Package 'fRNC'

January 10, 2021

Type Package

Title ncRNA module search functions

Version 0.1

Author leiming jiang

Maintainer leiming jiang <leiming8886@163.com>

Depends R (>= 3.6.0), igraph

Imports graph,
RBGL,
Biobase,
methods,
stats,
XML,
survival,
ggpubr,
limma,
edgeR

Suggests devtools,
roxygen2,
testthat

Description More about what it does (maybe more than one line)
Use four spaces when indenting paragraphs within the Description.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

R topics documented:

`fRNC-package` *fRNC: Search for modules in a node-weighted ncRNA network*

Description

fRNC constructs a node-weighted ncRNA network, performs module searching, generates simulation data from random networks, normalizes module scores using simulation data, removes unqualified modules, and orders resultant modules according to their significance.

Details

This package takes three types of data as input: a list of genes with association p-values and logFC, a human ncRNA network. generate_graph constructs a node-weighted ncRNA network. runmodule performs module search upon the node-weighted ncRNA network.

References

Hongbo Shi, Jiayao Li, Qiong Song et al. (2019) Systematic identification and analysis of dysregulated miRNA and transcription factor feed-forward loops in hypertrophic cardiomyopathy

Peilin Jia, Siyuan Zheng, Jirong Rong, Wei Zheng, Zhongming Zhao. (2011) Bioformatics. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks.

case.exp_miRNA *miRNA-seq expression data in the ECSA cancer samples*

Description

miRNA-seq expression data in the ECSA cancer samples, the row names are miRNA mirbase IDS, the column names are sample names in which last two sign is 01

Usage

```
data("case.exp_miRNA")
```

Examples

```
data("case.exp_miRNA")
case.exp_miRNA[1:100,]
```

case.exp_rna *RNA-seq expression data in the ECSA cancer samples*

Description

RNA-seq expression data in the ECSA cancer samples, the row names are RNA mirbase IDS, the column names are sample names in which last two sign is 01

Usage

```
data("case.exp_rna")
```

Examples

```
data("case.exp_rna")
case.exp_rna[1:100,]
```

cepair *extract lncRNA-circRNA ceRNA pair*

Description

extract lncRNA-circRNA ceRNA pair

Usage

```
cepair(interac, N_mi = 50)
```

Arguments

interac	the interaction data matrix, in which the column name contain "node_gene_ID", "type" and "target_gene_ID"
N_mi	a numeric value, shared miRNAs by a lncRNA and circRNA

Value

interac_temp the interaction data matrix, in which the column name contain "node_gene_ID", "type" and "target_gene_ID"

Examples

```
## Not run:
interac <- interStringency(type = "ncRNA", stringency = "high")
interac_high_p_50 <- cepair(interac, N_mi= 50)

## End(Not run)
```

---

clinData *clinical data in the ECSA cancer patients*

Description

clinical data in the ECSA cancer patients, the column names are TCGA sample ID(cancer sample ID),survival time (month),survival status

Usage

```
data("clinData")
```

Examples

```
data("clinData")
clinData[1:100,]
```

---

combinp *Generate corrected p-value based on p-value and logFC in the expression matrix*

Description

based on the formula: corrected p-value = 2*(1-pnorm((-log10(p-value)) * abs(log2FC))), corrected p-value was generated

Usage

```
combinp(node_attr = NULL, islog = T)
```

Arguments

node_attr	A data frame containing three columns: type, logFC and p value, and the row name is the gene identifier.
islog	Boolean value, whether to use the logFC, if FALSE, the weight is the p-value, or "TRUE", the corrected p-value is used

Value

A data matrix containing three columns: type, gene, weight(corrected p-value), the row name is the gene identifier.

References

Hongbo Shi, Jiayao Li, Qiong Song et al. (2019) Systematic identification and analysis of dysregulated miRNA and transcription factor feed-forward loops in hypertrophic cardiomyopathy

Examples

```
data("dataN")
result <- combinp(dataN[,c("type","logFC","PValue")])
```

---

control.exp_miRNA *miRNA-seq expression data in the ECSA normal samples*

Description

miRNA-seq expression data in the ECSA normal samples, the row names are miRNA mirbase IDS, the column names are sample names in which last two sign is 11

Usage

```
data("control.exp_miRNA")
```

Examples

```
data("control.exp_miRNA")
control.exp_miRNA[1:100,]
```

---

control.exp_rna *RNA-seq expression data in the ECSA normal samples*

Description

RNA-seq expression data in the ECSA normal samples, the row names are RNA mirbase IDS, the column names are sample names in which last two sign is 11

Usage

```
data("control.exp_rna")
```

Examples

```
data("control.exp_rna")
control.exp_rna[1:100,]
```

dataM2C *Example network for ncRNA network on miRNA-circRNA*

Description

Data example consisting of a matrix of network.

