Package ‘pathlinkR’

May 30, 2024

Type Package
Title Analyze and interpret RNA-Seq results
Version 1.0.0
Description pathlinkR is an R package designed to facilitate analysis of RNA-Seq results. Specifically, our aim with pathlinkR was to provide a number of tools which take a list of DE genes and perform different analyses on them, aiding with the interpretation of results. Functions are included to perform pathway enrichment, with multiple databases supported, and tools for visualizing these results. Genes can also be used to create and plot protein-protein interaction networks, all from inside of R.

biocViews GeneSetEnrichment, Network, Pathways, Reactome, RNASeq, NetworkEnrichment

BiocType Software

BugReports https://github.com/hancockinformatics/pathlinkR/issues
License GPL-3 + file LICENSE
URL https://github.com/hancockinformatics/pathlinkR
Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

VignetteBuilder knitr

Config/testthat/edition 3

Depends R (>= 4.3.0)
Imports circlize, clusterProfiler, ComplexHeatmap, dplyr, ggforce, ggplot2, ggpurbr, ggraph, ggrepel, grid, igraph, purrr, sigora, stringr, tibble, tidygraph, tidyr, vegan, visNetwork

Suggests AnnotationDbi, BiocStyle, biomaRt, covr, DESeq2, jsonlite, knitr, org.Hs.eg.db, rmarkdown, scales, testthat (>= 3.0.0), vdiff

git_url https://git.bioconductor.org/packages/pathlinkR
git_branch RELEASE_3_19
## Contents

<table>
<thead>
<tr>
<th>Function</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>pathlinkR-package</td>
<td>3</td>
</tr>
<tr>
<td>.eruptionBreaks</td>
<td>4</td>
</tr>
<tr>
<td>.plotFoldChangeLegend</td>
<td>4</td>
</tr>
<tr>
<td>.runSigora</td>
<td>5</td>
</tr>
<tr>
<td>.truncNeatly</td>
<td>5</td>
</tr>
<tr>
<td>eruption</td>
<td>6</td>
</tr>
<tr>
<td>exampleDESeqResults</td>
<td>8</td>
</tr>
<tr>
<td>getPathwayDistances</td>
<td>9</td>
</tr>
<tr>
<td>groupedPathwayColours</td>
<td>10</td>
</tr>
<tr>
<td>hallmarkDatabase</td>
<td>10</td>
</tr>
<tr>
<td>innateDbPPI</td>
<td>11</td>
</tr>
<tr>
<td>mappingFile</td>
<td>12</td>
</tr>
<tr>
<td>pathnetCreate</td>
<td>12</td>
</tr>
<tr>
<td>pathnetFoundation</td>
<td>15</td>
</tr>
<tr>
<td>pathnetGGraph</td>
<td>16</td>
</tr>
<tr>
<td>pathnetVisNetwork</td>
<td>18</td>
</tr>
<tr>
<td>pathwayCategories</td>
<td>20</td>
</tr>
<tr>
<td>pathwayEnrichment</td>
<td>21</td>
</tr>
<tr>
<td>pathwayPlots</td>
<td>23</td>
</tr>
<tr>
<td>plotFoldChange</td>
<td>25</td>
</tr>
<tr>
<td>ppiBuildNetwork</td>
<td>28</td>
</tr>
<tr>
<td>ppiCleanNetwork</td>
<td>30</td>
</tr>
<tr>
<td>ppiEnrichNetwork</td>
<td>32</td>
</tr>
<tr>
<td>ppiExtractSubnetwork</td>
<td>33</td>
</tr>
<tr>
<td>ppiPlotNetwork</td>
<td>35</td>
</tr>
<tr>
<td>ppiRemoveSubnetwork</td>
<td>38</td>
</tr>
<tr>
<td>reactomeDatabase</td>
<td>38</td>
</tr>
<tr>
<td>sigoraDatabase</td>
<td>39</td>
</tr>
<tr>
<td>sigoraExamples</td>
<td>40</td>
</tr>
</tbody>
</table>

## Index

41
Description

pathlinkR a package for analyzing RNA-Seq data

Details

The pathlinkR package is a suite of functions designed to facilitate the analysis and visualization of RNA-Seq results. The main functions are:

- **eruption** - Create volcano plots from RNA-Seq results
- **plotFoldChange** - Heatmaps to visualize and compare gene expression across multiple conditions
- **pathwayEnrichment** - Test DE genes for enriched Reactome pathways or Hallmark terms, with different methods supported. Results can be visualized with **pathwayPlots**
- **ppiBuildNetwork** - Construct PPI networks from DE genes, using interaction data from InolateDB. Networks can be plotted with **ppiPlotNetwork**, tested for enriched pathways with **ppiEnrichNetwork**, or subnetworks extracted using **ppiExtractSubnetwork**
- **pathnetCreate** - Turn pathway enrichment results into a network of connected pathways, and create static plots with **pathnetGGGraph** or interactive plots with **pathnetVisNetwork**

For more details, please see the package vignette by entering `vignette("pathlinkR")` into the console. Another document with more examples is linked near the top of the included vignette.

Any software-related questions can be posted on the Bioconductor Support site: [https://support.bioconductor.org](https://support.bioconductor.org)

The code is made publicly available on our Github page: [https://github.com/hancockinformatics/pathlinkR](https://github.com/hancockinformatics/pathlinkR)

Author(s)

Travis Blimkie, Andy An

See Also

Useful links:

- [https://github.com/hancockinformatics/pathlinkR](https://github.com/hancockinformatics/pathlinkR)
- Report bugs at [https://github.com/hancockinformatics/pathlinkR/issues](https://github.com/hancockinformatics/pathlinkR/issues)
.eruptionBreaks

INTERNAL Create manual breaks/labels for volcano plots

Description

Internal function which is used to create even breaks for volcano plots produced by eruption.

Usage

.eruptionBreaks(x)

Arguments

x Length-two numeric vector to manually specify limits of the x-axis in log2 fold change; defaults to NA which lets ggplot2 determine the best values.

Value

ggplot scale object

See Also

https://github.com/hancockinformatics/pathlinkR

.plotFoldChangeLegend

INTERNAL Construct heatmap legend

Description

Helper function to handle heatmap legends without cluttering up the main function.

Usage

.plotFoldChangeLegend(.matFC, .log2FoldChange, .cellColours)

Arguments

.matFC Matrix of fold change values
.log2FoldChange Boolean denoting if values will be in log2
.cellColours Colours for fold change values

Value

A list containing heatmap legend parameters and colour function
.runSigora

Description

Internal wrapper function to run Sigora and return the results with desired columns

Usage

.runSigora(enrichGenes, gpsRepo, pValFilter = NA)

Arguments

enrichGenes Vector of genes to enrich
gpsRepo GPS object to use for testing pathways
pValFilter Desired threshold for filtering results

Value

A "data.frame" (tibble) of results from Sigora

References

https://cran.r-project.org/package=sigora

See Also

https://github.com/hancockinformatics/pathlinkR

.truncNeatly

Description

Trims a character string to the desired length, without breaking in the middle of a word (i.e. chops at the nearest space). Appends an ellipsis at the end to indicate some text has been removed.

Usage

.truncNeatly(x, l = 60)
Arguments

x  Character to be truncated

1  Desired maximum length for the output character

Value

Character vector

See Also

https://github.com/hancockinformatics/pathlinkR

---

eruption  
Create a volcano plot of RNA-Seq results

Description

Creates a volcano plot of genes from RNA-Seq results, with various options for tweaking the appearance. Ensembl gene IDs should be the rownames of the input object.

