDA implementation in Python Library "netZooPy" netZooPy.

Usage

```
pandaPy(
  expr_file,
  motif_file = NULL,
  ppi_file = NULL,
  computing = "cpu",
  precision = "double",
  save_memory = FALSE,
  save_tmp = TRUE,
  keep_expression_matrix = FALSE,
  modeProcess = "union",
  remove_missing = FALSE,
  with_header = FALSE
)
```

Arguments

expr_file	Character string indicating the file path of expression values file, with each gene(in rows) across samples(in columns).

`motif_file` An optional character string indicating the file path of a prior transcription factor binding motifs dataset. When this argument is not provided, analysis will continue with Pearson correlation matrix.

`ppi_file` An optional character string indicating the file path of protein-protein interaction edge dataset. Also, this can be generated with a list of proteins of interest by `sourcePPI`.

`computing` 'cpu' uses Central Processing Unit (CPU) to run PANDA; 'gpu' use the Graphical Processing Unit (GPU) to run PANDA. The default value is "cpu".

`precision` 'double' computes the regulatory network in double precision (15 decimal digits); 'single' computes the regulatory network in single precision (7 decimal digits) which is fastaer, requires half the memory but less accurate. The default value is 'double'.

`save_memory` 'TRUE' removes temporary results from memory. The result network is weighted adjacency matrix of size (nTFs, nGenes); 'FALSE' keeps the temporary files in memory. The result network has 4 columns in the form gene - TF - weight in motif prior - PANDA edge. PANDA indegree/outdegree of panda network, only if save_memory = FALSE. The default value is 'FALSE'.

`save_tmp` 'TRUE' saves middle data like expression matrix and normalized networks; 'FALSE' deletes the middle data. The default value is 'TURE'.

`keep_expression_matrix` 'TRUE' keeps the input expression matrix as an attribute in the result Panda object.'FALSE' deletes the expression matrix attribute in the Panda object. The default value is 'FALSE'.

`modeProcess` 'legacy' refers to the processing mode in netZooPy<=0.5, 'union': takes the union of all TFs and genes across priors and fills the missing genes in the priors with zeros; 'intersection': intersects the input genes and TFs across priors and removes the missing TFs/genes. Default values is 'union'.

`remove_missing` Only when modeProcess='legacy': remove_missing='TRUE' removes all unmatched TF and genes; remove_missing='FALSE' keeps all tf and genes. The default value is 'FALSE'.

`with_header` Boolean to read gene expression file with a header for sample names

Value

When save_memory=FALSE(default), this function will return a list of three items: Use `$panda` to access the standard output of PANDA as data frame, which consists of four columns: "TF", "Gene", "Motif" using 0 or 1 to indicate if this edge belongs to prior motif dataset, and "Score".

Use `$indegree` to access the indegree of PANDA network as data frame, which consists of two columns: "Gene", "Score".

Use `$outdegree` to access the outdegree of PANDA network as data frame, which consists of two columns: "TF", "Score".

When save_memory=TRUE, this function will return a weigheted adjacency matirx of size (nTFs, nGenes), use `$WAMpanda` to access.

Examples

```
# take the treated TB dataset as example here.
# refer to the datasets files path in inst/extdat

treated_expression_file_path <- system.file("extdata", "expr4_matched.txt",
package = "netZooR", mustWork = TRUE)
treated_expression_file_path <- system.file("extdata", "expr4_matched.txt",
 package = "netZooR", mustWork = TRUE)
motif_file_path <- system.file("extdata", "chip_matched.txt", package = "netZooR", mustWork = TRUE)
ppi_file_path <- system.file("extdata", "ppi_matched.txt", package = "netZooR", mustWork = TRUE)


# Run PANDA for treated and control network

treated_all_panda_result <- pandaPy(expr_file = treated_expression_file_path,
motif_file = motif_file_path, ppi_file = ppi_file_path,
modeProcess="legacy", remove_missing = TRUE )

# access PANDA regulatory network
treated_net <- treated_all_panda_result$panda

# access PANDA regulatory indegree network.
indegree_net <- treated_all_panda_result$indegree

# access PANDA regulatory outdegree networks
outdegree_net <- treated_all_panda_result$outdegree
```

---

pandaToAlpaca *Use two PANDA network to generate an ALPACA result*

Description

ALPACA(ALtered Partitions Across Community Architectures) is a method for comparing two genome-scale networks derived from different phenotypic states to identify condition-specific modules. [(Padi and Quackenbush 2018)]) This function compares two networks generate by pandaPy in this package and finds the sets of nodes that best characterize the change in modular structure.

Usage

```
pandaToAlpaca(panda.net1, panda.net2, file.stem = "./alpaca", verbose = FALSE)
```

Arguments

panda.net1	data.frame indicating an complete network of one condition generated by pandaPy
panda.net2	data.frame indicating an complete network of another condition generated by pandaPy

file.stem	String indicating the folder path and prefix of result files, where all results will be stored.
verbose	Boolean vector indicating whether the full differential modularity matrix should also be written to a file. The default values is 'FALSE'.

