Package 'LncPath'

October 12, 2022

Type Package

Title Identifying the Pathways Regulated by LncRNA Sets of Interest

Version 1.1

Date 2018-09-26

Author Junwei Han, Zeguo Sun

Maintainer Junwei Han <hanjunwei1981@163.com>

Description Identifies pathways synergisticly regulated by the interested lncRNA(long non-coding RNA) sets based on a lncRNA-mRNA(messenger RNA) interaction network. 1) The lncRNA-mRNA interaction network was built from the protein-protein interactions and the lncRNA-mRNA co-expression relationships in 28 RNA-Seq data sets. 2) The interested lncRNAs can be mapped into networks as seed nodes and a random walk strategy will be performed to evaluate the rate of each coding genes influenced by the seed lncRNAs. 3) Pathways regulated by the lncRNA set will be evaluated by a weighted Kolmogorov-Smirnov statistic as an ES Score. 4) The p value and false discovery rate value will also be calculated through a permutation analysis. 5) The running score of each pathway can be plotted and the heat map of each pathway can also be plotted if an expression profile is provided. 6) The rank and scores of the gene list of each pathway can be printed.

Imports stats, graphics, utils, grDevices

Depends R (>= 3.2.1), igraph

Suggests Matrix,graph

License GPL (>= 2)

LazyData Yes

NeedsCompilation no

Repository CRAN

Date/Publication 2018-09-26 14:20:06 UTC

R topics documented:

`drawAHeatMap` *Draw a heatmap for the genes of a pathway*

Description

Draw a heatmap for the genes of a certain pathway based on the expression profile user specified.

Usage

```
drawAHeatMap(Result, Name, PCExpr, Labels)
```

Arguments

`Result`	A lncPath object come from the lncPath function.
`Name`	A string, the name of the pathway to be plot.
`PCExpr`	A data frame, the expression profile to be plotted.
`Labels`	A vector of 0 and 1, 0 indicates control and 1 indicates case.

Details

Draw a heatmap of the genes of a pathway based on the expression profile. The rows of heatmap are genes ranked by their weights and the columns of heatmap are samples ordered the same as the expression profile.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

Result <- getExampleData("Result")
Profile <- getExampleData("Profile")
Labels <- getExampleData("Labels")
drawAHeatMap(Result, "KEGG_RIBOSOME", Profile, Labels)
```

`findSigGenes` *Find genes significantly differentially expressed between two conditions.*

Description

For a given expression profile of two conditions, find the genes differencially expressed using T-test, fold change or SAM algorithm.

Usage

```
findSigGenes(Expr, Label, Method = "tTest", Directed = TRUE,
FdrCut = 0.01, FDCut = 1)
```

Arguments

`Expr`	A data frame, the expression profile to find differentially expressed genes, the rownames should be the ID of genes.
`Label`	A vector of 0/1s, indicating the class of samples in the expression profile, 0 represents case, 1 represents control.
`Method`	A string, specifying the method to calculate the differentially expressed genes, should be one of the "tTest"or"foldChange".
`Directed`	Logical, if the the up or down regulated set should be distinguished.
`FdrCut`	Numeric, the fdr cutoff for T test, can be ignored if not using t-test.
`FDCut`	Numeric, the cutoff for fold change, can be ignored if not using fold change.

Details

For a given expression profile of two conditions, lncPath package provide two method to find differentially expressed genes: t-text and fold change. The row of the expression profile should be gene IDs and the column of the expression profile should be names of samples. Samples should be under two conditions and the label should be given as 0 and 1. For t-test, fold change and SAM, different threshold can be set for significant differentially expressed genes.

Value

A vector of strings, the IDs of differentially expressed genes.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.
Profile <- getExampleData("Profile")
Labels <- getExampleData("Labels")

SigGenes <- findSigGenes(Profile, Labels)
head(SigGenes)
```

`geneSetDetail` *Gain insight into the detail of the genes in a certain pathway*

Description

Gain insight into the detail of the genes in a certain pathway, inculding the ranks, weights and cummulative running scores of each gene.

