

Package ‘GENEMABR’

May 24, 2019

Title Gene module/list annotation

Version 0.99.13

Author Tao Fang and Jitao David Zhang, with inputs from Daniel Marbach

Maintainer Tao Fang <talonvonfang@gmail.com>

Description

Gene-set module annotation or gene-set enrichment within a regression based framework.

Imports glmnet, igraph, Matrix, utils

Depends R (≥ 3.5.0)

biocViews GeneSetEnrichment,Regression,Pathways,GO, Reactome

License Artistic-2.0

BugReports <https://github.com/TaoDFang/GENEMABR/issues>

LazyData false

Encoding UTF-8

VignetteBuilder knitr

Suggests BiocStyle, knitr, rmarkdown, testthat (≥ 2.1.0)

RoxygenNote 6.1.1

NeedsCompilation no

R topics documented:

find_root_ids	2
fisher_exact_test	2
from_id2name	3
gene_pathway_matrix	4
get_steps	4
human_go_ontology	5
human_go_roots	5
human_go_sub_roots	6
human_reactome_ontology	6
human_reactome_roots	7
mydata	7
regression_selected_pathways	8

Index	10
--------------	-----------

find_root_ids	<i>find_root_ids</i>
---------------	----------------------

Description

If you use the default pathway databases(GO Ontology and REACTOME), this function allows you to extract GO sub-roots or REACTOME roots for certain pathways from GO or REACTOME to help you better understanding their biological meanings

Usage

```
find_root_ids(selected_pathways)
```

Arguments

selected_pathways

A vector of GO and/or REACTOME pathways IDs.

Value

A list of GO sub-root or REACTOME root ids for provided pathways. If a certain pathway has more than one GO sub-roots or REACTOME roots, they will be separated by "#".

Examples

```
find_root_ids(selected_pathways=c("GO:0005834", "R-HSA-111469"))
```

fisher_exact_test	<i>fisher_exact_test</i>
-------------------	--------------------------

Description

This function allows you to compute two sided fisher exact pvalue of gene list for selected pathways. To know more about this method, I recommend you to read the paper (Enrichment or depletion of a GO category within a class of genes: which test?) for more details

Usage

```
fisher_exact_test(selected_pathways, gene_input,
  gene_pathway_matrix = NULL)
```

Arguments

selected_pathways

A vector of pathways to be used for enrichment analysis for genes in *gene_input*. It should have same ID types (E.g. pathway ID, pathway names) as the pathways in *gene_pathway_matrix*.

gene_input

A vector of genes to be annotated. It should have same ID types (E.g. Ensembl ID, HUGO gene symbol) as the genes in *gene_pathway_matrix*.

gene_pathway_matrix

A binary background matrix whose columns are the pathways/gene sets and whose rows are all the genes from pathways/gene sets . It could be in sparse matrix format ((inherit from class "sparseMatrix" as in package Matrix) to save memory. For gene i and pathway j, the value of matrix(i,j) is 1 is gene i belonging to pathway j otherwise 0. Users could leave it as default value then it will use pre-collected gene_pathway_matrix from GO Ontology and REACTOME databaase. Otherwise, they could use their own customized gene_pathway_matrix

Value

A list of two elements:

- selected_pathways_fisher_pvalue - Fisher exact pvalue for selected pathways
- selected_pathways_num_genes - The number of genes for selected pathways in background

Examples

```
fetRes <- fisher_exact_test(selected_pathways=c("GO:0007250","GO:0008625"),
  gene_input=c("TRPC4AP","CDC37","TNIP1","IKBKB","NKIRAS2","NFKBIA","TIMM50",
    "RELB","TNFAIP3","NFKBIB","HSPA1A","NFKBIE","SPAG9","NFKB2","ERLIN1",
    "REL","TNIP2","TUBB6","MAP3K8"),gene_pathway_matrix=NULL)
```

from_id2name

from_id2name

Description

If you use the default pathway databases(GO Ontologyand REACTOME),this function can help you to get pathways names from pathways IDs.

