Package 'PathQuant'

May 26, 2017

Type Package

Title Annotation of gene-metabolite pairs using topology of KEGG metabolic pathway maps

Version 0.1.1

Date 2015-07-22

Author Sandra Therrien-Laperriere <sandra.therrien.laperriere@gmail.com>
Sarah Cherkaoui <cherkaos@gmail.com>

Maintainer Sandra Therrien-Laperriere <sandra.therrien.laperriere@gmail.com>

Description Modeling of KEGG metabolism pathway maps into compound graph format to compute shortest reactional distance (srd) for gene-metabolite pairs to assess their biological relevance and output results in data.frame and graphical outputs.

License GNU-3

Imports igraph, XML, ggplot2, data.table, knitr, rmarkdown, scales, stringr, gdata, RCurl, reshape2, stringr, dplyr, plyr, doParallel, parallel, foreach

VignetteBuilder knitr

Suggests testthat, knitr, rmarkdown

LazyData TRUE

RoxygenNote 6.0.1

R topics documented:

annotateAssociationData *Gene annotation classification of associations*

Description

Gene annotation involve EC number, kegg brite annotation High level classification made by involved brite trnasporter, enzyme or protein

Usage

```
annotateAssociationData(associations)
```

Arguments

association	Dataframe with 2 columns, where each line reprensents an association. First column are the genes and the sencond column are the metabolites. Only use KEGG Ids.

Examples

```
annotateAssociationData(associations)
```

exportDFtoTxt *Export results data of a data frame into a text files, for which each lines separates the data from every data frame colum by tabulation. This text file can be easily opened with excel.*

Description

Export results data of a data frame into a text files, for which each lines separates the data from every data frame colum by tabulation. This text file can be easily opened with excel.

Usage

```
exportDFtoTxt(export_data, exportNameFile)
```

Arguments

a	dataframe.

Examples

```
exportDFtoTxt(dataFrame, exportFileName)
```

gene.distribution *Distance distribution plots for single gene*

Description

Function ploting the distribution of distances between a gene and all measured metabolite, highlithing the distance of its associated metabolites.

Usage

```
gene.distribution(pathway, gene, pairs, metabolites)
```

Arguments

pathway	Selected pathway. Only use KEGG Ids.
gene	Gene selected. Only use KEGG Ids.
pairs	Dataframe with 2 columns, where each line reprensents an association. First column are the genes and the sencond column are the metabolites. Only use KEGG Ids.
metabolites	Dataframe of 1 column containing a lsit of selected metabolites. Default = "All". All metabolite of the selected pathway map. Only use KEGG Ids.

Details

The plot is depicted as frequency bars, which represent the number of metabolites at a given distance for the selected gene. Frequency bars are shown in grey for metabolites that are not associated with the selected gene and in red if there is at least one metabolite associated with this gene.

If a gene or a metabolite is present on multiple edges or nodes, then the shortest distance is selected.

Examples

```
gene.distribution(hsa01100, "hsa:1373" , pairs.df, metabolite.df)
```

get.srd *isGeneInMap, metaboliteCommonName, metaboliteKEGGId, isMetaboliteInMap, distance*

Description

isGeneInMap, metaboliteCommonName, metaboliteKEGGId, isMetaboliteInMap, distance

Usage

```
get.srd(pairs, organism_code, pathway = "All", ordered = FALSE,
  commonNames = FALSE, path = FALSE)
```

Arguments

pairs	dataframe with 2 columns, where each line reprensents a pair. The first column is genes and the sencond column is metabolites. Only use KEGG Ids.
pathway	list of selected metabolic pathway in KEGG. Default = "All". Only use KEGG Ids.
ordered	[option] ascendent ordering of distance
commonNames	[option] get KEGG's Common Names of the KEGG Id in the results.
path	[option] get reactional path used for srd computation.

Examples

```
get.srd(pairs.df,"All")
```

---

heatmap *Function that output a heatmap to visualize distance computed between every gene-metabolite associations in input using the overview map only*

Description

Function computing shortest distance between every genes and metabolites in a gene-metabolite pairs of your association dataset on a graph model of KEGG map selected, where nodes are metabolites and reactions/genes are edges.

Usage

```
heatmap(association, pathway, commonNames = FALSE)
```

Arguments

association	dataframe with 2 columns, where each line reprensents a uniq association. First column are genes and sencond column are metabolites. Only use KEGG Ids.
pathway	KEGG Id.
display	common names of KEGG ids input (TRUE) or display KEGG Ids for axis description.

Details

If a gene or a metabolite is present on multiple edges or nodes, then the shortest distance is selected within the group

Output: Heatmap of distances calculated between associated genes-metabolites. Columns represent genes and rows represent metabolites. The calculated distance is shown in each cell with the corresponding color code (from red - closest; to yellow - farthest).

Examples

```
heatmap(shinAndAlDF, TRUE)
```

importTXTtoDF *Importing data from an txt file (with table separators), in the context of this package, one of the data to import is the association file.*

Description

Importing data from an txt file (with table separators), in the context of this package, one of the data to import is the association file.

Usage

```
importTXTtoDF(file)
```

Arguments

path	to txt file.

Examples

```
inputXLSXtoDF("file_path/fileName.xlsx")
```

importXLSXtoDF *Importing data from an XSLX file, in the context of this package, one of the data to import is the association file.*

Description

Importing data from an XSLX file, in the context of this package, one of the data to import is the association file.

Usage

```
importXLSXtoDF(file, sheet.num = 1)
```

Arguments

path	to XSLX file.

Examples

```
inputXLSXtoDF("file_path/fileName.xlsx")
```

`permutationTest` *Statistical permutation test to asses the relevance of associated gene-metabolite pairs vs. randomly selected pairs*

Description

Permutation test to evauluate if gene-metabolite associations are significantly closer than randomly selected gene-metabolite pairs.

Usage

```
permutationTest(pathwayId, association, gene, metabolite, permutation,
  output = c("medians", "pvalue", "histogram"))
```

Arguments

`pathwayId`	KEGG Id of selected pathway.
`association`	Dataframe with 2 columns, where each line reprensents an association. First column are genes and the sencond column are metabolites. Only use KEGG Ids.
`gene`	Dataframe of 1 column, representing all genes reported. Only use KEGG Ids.
`metabolite`	Dataframe of 1 column, representing all the measured metabolites. Only use KEGG Ids.
`permutation`	Number desired permutations
`output`	'medians' of all permutations, 'pvalue' of the permutation test, 'histogram' representing the distribution of all permutations' median

Details

If a gene or a metabolite is present on multiple edges or nodes, then the shortest distance is selected.

Examples

```
permutationFunction("hsa01100", shinAndAlDF,
            completeGeneDF,completeMetaboDF, 1000, "histogram")
```

`srd.distribution` *srd.distribution*

Description

Function ploting the distribution of srds computed by get.srd function.

Usage

```
srd.distribution(distance)
```

Arguments

col	of srd values of the data.frame obtained by get.srd function

Details

The plot is depicted as frequency bars, which represent the number of pairs at a given srd value. Frequency bars are shown in heatmap colors to depict the small from larger srd values.

Examples

```
srd.distribution(res[,7])
```

Index