Usage

eruption(
  rnaseqResult,
  columnFC = NA,
  columnP = NA,
  pCutoff = 0.05,
  fcCutoff = 1.5,
  baseColour = "steelblue4",
  nonsigColour = "lightgrey",
  alpha = 0.5,
  pointSize = 1,
  title = NA,
  nonlog2 = FALSE,
  xaxis = NA,
  yaxis = NA,
  highlightGenes = c(),
  highlightColour = "red",
  highlightName = "Selected",
  label = "auto",
  n = 10,
  manualGenes = c(),
  removeUnannotated = TRUE,
  labelSize = 3.5,
  pad = 1.4
)
Arguments

rnaseqResult Data frame of RNASeq results, with Ensembl gene IDs as rownames. Can be a "DESeqResults" or "TopTags" object, or a simple data frame. See "Details" for more information.

columnFC Character; Column to plot along the x-axis, typically log2 fold change values. Only required when rnaseqResult is a simple data frame. Defaults to NA.

columnP Character; Column to plot along the y-axis, typically nominal or adjusted p values. Only required when rnaseqResult is a simple data frame. Defaults to NA.

pCutoff Adjusted p value cutoff, defaults to < 0.05

fcCutoff Absolute fold change cutoff, defaults to > 1.5

baseColour Colour of points for all significant DE genes ("steelblue4")
nonsigColour Colour of non-significant DE genes ("lightgrey")

alpha Transparency of the points (0.5)

pointSize Size of the points (1)

title Title of the plot

nonlog2 Show non-log2 fold changes instead of log2 fold change (FALSE)

xaxis Length-two numeric vector to manually specify limits of the x-axis in log2 fold change; defaults to NA which lets ggplot2 determine the best values.

yaxis Length-two numeric vector to manually specify limits of the y-axis (in -log10). Defaults to NA which lets ggplot2 determine the best values.

highlightGenes Vector of genes to emphasize by colouring differently (e.g. genes of interest). Must be Ensembl IDs.

highlightColour Colour for the genes specified in highlightGenes

highlightName Optional name to call the highlightGenes (e.g. Unique, Shared, Immune related, etc.)

label When set to "auto" (default), label the top n up- and down-regulated DE genes. When set to "highlight", label top n up- and down-regulated genes provided in highlightGenes. When set to "manual" label a custom selection of genes provided in manualGenes.

n number of top up- and down-regulated genes to label. Applies when label is set to "auto" or "highlight".

manualGenes If label="manual", these are the genes to be specifically label. Can be HGNC symbols or Ensembl gene IDs.

removeUnannotated Boolean (TRUE): Remove genes without annotations (no HGNC symbol).

labelSize Size of font for labels

pad Padding of labels; adjust this if the labels overlap
Details

The input to eruption() can be of class "DESeqResults" (from DESeq2), "TopTags" (edgeR), or a simple data frame. When providing either of the former, the columns to plot are automatically pulled ("log2FoldChange" and "padj" for DESeqResults, or "logFC" and "FDR" for TopTags). Otherwise, the arguments "columnFC" and "columnP" must be specified. If one wishes to override the default behaviour for "DESeqResults" or "TopTags" (e.g. plot nominal p values on the y-axis), convert those objects to data frames, then supply "columnFC" and "columnP".

The argument highlightGenes can be used to draw attention to a specific set of genes, e.g. those from a pathway of interest. Setting the argument label="highlight" will also mean those same genes (at least some of them) will be given labels, further emphasizing them in the volcano plot.

Since this function returns a ggplot object, further custom changes could be applied using the standard ggplot2 functions (labs(), theme(), etc.).

Value

Volcano plot of genes from an RNA-Seq experiment; a "ggplot" object

See Also

https://github.com/hancockinformatics/pathlinkR

Examples

data("exampleDESeqResults")
eruption(rnaseqResult=exampleDESeqResults[[1]])

exampleDESeqResults  List of example results from DESeq2

Description

List of example results from DESeq2

Usage

data(exampleDESeqResults)

Format

A list of two "DESeqResults" objects, each with 5000 rows and 6 columns:

baseMean A combined score for the gene
log2FoldChange Fold change value for the gene
lfcSE Standard error for the fold change value
stat The statistic value
pvalue The nominal p value for the gene
padj The adjusted p value for the gene
getPathwayDistances

Value

An object of class "list"

Source

For details on DESeq2 and its data structures/methods, please see https://bioconductor.org/packages/DESeq2/

Description

Given a data frame of pathways and their member genes, calculate the pairwise distances using a constructed identity matrix. Zero means two pathways are identical, while one means two pathways share no genes in common.

Usage

getPathwayDistances(pathwayData = sigoraDatabase, distMethod = "jaccard")

Arguments

pathwayData  Three column data frame of pathways and their constituent genes. Defaults to the provided sigoraDatabase object, but can be any set of Reactome pathways. Must contain Ensembl gene IDs in the first column, human Reactome pathway IDs in the second, and pathway descriptions in the third.

distMethod  Character; method used to determine pairwise pathway distances. Can be any option supported by vegan::vegdist().

Value

Matrix of the pairwise pathway distances (dissimilarity) based on overlap of their constituent genes; object of class "matrix".

References

None.

See Also

https://github.com/hancockinformatics/pathlinkR
Examples

# Here we'll use a subset of all the pathways, to save time
data("sigoraDatabase")

getPathwayDistances(
  pathwayData=dplyr::slice_head(
    dplyr::arrange(sigoraDatabase, pathwayId),
    prop=0.05
  ),
  distMethod="jaccard"
)

groupedPathwayColours  Colour assignments for grouped pathways

Description

Colour assignments for grouped pathways

Usage

data(groupedPathwayColours)

Format

A length 8 named vector of hex colour values

Value

An object of class "character"

hallmarkDatabase  Table of Hallmark gene sets and their genes

Description

Table of Hallmark gene sets and their genes

Usage

data(hallmarkDatabase)
**Format**

A data frame (tibble) with 8,209 rows and 2 columns

- `pathwayId` Name of the Hallmark Gene Set
- `ensemblGeneId` Ensembl gene IDs

**Value**

An object of class "tbl", "tbl.df", "data.frame"

**Source**

For more information on the MSigDB Hallmark gene sets, please see [https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp](https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp)

---

### innateDbPPI

**innateDbPPI**

**InnateDB PPI data**

**Description**


A few important steps have been taken to filter the data, namely the removal of duplicate interactions, and removing interactions that have the same components but are swapped between A and B.

**Usage**

```r
data(innateDbPPI)
```

**Format**

A data frame (tibble) with 152,256 rows and 2 columns:

- `ensemblGeneA` Ensembl gene ID for the first gene/protein in the interaction
- `ensemblGeneB` Ensembl gene ID for the second gene/protein in the interaction

**Value**

An object of class "tbl", "tbl.df", "data.frame"

**Source**

For more details on the data sourced from InnateDB, please see their website: [https://www.innatedb.com](https://www.innatedb.com)
Table of human gene ID mappings

Description
A data frame to aid in mapping human gene IDs between different formats, including Ensembl IDs, HGNC symbols, and Entrez IDs. Mapping information was sourced using biomaRt and AnnotationDbi.

Usage
data(mappingFile)

Format
A data frame (tibble) with 43,993 rows and 3 columns

- ensemblGeneId Ensembl IDs
- hgncSymbol HGNC symbols
- entrezGeneId NCBI Entrez IDs

Value
An object of class "tbl", "tbl.df", "data.frame"

Source
See https://bioconductor.org/packages/biomaRt/ and https://bioconductor.org/packages/AnnotationDbi/ for information on each of the utilized packages and functions.