Value

A string message showing the location of output file if file.stem is given, and a List where the first element is the membership vector and second element is the contribution score of each node to its module's total differential modularity

Examples

```
# refer to four input datasets files in inst/extdat
treated_expression_file_path <- system.file("extdata", "expr4_matched.txt",
package = "netZooR", mustWork = TRUE)
control_expression_file_path <- system.file("extdata", "expr10_matched.txt",
package = "netZooR", mustWork = TRUE)
motif_file_path <- system.file("extdata", "chip_matched.txt", package = "netZooR", mustWork = TRUE)
ppi_file_path <- system.file("extdata", "ppi_matched.txt", package = "netZooR", mustWork = TRUE)


# Run PANDA for treated and control network

treated_panda_net <- pandaPy(expr_file = treated_expression_file_path,
motif_file = motif_file_path, ppi_file = ppi_file_path,
modeProcess="legacy", remove_missing = TRUE )$panda
control_panda_net <- pandaPy(expr_file = control_expression_file_path,
motif_file = motif_file_path, ppi_file = ppi_file_path,
modeProcess="legacy", remove_missing = TRUE )$panda

# Run ALPACA
alpaca<- pandaToAlpaca(treated_panda_net, control_panda_net, "./TB", verbose=TRUE)

# Delete files.
file.remove("TB_ALPACA_ctrl_memb.txt")
file.remove("TB_ALPACA_final_memb.txt")
file.remove("TB_ALPACA_scores.txt")
file.remove("TB_DWBM.txt")
file.remove("TB_DWBM_colnames.txt")
file.remove("TB_DWBM_rownames.txt")
```

---

pandaToCondorObject *Turn PANDA network into a CONDOR object*

Description

CONDOR (COmplex Network Description Of Regulators) implements methods for clustering bi-apartite networks and estimatiing the contribution of each node to its community's modularity, [(Platig et al. 2016)]) This function uses the result of PANDA algorithm as the input dataset to run CONDOR algorithm. More about condor package and usage.

Usage

```
pandaToCondorObject(panda.net, threshold)
```

Arguments

panda.net	Data Frame indicating the result of PANDA regulatory network, created by pandaPy
threshold	Numeric vector of the customered threshold to select edges. Default value is the the midpoint between the median edge-weight of prior (3rd column "Motif" is 1.0) edges and the median edge-weight of non-prior edges (3rd column "Motif" is 0.0) in PANDA network. and the median edge-weight of non-prior edges (3rd column "Motif" is 0.0) in PANDA network.

Value

a CONDOR object, see createCondorObject.

Examples

```
# refer to three input datasets files in inst/extdat
treated_expression_file_path <- system.file("extdata", "expr4_matched.txt",
package = "netZooR", mustWork = TRUE)
motif_file_path <- system.file("extdata", "chip_matched.txt", package = "netZooR", mustWork = TRUE)
ppi_file_path <- system.file("extdata", "ppi_matched.txt", package = "netZooR", mustWork = TRUE)


# Run PANDA to construct the treated network

treated_all_panda_result <- pandaPy(expr_file = treated_expression_file_path,
motif_file= motif_file_path, ppi_file = ppi_file_path,
modeProcess="legacy", remove_missing = TRUE )

# access PANDA regulatory network
treated_net <- treated_all_panda_result$panda

# Obtain the condor.object from PANDA network
treated_condor_object <- pandaToCondorObject(treated_net, threshold = 0)

# cluster condor.object
treated_condor_object <- condorCluster(treated_condor_object, project = FALSE)

# package igraph and package viridisLite are already loaded with this package.
library(viridisLite)
treated_color_num <- max(treated_condor_object$red.memb$com)
```

```
treated_color <- viridis(treated_color_num, alpha = 1, begin = 0, end = 1,
direction = 1, option = "D")
condorPlotCommunities(treated_condor_object, color_list=treated_color,
point.size=0.04, xlab="Target", ylab="Regulator")
```

`plotCMDS` *Plot classical MDS of dataset*

Description

This function plots the MDS coordinates for the "n" features of interest. Potentially uncovering batch effects or feature relationships.

Usage

```
plotCMDS(
  obj,
  comp = 1:2,
  normalized = FALSE,
  distFun = dist,
  distMethod = "euclidian",
  n = NULL,
  samples = TRUE,
  log = TRUE,
  plotFlag = TRUE,
  ...
)
```

Arguments

`obj`	ExpressionSet object or objrix.
`comp`	Which components to display.
`normalized`	TRUE / FALSE, use the normalized matrix or raw counts.
`distFun`	Distance function, default is dist.
`distMethod`	The distance measure to be used. This must be one of "euclidean", "maximum", "manhattan", "canberra", "binary" or "minkowski". Any unambiguous substring can be given.
`n`	Number of features to make use of in calculating your distances.
`samples`	Perform on samples or genes.
`log`	TRUE/FALSE log2-transform raw counts.
`plotFlag`	TRUE/FALSE whether to plot or not.
`...`	Additional plot arguments.

