Package 'SBGNview.data'

January 19, 2022

Title Supporting datasets for SBGNview package

Description This package contains:
1. A microarray gene expression dataset from a human breast cancer study.
2. A RNA-Seq gene expression dataset from a mouse study on IFNG knockout.
3. ID mapping tables between gene IDs and SBGN-ML file glyph IDs.
4. Percent of orthologs detected in other species of the genes in a pathway. Cutoffs of this percentage for defining if a pathway exists in another species.
5. XML text of SBGN-ML files for all pre-collected pathways.

Version 1.8.0

Author Xiaoxi Dong*, Kovidh Vegesna*, Weijun Luo

Maintainer Weijun Luo <luo_weijun@yahoo.com>

Depends R (>= 3.6)

License AGPL-3

Collate data.R

LazyData FALSE

Imports knitr, rmarkdown, bookdown

Suggests SummarizedExperiment

RoxygenNote 7.1.1

VignetteBuilder knitr

biocViews ExperimentData, CancerData, BreastCancerData, MicroarrayData, GEO, RNASeqData

git_url https://git.bioconductor.org/packages/SBGNview.data

git_branch RELEASE_3_14

git_last_commit f29ff8b

git_last_commit_date 2021-10-26

Date/Publication 2022-01-19

R topics documented:

`cancer.ds` *A demo microarray dataset from a cancer study*

Description

A demo microarray dataset from a cancer study

Usage

```
data(cancer.ds)
```

Format

A SummarizedExperiment object.

Details

This dataset is constructed using the first three columns of data **gse16873.d** in package **pathview** (i.e. columns "DCIS_1","DCIS_2" and "DCIS_3"). the original values were used without additional processing. It is constructed for showing SBGNview's visualization ability, not for data analysis. Each column in the assay table is a pair of cancer-v.s.-control samples. The value of a gene in a column is the log fold change of this gene in the corresponding pair of cancer-v.s.-control samples.

`IFNg` *RNA-Seq result from a mouse IFNG knockout experiment*

Description

RNA-Seq result from a mouse IFNG knockout experiment

Usage

```
data(IFNg)
```

Format

A SummarizedExperiment object.

Details

This RNA-Seq dataset contains RNA abundance table of two groups: IFNG knockout mice and wild type mice. RNA abundance values were log2 transformed. For demo purpose, data of 4 IFNG knockout mice and 4 wild type mice were included. The experiment and data processing was described in this work: Greer, Renee L., Xiaoxi Dong, et al. "Akkermansia muciniphila mediates negative effects of IFNG on glucose metabolism." Nature communications 7 (2016): 13329.

`mapping` *Mapping table between two types of IDs*

Description

Mapping table between two types of IDs

Usage

```
data(ENZYME_pathway.id)

data(hsa_KO_ENTREZID)

data(mmu_KO_ENSEMBL)

data(chebi_pathway.id)

data(mmu_KO_ENTREZID)

data(chebi_compound.name)

data(compound.name_pathwayCommons)

data(kegg_pathwayCommons)

data(hsa_pathwayCommons_ENSEMBL)

data(mmu_pathwayCommons_ENTREZID)

data(KO_pathway.id)

data(KO_pathwayCommons)

data(SYMBOL_pathway.id)

data(pathwayCommons_SYMBOL)

data(chebi_pathwayCommons)
```

```
data(mmu_pathwayCommons_ENSEMBL)

data(hsa_ENTREZID_pathwayCommons)

data(chebi_kegg)

data(chebi_metacyc.SBGN)

data(compound.name_pathway.id)

data(compound.name_kegg)
```

Format

A matrix/data.frame with two columns: the ID mapping between two types of IDs.

An object of class `data.frame` with 16601 rows and 2 columns.

An object of class `data.frame` with 10446 rows and 3 columns.

An object of class `matrix` (inherits from `array`) with 11016 rows and 2 columns.

An object of class `data.frame` with 79911 rows and 2 columns.