Usage

```
data("dataM2C")
```

Examples

```
data("dataM2C")
dataM2C[1:100,]
```

dataM2L *Example network for ncRNA network on miRNA-lncRNA*

Description

Data example consisting of a matrix of network.

Usage

```
data("dataM2L")
```

Examples

```
data("dataM2L")
dataM2L[1:100,]
```

dataM2R *Example network for ncRNA network on miRNA-RPB*

Description

Data example consisting of a matrix of network.

Usage

```
data("dataM2R")
```

Examples

```
data("dataM2R")
dataM2R[1:100,]
```

dataN *Example node_attr for ncRNA network*

Description

Data example consisting of a matrix of the colname of genes, type, logFC, logCPM, PValue, FDR.

Usage

```
data("dataN")
```

Examples

```
data("dataN")
dataN[1:100,]
```

dataR2C *Example network for ncRNA network on RPB-circRNA*

Description

Data example consisting of a matrix of network.

Usage

```
data("dataR2C")
```

Examples

```
data("dataR2C")
dataR2C[1:100,]
```

dataR2L *Example network for ncRNA network on RBP-lncRNA*

Description

Data example consisting of a matrix of network.

Usage

```
data("dataR2L")
```

Examples

```
data("dataR2L")
dataR2L[1:100,]
```

DEGs *Performe differential expression analysis*

Description

Differential expression analysis using edgeR or limma for two group comparison

Usage

```
DEGs(case.exp, control.exp, geneid, data_type)
```

Arguments

case.exp	the case expression matrix, in which the row name is gene id and the column name is sample id
control.exp	the control expression matrix, in which the row name is gene id (the same with the case) and the column name is sample id
geneid	gene id in the case or control expression matrix
data_type	a character string indicating which date type to deal with is to be choosed, One of "RNAseq_counts" , "fpkm" and "microarray": can be abbreviated

Value

DEGlist list contain four elements, DEGs the differential expression matrix, Nor_expr the normalized expression matrix, data_type a character string of the data type algorithm a character string indicating which algorithm was used, One of "edgeR" , "limma"

Examples

```
## Not run:
data("case.exp_miRNA")
data("control.exp_miRNA")
result_miR <- DEGs(case.exp_miRNA,control.exp_miRNA,
geneid= rownames(control.exp_miRNA), data_type = "RNAseq_counts")


## End(Not run)
```

gene_type *Table of the gene identifier and the type for ensembl ID*

Description

Data consisting of a matrix, the colname contain gene_ID and gene_Name

Usage

```
data("gene_type")
```

Examples

```
data("gene_type")
gene_type[1:100,]
```

`IDsymbol` *Table of the gene identifier and the gene symbol*

Description

Data consisting of a matrix of the colname of gene_ID and gene_Name

Usage

```
data("IDsymbol")
```

Examples

```
data("IDsymbol")
IDsymbol[1:100,]
```

`integPvals` *Integrate multiple p-values into joint p-values*

Description

The function integrate multiple p-values into the joint p-value of p-values based on the order statistics of p-values. An joint p-value #is given by the kth order statistic.

Usage

```
integPvals(pvalmatrix)
```

Arguments

`pvalmatrix`	Numeric matrix of p-values, columns represent different sets of p-values

Value

matrix The matrix of two columns: gene, weight(p-value)

interStringency *Extract interactions according to stringency and interaction type*

Description

interactions were extracted according to stringency and interaction type in the database of ENCORI

Usage

```
interStringency(
  type = c("RBP", "ncRNA"),
  stringency = c("low", "medium", "high", "strict")
)
```

Arguments

type	a character string indicating which interaction type is to be choosed, . One of "RBP" (RBP-circRNA,RBP-lncRNA,miRNA-circRNA,miRNA-lncRNA,miRNA-RBP), "ncRNA (miRNA-circRNA,miRNA-lncRNA)": can be abbreviated
stringency	a character string indicating which interaction stringency is to be choosed, . One of "low" (number of supported experiments > = 1), "medium (> = 2)","high (> = 3)","strict (> = 5)"