Value

coordinates

Examples

```
data(skin)
res <- plotCMDS(skin,pch=21,bg=factor(pData(skin)$SMTSD))

# library(calibrate)
# textxy(X=res[,1],Y=res[,2],labs=rownames(res))
```

`plotDensity` *Density plots of columns in a matrix*

Description

Plots the density of the columns of a matrix. Wrapper for `matdensity`.

Usage

```
plotDensity(obj, groups = NULL, normalized = FALSE, legendPos = NULL, ...)
```

Arguments

`obj`	ExpressionSet object
`groups`	Vector of labels for each sample or a column name of the phenoData slot for the ids to filter. Default is the column names.
`normalized`	TRUE / FALSE, use the normalized matrix or log2-transformed raw counts
`legendPos`	Legend title position. If null, does not create legend by default.
`...`	Extra parameters for matdensity.

Value

A density plot for each column in the ExpressionSet object colored by groups

Examples

```
data(skin)
filtData <- filterLowGenes(skin,"SMTSD")
plotDensity(filtData,groups="SMTSD",legendPos="topleft")
# to remove the legend
plotDensity(filtData,groups="SMTSD")
```

plotHeatmap *Plot heatmap of most variable genes*

Description

This function plots a heatmap of the gene expressions forthe "n" features of interest.

Usage

```
plotHeatmap(obj, n = NULL, fun = stats::sd, normalized = TRUE, log = TRUE, ...)
```

Arguments

obj	ExpressionSet object or objrix.
n	Number of features to make use of in plotting heatmap.
fun	Function to sort genes by, default sd.
normalized	TRUE / FALSE, use the normalized matrix or raw counts.
log	TRUE/FALSE log2-transform raw counts.
...	Additional plot arguments for heatmap.2.

Value

coordinates

Examples

```
data(skin)
tissues <- pData(skin)$SMTSD
plotHeatmap(skin,normalized=FALSE,log=TRUE,trace="none",n=10)
# Even prettier

# library(RColorBrewer)
data(skin)
tissues <- pData(skin)$SMTSD
heatmapColColors <- RColorBrewer::brewer.pal(12,"Set3")[as.integer(factor(tissues))]
heatmapCols <- colorRampPalette(RColorBrewer::brewer.pal(9, "RdBu"))(50)
plotHeatmap(skin,normalized=FALSE,log=TRUE,trace="none",n=10,
 col = heatmapCols,ColSideColors = heatmapColColors,cexRow = 0.6,cexCol = 0.6)
```

PlotNetwork *Plot the networks, using different colors for transcription factors, genes of interest, and additional genes.*

Description

Plot the networks, using different colors for transcription factors, genes of interest, and additional genes.

Usage

```
PlotNetwork(
  network,
  genesOfInterest,
  tfColor = "blue",
  nodeSize = 1,
  edgeWidth = 0.5,
  vertexLabels = NA,
  vertexLabelSize = 0.7,
  vertexLabelOffset = 0.5,
  layoutBipartite = TRUE,
  geneColorMapping = NULL
)
```

Arguments

network	A data frame with the following format: tf,gene
genesOfInterest	Which genes of interest to highlight
tfColor	Color for the transcription factors
nodeSize	Size of node
edgeWidth	Width of edges
vertexLabels	Which vertex labels to include. By default, none are included.
vertexLabelSize	The size of label to use for the vertex, as a fraction of the default.
vertexLabelOffset	Number of pixels in the offset when plotting labels. Default is TRUE.
layoutBipartite	Whether or not to layout as a bipartite graph.
geneColorMapping	Color mapping from a set of genes to a color. The nodes and edges connected to them will be this color. If NULL, all genes and their edges will be gray. The format is a data frame, where the first column ("gene") is the name of the gene and the second ("color") is the color.

Value

A bipartite plot of the network

---

priorPp *Filter low confident edge signs in the prior network using GeneNet*

Description

Filter low confident edge signs in the prior network using GeneNet

Usage

```
priorPp(prior, expr)
```

Arguments

prior	A prior network (adjacency matrix) with rows as TFs and columns as genes.
expr	A normalized log-transformed gene expression matrix.

Value

A filtered prior network (adjacency matrix).

