Package ‘pathfindR.data’

December 11, 2022

Title Data Package for 'pathfindR'

Version 1.1.3

Maintainer Ege Ulgen <egeulgen@gmail.com>

Description This is a data-only package, containing data needed to run the CRAN package 'pathfindR', a package for enrichment analysis utilizing active subnetworks. This package contains protein-protein interaction network data, data related to gene sets and example input/output data.

Encoding UTF-8

LazyData true

Depends R (>= 4.0)

RoxygenNote 7.2.2

URL https://github.com/egeulgen/pathfindR.data

BugReports https://github.com/egeulgen/pathfindR.data/issues

NeedsCompilation no

Author Ege Ulgen [cre, cph] (<https://orcid.org/0000-0003-2090-3621>),
Ozan Ozisik [aut] (<https://orcid.org/0000-0001-5980-8002>)

Repository CRAN

Date/Publication 2022-12-11 10:10:02 UTC

R topics documented:

`biocarta_descriptions` *BioCarta Pathways - Descriptions*

Description

A named vector containing the descriptions for each human BioCarta pathway. *Generated on Dec 10, 2022.*

Usage

```
biocarta_descriptions
```

Format

named vector containing 292 character values, the descriptions for the given pathways.

`biocarta_genes` *BioCarta Pathways - Gene Sets*

Description

A list containing the genes involved in each human BioCarta pathway. Each element is a vector of gene symbols located in the given pathway. *Generated on Dec 10, 2022.*

Usage

```
biocarta_genes
```

Format

list containing 292 vectors of gene symbols. Each vector corresponds to a gene set.

`cell_markers_descriptions` *Human Cell Markers - Descriptions*

Description

A named vector containing descriptions of different cell types from different tissues in human. Names of the vectors are Cell Ontology IDs (if available) of the cell types in the following format: "tissue type, cancer type, cell name" For more information, refer to the article: Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic Acids Res. 2022 Oct 27;gkac947. *Generated on Dec 10, 2022.*

Usage

```
cell_markers_descriptions
```

Format

named vector containing 1986 character values, the descriptions for the given human cell types.

`cell_markers_gsets` *Human Cell Markers - Gene Sets*

Description

A list containing the sets of genes that are cell markers of different cell types from different tissues in human. Each element is a vector of cell marker gene symbols for the given cell type. Names correspond to the Cell Ontology ID (if available) of the cell type. For more information, refer to the article: Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic Acids Res. 2022 Oct 27;gkac947. *Generated on Dec 10, 2022.*

Usage

```
cell_markers_gsets
```

Format

list containing 1986 vectors. Each vector corresponds to a cell marker gene set for a given human cell type.

`custom_result` *Custom Gene Set Enrichment Results*

Description

A data frame consisting of pathfindR enrichment analysis results on the example TF target genes data (target gene sets of CREB and MYC). *Generated on Aug 20, 2021.*

Usage

```
custom_result
```

Format

data frame containing 2 rows and 8 columns. Each row is a gene set (the TF target gene sets).

`example_active_snws` *Example Active Subnetworks*

Description

A list of vectors containing genes for each active subnetwork that passed the filtering step. *Generated on Nov 1, 2019.*

Usage

```
example_active_snws
```

Format

list containing 112 vectors. Each vector is the set of genes for the given active subnetwork.

`go_all_genes` *Gene Ontology - All Gene Ontology Gene Sets*

Description

A list containing the genes involved in each GO ontology term. Each element is a vector of gene symbols located in the given gene set. *Generated on Dec 10, 2022.*

Usage

```
go_all_genes
```

Format

list containing 15508 vectors of gene symbols. Each vector corresponds to a GO gene set.

`kegg_descriptions` *KEGG Pathways - Descriptions*

Description

A named vector containing the descriptions for each Homo sapiens KEGG pathway. Names of the vector correspond to the KEGG ID of the pathway. Pathways that did not contain any genes were discarded. *Generated on Dec 10, 2022.*

Usage

```
kegg_descriptions
```

Format

named vector containing 341 character values, the descriptions for the given pathways.

`kegg_genes` *KEGG Pathways - Gene Sets*

Description

A list containing the genes involved in each Homo sapiens KEGG pathway. Each element is a vector of gene symbols located in the given pathway. Names correspond to the KEGG ID of the pathway. Pathways that did not contain any genes were discarded. *Generated on Dec 10, 2022.*

Usage

```
kegg_genes
```

Format

list containing 341 vectors of gene symbols. Each vector corresponds to a pathway.

`mmu_kegg_descriptions` *Mus Musculus KEGG Pathways - Descriptions*

Description

A named vector containing the descriptions for each Mus musculus KEGG pathway. Names of the vector correspond to the KEGG ID of the pathway. Pathways that did not contain any genes were discarded. *Generated on Dec 10, 2022.*

Usage

```
mmu_kegg_descriptions
```

Format

named vector containing 336 character values, the descriptions for the given pathways.

