Package ‘gsean’

January 20, 2025

Type Package

Title Gene Set Enrichment Analysis with Networks

Description Biological molecules in a living organism seldom work individually. They usually interact each other in a cooperative way. Biological process is too complicated to understand without considering such interactions. Thus, network-based procedures can be seen as powerful methods for studying complex process. However, many methods are devised for analyzing individual genes. It is said that techniques based on biological networks such as gene co-expression are more precise ways to represent information than those using lists of genes only. This package is aimed to integrate the gene expression and biological network. A biological network is constructed from gene expression data and it is used for Gene Set Enrichment Analysis.

Version 1.26.0

Date 2023-05-24

Author Dongmin Jung

Maintainer Dongmin Jung `<dmdmjung@gmail.com>`

Depends R (>= 3.5), fgsea, PPInfer

Suggests SummarizedExperiment, pasilla, org.Dm.eg.db, AnnotationDbi, knitr, plotly, WGCNA, rmarkdown

License Artistic-2.0

biocViews Software, StatisticalMethod, Network, GraphAndNetwork, GeneSetEnrichment, GeneExpression, NetworkEnrichment, Pathways, DifferentialExpression

NeedsCompilation no

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/gsean

git_branch RELEASE_3_20

git_last_commit b47f1ee

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2025-01-20

Contents

`gsean-package` *Gene Set Enrichment Analysis with Networks*

Description

Biological molecules in a living organism seldom work individually. They usually interact each other in a cooperative way. Biological process is too complicated to understand without considering such interactions. Thus, network-based procedures can be seen as powerful methods for studying complex process. However, many methods are devised for analyzing individual genes. It is said that techniques based on biological networks such as gene co-expression are more precise ways to represent information than those using lists of genes only. This package is aimed to integrate the gene expression and biological network. A biological network is constructed from gene expression data and it is used for Gene Set Enrichment Analysis.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

Author(s)

Dongmin Jung

Maintainer: Dongmin Jung <dmdmjung@gmail.com>

`centrality_gsea` *Gene Set Enrichment Analysis with centrality measure*

Description

GSEA is performed with centrality measure

Usage

```
centrality_gsea(geneset, x, adjacency, pseudo = 1, nperm = 1000,
                centrality = function(x) rowSums(abs(x)),
                weightParam = 1, minSize = 1, maxSize = Inf,
                gseaParam = 1, nproc = 0, BPPARAM = NULL)
```

Arguments

`geneset`	list of gene sets
`x`	Named vector of gene-level statistics. Names should be the same as in gene sets.
`adjacency`	adjacency matrix
`pseudo`	pseudo number for log2 transformation (default: 1)
`nperm`	number of permutations (default: 1000)
`centrality`	centrality measure, degree centrality or node strength is default
`weightParam`	weight parameter value for the centrality measure, equally weight if weightParam = 0 (default: 1)
`minSize`	minimal size of a gene set (default: 1)
`maxSize`	maximal size of a gene set (default: Inf)
`gseaParam`	GSEA parameter value (default: 1)
`nproc`	see fgsea::fgsea
`BPPARAM`	see fgsea::fgsea

Value

GSEA result

Author(s)

Dongmin Jung

See Also

fgsea::fgsea

Examples

```
data(examplePathways)
data(exampleRanks)
exampleRanks <- exampleRanks[1:100]
adjacency <- diag(length(exampleRanks))
rownames(adjacency) <- names(exampleRanks)
set.seed(1)
result.GSEA <- centrality_gsea(examplePathways, exampleRanks, adjacency)
```

`exprs2adj` *Convert gene expression data to adjacency matrix by using correlation coefficients*

Description

A biological network is constructed from gene expression data and it is used for Gene Set Enrichment Analysis.

Usage

```
exprs2adj(x, pseudo = 1, ...)
```

Arguments

`x`	gene expression data
`pseudo`	pseudo number for log2 transformation (default: 1)
`...`	additional parameters for correlation; see WGCNA::cor

Value

adjacency matrix

Author(s)

Dongmin Jung

See Also

fgsea::fgsea, WGCNA::cor

Examples

```
data(exampleRanks)
Names <- names(exampleRanks)
exprs <- matrix(rnorm(10*length(exampleRanks)), ncol = 10)
adjacency <- exprs2adj(exprs)
```

`GO_dme` *Gene Ontology terms with gene ID for Drosophila melanogaster*

Description

The data set contains all Gene Ontology terms for Drosophila melanogaster and genes are identified by gene ID. There are 2823 categories.