---

puma *PANDA using microRNA associations*

Description

This function runs the PUMA algorithm to predict a miRNA-gene regulatory network

Usage

```
puma(
  motif,
  expr = NULL,
  ppi = NULL,
  alpha = 0.1,
  mir_file,
  hamming = 0.001,
  iter = NA,
  output = c("regulatory", "coexpression", "cooperative"),
  zScale = TRUE,
  progress = FALSE,
  randomize = c("None", "within.gene", "by.gene"),
  cor.method = "pearson",
```

```
  scale.by.present = FALSE,
  edgelist = FALSE,
  remove.missing.ppi = FALSE,
  remove.missing.motif = FALSE,
  remove.missing.genes = FALSE,
  mode = "union"
)
```

Arguments

motif	A miRNA target dataset, a data.frame, matrix or exprSet containing 3 columns. Each row describes the association between a miRNA (column 1) its target gene (column 2) and a score (column 3) for the association from TargetScan or miRanda
expr	An expression dataset, as a genes (rows) by samples (columns) data.frame
ppi	This can be set to 1) NULL which will be encoded as an identity matrix between miRNAs in PUMA for now. Or 2) it can include a set of TF interactions, or 3) a mix of TFs and miRNAs.
alpha	value to be used for update variable, alpha (default=0.1)
mir_file	list of miRNA to filter the PPI matrix and prevent update of miRNA edges.
hamming	value at which to terminate the process based on hamming distance (default 10^-3)
iter	sets the maximum number of iterations PUMA can run before exiting.
output	a vector containing which networks to return. Options include "regulatory", "coregulatory", "cooperative".
zScale	Boolean to indicate use of z-scores in output. False will use [0,1] scale.
progress	Boolean to indicate printing of output for algorithm progress.
randomize	method by which to randomize gene expression matrix. Default "None". Must be one of "None", "within.gene", "by.genes". "within.gene" randomization scrambles each row of the gene expression matrix, "by.gene" scrambles gene labels.
cor.method	Correlation method, default is "pearson".
scale.by.present	Boolean to indicate scaling of correlations by percentage of positive samples.
edgelist	Boolean to indicate if edge lists instead of matrices should be returned.
remove.missing.ppi	Boolean to indicate whether miRNAs in the PPI but not in the motif data should be removed. Only when mode=='legacy'.
remove.missing.motif	Boolean to indicate whether genes targeted in the motif data but not the expression data should be removed. Only when mode=='legacy'.
remove.missing.genes	Boolean to indicate whether genes in the expression data but lacking information from the motif prior should be removed. Only when mode=='legacy'.

mode — The data alignment mode. The mode 'union' takes the union of the genes in the expression matrix and the motif and the union of TFs in the ppi and motif and fills the matrics with zeros for nonintersecting TFs and gens, 'intersection' takes the intersection of genes and TFs and removes nonintersecting sets, 'legacy' is the old behavior with version 1.19.3. #' Parameters remove.missing.ppi, remove.missingmotif, remove.missing.genes work only with mode=='legacy'.

Value

An object of class "panda" containing matrices describing networks achieved by convergence with PUMA algorithm.
"regNet" is the regulatory network
"coregNet" is the coregulatory network
"coopNet" is the cooperative network which is not updated for miRNAs

References

Kuijjer, Marieke L., et al. "PUMA: PANDA using microRNA associations." Bioinformatics 36.18 (2020): 4765-4773.

Examples

```
data(pandaToyData)
mirs = c("AHR","AR","ARID3A","ARNT","BRCA1","CEBPA","CREB1","DDIT3")
pumaRes <- puma(pandaToyData$motif,
           pandaToyData$expression,NULL,mir_file=mirs,hamming=.1,progress=TRUE)
```

---

qsmooth *Quantile shrinkage normalization*

Description

This function was modified from github user kokrah.

Usage

```
qsmooth(
  obj,
  groups,
  norm.factors = NULL,
  plot = FALSE,
  window = 0.05,
  log = TRUE
)
```

Arguments

obj	for counts use log2(raw counts + 1)), for MA use log2(raw intensities)
groups	groups to which samples belong (character vector)
norm.factors	scaling normalization factors
plot	plot weights? (default=FALSE)
window	window size for running median (a fraction of the number of rows of exprs)
log	Whether or not the data should be log transformed before normalization, TRUE = YES.

Value

Normalized expression

Source

Kwame Okrah's qsmooth R package

Examples

```
data(skin)
head(netZooR:::qsmooth(skin,groups=pData(skin)$SMTSD))
```

---

qstats *Compute quantile statistics*

Description

This function was directly borrowed from github user kokrah.

Usage

```
qstats(exprs, groups, window)
```

Arguments

exprs	for counts use log2(raw counts + 1)), for MA use log2(raw intensities)
groups	groups to which samples belong (character vector)
window	window size for running median as a fraction on the number of rows of exprs

Value

list of statistics

Source

Kwame Okrah's qsmooth R package Compute quantile statistics

RunBLOBFISH — *Given a set of genes of interest, full bipartite networks with scores (one network for each sample), a significance cutoff for statistical testing, and a hop constraint, BLOBFISH finds a subnetwork of significant edges connecting the genes.*

Description

Given a set of genes of interest, full bipartite networks with scores (one network for each sample), a significance cutoff for statistical testing, and a hop constraint, BLOBFISH finds a subnetwork of significant edges connecting the genes.