Description
Create a pathway network from enrichment results and a pathway interaction foundation.

Creates a tidygraph network object from the provided pathway information, ready to be visualized with pathnetGGraph or pathnetVisNetwork.
Usage

pathnetCreate(
  pathwayEnrichmentResult,
  columnId = "pathwayId",
  columnP = "pValueAdjusted",
  foundation,
  trim = TRUE,
  trimOrder = 1
)

Arguments

- pathwayEnrichmentResult: Data frame of results from pathwayEnrichment run with Sigora or ReactomePA (should be based on Reactome data).
- columnId: Character; column containing the Reactome pathway IDs. Defaults to "pathwayID".
- columnP: Character; column containing the adjusted p values. Defaults to "pValueAdjusted".
- foundation: List of pathway pairs to use in constructing a network. Typically this will be the output from createFoundation.
- trim: Remove independent subgraphs which don’t contain any enriched pathways (default is TRUE).
- trimOrder: Order to use when removing subgraphs; Higher values will keep more non-enriched pathway nodes. Defaults to 1.

Details

With the "trim" option enabled, nodes (pathways) and subgraphs which are not sufficiently connected to enriched pathways will be removed. How aggressively this is done can be controlled via the trimOrder argument, and the optimal value will depend on the number of enriched pathways and the number of interacting pathways (i.e. number of rows in "foundation").

Value

A pathway network as a "tidygraph" object, with the following columns for nodes:

- pathwayId: Reactome pathway ID
- pathwayName: Reactome pathway name
- comparison: Name of source comparison, if this pathway was enriched
- direction: Whether an enriched pathway was found in all genes or up- or down-regulated genes
- pValue: Nominal p-value from the enrichment result
- pValueAdjusted: Corrected p-value from the enrichment
- genes: Candidate genes for the given pathway if it was enriched
numCandidategenes
Number of candidate genes
dnumBggenes
Number of background genes
geneRatio
Ratio of candidate and background genes
totalGenes
Total number of DE genes tested, for an enriched pathway
topLevelPathway
Highest level Reactome term for a given pathway
groupedPathway
Custom pathway category used in visualizations
For edges, the following information is also included:
from
Starting node (row number) for the edge
to
Ending node (row number) for the edge
similarity
Similarity of two nodes/pathways
distance
Inverse of similarity
See Also
https://github.com/hancockinformatics/pathlinkR
Examples
data("sigoraDatabase", "sigoraExamples")

pathwayDistancesJaccard <- getPathwayDistances(
  pathwayData=dplyr::slice_head(
    dplyr::arrange(sigoraDatabase, pathwayId),
    prop=0.05
  ),
  distMethod="jaccard"
)

startingPathways <- pathnetFoundation(
  mat=pathwayDistancesJaccard,
  maxDistance=0.8
)

pathnetCreate(
  pathwayEnrichmentResult=sigoraExamples[grepl("Pos",
    sigoraExamples$comparison
  ), ],
  foundation=startingPathways,
  trim=TRUE,
  trimOrder=1
)
Description

From a "n by n" distance matrix, generate a table of interacting pathways to use in constructing a pathway network. The cutoff can be adjusted to have more or fewer edges in the final network, depending on the number of pathways involved, i.e. the number of enriched pathways you’re trying to visualize.

The desired cutoff will also vary based on the distance measure used, so some trial-and-error may be needed to find an appropriate value.

Usage

pathnetFoundation(mat, maxDistance = NA, propToKeep = NA)

Arguments

- **mat**: Matrix of distances between pathways, i.e. 0 means two pathways are identical. Should match the output from getPathwayDistances.
- **maxDistance**: Numeric distance cutoff (less than or equal) used to determine if two pathways should share an edge. Pathway pairs with a distance of 0 are always removed. One of maxDistance or propToKeep must be provided.
- **propToKeep**: Top proportion of pathway pairs to keep as edges, ranked based distance. One of maxDistance or propToKeep must be provided.

Value

A "data.frame" (tibble) of interacting pathway pairs with the following columns:

- **pathwayName1**: Name of the first pathway in the pair
- **pathwayName2**: Name of the second pathway in the pair
- **distance**: Distance measure for the two pathways
- **pathway1**: Reactome ID for the first pathway in the pair
- **pathway2**: Reactome ID for the first pathway in the pair

References

None.

See Also

https://github.com/hancockinformatics/pathlinkR
Examples

data("sigoraDatabase")

pathwayDistancesJaccard <- getPathwayDistances(
    pathwayData = dplyr::slice_head(
        dplyr::arrange(sigoraDatabase, pathwayId),
        prop = 0.05
    ),
    distMethod = "jaccard"
)

startingPathways <- pathnetFoundation(
    mat = pathwayDistancesJaccard,
    maxDistance = 0.8
)

pathnetGGraph

Visualize enriched Reactome pathways as a static network

Description

Plots the network object generated from createPathnet, creating a visual representation of pathway similarity/interactions based on overlapping genes.

Usage

pathnetGGraph(
    network,
    networkLayout = "nicely",
    nodeSizeRange = c(4, 8),
    nodeBorderWidth = 1.5,
    nodeLabelSize = 5,
    nodeLabelColour = "black",
    nodeLabelAlpha = 0.67,
    nodeLabelOverlaps = 6,
    labelProp = 0.25,
    segColour = "black",
    edgeColour = "grey30",
    edgeWidthRange = c(0.33, 3),
    edgeAlpha = 1,
    themeBaseSize = 16
)

Arguments

network Tidygraph network object, output from createPathnet.
networkLayout Desired layout for the network visualization. Defaults to "nicely", but supports any method found in ?layout_tbl_graph_igraph

nodeSizeRange Size range for nodes, mapped to significance (Bonferroni p-value). Defaults to c(4, 8).

nodeBorderWidth Width of borders on nodes, defaults to 1.5

nodeLabelSize Size of node labels; defaults to 5.

nodeLabelColour Colour of the node labels; defaults to "black".

nodeLabelAlpha Transparency of node labels. Defaults to 0.67.

nodeLabelOverlaps Max overlaps for node labels, from ggrepel. Defaults to 6.

labelProp Proportion of "interactor" (i.e. non-enriched) pathways that the function will attempt to label. E.g. setting this to 0.5 (the default) means half of the non-enriched pathways will potentially be labeled - it won't be exact because the node labeling is done with ggrepel.

segColour Colour of line segments connecting labels to nodes. Defaults to "black".

edgeColour Colour of network edges; defaults to "grey30".

edgeWidthRange Range of edge widths, mapped to log10(similarity). Defaults to c(0.33, 3).

edgeAlpha Alpha value for edges; defaults to 1.

themeBaseSize Base font size for all plot elements. Defaults to 16.

Details

A note regarding node labels: The function tries to prioritize labeling enriched pathways (filled nodes), with the labelProp argument determining roughly how many of the remaining interactor pathways might get labels. You'll likely need to tweak this value, and try different seeds, to get the desired effect.

Value

A pathway network or "pathnet"; a plot object of class "ggplot"

References

None.

See Also

https://github.com/hancockinformatics/pathlinkR
Examples

data("sigoraDatabase", "sigoraExamples")

pathwayDistancesJaccard <- getPathwayDistances(
  pathwayData=dplyr::slice_head(
    dplyr::arrange(sigoraDatabase, pathwayId),
    prop=0.05,
    distMethod="jaccard"
  )
)

startingPathways <- pathnetFoundation(
  mat=pathwayDistancesJaccard,
  maxDistance=0.8
)

exPathnet <- pathnetCreate(
  pathwayEnrichmentResult=sigoraExamples[grepl("Pos",
    sigoraExamples$comparison),],
  foundation=startingPathways,
  trim=TRUE,
  trimOrder=1
)

pathnetGGraph(
  exPathnet,
  labelProp=0.1,
  nodeLabelSize=4,
  nodeLabelOverlaps=8,
  segColour="red"
)

---

**pathnetVisNetwork**  
Visualize enriched Reactome pathways as an interactive network

**Description**

Plots the network object generated from createPathnet, creating a visual and interactive representation of similarities/interactions between pathways using their overlapping genes.

