Package 'NetActivityData'

March 13, 2025

Type Package

Title Data required for getting the gene set scores with NetActivity package

Version 1.9.0

Description This package contains the weights from pre-trained shallow sparsely-connected autoencoders. This data is required for getting the gene set scores with NetActivity package.

License MIT + file LICENSE

Encoding UTF-8

LazyData false

Depends R (>= 4.2.0)

RoxygenNote 7.2.1

Suggests BiocStyle, knitr

VignetteBuilder knitr

biocViews ExperimentData, RepositoryData

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

git_branch devel

git_last_commit dc59833

git_last_commit_date 2024-10-29

Repository Bioconductor 3.21

Date/Publication 2025-03-13

Author Carlos Ruiz-Arenas [aut, cre]

Maintainer Carlos Ruiz-Arenas <`carlos.ruiza@upf.edu`>

Contents

gtex_gokegg 2
gtex_gokegg_annot 2
NetActivityData 3
tcga_gokegg 3
tcga_gokegg_annot 3

Index 4

`gtex_gokegg` *Weights of GO_KEGG model trained with GTEx samples*

Description

We trained a shallow sparsely-connected autoencoder using GO biological processes and KEGG pathways as gene sets. The model was trained using all the GTEx samples (in total 7,729 samples for training). The matrix contains the weights obtained after training the model. Dimensions of the matrix correspond to the number of genes in the original matrix (8,758 genes) and the selected gene sets (1,518 gene sets). Although the weights sign is arbitrary, genes not included in a gene set have a weight of zero. Rownames correspond to gene set names (either GO term id or KEGG pathway id) while colnames correspond to the genes' ENSEMBL id.

Format

A matrix with 1518 rows (gene sets) and 8758 columns (genes). Colnames and rownames include gene set and gene names.

`gtex_gokegg_annot` *Annotation of GO_KEGG model trained with GTEx samples*

Description

We provide the annotation for the gene sets present in the GO_KEGG model trained with the GTEx samples. This object contains the GO terms and KEGG pathways names as well as the genes weights.

Format

A data.frame with 1518 rows corresponding to (gene sets).

- Term Gene set full name
- GeneSet Gene set ID. This ID matches the gene set id in the weights' matrix.
- Weights List with gene weights used to compute the gene set score. Gene names correspond to ENSEMBL ids.
- Weights_SYMBOL List with gene weights used to compute the gene set score. Gene names correspond to SYMBOL ids.

`NetActivityData` *NetActivityData: data required for getting the gene set scores with NetActivity package*

Description

NetActivityData contains the weights from pre-trained shallow sparsely-connected autoencoders. This data is required for getting the gene set scores with NetActivity package.

`tcga_gokegg` *Weights of GO_KEGG model trained with TCGA samples*

Description

We trained a shallow sparsely-connected autoencoder using GO biological processes and KEGG pathways as gene sets. The model was trained using TCGA samples from multiple tumors (in total 8,874 samples for training). The matrix contains the weights obtained after training the model. Dimensions of the matrix correspond to the number of genes in the original matrix (8,758 genes) and the selected gene sets (1,518 gene sets). Although the weights sign is arbitrary, genes not included in a gene set have a weight of zero. Rownames correspond to gene set names (either GO term id or KEGG pathway id) while colnames correspond to the genes' ENSEMBL id.

Format

A matrix with 1518 rows (gene sets) and 8758 columns (genes). Colnames and rownames include gene set and gene names.

`tcga_gokegg_annot` *Annotation of GO_KEGG model trained with TCGA samples*

Description

We provide the annotation for the gene sets present in the GO_KEGG model trained with the TCGA samples. This object contains the GO terms and KEGG pathways names as well as the genes weights.

Format

A data.frame with 1518 rows corresponding to (gene sets).

- Term Gene set full name
- GeneSet Gene set ID. This ID matches the gene set id in the weights' matrix.
- Weights List with gene weights used to compute the gene set score. Gene names correspond to ENSEMBL ids.
- Weights_SYMBOL List with gene weights used to compute the gene set score. Gene names correspond to SYMBOL ids.

Index

gtex_gokegg, 2
gtex_gokegg_annot, 2

NetActivityData, 3

tcga_gokegg, 3
tcga_gokegg_annot, 3