Usage

```
geneSetDetail(Result, Name)
```

Arguments

`Result`	A lncPath object come from the lncPath function.
`Name`	A string, the name of the pathway to be print.

Details

List all the genes of pathways ranked by the weights. The table also contains the gene name, the rank of genes in the whole gene list, the cumulative ES score and whether the gene is in the core gene sets which contribute to the score of the pathway.

Value

A data frame, the rows are gene names and the columns are detail of genes including gene name, rank, weight, cumulative ES score and core erichment.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

Result <- getExampleData("Result")
Detail <- geneSetDetail(Result, "KEGG_RIBOSOME")
head(Detail)
```

`getExampleData` *Get the example data*

Description

Get the example data of LncPath package for litte trials.

Usage

```
getExampleData(ExampleData)
```

Arguments

`ExampleData`	A character, should be one of "SigLncs", "ExampleNet", "Labels", "Profile", "Result" and "Table".

Details

The function getExampleData(ExampleData = "SigLncs") obtains a vector of lncRNAs confirmed to be related with breast cancer. The function getExampleData(ExampleData = "Profile") obtains the expression profile as a data frame. The function getExampleData(ExampleData = "Labels") obtains a vector of 0/1s describing the class of samples in the expression profile. The function getExampleData(ExampleData = "Result") obtains a lncPath object come from the lncPath function. The function getExampleData(ExampleData = "Table") obtains a data frame as the summary of lncPath object. The function getExampleData(ExampleData = "ExampleNet") obtains a data frame as the edges of lncRNA-mRNA interaction net.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

`getNet` *Get the background lncRNA-mRNA interaction network*

Description

Get the background lncRNA-mRNA interaction network.

Usage

```
getNet()
```

Details

Get the background lncRNA-mRNA interaction network, it was built by intergrating an lncRNA-mRNA co-expression network and the protein-protein interaction network.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.
LncPathNet <- getNet();
```

`lncPath` *Identify pathways synergisticly regulated by lncRNA sets.*

Description

Identify pathways synergisticly regulated by lncRNA sets by combining the random walk strategy and weighted Kolmogorov-Smirnov statistic based on a huge lncRNA-mRNA interaction network.

Usage

```
lncPath(LncRNAList, Network, Weighted = TRUE, PathwayDataSet = "KEGG",
minPathSize = 15, maxPathSize = 500, nperm = 1000)
```

Arguments

`LncRNAList`	A character vector, contains the user interested lncRNAs, the ID of lncRNAs should be the Ensembl ID .
`Network`	A dataframe with two columns, describing the edges of the network to perform the random walk.
`Weighted`	Logical, tell if a weighted analysis to be performed, see detail.
`PathwayDataSet`	A character, tells which pathway database is to be used, should be one of "KEGG", "Reactome" and "BioCarta".
`minPathSize`	An integer, the lower limit of the mapped genes in pathway.
`maxPathSize`	An integer, the upper limit of the mapped genes in pathway.
`nperm`	An integer, how manny times of perturbation to be performed in the perturbation analysis.

Details

lncPath is the main function of lncPath package, it takes a list of interested lncRNAs and a lncRNA-mRNA interaction network as input. Then it maps the lncRNAs into the lncRNA-mRNA interaction network as seed nodes and performs a random walk strategy to evaluate the rate of noedes effected by the seed nodes. A weighted Kolmogorov-Smirnov statistic was finnally used to evaluate the pathways related to the lncRNA sets. If the Weighted parameter is set to TRUE, the scores of mRNAs generated from random walk will be treated as the weight in Kolmogorov-Smirnov statistic.If the Weighted parameter is set to FALSE, only the ranks of mRNAs will be taken into consideration. Now three pathway data sets are surpported, includeing the KEGG, Reactome and BioCarta. And pathways with number of genes out of the limit will be filtered.