**Usage**

pathnetVisNetwork(
  network,
  networkLayout = "layout_nicely",
  nodeSizeRange = c(20, 50),
  ...)
pathnetVisNetwork

```r
nodeBorderWidth = 2.5,
labelNodes = TRUE,
nodeLabelSize = 60,
nodeLabelColour = "black",
edgeColour = "#848484",
edgeWidthRange = c(5, 20),
highlighting = TRUE
```

Arguments

- **network**: Tidygraph network object as output by `createPathnet`
- **networkLayout**: Desired layout for the network visualization. Defaults to "layout_nicely", and should support most igraph layouts. See ?visIgraphLayout for more details.
- **nodeSizeRange**: Node size is mapped to the negative log of the Bonferroni-adjusted p value, and this length-two numeric vector controls the minimum and maximum. Defaults to c(20, 50).
- **nodeBorderWidth**: Size of the node border, defaults to 2.5
- **labelNodes**: Boolean determining if nodes should be labeled. Note it will only ever label enriched nodes/pathways.
- **nodeLabelSize**: Size of the node labels in pixels; defaults to 60.
- **nodeLabelColour**: Colour of the node labels; defaults to "black".
- **edgeColour**: Colour of network edges; defaults to "#848484".
- **edgeWidthRange**: Edge width is mapped to the similarity measure (one over distance). This length-two numeric vector controls the minimum and maximum width of edges. Defaults to c(5, 20).
- **highlighting**: When clicking on a node, should directly neighbouring nodes be highlighted (other nodes are dimmed)? Defaults to TRUE.

Details

This function makes use of the visNetwork library, which allows for various forms of interactivity, such as including text when hovering over nodes, node selection and dragging (including multiple selections), and highlighting nodes belonging to a larger group (e.g. top-level Reactome category).

Value

An interactive pathway, network or "pathnet"; object of class "visNetwork"

References

https://datastorm-open.github.io/visNetwork/

See Also

https://github.com/hancockinformatics/pathlinkR
**pathwayCategories**

**Examples**

```r
data("sigoraDatabase", "sigoraExamples")

pathwayDistancesJaccard <- getPathwayDistances(
  pathwayData=dplyr::slice_head(
    dplyr::arrange(sigoraDatabase, pathwayId),
    prop=0.05
  ),
  distMethod="jaccard"
)

startingPathways <- pathnetFoundation(
  mat=pathwayDistancesJaccard,
  maxDistance=0.8
)

exPathnet <- pathnetCreate(
  pathwayEnrichmentResult=sigoraExamples[grepl("Pos", sigoraExamples$comparison),],
  foundation=startingPathways,
  trim=TRUE,
  trimOrder=1
)

pathnetVisNetwork(exPathnet)
```

---

**pathwayCategories**

**Top-level pathway categories**

**Description**

A data frame containing all Reactome pathways and Hallmark terms, along with a manually-curated top-level category for each entry.

**Usage**

```r
data(pathwayCategories)
```

**Format**

A data frame (tibble) with 2685 rows and 5 columns

- **pathwayId** Reactome or Hallmark pathway identifier
- **pathwayName** Pathway name
- **topLevelPathway** Top hierarchy pathway term, shortened in some cases
- **groupedPathway** Top grouped pathway, 8 for Reactome
- **topLevelOriginal** Original top pathway name
pathwayEnrichment

Value

An object of class "tbl", "tbl.df", "data.frame"

Source

See https://reactome.org/ and https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp for information on each of these databases.

Description

This function provides a simple and consistent interface to three different pathway enrichment tools: Sigora and ReactomePA (which both test for Reactome pathways), and MSigDB Hallmark gene set enrichment.

Usage

pathwayEnrichment(
  inputList,
  columnFC = NA,
  columnP = NA,
  filterInput = TRUE,
  pCutoff = 0.05,
  fcCutoff = 1.5,
  split = TRUE,
  analysis = "sigora",
  filterResults = "default",
  gpsRepo = "default",
  geneUniverse = NULL,
  verbose = FALSE
)

Arguments

inputList A list, with each element containing RNA-Seq results as a "DESeqResults", "TopTags", or "data.frame" object. Rownames of each table must contain Ensembl Gene IDs. The list names are used as the comparison name for each element (e.g. "COVID vs Healthy"). See Details for more information on supported input types.

columnFC Character; Column to plot along the x-axis, typically log2 fold change values. Only required when rnaseqResult is a simple data frame. Defaults to NA.

columnP Character; Column to plot along the y-axis, typically nominal or adjusted p values. Only required when rnaseqResult is a simple data frame. Defaults to NA.
pathwayEnrichment

filterInput When providing list of data frames containing the unfiltered RNA-Seq results (i.e. not all genes are significant), set this to TRUE to remove non-significant genes using the thresholds set by the pCutoff and fcCutoff. When this argument is FALSE its assumed your passing a pre-filtered data in inputList, and no more filtering will be done.

pCutoff Adjusted p value cutoff when filtering. Defaults to < 0.05.

fcCutoff Minimum absolute fold change value when filtering. Defaults to > 1.5

split Boolean (TRUE); Split into up- and down-regulated DE genes using the fold change column, and do enrichment independently on each. Results are combined at the end, with an added "direction" column.

analysis Method/database to use for enrichment analysis. The default is "sigora", but can also be "reactomepa" or "hallmark"

filterResults Should the output be filtered for significance? Use 1 to return the unfiltered results, or any number less than 1 for a custom p-value cutoff. If left as default, the significance cutoff for Sigora is 0.001, or 0.05 for ReactomePA and Hallmark.

gpsRepo Only applies to analysis="sigora". Gene Pair Signature object for Sigora to use to test for enriched pathways. Leaving this set as "default" will use the "reaH" GPS object from Sigora, or you can provide your own custom GPS repository.

geneUniverse Only applies when analysis is "reactomepa" or "hallmark". The set of background genes to use when testing with ReactomePA or Hallmark gene sets. For ReactomePA this must be a character vector of Entrez genes. For Hallmark, it must be Ensembl IDs.

verbose Logical; If FALSE (the default), don’t print info/progress messages.

Details

inputList must be a named list of RNA-Seq results, with each element being of class "DESeqResults" from DESeq2, "TopTags" from edgeR, or a simple data frame. For the first two cases, column names are expected to be the standard defined by each class ("log2FoldChange" and "padj" for "DESeqResults", and "logFC" and "FDR" for "TopTags"). Hence for these two cases the arguments columnFC and columnP can be left as NA.

In the last case (elements are "data.frame"), both columnFC and columnP must be supplied when filterInput=TRUE, and columnFC must be given if split=TRUE.