An object of class `data.frame` with 11105 rows and 3 columns.

An object of class `data.frame` with 285207 rows and 2 columns.

An object of class `data.frame` with 298727 rows and 2 columns.

An object of class `data.frame` with 6448 rows and 2 columns.

An object of class `matrix` (inherits from `array`) with 383983 rows and 2 columns.

An object of class `matrix` (inherits from `array`) with 628413 rows and 2 columns.

An object of class `data.frame` with 47719 rows and 2 columns.

An object of class `data.frame` with 88148 rows and 2 columns.

An object of class `data.frame` with 252730 rows and 2 columns.

An object of class `data.frame` with 406940 rows and 3 columns.

An object of class `matrix` (inherits from `array`) with 60425 rows and 2 columns.

An object of class `matrix` (inherits from `array`) with 619656 rows and 2 columns.

An object of class `matrix` (inherits from `array`) with 384671 rows and 2 columns.

An object of class `data.frame` with 20692 rows and 2 columns.

An object of class `matrix` (inherits from `array`) with 60659 rows and 2 columns.

An object of class `data.frame` with 571283 rows and 2 columns.

An object of class `data.frame` with 61143 rows and 2 columns.

Details

Each dataset contains a mapping table. There are several types of ID pairs, such as molecule ID <=> pathway_glyph_ID, molecule ID <=> pathway ID, and molecule ID <=> KEGG ortholog ID. molecule ID <=> pathway_glyph_ID tables are extracted from Biopax files. For example:

http://www.pathwaycommons.org/archives/PC3/v10/PathwayCommons10.reactome.BIOPAX.owl.gz. Glyph IDs are extracted from the ID of each XML element "Protein". Its matching molecule ID is extracted from the corresponding XML child element "UnificationXref". See more details and examples in vignette 'SBGNview.data.vignette'

`pathway.completeness.cutoff.info`
Cutoffs of pathway completeness used for defining existance of pathway in a species

---

Description

Cutoffs of pathway completeness used for defining existance of pathway in a species

Usage

```
data(pathway.completeness.cutoff.info)
```

Format

A matrix

Details

PathwayCommons only annotated human pathways, we mapped pathwayCommons' genes to other species using KEGG ortholog annotation. As a result, not all of the genes have corresponding genes in another species. We call the percentage of mapped genes the "coverage or completeness" in the species. To determin if a pathway exists in a species, we use a cutoff for this completeness. This cutoff is selected using the following approach: 1. A pathway has different completeness in different species thus form a completeness vector across all species (vector C) . 2. Use a completeness cutoff we can define whether this pathway "exists" in a species, thus form a label vector E (a pathway "Exist" or "not Exist" across all species). 3. Use one way ANOVA to calculate F statistic of completeness between the two groups ("Exist" or "not Exist"), thus one cutoff will have one F statistic. 4. Try different cutoffs(unique completeness values in vector C) and select the one with the largest F statistic, i.e. the cutoff the can maximize the difference between Exist" and "not Exist" groups. This is not a perfect way to define if a pathway exists in a species, but can serve as a filter criteria.

pathway.species.pct_Mapped *Pathway completeness in a species*

Description

Pathway completeness in a species

Usage

```
data(pathway.species.pct_Mapped)
```

Format

A matrix

Details

PathwayCommons only annotated human pathways, we mapped pathwayCommons' genes to other species using KEGG ortholog annotation. As a result, not all of the genes have corresponding genes in another species. We call the percentage of mapped genes the "coverage or completeness" in the species.

sbgn.xmls *XML code of a SBGN-ML file*

Description

XML code of a SBGN-ML file

Usage

```
data(sbgn.xmls)
```

Format

A list of character strings

Details

Each string is the full XML code of a SBGN-ML file. It includes glyphs and arcs of a SBGN map. ***Note:** Please note that `sbgn.xmls` is a large R object and will take a few seconds to load. It is necessary to be loaded into the environment in order to access the pre-generated SBGN-ML files.

Index