`mmu_kegg_genes` *Mus Musculus KEGG Pathways - Gene Sets*

Description

A list containing the genes involved in each Mus musculus KEGG pathway. Each element is a vector of gene symbols located in the given pathway. Names correspond to the KEGG ID of the pathway. Pathways that did not contain any genes were discarded. *Generated on Dec 10, 2022.*

Usage

```
mmu_kegg_genes
```

Format

list containing 336 vectors of gene symbols. Each vector corresponds to a pathway.

`myeloma_input` *Example Input for Myeloma Analysis (Mus Musculus)*

Description

A dataset containing the differentially-expressed genes and adjusted p-values for the GEO dataset GSE99393. The RNA microarray experiment was perform to detail the global program of gene expression underlying polarization of myeloma-associated macrophages by CSF1R antibody treatment. The samples were 6 murine bone marrow derived macrophages co-cultured with myeloma cells (myeloma-associated macrophages), 3 of which were treated with CSF1R antibody (treatment group) and the rest were treated with control IgG antibody (control group). In this dataset, differentially-expressed genes with |logFC| >= 2 and FDR < 0.05 are presented. *Generated on Nov 1, 2019.*

Usage

```
myeloma_input
```

Format

A data frame with 45 rows and 2 variables:

Gene_Symbol MGI gene symbols of the differentially-expressed genes

FDR adjusted p values, via the Benjamini & Hochberg (1995) method

Source

`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99393`

See Also

`myeloma_output` for the example mmu enrichment output. `run_pathfindR` for details on the pathfindR enrichment analysis.

`myeloma_output` *Example Output for Myeloma Analysis (Mus Musculus)*

Description

A dataset containing the results of pathfindR's active-subnetwork-oriented enrichment workflow performed on the Mus musculus myeloma differential expression dataset `myeloma_input`. *Generated on Dec 11, 2022.*

Usage

```
myeloma_output
```

Format

A data frame with 26 rows and 9 columns:

ID ID of the enriched term

Term_Description Description of the enriched term

Fold_Enrichment Fold enrichment value for the enriched term

occurrence the number of iterations that the given term was found to enriched over all iterations

support the median support (proportion of active subnetworks leading to enrichment within an iteration) over all iterations

lowest_p the lowest adjusted-p value of the given term over all iterations

highest_p the highest adjusted-p value of the given term over all iterations

Up_regulated the up-regulated genes in the input involved in the given term, comma-separated

Down_regulated the down-regulated genes in the input involved in the given term, comma-separated

See Also

`myeloma_input` for the example mmu input. `run_pathfindR` for details on the pathfindR enrichment workflow.

`pathfindR.data_updates`

Table of Data for pathfindR

Description

Data frame containing all the data for pathfindR along with descriptions and last update dates.

Usage

```
pathfindR.data_updates
```

Format

A data frame with 30 rows and 4 columns:

Category Category of the data

Name Name of the data

Description Description of the data

Last_Update Last update date

`RA_clustered` *Example Output for the pathfindR Clustering Workflow - Rheumatoid Arthritis*

Description

A dataset containing the results of pathfindR's clustering and partitioning workflow performed on the rheumatoid arthritis enrichment results `RA_output`. The clustering and partitioning function `cluster_enriched_terms` was used with the default settings (i.e. hierarchical clustering was performed and the agglomeration method was "average"). *Generated on Dec, 2022.*

Usage

```
RA_clustered
```

Format

A data frame with 128 rows and 11 columns:

ID ID of the enriched term

Term_Description Description of the enriched term

Fold_Enrichment Fold enrichment value for the enriched term

occurrence the number of iterations that the given term was found to enriched over all iterations

support the median support (proportion of active subnetworks leading to enrichment within an iteration) over all iterationss

lowest_p the lowest adjusted-p value of the given term over all iterations

highest_p the highest adjusted-p value of the given term over all iterations

Up_regulated the up-regulated genes in the input involved in the given term, comma-separated

Down_regulated the down-regulated genes in the input involved in the given term, comma-separated

Cluster the cluster to which the enriched term is assigned

Status whether the enriched term is the "Representative" term in its cluster or only a "Member"

See Also

`RA_input` for the RA differentially-expressed genes data frame `RA_exp_mat` for the RA differentially-expressed genes expression matrix `run_pathfindR` for details on the pathfindR enrichment analysis `RA_output` for the RA example pathfindR enrichment output `cluster_enriched_terms` for details on clustering methods

---

`RA_comparison_output` *Second Example Output for the pathfindR Enrichment Workflow*

Description

The data frame containing the results of pathfindR's active-subnetwork-oriented enrichment workflow performed on the rheumatoid arthritis dataset GSE84074 `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84074`. Analysis via `run_pathfindR` was performed using the default settings. *Generated on Dec 11, 2022.*