Value

A "data.frame" (tibble) of pathway enrichment results for all input comparisons, with the following columns:

| comparison | Source comparison from the names of inputList |
| direction | Whether the pathway was enriched in all genes (split=FALSE), or up- or down-regulated genes (split=TRUE) |
| pathwayId | Pathway identifier |
| pathwayName | Pathway name |
pathwayPlots

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pValue</td>
<td>Nominal p value for the pathway</td>
</tr>
<tr>
<td>pValueAdjusted</td>
<td>p value, corrected for multiple testing</td>
</tr>
<tr>
<td>genes</td>
<td>Candidate genes, which were DE for the comparison and also in the pathway</td>
</tr>
<tr>
<td>numCandidateGenes</td>
<td>Number of candidate genes</td>
</tr>
<tr>
<td>numBgGenes</td>
<td>Number of background genes for the pathway</td>
</tr>
<tr>
<td>geneRatio</td>
<td>Ratio of candidate and background genes</td>
</tr>
<tr>
<td>totalGenes</td>
<td>Number of DE genes which were tested for enriched pathways</td>
</tr>
<tr>
<td>topLevelPathway</td>
<td>High level Reactome term which serves to group similar pathways</td>
</tr>
</tbody>
</table>

References

- Sigora: https://cran.r-project.org/package=sigora
- ReactomePA: https://www.bioconductor.org/packages/ReactomePA/
- MSigDB/Hallmark: https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp

See Also

https://github.com/hancockinformatics/pathlinkR

Examples

data("exampleDESeqResults")

pathwayEnrichment(
  inputList = exampleDESeqResults[1],
  filterInput = TRUE,
  split = TRUE,
  analysis = "hallmark",
  filterResults = "default"
)

pathwayPlots

Plot pathway enrichment results

Description

Creates a plot to visualize and compare pathway enrichment results from multiple DE comparisons. Can automatically assign each pathway into an informative top-level category.
Usage

`pathwayPlots(
  pathwayEnrichmentResults,
  columns = 1,
  specificTopPathways = "any",
  specificPathways = "any",
  colourValues = c("blue", "red"),
  nameWidth = 35,
  nameRows = 1,
  xAngle = "angled",
  maxPVal = 50,
  intercepts = NA,
  includeGeneRatio = FALSE,
  size = 5,
  legendMultiply = 1,
  showNumGenes = FALSE,
  pathwayPosition = "right",
  newGroupNames = NA
)
``

Arguments

`pathwayEnrichmentResults`

Data frame of results from the function `enrichPathway`

`columns`

Number of columns to split the pathways across, particularly relevant if there are many significant pathways. Can specify up to 3 columns, with a default of 1.

`specificTopPathways`

Only plot pathways from a specific vector of "topLevelPathway". Defaults to "any" which includes all pathway results, or see `unique(pathwayEnrichmentResults$topLevelPathway)` (i.e. the input) for possible values.

`specificPathways`

Only plot specific pathways. Defaults to "any".

`colourValues`

Length-two character vector of colours to use for the scale. Defaults to `c("blue", "red")`.

`nameWidth`

How many characters to show for pathway name before truncating? Defaults to 35.

`nameRows`

How much to rows to wrap across for the pathway name? Defaults to 1.

`xAngle`

Angle of x axis labels, set to "angled" (45 degrees), "horizontal" (0 degrees), or "vertical" (90 degrees).

`maxPVal`

P values below $10^{-\text{maxPVal}}$ will be set to that value.

`intercepts`

Add vertical lines to separate different groupings, by providing a vector of intercepts (e.g. `c(1.5, 2.5)`). Defaults to NA.

`includeGeneRatio`

Boolean (FALSE). Should the gene ratio be included as an aesthetic mapping? If so, then it is attributed to the size of the triangles.
plotFoldChange

size
Size of points if not scaling to gene ratio. Defaults to 5.

legendMultiply
Size of the legend, e.g. increase if there are a lot of pathways which makes the legend small and unreadable by comparison. Defaults to 1, i.e. no increase in legend size.

showNumGenes
Boolean, defaults to FALSE. Show the number of genes for each comparison as brackets under the comparison’s name.

pathwayPosition
Whether to have the y-axis labels (pathway names) on the left or right side. Default is "right".

newGroupNames
If you want to change the names of the comparisons to different names. Input a vector in the order as they appear.

Value
A plot of enriched pathways; a "ggplot" object

See Also
https://github.com/hancockinformatics/pathlinkR

Examples

data("sigoraExamples")
pathwayPlots(sigoraExamples, columns=2)

plotFoldChange

Create a heatmap of fold changes to visualize RNA-Seq results

Description
Creates a heatmap of fold changes values for results from RNA-Seq results, with various parameters to tweak the appearance.

Usage

plotFoldChange(
  inputList,
  columnFC = NA,
  columnP = NA,
  pathName = NA,
  pathId = NA,
  genesToPlot = NA,
  manualTitle = NA,
  titleSize = 14,
  geneFormat = "ensembl",
  pCutoff = 0.05,
)
plotFoldChange

fcCutoff = 1.5,
cellColours = c("blue", "white", "red"),
cellBorder = gpar(col = "grey"),
plotSignificantOnly = TRUE,
showStars = TRUE,
hideNonsigFC = TRUE,
vjust = 0.75,
rot = 0,
invert = FALSE,
log2FoldChange = FALSE,
colSplit = NA,
clusterRows = TRUE,
clusterColumns = FALSE,
colAngle = 90,
colCenter = TRUE,
rowAngle = 0,
rowCenter = FALSE)

Arguments

inputList A list, with each element containing RNA-Seq results as a "DESeqResults",
"TopTags", or "data.frame" object, with Ensembl gene IDs in the rownames. The
list names are used as the comparison name for each dataframe (e.g. "COVID
vs Healthy"). See Details for more information on supported input types.

columnFC Character; Column to plot along the x-axis, typically log2 fold change values.
Only required when rnaseqResult is a simple data frame. Defaults to NA.

columnP Character; Column to plot along the y-axis, typically nominal or adjusted p
values. Only required when rnaseqResult is a simple data frame. Defaults to NA.

pathName The name of a Reactome pathway to pull genes from, also used for the plot title.
Alternative to pathID.

pathId ID of a Reactome pathway to pull genes from. Alternative to pathName.

genesToPlot Vector of Ensembl gene IDs you want to plot, instead of pulling the genes from
a pathway, i.e. this option and pathName/pathID are mutually exclusive.

manualTitle Provide your own title, and override the use of a pathway name the title.

titleSize Font size for the title.

geneFormat Type of genes given in genesToPlot. Default is Ensembl gene IDs ("ensembl"),
but can also input a vector of HGNC symbols ("hgnc").

pCutoff P value cutoff, default is <0.05

fcCutoff Absolute fold change cutoff, default is >1.5

cellColours Vector specifying desired colours to use for the cells in the heatmap. Defaults to
c("blue", "white", "red").