Usage

```
RunBLOBFISH(
  geneSet,
  networks,
  alpha,
  hopConstraint,
  nullDistribution,
  verbose = FALSE,
  topX = NULL,
  doFDRAdjustment = TRUE,
  pValueChunks = 100,
  loadPValues = FALSE,
  pValueFile = "pvalues.RDS"
)
```

Arguments

geneSet	A character vector of genes comprising the targets of interest.
networks	A list of bipartite (PANDA-like) networks, where each network is a data frame with the following format: tf,gene,score
alpha	The significance cutoff for the statistical test.
hopConstraint	The maximum number of hops to be considered between gene pairs. Must be an even number.
nullDistribution	The null distribution, specified as a vector of values.
verbose	Whether or not to print detailed information about the run.
topX	Select the X lowest significant p-values for each gene. NULL by default.
doFDRAdjustment	Whether or not to perform FDR adjustment.
pValueChunks	The number of chunks to split when calculating the p-value. This parameter allows the edges to be split into chunks to prevent memory errors.

loadPValues	Whether p-values should be loaded from pValueFile or re-generated. Default is FALSE.
pValueFile	The file where the p-values should be saved. If NULL, they are not saved and need to be recalculated.

Value

A bipartite subnetwork in the same format as the original networks.

`runEgret` *Run EGRET in R*

Description

Description: NOTE: Beta version. EGRET infers individual-specific gene regulatory networks using inidividual level data - a genotype vcf file (v) and QBiC binding predictions (q) - as well as population/reference level data - eQTLs (b), a motif-gene prior (m), PPI network (p), and gene expression (e). An annotation file g is also used to map TF names to their corresponding ensemble ids.

Usage

```
runEgret(b, v, q, m, e, p, g, t)
```

Arguments

b	: Data frame of eQTL data, each row containing an eQTL which exist within motif regions adjacent to the eGene, with columns TF, gene, variant position,variant chromosome, eQTL beta value.
v	: Data frame of VCF file containing SNPs of the individual in question
q	: Data frame of QBiC predictions of the effect of eQTL variants on TF binding. Each row represents an eQTL variant with a predicted negative (disruptive) effect on the binding of the TF corresponding to the motif in which the eQTL variant resides. Colums are: eQTL variant as chr[chrNum]_position, TF, adjacent eGene, QBiC binding effect size and QBiC binding effect (should be negative)
m	: Motif prior data frame. Each row represents an edge in the bipartite motif prior, with columns TF, gene and edge weight. The edge weight should be 1 or 0 based on the presence/absence of the TF motif in the promoter region of the gene.
e	: Gene expression data frame in which each row represents a gene and each column represents the expression of that gene in a sample. The first column should contain gene IDs.
p	: PPI network data frame. Each row represents an edgem with columns TF, TF and interaction weight.
g	: Data frame mapping gene names to gene ids, with columns containing the gene ID the corresponding gene name.

t : A string containing a name for the EGRET run. Output files will be labelled with this tag.

Outputs:

Details

Inputs:

Value

EGRET : Predicted genotye-specific gene regulatory network saved as tag_egret.RData

BASELINE : A Baseline (PANDA) genotype-agnostic gene regulatory network saved as tag_panda.RData

Examples

```
# Run EGRET algorithm
toy_qbic_path <- system.file("extdata", "toy_qbic.txt", package = "netZooR",
mustWork = TRUE)
toy_genotype_path <- system.file("extdata", "toy_genotype.vcf",
package = "netZooR", mustWork = TRUE)
toy_motif_path <- system.file("extdata", "toy_motif_prior.txt",
package = "netZooR", mustWork = TRUE)
toy_expr_path <- system.file("extdata", "toy_expr.txt",
package = "netZooR", mustWork = TRUE)
toy_ppi_path <- system.file("extdata", "toy_ppi_prior.txt",
package = "netZooR", mustWork = TRUE)
toy_eqtl_path <- system.file("extdata", "toy_eQTL.txt",
package = "netZooR", mustWork = TRUE)
toy_map_path <- system.file("extdata", "toy_map.txt",
package = "netZooR", mustWork = TRUE)
qbic <- read.table(file = toy_qbic_path, header = FALSE)
vcf <- read.table(toy_genotype_path, header = FALSE, sep = "\t",
stringsAsFactors = FALSE,
colClasses = c("character", "numeric", "character", "character", "character",
"character", "character", "character", "character", "character"))
motif <- read.table(toy_motif_path, sep = "\t", header = FALSE)
expr <- read.table(toy_expr_path, header = FALSE, sep = "\t", row.names = 1)
ppi <- read.table(toy_ppi_path, header = FALSE, sep = "\t")
qtl <- read.table(toy_eqtl_path, header = FALSE)
nameGeneMap <- read.table(toy_map_path, header = FALSE)
tag <- "my_toy_egret_run"

runEgret(qtl,vcf,qbic,motif,expr,ppi,nameGeneMap,tag)

file.remove("my_toy_egret_run_egret.RData")
file.remove("my_toy_egret_run_panda.RData")
file.remove("priors_my_toy_egret_run.txt")
```

sambar *Main SAMBAR function.*

Description

Main SAMBAR function.