Usage

```
RA_comparison_output
```

Format

A data frame with 53 rows and 9 columns:

ID ID of the enriched term

Term_Description Description of the enriched term

Fold_Enrichment Fold enrichment value for the enriched term

occurrence the number of iterations that the given term was found to enriched over all iterations

support the median support (proportion of active subnetworks leading to enrichment within an iteration) over all iterations

lowest_p the lowest adjusted-p value of the given term over all iterations

highest_p the highest adjusted-p value of the given term over all iterations

Up_regulated the up-regulated genes in the input involved in the given term, comma-separated

Down_regulated the down-regulated genes in the input involved in the given term, comma-separated

See Also

`RA_input` for the RA differentially-expressed genes data frame `RA_output` for the RA example pathfindR enrichment output `RA_clustered` for the RA example pathfindR clustering output `RA_exp_mat` for the RA differentially-expressed genes expression matrix `run_pathfindR` for details on the pathfindR enrichment analysis

---

`RA_exp_mat` *Example Input for pathfindR - Enriched Term Scoring*

Description

A matrix containing the log_2-transformed and quantile-normalized expression values of the differentially-expressed genes for 18 rheumatoid arthritis (RA) patients and 15 healthy subjects. The matrix contains expression values of 572 significantly differentially-expressed genes (see `RA_input`) with adj.P.Val <= 0.05. *Generated on Sep 28, 2019.*

Usage

```
RA_exp_mat
```

Format

A matrix with 572 rows and 33 columns.

Source

`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15573`

See Also

`RA_input` for the RA differentially-expressed genes data frame `RA_output` for the RA example pathfindR enrichment output `score_terms` for details on calculating agglomerated scores of enriched terms

---

`RA_input` *Example Input for the pathfindR Enrichment Workflow - Rheumatoid Arthritis*

Description

A dataset containing the differentially-expressed genes along with the associated log_2(fold-change) values and FDR adjusted p-values for the GEO dataset GSE15573. This microarray dataset aimed to characterize gene expression profiles in the peripheral blood mononuclear cells of 18 rheumatoid arthritis (RA) patients versus 15 healthy subjects. Differentially-expressed genes with adj.P.Val < 0.05 are presented in this data frame. *Generated on Nov 1, 2019.*

Usage

```
RA_input
```

Format

A data frame with 572 rows and 3 variables:

Gene.symbol HGNC gene symbols of the differentially-expressed genes

logFC log_2(fold-change) values

adj.P.Val adjusted p values, via the Benjamini & Hochberg (1995) method

Source

`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15573`

See Also

`RA_output` for the RA example pathfindR enrichment output `RA_clustered` for the RA example pathfindR clustering output `RA_exp_mat` for the RA differentially-expressed genes expression matrix `run_pathfindR` for details on the pathfindR enrichment analysis

---

`RA_output` *Example Output for the pathfindR Enrichment Workflow - Rheumatoid Arthritis*

Description

The data frame containing the results of pathfindR's active-subnetwork-oriented enrichment workflow performed on the rheumatoid arthritis differential-expression data frame `RA_input`. Analysis via `run_pathfindR` was performed using the default settings. *Generated on Dec 11, 2022.*

Usage

```
RA_output
```

Format

A data frame with 128 rows and 9 columns:

ID ID of the enriched term

Term_Description Description of the enriched term

Fold_Enrichment Fold enrichment value for the enriched term

occurrence the number of iterations that the given term was found to enriched over all iterations

support the median support (proportion of active subnetworks leading to enrichment within an iteration) over all iterations

lowest_p the lowest adjusted-p value of the given term over all iterations

highest_p the highest adjusted-p value of the given term over all iterations

Up_regulated the up-regulated genes in the input involved in the given term, comma-separated

Down_regulated the down-regulated genes in the input involved in the given term, comma-separated

See Also

`RA_input` for the RA differentially-expressed genes data frame `RA_clustered` for the RA example pathfindR clustering outputs `RA_exp_mat` for the RA differentially-expressed genes expression matrix `run_pathfindR` for details on the pathfindR enrichment analysis

---

`reactome_descriptions` *Reactome Pathways - Descriptions*

Description

A named vector containing the descriptions for each human Reactome pathway. Names of the vector correspond to the Reactome ID of the pathway. *Generated on Dec 10, 2022.*

Usage

```
reactome_descriptions
```

Format

named vector containing 2593 character values, the descriptions for the given pathways.

`reactome_genes` *Reactome Pathways - Gene Sets*

Description

A list containing the genes involved in each human Reactome pathway. Each element is a vector of gene symbols located in the given pathway. Names correspond to the Reactome ID of the pathway. *Generated on Dec 10, 2022.*

Usage

```
reactome_genes
```

Format

list containing 2593 vectors of gene symbols. Each vector corresponds to a pathway.

Index