Package 'muleaData'

April 24, 2025

Title Genes Sets for Functional Enrichment Analysis with the 'mulea' R Package

Version 1.4.0

Date 2023-11-20

Description ExperimentHubData package for the 'mulea' comprehensive overrepresentation and functional enrichment analyser R package. Here we provide ontologies (gene sets) in a data.frame for 27 different organisms, ranging from Escherichia coli to human, all acquired from publicly available data sources. Each ontology is provided with multiple gene and protein identifiers. Please see the NEWS file for a list of changes in each version.

License MIT + file LICENSE

URL https://github.com/ELTEbioinformatics/muleaData

BugReports https://support.bioconductor.org/tag/muleaData/issues

biocViews ExperimentData, ExperimentHub, Arabidopsis_thaliana_Data, Bacillus_subtilis_Data, Caenorhabditis_elegans_Data, Danio_rerio_Data, Drosophila_melanogaster_Data, Escherichia_coli_Data, Homo_sapiens_Data, Pan_troglodytes_Data, Pseudomonas_aeruginosa_Data, Rattus_norvegicus_Data, Saccharomyces_cerevisiae_Data, Staphylococcus_aureus_Data, ChIPSeqData, DNASeqData, ExpressionData, miRNAData

BiocType ExperimentData

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.1

Suggests knitr, rmarkdown, ExperimentHub, dplyr

VignetteBuilder knitr

NeedsCompilation no

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

git_branch RELEASE_3_21

git_last_commit 6d230c1

git_last_commit_date 2025-04-15

Repository Bioconductor 3.21

Date/Publication 2025-04-24

Author Eszter Ari [aut, cre] (ORCID: <https://orcid.org/0000-0001-7774-1067>),
Márton Ölbei [aut] (ORCID: <https://orcid.org/0000-0002-4903-6237>),
Lejla Gul [aut],
Balázs Bohár [aut] (ORCID: <https://orcid.org/0000-0002-3033-5448>),
Tamás Stirling [aut] (ORCID: <https://orcid.org/0000-0002-8964-6443>)

Maintainer Eszter Ari <arieszter@gmail.com>

Contents

muleaData 2

Index **5**

`muleaData` *muleaData - GMT Datasets for the mulea Package*

Description

muleaData provides pre-processed gene set (ontology) data for use with the `mulea` R package, a comprehensive tool for overrepresentation and functional enrichment analysis. `mulea` leverages ontologies (gene and protein sets) stored in the standardized Gene Matrix Transposed (GMT) format. We provide these GMT files for 27 different model organisms, ranging from *Escherichia coli* to human. These files are compiled from publicly available sources and include various gene and protein identifiers like 'UniProt' protein IDs, 'Entrez', 'Gene Symbol', and 'Ensembl' gene IDs. The GMT files and the scripts we applied to create them are available at the 'ELTEbioinformatics/GMT_files_for_mulea' github repository. For the `muleaData` we read these GMT files with the `mulea::read_gmt()` function and saved them to *.rds* files with the standard R `saveRDS()` function.

Usage

```
query(ExperimentHub::ExperimentHub(), "muleaData")
```

Format

An 'ExperimentHub' object data frame with 879 records containing gene sets (ontologies) from 27 species, collected from 16 data sources. Each record contains a `data.frame` created from a gene set stored in Gene Matrix Transposed (GMT) format, read by the `mulea::read_gmt()` command.

These `data.frame` GMT objects contains 3 columns:

- 'ontology_id': Ontology identifier that uniquely identifies the element within the referenced ontology.

- 'ontology_name': Ontology name or description that provides a user-friendly label or textual description for the 'ontology_id'.
- 'list_of_values': Associated genes or proteins that is a vector of identifiers of genes or proteins belonging to the 'ontology_id'.

Details

The muleaData object is generated dynamically by querying the `ExperimentHub` package.

Source

The data were downloaded from these sources and formatted to GMT files:

- http://atrm.gao-lab.org/
- http://flyatlas.org/atlas.cgi
- http://geneontology.org/
- http://signalink.org/
- http://www.yeastract.com
- https://github.com/saezlab/dorothea
- https://mirtarbase.cuhk.edu.cn/~miRTarBase/miRTarBase_2022/php/index.php
- https://reactome.org/
- https://regulondb.ccg.unam.mx/
- https://tflink.net/
- https://wiki.flybase.org/wiki/FlyBase:ModENCODE_data_at_FlyBase
- https://www.ebi.ac.uk/interpro/
- https://www.ensembl.org
- https://www.grnpedia.org/trrust/
- https://www.pathwaycommons.org/
- https://www.wikipathways.org

See Also

`ExperimentHub`, `query`

Examples

```
# Calling the ExperimentHub library.
library(ExperimentHub)

# Downloading the metadata from ExperimentHub.
eh <- ExperimentHub()

# Creating the muleaData variable.
muleaData <- query(ExperimentHub::ExperimentHub(), "muleaData")

# Checking the muleaData variable.
```

```
muleaData

# Calling dplyr library to explore the data
library(dplyr)

# Looking for the ExperimentalHub ID of i.e. target genes of transcription
# factors from TFLink in Caenorhabditis elegans.
mcols(muleaData) %>%
  as.data.frame() %>%
  dplyr::filter(species == "Caenorhabditis elegans" &
                  sourceurl == "https://tflink.net/")

# Creating a variable for the GMT data.frame of EH8735.
# EH8735 contains small-scale measurement results, where the target genes are
# coded with Ensembl ID-s
Transcription_factor_TFLink_Caenorhabditis_elegans_SS_EnsemblID <-
  muleaData[["EH8735"]]
```

Index

ExperimentHub, *3*

muleaData, 2

query, *3*