cellBorder A call to grid::gpar() to specify borders between cells in the heatmap. The
default is gpar(col="grey"). To remove borders set to gpar(col=NA)
**plotFoldChange**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>plotSignificantOnly</code></td>
<td>Boolean (TRUE). Only plot genes that are differentially expressed (i.e. they pass <code>pCutoff</code> and <code>fcCutoff</code>) in any comparison from the provided list of data frames.</td>
</tr>
<tr>
<td><code>showStars</code></td>
<td>Boolean (TRUE) show significance stars on the heatmap</td>
</tr>
<tr>
<td><code>hideNonsigFC</code></td>
<td>Boolean (TRUE). If a gene is significant in one comparison but not in another, this will set the colour of the non-significant gene as grey to visually emphasize the significant genes. If set to FALSE, it will be set the colour to the fold change, and if the p value passes <code>pCutoff</code>, it will also display the p value (the asterisks will be grey instead of black).</td>
</tr>
<tr>
<td><code>vjust</code></td>
<td>Adjustment of the position of the significance stars. Default is 0.75. May need to adjust if there are many genes.</td>
</tr>
<tr>
<td><code>rot</code></td>
<td>Rotation of the position of the significance stars. Default is 0.</td>
</tr>
<tr>
<td><code>invert</code></td>
<td>Boolean (FALSE). The default setting plots genes as rows and comparisons as columns. Setting this to TRUE will place genes as columns and comparisons as rows.</td>
</tr>
<tr>
<td><code>log2FoldChange</code></td>
<td>Boolean (FALSE). Default plots the fold changes in the legend as the true fold change. Set to TRUE if you want log2 fold change.</td>
</tr>
<tr>
<td><code>colSplit</code></td>
<td>A vector, with the same length as <code>inputList</code>, which assigns each data frame in <code>inputList</code> to a group, and splits the heatmap on these larger groupings. The order of groups in the heatmap will be carried over, so one can alter the order of <code>inputList</code> and <code>colSplit</code> to affect the heatmap. This argument will be ignored if <code>clusterColumns</code> is set to TRUE. See Details for more information.</td>
</tr>
<tr>
<td><code>clusterRows</code></td>
<td>Boolean (TRUE). Whether to cluster the rows (genes). May need to change if <code>invert=TRUE</code>.</td>
</tr>
<tr>
<td><code>clusterColumns</code></td>
<td>Boolean (FALSE). Whether to cluster the columns (comparisons). Will override order of <code>colSplit</code> if set to TRUE. May need to change if <code>invert=TRUE</code>.</td>
</tr>
<tr>
<td><code>colAngle</code></td>
<td>Angle of column text. Defaults to 90.</td>
</tr>
<tr>
<td><code>colCenter</code></td>
<td>Whether to center column text. Default is TRUE, but it should be set to FALSE if the column name is angled (e.g. <code>colAngle=45</code>).</td>
</tr>
<tr>
<td><code>rowAngle</code></td>
<td>Angle of row text, defaults to 0.</td>
</tr>
<tr>
<td><code>rowCenter</code></td>
<td>Whether to center column text. The default is FALSE, but it should be set to TRUE if vertical column name (e.g. <code>rowAngle=90</code>).</td>
</tr>
</tbody>
</table>

**Details**

All elements of `inputList` should belong to one of the following classes: "DESeqResults" from DESeq2, "TopTags" from edgeR, or a simple "data.frame". In the first two cases, the proper columns for fold change and p values are detected automatically ("log2FoldChange" and "padj" for "DESeqResults", or "logFC" and "FDR" for "TopTags"). In the third case, the arguments `columnFC` and `columnP` must be supplied. Additionally, if one wished to override the default columns for either "DESeqResults" or "TopTags" objects, simply coerce the object to a simple "data.frame" and supply `columnFC` and `columnP` as desired.

The `cellColours` argument is designed to map a range of negative and positive values to the three provided colours, with zero as the middle colour. If the plotted matrix contains only positive (or negative) values, then it will become a two-colour scale, white-to-red (or blue-to-white).
The colSplit argument can be used to define larger groups represented in inputList. For example, consider an experiment comparing two different treatments to an untreated control, in both wild type and mutant cells. This would give the following comparisons: "wildtype_treatment1_vs_untreated", "wildtype_treatment2_vs_untreated", "mutant_treatment1_vs_untreated", and "mutant_treatment2_vs_untreated". One could then specify colSplit as c("Wild type", "Wild type", "Mutant", "Mutant") to make the wild type and mutant results more visually distinct.

Value

A heatmap of fold changes for genes of interest; an "ggplot" class object

References

https://bioconductor.org/packages/ComplexHeatmap/

See Also

https://github.com/hancockinformatics/pathlinkR

Examples

data("exampleDESeqResults")

plotFoldChange(
  exampleDEseqResults,
  pathName="Generation of second messenger molecules"
)

```r
ppiBuildNetwork
```

Construct a PPI network from input genes and InnateDB’s database

Description

Creates a protein-protein interaction (PPI) network using data from InnateDB, with options for network order, and filtering input.

Usage

```r
ppiBuildNetwork(
  rnaseqResult,
  filterInput = TRUE,
  columnFC = NA,
  columnP = NA,
  pCutoff = 0.05,
  fcCutoff = 1.5,
  order = "zero",
  hubMeasure = "betweenness",
  ppiData = innateDbPPI
)
```
Arguments

rnaseqResult  An object of class "DESeqResults", "TopTags", or a simple data frame. See Details for more information on input types.
filterInput If providing list of data frames containing the unfiltered output from DESeq2::results(), set this to TRUE to filter for DE genes using the thresholds set by the pCutoff and fcCutoff arguments. When FALSE it's assumed your passing the filtered results into inputList and no more filtering will be done.
columnFC Character; optional column containing fold change values, used only when filterInput=TRUE and the input is a data frame.
columnP Character; optional column containing p values, used only when filterInput=TRUE and the input is a data frame.
pCutoff Adjusted p value cutoff, defaults to <0.05
fcCutoff Absolute fold change cutoff, defaults to an absolute value of >1.5
order Desired network order. Possible options are "zero" (default), "first," "minSimple."
hubMeasure Character denoting what measure should be used in determining which nodes to highlight as hubs when plotting the network. Options include "betweenness" (default), "degree", and "hubscore". These represent network statistics calculated by their respective tidygraph::centrality_x.functions.
ppiData Data frame of PPI data; must contain rows of interactions as pairs of Ensembl gene IDs, with columns named "ensemblGeneA" and "ensemblGeneB". Defaults to pre-packaged InnateDB PPI data.

Details

The input to ppiBuildNetwork() can be a "DESeqResults" object (from DESeq2), "TopTags" (edgeR), or a simple data frame. When not providing a basic data frame, the columns for filtering are automatically pulled ("log2FoldChange" and "padj" for DESeqResults, or "logFC" and "FDR" for TopTags). Otherwise, the arguments "columnFC" and "columnP" must be specified.

The "hubMeasure" argument determines how ppiBuildNetwork assesses connectedness of nodes in the network, which will be used to highlight nodes when visualizing with ppiPlotNetwork. The options are "degree", "betweenness", or "hubscore". This last option uses the igraph implementation of the Kleinburg hub centrality score - details on this method can be found at ?igraph::hub_score.

Value

A Protein-Protein Interaction (PPI) network; a "tidygraph" object for plotting or further analysis, with the minimum set of columns for nodes (additional columns from the input will also be included):

name  Ensembl gene ID for the node
degree  Degree of the node, i.e. the number of interactions
betweenness  Betweenness measure for the node
seed  TRUE when the node was part of the input list of genes
hubScore  Special hubScore for each node. The suffix denotes the measure being used; e.g. "hubScoreBtw" is for betweenness
hgncSymbol  HGNC gene name for the node

Additionally the following columns are provided for edges:

from  Starting node for the interaction/edge as a row number
to  Ending node for the interaction/edge as a row number

References
InnateDB: https://www.innatedb.com/

See Also
https://github.com/hancockinformatics/pathlinkR/

Examples

data("exampleDESeqResults")

ppiBuildNetwork(
  rnaseqResult = exampleDESeqResults[[1]],
  filterInput = TRUE,
  order = "zero"
)

ppiCleanNetwork  Clean GraphML or JSON input

Description
Takes network file (GraphML or JSON) and process it into a tidygraph object, adding network statistics along the way.