Usage

```
sambar(
  mutdata = mut.ucec,
  esize = exon.size,
 signatureset = system.file("extdata", "h.all.v6.1.symbols.gmt", package = "netZooR",
    mustWork = TRUE),
  cangenes = genes,
  kmin = 2,
  kmax = 4
)
```

Arguments

mutdata	Mutation data in matrix format. The number of mutations should be listed for samples (rows) and genes (columns).
esize	A integer vector of gene lengths, with gene symbols as names.
signatureset	A file containing gene sets (signatures) in .gmt format. These gene sets will be used to de-sparsify the gene-level mutation scores.
cangenes	A vector of genes, for example of cancer-associated genes. This will be used to subset the gene-level mutation data to.
kmin	The minimum number of subtypes the user wants to assess. Defaults to 2.
kmax	The maximum number of subtypes the user wants to assess. Defaults to 4.

Value

A list of samples and the subtypes to which these samples are assigned, for each k.

Examples

```
data("exon.size")
data("mut.ucec")
data("genes")
sambar(mutdata=mut.ucec, esize=exon.size, signatureset=system.file("extdata",
"h.all.v6.1.symbols.gmt", package="netZooR", mustWork=TRUE),
       cangenes=genes, kmin=2, kmax=4)
```

sambarConvertgmt *Convert .gmt files into a binary matrix.*

Description

Convert .gmt files into a binary matrix.

Usage

```
sambarConvertgmt(signature, cagenes)
```

Arguments

signature	A file containing gene sets (signatures) in .gmt format. These gene sets will be used to de-sparsify the gene-level mutation scores.
cagenes	A vector of genes, for example of cancer-associated genes. This will be used to subset the gene-level mutation data to.

Value

A matrix containing gene set mutation scores.

sambarCorgenelength *Normalize gene mutation scores by gene length.*

Description

Normalize gene mutation scores by gene length.

Usage

```
sambarCorgenelength(x, cagenes, exonsize)
```

Arguments

x	Mutation data, in the format of a matrix, including the number of mutations for samples (rows) and genes (columns).
cagenes	A vector of genes, for example of cancer-associated genes. This will be used to subset the gene-level mutation data to.
exonsize	A vector of gene lengths. This will be used to normalize the gene mutation scores.

Value

Mutation rate-adjusted gene mutation scores.

`sambarDesparsify` *De-sparsify gene-level mutation scores into gene set-level mutation scores.*

Description

De-sparsify gene-level mutation scores into gene set-level mutation scores.

Usage

```
sambarDesparsify(edgx, mutratecorx)
```

Arguments

`edgx`	A binary matrix containing information on which genes belong to which gene sets. Output from the sambarConvertgmt function.
`mutratecorx`	Gene-level mutation scores corrected for the number of gene sets each gene belongs to (from sambar function).

Value

De-sparsified mutation data.

`SignificantBreadthFirstSearch` *Find all significant edges adjacent to the starting nodes, excluding the nodes specified.*

Description

Find all significant edges adjacent to the starting nodes, excluding the nodes specified.

Usage

```
SignificantBreadthFirstSearch(
  networks,
  pValues,
  startingNodes,
  nodesToExclude,
  startFromTF,
  verbose = FALSE,
  topX = NULL
)
```

Arguments

networks	A concatenation of n PANDA-like networks with the following format: tf,gene,score_net1, score_net2, ... , score_netn Edges must be specified as "tf__gene".
pValues	The p-values from the original network.
startingNodes	The list of nodes from which to start.
nodesToExclude	The list of nodes to exclude from the search.
startFromTF	Whether to start from transcription factors (TRUE) or genes (FALSE).
verbose	Whether or not to print detailed information about the run.
topX	Select the X lowest significant p-values for each gene. NULL by default.

Value

A bipartite subnetwork in the same format as the original networks.

---

skin *Skin RNA-seq data from the GTEx consortium*

Description

Skin RNA-seq data from the GTEx consortium. V6 release. Random selection of 20 skin samples. 13 of the samples are fibroblast cells, 5 Skin sun exposed, 2 sun unexposed.

Usage

```
data(skin)
```

Format

An object of class "ExpressionSet", see ExpressionSet.

Value

ExpressionSet object

Source

GTEx Portal

References

GTEx Consortium, 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235), pp.648-660. (PubMed)

Examples

```
data(skin);
checkMisAnnotation(skin,"GENDER");
```

small1976 *Pollinator-plant interactions*

Description

A dataset containing the number of interactions 34 plants and 13 pollinators. The variables are as follows:

Usage

```
data(small1976)
```

Format

A data frame with 442 rows and 3 variables

Details

- pollinator. Species name of insect pollinator
- plant. Species name of plant
- interactions. Number of visitors caught on each plant species

References

https://www.nceas.ucsb.edu/interactionweb/html/small_1976.html

sourcePPI *Source the Protein-Protein interaction in STRING database*

Description

This function uses a list of Transcription Factors (TF) of interest to source the Protein-Protein interactions (PPI) from STRING database using all types of interactions not only the physical subnetwork Important: this function produces a simple unweighted network for tutorial purposes, and does not support weighted PPI edges for the moment. For more complex PPI network modeling, consider pulling the PPI network directly from STRINGdb directly or through their R package.

