Package 'dorothea'

October 16, 2021

Type Package

Title Collection Of Human And Mouse TF Regulons

Version 1.4.2

Description This package contains human and mouse TF regulons.
The human regulons were curated and collected from different types of
evidence such as literature curated resources, ChIP-seq peaks, TF binding
site motifs and interactions inferred directly from gene expression. The
mouse regulons were constructed by mapping the human gene symbols to their
orthologs in mice. Those regulons can be coupled with any statistical method
that aims to analyse gene sets to infer TF activity from gene expression
data. Preferably the statistical method viper is used.

URL https://saezlab.github.io/dorothea/,
https://github.com/saezlab/dorothea

BugReports https://github.com/saezlab/dorothea/issues

Depends R (>= 4.0)

License GPL-3 + file LICENSE

Encoding UTF-8

LazyData true

LazyDataCompression xz

RoxygenNote 7.1.1

VignetteBuilder knitr

biocViews ExperimentData, Homo_sapiens_Data, Mus_musculus_Data

Imports dplyr, magrittr, bcellViper, viper

Suggests Biobase, BiocStyle, knitr, pheatmap, pkgdown, rmarkdown,
Seurat, SingleCellExperiment, SummarizedExperiment, testthat
(>= 2.1.0), tibble, tidyr, utils

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

git_branch RELEASE_3_13

git_last_commit cd0c31d

git_last_commit_date 2021-10-08

Date/Publication 2021-10-16

Author Christian H. Holland [cre, aut]
  (<<https://orcid.org/0000-0002-3060-5786>>),
  Luz Garcia-Alonso [aut],
  Alberto Valdeolivas [ctb],
  Minoo Ashtiani [ctb],
  Attila Gabor [ctb]

Maintainer Christian H. Holland <cholland2408@gmail.com>

R topics documented:

---

df2regulon   *Construction of dorothea regulons for viper analysis*

Description

This function converts DoRothEA's regulons that are stored in a table to the format required by the [viper](#) function.

Usage

```
df2regulon(df)
```

Arguments

df   A regulon table from dorothea package.

Value

Regulons in the viper format.

Examples

```
# acessing (human) dorothea regulons
# for mouse regulons: data(dorothea_mm, package = "dorothea")
data(dorothea_hs, package = "dorothea")
# convert to the format required by viper
viper_regulons = df2regulon(dorothea_hs)
```

---

dorothea_hs *Human DoRothEA*

Description

A table reporting signed human TF-target interactions. This database covers in total 1395 TFs targeting 20,244 genes with 486,676 unique interactions. In addition, each TF is accompanied with an empirical confidence level that was derived from the number of supporting evidences for this TF/interaction. The range is from A (high quality) to E (low quality).

Format

A table of human TF-target interactions:

tf TF identifier as HGNC symbols

confidence Summary confidence score classifying regulons based on their quality

target target identifier as HGNC symbols

mor mode of regulation indicating the effect of a TF on the target

Source

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6673718/

---

dorothea_hs_pancancer *Human DoRothEA (pancancer)*

Description

A table reporting signed human TF-target interactions for cancer application. The difference to "dorothea_hs" is that TGCA gene expression data was used instead of GTEx to infer the networks with ARACNE. This database covers in total 1344 TFs targeting 20,582 genes with 213,230 unique interactions. In addition, each TF is accompanied with an empirical confidence level that was derived from the number of supporting evidences for this TF/interaction. The range is from A (high quality) to E (low quality).

Format

A table of human TF-target interactions:

tf TF identifier as HGNC symbols

confidence Summary confidence score classifying regulons based on their quality

target target identifier as HGNC symbols

mor mode of regulation indicating the effect of a TF on the target

Source

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6673718/

---

dorothea_mm *Mouse DoRothEA*

Description

A table reporting signed human TF-target interactions. This database covers in total 1179 TFs targeting 17,410 genes with 410,455 unique interactions. In addition, each TF is accompanied with an empirical confidence level that was derived from the number of supporting evidences for this TF/interaction. The range is from A (high quality) to E (low quality).

Format

A table of mouse TF-target interactions:

tf TF identifier as MGI symbols

confidence summary confidence score classifying regulons based on their quality

target target identifier as MGI symbols

mor mode of regulation indicating the effect of a TF on the target

Source

https://www.ncbi.nlm.nih.gov/pubmed/31525460

---

dorothea_mm_pancancer *Mouse DoRothEA (pancancer)*

Description

A table reporting signed human TF-target interactions. The difference to "dorothea_mm" is that TGCA gene expression data was used instead of GTEx to infer the networks with ARACNE.This database covers in total 1096 TFs targeting 17,695 genes with 187,955 unique interactions. In addition, each TF is accompanied with an empirical confidence level that was derived from the number of supporting evidences for this TF/interaction. The range is from A (high quality) to E (low quality).

Format

A table of mouse TF-target interactions:

tf TF identifier as MGI symbols

confidence summary confidence score classifying regulons based on their quality

target target identifier as MGI symbols

mor mode of regulation indicating the effect of a TF on the target

Source

https://www.ncbi.nlm.nih.gov/pubmed/31525460

---

entire_database | *Entire database with associated meta data*

Description

This table lists all human TF-target interactions that were derived from the four lines of evidences. Each interaction is assigned a confidence score based on the number of supporting evidences. The table provides also all required information to trace back the origin of the interaction.

Source

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6673718/

---

run_viper | *VIPER wrapper*

Description

This function is a convenient wrapper for the viper function using DoRothEA regulons.

Usage

```
run_viper(input, regulons, options = list(), tidy = FALSE, assay_key = "RNA")
```

Arguments

input
: An object containing a gene expression matrix with genes (HGNC/MGI symbols) in rows and samples in columns. The object can be a simple matrix/data frame or more complex objects such as ExpressionSet, Seurat or SingleCellExperiment objects.

regulons
: DoRothEA regulons in table format.

options
: A list of named options to pass to viper such as minsize or method. These options should not include, eset or regulon.

tidy
: Logical, whether computed TF activities scores should be returned in a tidy format.

assay_key
: Only applies if the input is a Seurat object. It selects the name of the assay from where to extract the normalized expression data.

Value

A matrix of normalized enrichment scores for each TF across all samples. Of note, if you provide Bioconductor objects as input the function will return this object with added TF activities at appropriate slots. e.g. Seurat object with a new assay called dorothea. For all other inputs the function will return a matrix. If tidy is TRUE the normalized enrichment scores are returned in a tidy format (not supported for Bioconductor objects).

Examples

```
# use example gene expression matrix from bcellViper package
library(bcellViper)
data(bcellViper, package = "bcellViper")
# acessing (human) dorothea regulons
# for mouse regulons: data(dorothea_mm, package = "dorothea")
data(dorothea_hs, package = "dorothea")
# run viper
tf_activities <- run_viper(dset, dorothea_hs,
                           options =  list(method = "scale", minsize = 4,
                           eset.filter = FALSE, cores = 1,
                           verbose = FALSE))
```

Index