Usage

ppiCleanNetwork(network)

Arguments

network  tidygraph object from a GraphML or JSON file

Details
This function was designed so that networks created by other packages or websites (e.g. https://networkanalyst.ca) could be imported and visualized with ppiPlotNetwork.
Value

A Protein-Protein Interaction (PPI) network; a "tidygraph" object, with the minimal set of columns (other from the input are also included):

- **name**: Identifier for the node
- **degree**: Degree of the node, i.e. the number of interactions
- **betweenness**: Betweenness measure for the node
- **seed**: TRUE when the node was part of the input list of genes
- **hubScore**: Special hubScore for each node. The suffix denotes the measure being used; e.g. "hubScoreBtw" is for betweenness
- **hgncSymbol**: HGNC gene name for the node

Additionally the following columns are provided for edges:

- **from**: Starting node for the interaction/edge as a row number
- **to**: Ending node for the interaction/edge as a row number

See Also

https://github.com/hancockinformatics/pathlinkR/

Examples

```r
tj1 <- jsonlite::read_json(  
  system.file("extdata/networkAnalystExample.json", package="pathlinkR"),  
  simplifyVector=TRUE  
)

tj2 <- igraph::graph_from_data_frame(  
  d=dplyr::select(tj1$edges, source, target),  
  directed=FALSE,  
  vertices=dplyr::select(  
    tj1$nodes,  
    id,  
    label,  
    x,  
    y,  
    "types"=molType,  
    expr  
  )  
)

tj3 <- ppiCleanNetwork(tidygraph::as_tbl_graph(tj2))
```
ppiEnrichNetwork  Test a PPI network for enriched pathways

Description
Test a PPI network for enriched pathways

Usage
ppiEnrichNetwork(
  network,
  analysis = "sigora",
  filterResults = "default",
  gpsRepo = "default",
  geneUniverse = NULL
)

Arguments

- network: A "tidygraph" network object, with Ensembl IDs in the first column of the node table.
- analysis: Default is "sigora", but can also be "reactomepa" or "hallmark".
- filterResults: Should the output be filtered for significance? Use 1 to return the unfiltered results, or any number less than 1 for a custom p-value cutoff. If left as default, the significance cutoff for Sigora is 0.001, or 0.05 for ReactomePA and Hallmark.
- gpsRepo: Only applies to analysis="sigora". Gene Pair Signature object for Sigora to use to test for enriched pathways. Leaving this set as "default" will use the "reaH" GPS object from Sigora, or you can provide your own custom GPS repository.
- geneUniverse: Only applies when analysis is "reactomepa" or "hallmark". The set of background genes to use when testing with ReactomePA or Hallmark gene sets. For ReactomePA this must be a character vector of Entrez genes. For Hallmark, it must be Ensembl IDs.

Value
A "data.frame" (tibble) of enriched pathways, with the following columns:

- pathwayId: Pathway identifier
- pathwayName: Pathway name
- pValue: Nominal p value for the pathway
- pValueAdjusted: p value corrected for multiple testing
- genes: Candidate genes, which were DE for the comparison and also in the pathway
ppiExtractSubnetwork

Number of candidate genes
numCandidateGenes

Number of background genes for the pathway
genes

totalGenes

topLevelPathway

Number of DE genes which were tested for enriched pathways

High level Reactome term which serves to group similar pathways

References

Sigora: https://cran.r-project.org/package=sigora
ReactomePA: https://www.bioconductor.org/packages/ReactomePA/
MSigDB/Hallmark: https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp

See Also

https://github.com/hancockinformatics/pathlinkR

Examples

data("exampleDESeqResults")

exNetwork <- ppiBuildNetwork(
  rnaseqResult=exampleDESeqResults[[1]],
  filterInput=TRUE,
  order="zero"
)

ppiEnrichNetwork(
  network=exNetwork,
  analysis="hallmark"
)

ppiExtractSubnetwork Extract a subnetwork based on pathway genes

Description

Extract a subnetwork based on pathway genes

Usage

ppiExtractSubnetwork(
  network,
  genes = NULL,
  pathwayEnrichmentResult = NULL,
  pathwayToExtract
)
**Arguments**

- **network**: Input network object; output from `ppiBuildNetwork()`.
- **genes**: List of Ensembl gene IDs to use as the starting point to extract a subnetwork from the initial network. You must provide either the `genes` or `pathwayEnrichmentResult` argument.
- **pathwayEnrichmentResult**: Pathway enrichment result, output from `ppiEnrichNetwork`. You must provide either `genes` or `pathwayEnrichmentResult` argument.
- **pathwayToExtract**: Name of the pathway determining what genes (nodes) are pulled from the input network. Must be present in the "pathwayName" column of `pathwayEnrichmentResults`.

**Details**

Uses functions from the igraph package to extract a minimally connected subnetwork from the starting network, using either a list of Ensembl genes or genes from an enriched pathway as the basis. To see what genes were pulled out for the pathway, see the "starters" attribute of the output network.

**Value**

A Protein-Protein Interaction (PPI) network; a "tidygraph" object for plotting or further analysis, with the minimum set of columns for nodes (additional columns from the input will also be included):

- **name**: Ensembl gene ID for the node
- **degree**: Degree of the node, i.e. the number of interactions
- **betweenness**: Betweenness measure for the node
- **seed**: TRUE when the node was part of the input list of genes
- **hubScore**: Special hubScore for each node. The suffix denotes the measure being used; e.g. "hubScoreBtw" is for betweenness
- **hgncSymbol**: HGNC gene name for the node

Additionally the following columns are provided for edges:

- **from**: Starting node for the interaction/edge as a row number
- **to**: Ending node for the interaction/edge as a row number

**References**

Code for network module (subnetwork) extraction was based off of that used in "jboktor/NetworkAnalystR" on Github.

**See Also**

[https://github.com/hancockinformatics/pathlinkR](https://github.com/hancockinformatics/pathlinkR)
Examples

```r
data("exampleDESeqResults")

exNetwork <- ppiBuildNetwork(
  rnaseqResult = exampleDESeqResults[[1]],
  filterInput = TRUE,
  order = "zero"
)

exPathways <- ppiEnrichNetwork(
  network = exNetwork,
  analysis = "hallmark"
)

ppiExtractSubnetwork(
  network = exNetwork,
  pathwayEnrichmentResult = exPathways,
  pathwayToExtract = "INTERFERON ALPHA RESPONSE"
)
```

**ppiPlotNetwork**  
Plot an undirected PPI network using ggraph

**Description**

Visualize a protein-protein interaction (PPI) network using ggraph functions, output from ppiBuildNetwork.