Usage

```
sourcePPI(TF, STRING.version = "10", species.index, ...)
```

Arguments

TF	a data frame with one column indicating the TF of interest
STRING.version	a numeric vector indicating the STRING version. Default valuve is 10
species.index	a numeric vector indicating NCBI taxonomy identifiers
...	any ddditional arguments passed to

Value

A PPI data.frame which contains three columns: "from" and "to" indicating the direction of protein-protein interaction, and "score" indicating the interaction score between two proteins.

Examples

```
# the example motif file
motif_file_path <- system.file("extdata", "chip_matched.txt", package = "netZooR", mustWork = TRUE)
motif <- read.table(motif_file_path, sep="\t")
# create a TF data frame with one column
TF <-data.frame(motif[,1])
# create PPI data frame by searching in STRING version 10
# and specifying specie to "Mycobacterium tuberculosis H37Rv".
# STRING verison 11 is only accessible to R 4.0.

if(R.Version()$major=="3"){PPI <- sourcePPI(TF, STRING.version="10",
species.index=83332, score_threshold=0)}
if(R.Version()$major=="4"){PPI <- sourcePPI(TF, STRING.version="11",
species.index=83332, score_threshold=0)}

# write out locally then can be used in \code{\link{pandaPy}}.
```

---

spider *Seeding PANDA Interactions to Derive Epigenetic Regulation*

Description

This function runs the SPIDER algorithm

Usage

```
spider(
  motif,
  expr = NULL,
  epifilter = NULL,
  ppi = NULL,
  alpha = 0.1,
  hamming = 0.001,
  iter = NA,
  output = c("regulatory", "coexpression", "cooperative"),
  zScale = TRUE,
  progress = FALSE,
  randomize = c("None", "within.gene", "by.gene"),
  cor.method = "pearson",
  scale.by.present = FALSE,
  edgelist = FALSE,
  remove.missing.ppi = FALSE,
```

```
  remove.missing.motif = FALSE,
  remove.missing.genes = FALSE,
  mode = "union"
)
```

Arguments

motif	A motif dataset, a data.frame, matrix or exprSet containing 3 columns. Each row describes an motif associated with a transcription factor (column 1) a gene (column 2) and a score (column 3) for the motif.
expr	An expression dataset, as a genes (rows) by samples (columns) data.frame
epifilter	A binary matrix that is of the same size as motif that will be used as a mask to filter motif for open chromatin region. Motif interactions that fall in open chromatin region will be kept and the others are removed.
ppi	A Protein-Protein interaction dataset, a data.frame containing 3 columns. Each row describes a protein-protein interaction between transcription factor 1(column 1), transcription factor 2 (column 2) and a score (column 3) for the interaction.
alpha	value to be used for update variable, alpha (default=0.1)
hamming	value at which to terminate the process based on hamming distance (default 10^-3)
iter	sets the maximum number of iterations SPIDER can run before exiting.
output	a vector containing which networks to return. Options include "regulatory", "coregulatory", "cooperative".
zScale	Boolean to indicate use of z-scores in output. False will use [0,1] scale.
progress	Boolean to indicate printing of output for algorithm progress.
randomize	method by which to randomize gene expression matrix. Default "None". Must be one of "None", "within.gene", "by.genes". "within.gene" randomization scrambles each row of the gene expression matrix, "by.gene" scrambles gene labels.
cor.method	Correlation method, default is "pearson".
scale.by.present	Boolean to indicate scaling of correlations by percentage of positive samples.
edgelist	Boolean to indicate if edge lists instead of matrices should be returned.
remove.missing.ppi	Boolean to indicate whether TFs in the PPI but not in the motif data should be removed. Only when mode=='legacy'.
remove.missing.motif	Boolean to indicate whether genes targeted in the motif data but not the expression data should be removed. Only when mode=='legacy'.
remove.missing.genes	Boolean to indicate whether genes in the expression data but lacking information from the motif prior should be removed. Only when mode=='legacy'.

mode — The data alignment mode. The mode 'union' takes the union of the genes in the expression matrix and the motif and the union of TFs in the ppi and motif and fills the matrics with zeros for nonintersecting TFs and gens, 'intersection' takes the intersection of genes and TFs and removes nonintersecting sets, 'legacy' is the old behavior with PANDAR version 1.19.3. #' Parameters remove.missing.ppi, remove.missingmotif, remove.missing.genes work only with mode=='legacy'.