Usage

```
GO_dme
```

Format

a list of gene sets

Value

GO gene sets

Author(s)

Dongmin Jung

Source

http://www.go2msig.org/cgi-bin/prebuilt.cgi?taxid=7227

Examples

```
load(system.file("data", "GO_dme.rda", package = "gsean"))
```

`gsean` *Gene Set Enrichment Analysis with Networks*

Description

GSEA or ORA is performed with networks from gene expression data

Usage

```
gsean(geneset, x, exprs, pseudo = 1, threshold = 0.99, nperm = 1000,
      centrality = function(x) rowSums(abs(x)), weightParam = 1,
      minSize = 1, maxSize = Inf, gseaParam = 1, nproc = 0,
      BPPARAM = NULL, corParam = list(), tmax = 10, ...)
```

Arguments

`geneset`	list of gene sets
`x`	Named vector of gene-level statistics for GSEA or set of genes for ORA. Names should be the same as in gene sets.
`exprs`	gene expression data
`pseudo`	pseudo number for log2 transformation (default: 1)
`threshold`	threshold of correlation for nodes to be considered neighbors for ORA (default: 0.99)
`nperm`	number of permutations (default: 1000)
`centrality`	centrality measure, degree centrality or node strength is default
`weightParam`	weight parameter value for the centrality measure, equally weight if weightParam = 0 (default: 1)
`minSize`	minimal size of a gene set (default: 1)
`maxSize`	maximal size of a gene set (default: Inf)
`gseaParam`	GSEA parameter value (default: 1)
`nproc`	see fgsea::fgsea
`BPPARAM`	see fgsea::fgsea
`corParam`	additional parameters for correlation; see WGCNA::cor
`tmax`	maximum number of iterations for label propagtion (default: 10)
`...`	additional parameters for label propagation; see RANKS::label.prop

Value

GSEA result

Author(s)

Dongmin Jung

See Also

exprs2adj, label_prop_gsea, centrality_gsea

Examples

```
data(examplePathways)
data(exampleRanks)
exampleRanks <- exampleRanks[1:100]
Names <- names(exampleRanks)
exprs <- matrix(rnorm(10*length(exampleRanks)), ncol = 10)
rownames(exprs) <- names(exampleRanks)
set.seed(1)
result.GSEA <- gsean(examplePathways, exampleRanks, exprs)
```

`KEGG_hsa` *KEGG pathways with gene symbol for human*

Description

The data set contains 186 KEGG pathways for Drosophila melanogaster and genes are identified by gene symbol.

Usage

```
KEGG_hsa
```

Format

a list of gene sets

Value

KEGG gene sets

Author(s)

Dongmin Jung

Source

http://software.broadinstitute.org/gsea/msigdb/collections.jsp

Examples

```
load(system.file("data", "KEGG_hsa.rda", package = "gsean"))
```

`label_prop_gsea` *Over-representaion analysis with the label propagation algorithm*

Description

ORA is performed by GSEA with the label propagation algorithm

Usage

```
label_prop_gsea(geneset, x, adjacency, threshold = 0.99, nperm = 1000,
                minSize = 1, maxSize = Inf, gseaParam = 1, nproc = 0,
                BPPARAM = NULL, ...)
```

Arguments

`geneset`	list of gene sets
`x`	set of genes
`adjacency`	adjacency matrix
`threshold`	threshold of correlation for nodes to be considered neighbors (default: 0.99)
`nperm`	number of permutations (default: 1000)
`minSize`	minimal size of a gene set (default: 1)
`maxSize`	maximal size of a gene set (default: Inf)
`gseaParam`	GSEA parameter value (default: 1)
`nproc`	see fgsea::fgsea
`BPPARAM`	see fgsea::fgsea
`...`	additional parameters for label propagation; see RANKS::label.prop

Value

GSEA result

Author(s)

Dongmin Jung

See Also

fgsea::fgsea

Examples

```
data(examplePathways)
data(exampleRanks)
exampleRanks <- exampleRanks[1:100]
geneNames <- names(exampleRanks)
set.seed(1)
x <- sample(geneNames, 10)
adjacency <- diag(length(exampleRanks))
rownames(adjacency) <- geneNames
result.GSEA <- label_prop_gsea(examplePathways, x, adjacency)
```

Index