**Usage**

```r
ppiPlotNetwork(
  network,
  networkLayout = "nicely",
  title = NA,
  nodeSize = c(2, 6),
  fillColumn,
  fillType,
  catFillColours = "Set1",
  foldChangeColours = c("firebrick3", "#188119"),
  intColour = "grey70",
  nodeBorder = "grey30",
  hubColour = "blue2",
  subnetwork = TRUE,
  legend = FALSE,
  legendTitle = NULL,
  edgeColour = "grey40",
  edgeAlpha = 0.5,
  edgeWidth = 0.5,
)```
label = FALSE,
labelColumn,
labelFilter = 8,
labelSize = 4,
labelColour = "black",
labelFace = "bold",
labelPadding = 0.25,
minSegLength = 0.25
)

Arguments

network A tidygraph object, output from ppiBuildNetwork
networkLayout Layout of nodes in the network. Supports all layouts from ggraph/igraph, or a data frame of x and y coordinates for each node (order matters!).
title Optional title for the plot (NA)
nodeSize Length-two numeric vector, specifying size range of node sizes (maps to node degree). Default is c(2, 6).
fillColumn Tidy-select column for mapping node colour. Designed to handle continuous numeric mappings (either positive/negative only, or both), and categorical mappings, plus a special case for displaying fold changes from, for example, RNA-Seq data. See fillType for more details on how to set this up.
fillType String denoting type of fill mapping to perform for nodes. Options are: "foldChange", "twoSided", "oneSided", or "categorical".
catFillColours Colour palette to be used when fillType is set to "categorical." Defaults to "Set1" from RColorBrewer. Will otherwise be passed as the "values" argument in scale_fill_manual().
foldChangeColours A two-length character vector containing colours for up and down regulated genes. Defaults to c("firebrick3", "#188119").
intColour Fill colour for non-seed nodes, i.e. interactors. Defaults to "grey70".
nodeBorder Colour (stroke or outline) of all nodes in the network. Defaults to "grey30".
hubColour Colour of node labels for hubs. The top 2% of nodes (based on calculated hub score) are highlighted with this colour, if label=TRUE.
subnetwork Logical determining if networks from ppiExtractSubnetwork() should be treated as such. Defaults to TRUE.
legend Should a legend be included? Defaults to FALSE.
legendTitle Optional title for the legend, defaults to NULL.
edgeColour Edge colour, defaults to "grey40"
edgeAlpha Transparency of edges, defaults to 0.5
edgeWidth Thickness of edges connecting nodes. Defaults to 0.5
label Boolean, whether labels should be added to nodes. Defaults to FALSE.
**labelColumn**: Tidy-select column of the network/data to be used in labeling nodes. Recommend setting to `hgncSymbol`, which contains HGNC symbols mapped from the input Ensembl IDs via biomaRt.

**labelFilter**: Degree filter used to determine which nodes should be labeled. Defaults to 0. This value can be increased to reduce the number of node labels, to prevent the network from being too crowded.

**labelSize**: Size of node labels, defaults to 5.

**labelColour**: Colour of node labels, defaults to "black"

**labelFace**: Font face for node labels, defaults to "bold"

**labelPadding**: Padding around the label, defaults to 0.25 lines.

**minSegLength**: Minimum length of lines to be drawn from labels to points. The default specified here is 0.25, half of the normal default value.

**Details**

Any layout supported by ggraph can be specified here - see ?layout_tbl_graph_igraph for a list of options. Or you can supply a data frame containing coordinates for each node. The first and second columns will be used for x and y, respectively. Note that having columns named "x" and "y" in the input network will generate a warning message when supplying custom coordinates.

Since this function returns a standard ggplot object, you can tweak the final appearance using the normal array of ggplot2 function, e.g. `labs()` and `theme()` to further customize the final appearance.

The `fillType` argument will determine how the node colour is mapped to the desired column. "foldChange" represents a special case, where the fill column is numeric and whose values should be mapped to up (> 0) or down (< 0). "twoSided" and "oneSided" are designed for numeric data that contains either positive and negative values, or only positive/negative values, respectively. "categorical" handles any other non-numeric colour mapping, and uses "Set1" from RColorBrewer.

Node statistics (degree, betweenness, and hub score) are calculated using the respective functions from the tidygraph package.

**Value**

A Protein-Protein Interaction (PPI) network plot; an object of class "ggplot"

**See Also**

[https://github.com/hancockinformatics/pathlinkR/](https://github.com/hancockinformatics/pathlinkR/)

**Examples**

data("exampleDESeqResults")

exNetwork <- ppiBuildNetwork(  
  rnaseqResult=exampleDESeqResults[[1]],  
  filterInput=TRUE,  
  order="zero"  )
ppiPlotNetwork(
    network=exNetwork,
    title="COVID positive over time",
    fillColor=LogFoldChange,
    fillType="foldChange",
    legend=TRUE,
    label=FALSE
)

ppiRemoveSubnetworks

**Description**

INTERNAL Find and return the largest subnetwork

**Usage**

ppiRemoveSubnetworks(network)

**Arguments**

- **network**  
  Graph object

**Value**

Largest subnetwork from the input network list as an "igraph" object

**See Also**

https://github.com/hancockinformatics/pathlinkR/

---

**reactomeDatabase**

**Table of all Reactome pathways and genes**

**Description**

Table of all Reactome pathways and genes

**Usage**

data(reactomeDatabase)
sigoraDatabase

**Format**

A data frame (tibble) with 123574 rows and 3 columns

- **pathwayId**  Reactome pathway ID
- **entrezGeneId**  Entrez gene ID
- **pathwayName**  Name of the Reactome pathway

**Value**

An object of class "tbl", "tbl.df", "data.frame"

**Source**

See [https://reactome.org/](https://reactome.org/) for information on each of this pathway resource.

---

### sigoraDatabase

**Table of all Sigora pathways and their constituent genes**

**Description**

Table of all Sigora pathways and their constituent genes

**Usage**

```r
data(sigoraDatabase)
```

**Format**

A data frame (tibble) with 60775 rows and 4 columns

- **pathwayId**  Reactome pathway identifier
- **pathwayName**  Reactome pathway description
- **ensemblGeneId**  Ensembl gene identifier
- **hgncSymbol**  HGNC gene symbol

**Value**

An object of class "tbl", "tbl.df", "data.frame"

**Source**

Please refer to the Sigora package for more details: [https://cran.r-project.org/package=sigora](https://cran.r-project.org/package=sigora)
sigoraExamples  Sigora enrichment example

Description
Example Sigora output from running `pathwayEnrichment()` on "exampleDESeqResults"

Usage
```r
data(sigoraExamples)
```

Format
A data frame (tibble) with 66 rows and 12 columns

- **comparison**: Comparison from which results are derived; names of the input list
- **direction**: Was the pathway enriched in up or down regulated genes
- **pathwayId**: Reactome pathway identifier
- **pathwayName**: Description of the pathway
- **pValue**: Nominal p value for the enrichment
- **pValueAdjusted**: p value adjusted for multiple testing
- **genes**: Genes in the pathway/input
- **numCandidateGenes**: Analyzed genes found in the pathway of interest
- **numBgGenes**: All genes from the pathway database
- **geneRatio**: Quotient of the number of candidate and background genes
- **totalGenes**: Total number of input genes
- **topLevelPathway**: Pathway category

Value
An object of class "tbl", "tbl.df", "data.frame"

Source
Please refer to the Sigora package for more details on that method: [https://cran.r-project.org/package=sigora](https://cran.r-project.org/package=sigora)
Index

**datasets**
exampeldonDESeqResults, 8
 groupedPathwayColours, 10
 hallmarkDatabase, 10
 innateDbPPI, 11
 mappingFile, 12
 pathwayCategories, 20
 reactomeDatabase, 38
 sigoraDatabase, 39
 sigoraExamples, 40

**package**
pathlinkR-package, 3
 .eruptionBreaks, 4
 .plotFoldChangeLegend, 4
 .runSigora, 5
 .truncNeatly, 5

eruption, 3, 6
 exampleDESeqResults, 8

getPathwayDistances, 9
 groupedPathwayColours, 10

hallmarkDatabase, 10

innateDbPPI, 11

mappingFile, 12

pathlinkR (pathlinkR-package), 3
 pathlinkR-package, 3
 pathnetCreate, 3, 12
 pathnetFoundation, 15
 pathnetGGraph, 3, 16
 pathnetVisNetwork, 3, 18
 pathwayCategories, 20
 pathwayEnrichment, 3, 21
 pathwayPlots, 3, 23
 plotFoldChange, 3, 25
 ppiBuildNetwork, 3, 28
 ppiCleanNetwork, 30

ppiEnrichNetwork, 3, 32
 ppiExtractSubnetwork, 3, 33
 ppiPlotNetwork, 3, 35
 ppiRemoveSubnetworks, 38
 reactomeDatabase, 38
 sigoraDatabase, 39
 sigoraExamples, 40