Usage

```
from_id2name(selected_pathways)
```

Arguments

selected_pathways

A list of GO and/or REACTOME pathways IDs. Each elmment is this list can be a single id or multi-ids seperated "#"

Value

A list of GO sub-root or REACTOME root names for provided pathways.

Examples

```
from_id2name((selected_pathways=list(c("GO:0032991#GO:0044425#GO:0044464"),"R-HSA-5357801")))
```

gene_pathway_matrix	<i>Homo sapiens GO ontology and REACTOME ontology gene-pathway relationship</i>
---------------------	---

Description

A rds R object contains GO ontology and REACTOME ontology gene-pathway relationship

Usage

```
gene_pathway_matrix
```

Format

Formal class 'dgCMatrix' [package "Matrix"]

Source

<http://geneontology.org/docs/download-ontology/>, <https://reactome.org/download-data> A binary matrix whose columns are the pathways/gene sets from GO ontology and REACTOME database and whose rows are all the genes (represented by gene HUGO gene symbols) from GO ontology and REACTOME database. For gene i and pathway j, the value of matrix(i,j) is 1 if gene i belongs to pathway j otherwise 0

get_steps	<i>get_steps</i>
-----------	------------------

Description

If you use the default pathway databases (GO Ontology and REACTOME), this function allows you to extract the distances from certain pathways to GO roots or REACTOME root nodes.

Usage

```
get_steps(selected_pathways)
```

Arguments

selected_pathways

A vector of GO and/or REACTOME pathway IDs.

Value

A list contains distances from pathways to GO root or REACTOME root nodes

Examples

```
get_steps(selected_pathways=c("GO:0005834", "R-HSA-111469"))
```

human_go_ontology	<i>Homo sapiens GO ontology tree</i>
-------------------	--------------------------------------

Description

A rds R object contains GO ontology relationships (tree structure)

Usage

```
human_go_ontology
```

Format

Directed igraph format

Details

The igraph format tree was constructed by using data from <http://geneontology.org/docs/download-ontology/> (May 2108) It has three root notes representing Molecular Function, Cellular Component and Biological Process (<http://geneontology.org/docs/ontology-documentation/>)

Source

<http://geneontology.org/docs/download-ontology/>

human_go_roots	<i>human_go_roots</i>
----------------	-----------------------

Description

A rds R object contains GO ontology tree root nodes

Usage

```
human_go_roots
```

Format

A vector of GO ontology root notes (ID)

Details

You can view tree stuctor of GO ontology at <https://www.ebi.ac.uk/QuickGO/> Thre are three roots notes in GO ontology tree: GO:0008150 (biological_process), GO:0003674(molecular_function), GO:0005575(cellular_component)

Source

<http://geneontology.org/docs/download-ontology/>

human_go_sub_roots	<i>human_go_sub_roots</i>
--------------------	---------------------------

Description

A rds R object contains GO ontology tree sub-root nodes (The children of root nodes).

Usage

```
human_go_sub_roots
```

Format

A list of three elements contains GO ontology sub-root notes (ID)/the children of three root notes

Details

You can view tree stuctor of GO ontology at <https://www.ebi.ac.uk/QuickGO/>

Source

<http://geneontology.org/docs/download-ontology/>

human_reactome_ontology	<i>Homo sapiens REACTOME ontology tree</i>
-------------------------	--

Description

A rds R object contains Reactome ontology relationships (tree structure)

Usage

```
human_reactome_ontology
```

Format

Directed igraph format

Details

The igraph format tree was constructed by using data from <https://reactome.org/download-data> (May 2108) It has several root nodes representing REACTOME pathway categories (<https://reactome.org/PathwayBrowser/>)

Source

<https://reactome.org/download-data>

human_reactome_roots	<i>human_reactome_roots</i>
----------------------	-----------------------------