Value

A lncPath object, containing the details of each pathways: pathway ID, pathway name, number of genes, gene names, score of genes etc. It can be summarized by function by function lncPath2Table and can be visualized by function plotRunningES.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.
## get example data
SigLncs <- getExampleData("SigLncs")
head(SigLncs)

ExampleNet <- getExampleData("ExampleNet")
head(ExampleNet)

##run lncPath
Result <- lncPath(SigLncs, ExampleNet,  Weighted = TRUE, PathwayDataSet = "KEGG", nperm = 100,
 minPathSize = 0, maxPathSize = 500)

## Print to table
Table <- lncPath2Table(Result)
head(Table)
```

---

lncPath2Table *Simplify the lncPath object into table*

Description

Simplify the LncPath object into a data frame, which discribes the detail imformation of each pathway.

Usage

```
lncPath2Table(Result)
```

Arguments

Result	The lncPath object come from the lncPath function.

Details

The lncPath object come from the lncPath function may be too complicated for user to view. This function can simplify it into a data frame. Each row of the data frame describe the detail of one pathway, including informations of pathway name, number of genes in the pathway, enrichment scores, normalized enrichment scores, p value and false discovery rate.

Value

A data frame, rows are pathways and columns are details of each pathway.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
Result <- getExampleData("Result")
Table <- lncPath2Table(Result)
head(Table)
```

LncPathEnvir *The variables in the environment variable* LncPathEnvir *of the system*

Description

The variables in the environment variable LncPathEnvir of the system.

Format

An environment variable

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

`plotRunningES` *Visualize the Kolmogorov-Smirnov running score of pathway*

Description

Visualize the Kolmogorov-Smirnov running score of each gene of a certain pathway

Usage

```
plotRunningES(Result, Name)
```

Arguments

`Result`	A lncPath object come from the lncPath function.
`Name`	A string, the name of the pathway to be plot.

Details

Plot the KS-statistic running score of certain pathway. The plot has three sections, the top section is a curve describes the cumulative ES score of pathway through all coding genes. The middle section contains signals telling which gene is in the pathway. The bottom section describes the weight distribution of genes.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

Result <- getExampleData("Result")
plotRunningES(Result, "KEGG_RIBOSOME")
```

printSignifResult *Output the details of significant pathways*

Description

Export all of the significant pathways into a specified location.

Usage

```
printSignifResult(Result, Threshold = 0.01, Path = ".", HeatPlot = FALSE,
PCExpr = "", Labels = "", Top = 0)
```

Arguments

Result	A lncPath object come from the lncPath function.
Threshold	Numeric, the FDR threshold for selecting signifcant pathways.
Path	String, the output directory.
HeatPlot	Logical, should the heatmaps be plotted.
PCExpr	A data frame, represents the expression profile of genes, the rownames must be gene names, must be set if HeatPlot is TRUE.
Labels	A vector of 0 and 1, 0 indicates control and 1 indicates case.
Top	An integer, indicates the number of the most significant pathways to be print, the Threshold will be ignored.

Details

For a result from the lncPath function, pritSignifResult will output all the details of significant pathways. Significant pathways can be defined by the threshold user submit or by ranks. The detail of pathways contains the running score plot , the gene sets detail and the heatmap of each pathway. For heatmap plot , the corresponding expression profile is needed. Considering a lot of files will be output, the output directory can be specified.

Author(s)

Junwei Han <hanjunwei1981@163.com>, Zeguo Sun <zeguo.sun@163.com>

References

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A, 102, 15545-15550.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.
## Not run:
Result <- getExampleData("Result")
Profile <- getExampleData("Profile")
Labels <- getExampleData("Labels")
dir.create("Signif")
SignifReport(Result, Threshold = 0.01, Path = "Signif", HeatPlot = TRUE, Profile, Labels, Top = 30)

## End(Not run)
```

Index