Value

interaction of setting

Examples

```
## Not run:
  data("dataM2C")
data("dataM2L")
data("dataM2R")
data("dataR2C")
data("dataR2L")
 interac <- interStringency(type = "ncRNA",stringency = "strict")
 interac <- interac[,c("node_gene_ID","target_gene_ID")]

## End(Not run)
```

plotSub *Plot of the subnetwork*

Description

The function plots a network from graphNEL or igraph format. It is used to visualize the modules. For further plotting options use the plot.igraph function of the igraph package. The shapes of the nodes can be changed according to the scores argument, then negative scores appear squared The color of the nodes can be changed according to the diff.expr argument. Negative(positive) values lead to green(red) nodes.

Usage

```
plotSub(
  network,
  layout = layout.fruchterman.reingold,
  labels = NULL,
  diff.expr = NULL,
  scores = NULL,
  main = NULL,
  vertex.size = NULL
)
```

Arguments

network	A graph in igraph or graphNEL format.
layout	Layout algorithm, e.g. layout.fruchterman.reingold or layout.kamada.kawai.
labels	Labels for the nodes of the network
diff.expr	Named numerical vector of log2FC of the nodes in the network for coloring of the nodes.
scores	Named numerical vector of scores of the nodes for the shape of the node in the network.
main	Main title of the plot.
vertex.size	Numerical value or verctor for the size of the vertices.

References

Daniela Beisser, Gunnar W. Klau, Thomas Dandekar et al. (2010) BioNet: an R-Package for the functional analysis of biological networks

Examples

```
library(igraph)
edgel <- cbind(c("1", "2", "3", "4", "5", "6", "7"), c("b", "c", "d", "e", "f", "a", "b"))
g <- graph.edgelist(edgel, directed=TRUE)
V(g)$type <- c(rep("lncRNA",4),rep("miRNA",4),rep("circRNA",5))
plotSub(g)
```

---

runmodule *Run module search function*

Description

runmodule constructs a node-weighted ncRNA network, performs module searching, generates simulation data from random networks,

normalizes module scores using simulation data, removes un-qualified modules, and orders resultant modules according to their significance.

Usage

```
runmodule(
  network,
  gene2weight,
  maxsize = 15,
  method = c("global", "local"),
  d = 2,
  r = 0.1,
  seletN = NULL,
  FDR = 1e-14,
  issymbol = TRUE
)
```

Arguments

network	A data frame containing a symbolic edge list of the ncRNA network in which the columns must contain "node_gene_ID", "type", "target_gene_Name"
gene2weight	A weigth data frame containing three columns:"type","gene", "weight" the first "type" the type of the gene identifier; lncRNA, miRNA, circRNA the second gene is unique, gene identifier (should be coordinate with the node symbol used in ncRNA network); the third weight is gene-based p-value or corrected p-value derived from differentially gene analysis or survival analysis
maxsize	An integer: the numbel of size of the module for user settings in the method of "global", default 15.
method	a character string indicating which the search method is to be computed . One of "global" (default, refer to Heinz method), "local (refer to GS method)": can be abbreviated
d	An integer used to define the order of neighbour genes to be searched in the method of the method "local" . This parameter is default set up as 2
r	A float indicating the cut-off for increment during module expanding process in the method of the method "local". Greater r will generate smaller module. Default is 0.1.
seletN	a vector: gene identifier IDs, or a gene identifier ID, for example "MIMAT0000461",c("MIMAT00004 "ENSG00000250742")
FDR	Numeric value, from the false discovery rate a p-value threshold is calculated. P-values below this threshold are considered to be significant The FDR can be used to control the size of the maximum scoring module
issymbol	Boolean value, whether to set the node attribute "symbol"(gene symbol) in the network. @return runmodule returns a list containing relevant data and results, including:

GNCW	the node-weighted network used for searching
module	list of genes comprising each module, named for the seed gene if the method is "local" or the ig
module.score.matrix	contains Zm, Zn

References

Hongbo Shi, Jiayao Li, Qiong Song et al. (2019) Systematic identification and analysis of dysregulated miRNA and transcription factor feed-forward loops in hypertrophic cardiomyopathy

Peilin Jia, Siyuan Zheng, Jirong Rong, Wei Zheng, Zhongming Zhao. (2011) Bioformatics. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks.