Value

An object of class "panda" containing matrices describing networks achieved by convergence with SPIDER algorithm.
"regNet" is the regulatory network
"coregNet" is the coregulatory network
"coopNet" is the cooperative network

References

Sonawane, Abhijeet Rajendra, et al. "Constructing gene regulatory networks using epigenetic data." npj Systems Biology and Applications 7.1 (2021): 1-13.

Examples

```
data(pandaToyData)
pandaToyData$epifilter = pandaToyData$motif
nind=floor(runif(5000, min=1, max=dim(pandaToyData$epifilter)[1]))
pandaToyData$epifilter[nind,3] = 0
spiderRes <- spider(pandaToyData$motif,pandaToyData$expression,
                    pandaToyData$epifilter,pandaToyData$ppi,hamming=.1,progress=TRUE)
```

---

tiger *TIGER main function*

Description

TIGER main function

Usage

```
tiger(
  expr,
  prior,
  method = "VB",
  TFexpressed = TRUE,
  signed = TRUE,
  baseline = TRUE,
  psis_loo = FALSE,
  seed = 123,
```

```
  out_path = NULL,
  out_size = 300,
  a_sigma = 1,
  b_sigma = 1,
  a_alpha = 1,
  b_alpha = 1,
  sigmaZ = 10,
  sigmaB = 1,
  tol = 0.005
)
```

Arguments

expr	A normalized log-transformed gene expressison matrix. Rows are genes and columns are sampeles (cells).
prior	A prior regulatory network in adjacency matrix format. Rows are TFs and columns target genes.
method	Method used for Bayesian inference. "VB" or "MCMC". Defaults to "VB".
TFexpressed	TF mRNA needs to be expressed or not. Defaults to TRUE.
signed	Prior network is signed or not. Defaults to TRUE.
baseline	Include baseline or not. Defaults to TRUE.
psis_loo	Use pareto smoothed importance sampling leave-one-out cross validation to check model fitting or not. Defaults to FALSE.
seed	Seed for reproducible results. Defaults to 123.
out_path	(Optional) output path for CmdStanVB or CmdStanMCMC object. Defaults to NULL.
out_size	Posterior sampling size. Default = 300.
a_sigma	Hyperparameter of error term. Default = 1.
b_sigma	Hyperparameter of error term. Default = 1.
a_alpha	Hyperparameter of edge weight W. Default = 1.
b_alpha	Hyperparameter of edge weight W. Default = 1.
sigmaZ	Standard deviation of TF activity Z. Default = 10.
sigmaB	Standard deviation of baseline term. Default = 1.
tol	Convergence tolerance on ELBO.. Default = 0.005.

Value

A TIGER list object. * W is the estimated regulatory network, but different from prior network, rows are genes and columns are TFs. * Z is the estimated TF activities, rows are TFs and columns are samples. * TF.name, TG.name, and sample.name are the used TFs, target genes and samples. * If psis_loo is TRUE, loocv is a table of psis_loo result for model checking. * If psis_loo is TRUE, elpd_loo is the Bayesian LOO estimate of the expected log pointwise predictive density, which can be used for Bayesian stacking to handle multi-modality later.

Examples

```
data(TIGER_expr)
data(TIGER_prior)
tiger(TIGER_expr,TIGER_prior)
```

`TIGER_expr` *TIGER example expression matrix*

Description

TIGER example expression matrix

Usage

```
data(TIGER_expr)
```

Format

'TIGER_expr' A gene expression matrix with 1780 rows (genes) and 16 columns (samples)

Source

<https://zenodo.org/record/7425777>

`TIGER_prior` *TIGER example prior network*

Description

TIGER example prior network

Usage

```
data(TIGER_prior)
```

Format

'TIGER_prior' A prior network matrix with 14 rows (TFs) and 1772 columns (genes)

Source

<https://zenodo.org/record/7425777>

visPandaInCytoscape *Plot PANDA network in Cytoscape*

Description

This function is able to modify PANDA network and plot in Cytoscape. Please make sure that Cytoscape is installed and open it before calling this function.

Usage

```
visPandaInCytoscape(panda.net, network_name = "PANDA")
```

Arguments

panda.net	Character string indicating the input PANDA network in data frame structure type.
network_name	Character string indicating the name of Cytoscape network.

Value

PANDA network in Cytoscape

yeast *Toy data derived from three gene expression datasets and a mapping from transcription factors to genes.*

Description

This data is a list containing gene expression data from three separate yeast studies along with data mapping yeast transcription factors with genes based on the presence of a sequence binding motif for each transcription factor in the vicinity of each gene. The motif data.frame, yeast$motif, describes a set of pairwise connections where a specific known sequence motif of a transcription factor was found upstream of the corresponding gene. The expression data, yeast$exp.ko, yeast$exp.cc, and yeast$exp.sr, are three gene expression datasets measured in conditions of gene knockout, cell cycle, and stress response, respectively.

Usage

```
data(yeast)
```

Format

A list containing 4 data.frames

Value

A list of length 4

References

Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing Messages Between Biological Networks to Refine Predicted Interactions. PLoS One. 2013 May 31;8(5):e64832.

Index