Description

A rds R object contains REACTOME tree root nodes

Usage

```
human_reactome_roots
```

Format

A vector of REACTOME root notes (ID)

Details

You can view tree stuctor of REACTOME at <https://reactome.org/PathwayBrowser/>

Source

<https://reactome.org/download-data>

mydata	<i>Helper function to load file</i>
--------	-------------------------------------

Description

Helper function to load file

Usage

```
mydata(objName, package)
```

Arguments

objName	Object name to be loaded
package	Package name to be loaded

Value

The object This function is temporarily set up to help the transition from .rds to .RData file

```
regression_selected_pathways
      regression_selected_pathways
```

Description

This function allows you to extract enriched pathways for gene module/list via regression (elastic net) based method

Usage

```
regression_selected_pathways(gene_input, gene_pathway_matrix = NULL,
                             lambda = 0.007956622, alpha = 0.5, ...)
```

Arguments

- | | |
|----------------------------|---|
| gene_input | A vector of genes to be annotated. It should have same ID types (Ensembl ID, HUGO gene symbol) as the genes in <i>gene_pathway_matrix</i> . |
| gene_pathway_matrix | A binary background matrix whose columns are the pathways/gene sets and whose rows are all the genes from pathways/gene sets. It could be in sparse matrix format ((inherit from class "sparseMatrix" as in package Matrix) to save memory. For gene i and pathway j, the value of matrix(i,j) is 1 if gene i belongs to pathway j otherwise 0. Users could leave it as default value then it will use pre-collected <i>gene_pathway_matrix</i> from GO Ontology and REACTOME database. Otherwise, they could use their own customized <i>gene_pathway_matrix</i> . |
| lambda | We use glmnet function to do regression. <i>lambda</i> is an argument in glmnet . See glmnet function for more details. Here we use default value 0.007956622 after preliminary study. It can be overridden by giving <i>nlambda</i> and <i>lambda.min.ratio</i> arguments. |
| alpha | The elasticnet mixing parameter, with $0 \leq \alpha \leq 1$. The penalty is defined as $(1 - \alpha)/2 \ \beta\ _2^2 + \alpha \ \beta\ _1$. $\alpha=1$ is the lasso penalty, and $\alpha=0$ the ridge penalty. Default value: 0.5. |
| ... | Other parameters for glmnet function. |

Value

A list of four elements:

- *selected_pathways_names* - Pathways names for selected pathways
- *selected_pathways_coef* - Regression coefficients value for selected pathways
- *selected_pathways_fisher_pvalue* - Fisher exact pvalue for selected pathways
- *selected_pathways_num_genes* - The number of genes for selected pathways in background

Examples

```
rspResults <- regression_selected_pathways(gene_input=c("TRPC4AP", "CDC37",  
  "TNIP1", "IKBKB", "NKIRAS2", "NFKBIA", "TIMM50", "RELB", "TNFAIP3", "NFKBIB",  
  "HSPA1A", "NFKBIE", "SPAG9", "NFKB2", "ERLIN1", "REL", "TNIP2",  
  "TUBB6", "MAP3K8"),  
gene_pathway_matrix=NULL, lambda=0.007956622, alpha=0.5)
```

Index

*Topic **datasets**

- gene_pathway_matrix, [4](#)
- human_go_ontology, [5](#)
- human_go_roots, [5](#)
- human_go_sub_roots, [6](#)
- human_reactome_ontology, [6](#)
- human_reactome_roots, [7](#)

- find_root_ids, [2](#)
- fisher_exact_test, [2](#)
- from_id2name, [3](#)

- gene_pathway_matrix, [4](#)
- get_steps, [4](#)

- human_go_ontology, [5](#)
- human_go_roots, [5](#)
- human_go_sub_roots, [6](#)
- human_reactome_ontology, [6](#)
- human_reactome_roots, [7](#)

- mydata, [7](#)

- regression_selected_pathways, [8](#)