Daniela Beisser, Gunnar W. Klau, Thomas Dandekar et al. (2019) BioNet: an R-Package for the functional analysis of biological networks

Examples

```
## Not run:
data("dataN")
gene2weight <- combinp(dataN[,c("type","logFC","PValue")])
interac <- interStringency(type = "ncRNA",stringency = "strict")
interac <- interac[,c("node_gene_ID","type","target_gene_ID")]
 res.list_global <- runmodule(network = interac, gene2weight, method = "global")
 res.list_local <- runmodule(network = interac, gene2weight, method = "local",
 maxsize=15, seletN = "MIMAT0000461")

## End(Not run)
```

---

savelocalM *save and plot module*

Description

save and plot module for the methd "local" result

Usage

```
savelocalM(res.list_local)
```

Arguments

res.list_local	the methd "local" result

Value

the plot and the format "XGMML" of the each module, filenames is the seed node

---

saveNetwork *save of the subnetwork*

Description

The function plots a network from graphNEL or igraph format. It is used to visualize the modules. For further plotting options use the plot.igraph function of the igraph package. The shapes of the nodes can be changed according to the scores argument, then negative scores appear squared The color of the nodes can be changed according to the diff.expr argument. Negative(positive) values lead to green(red) nodes.

Usage

```
saveNetwork(
  network,
  name = "network",
  file,
  type = c("table", "XGMML", "sif", "tab")
)
```

Arguments

network	A graph in igraph or graphNEL format.
name	Name of the network, only needed for the XGMML format.
file	File name to save.
type	Type in which graph shall be saved.

References

Daniela Beisser, Gunnar W. Klau, Thomas Dandekar et al. (2010) BioNet: an R-Package for the functional analysis of biological networks

Examples

```
library(igraph)
edgel <- cbind(c("1", "2", "3", "4", "5", "6", "7"), c("b", "c", "d", "e", "f", "a", "b"))
g <- graph.edgelist(edgel, directed=TRUE)
V(g)$type <- c(rep("lncRNA",4),rep("miRNA",4),rep("circRNA",5))
saveNetwork(g,file ="g", type = "XGMML")
```

---

subNetwork_only *Create a subGraph*

Description

The function creates a subgraph with the nodes given in the nodeList

Usage

```
subNetwork_only(nodeList, network)
```

Arguments

nodeList	Character vector of nodes, contained in the subgraph.
network	Graph that is used for subgraph extraction

Value

A graph object.

Examples

```
library(igraph)
edgel <- cbind(c("a1", "a2", "a3", "a4", "a5", "a6", "a7"), c("b", "c", "d", "e", "f", "a", "b"))
g <- graph.edgelist(edgel, directed=TRUE)
node.list <- c("a1", "b", "c", "a1")
graph <- subNetwork_only(nodeList=node.list, network=g)
```

`survival.miR_rna` *survival analysis*

Description

The function calculate the P-values from a univariable Cox proportional hazards regression model between survival data and corresponding expression of node. First, the overlap of the sample IDs among the survival data and the two expression data is obtained, then based on the overlap sample IDs, the p-value of every node from a univariable Cox proportional hazards regression model is calculated on the R package survival.

Usage

```
survival.miR_rna(miRNA_profile = NULL, gene_profile = NULL, clinData = NULL)
```

Arguments

`miRNA_profile`	the miRNA expressin data matrix,in which the row name is gene id and the column name is sample id.
`gene_profile`	the mRNA expressin data matrix,in which the row name is gene id and the column name is sample id.
`clinData`	the data matrix of survival data,in which the column name is "ID"(sample IDs), "Survival"(months) and "Status"(0,1)

Value

surivallis list contain three elements, miR_p the p-value matrix of IDs, rna_p the p-value matrix of IDs, algorithm a character string indicating which algorithm was used

Examples

```
## Not run:
result_miR <- DEGs(case.exp_miRNA,control.exp_miRNA,
geneid= rownames(control.exp_miRNA),
result_rna <- DEGs(case.exp_rna,control.exp_rna,
geneid= rownames(control.exp_rna), data_type = "RNAseq_counts")
interac <- interac[,c("node_gene_ID","type","target_gene_ID")]
survival.miR_rna(miRNA_profile=result_miR$Nor_expr,
gene_profile = result_rna$Nor_expr, clinData = clinData)

## End(Not run)
```

Index