Package ‘MOFAdata’

March 19, 2024

Type Package
Title Data package for Multi-Omics Factor Analysis (MOFA)
Version 1.18.0
Maintainer Britta Velten <britta.velten@gmail.com>
Author Ricard Argelaguet, Britta Velten, Damien Arnol, Florian Buettner, Wolfgang Huber, Oliver Stegle
Date 2018-06-26
License LGPL-3
Depends R (>= 3.5)
Description A collection of datasets to accompany the R package MOFA and illustrate running and analysing MOFA models.
Encoding UTF-8
Suggests knitr, MultiAssayExperiment, rmarkdown, BiocStyle
biocViews ReproducibleResearch
VignetteBuilder knitr
LazyData false
NeedsCompilation yes
RoxygenNote 6.1.1
git_url https://git.bioconductor.org/packages/MOFAdata
git_branch RELEASE_3_18
git_last_commit f0c52b3
git_last_commit_date 2023-10-24
Repository Bioconductor 3.18
Date/Publication 2024-03-19
## R topics documented:

- **CLL_covariates**
  - Description: Data frame containing additional information on the patient samples, i.e. diagnosis and gender.
  - Usage: `data(CLL_covariates)`
  - Format: A data frame diagnosis and gender for the n=200 patient samples in CLL_data

- **CLL_data**
  
  **Description**: A list containing different omic measurements for Chronic lymphocytic leukaemia (CLL) patient samples.
  
  The data set was originally published in [https://www.ncbi.nlm.nih.gov/pubmed/29227286](https://www.ncbi.nlm.nih.gov/pubmed/29227286). The MOFA analysis for this data set can be found in [http://msb.embopress.org/content/14/6/e8124](http://msb.embopress.org/content/14/6/e8124)
  
  **Usage**: `data(CLL_data)`
**Format**

A list of matrices, with the following -omics:

- mRNA: normalized expression values for the 5000 most variable genes
- Methylation: methylation M-values for the 4248 most variable CpG sites
- Drugs: viability values in response to 310 different drugs and concentrations
- Mutations: Mutation status for 69 selected genes

**Source**


---

**Description**

The MOFAdata package provides a collection of datasets to accompany the R package MOFA, where they are used to illustrate how to run MOFA and analyse its results. Briefly, it contains multi-omics data for a bulk study on chronic lymphocytic leukemia and a single cell scMT-seq study. For these two data sets, we also provide pretrained MOFA objects to be used in downstream analysis with MOFA. In addition, various gene sets are stored here, that can be useful when performing feature set enrichment analysis in MOFA.

**Details**

An overview of the datasets contained in this package can be found in the vignette "MOFAdata".

---

**Description**

A matrix containing human gene sets from the MSigDB 6.0 data base (http://software.broadinstitute.org/gsea/msigdb).

- **species**: Homo sapiens
- **Gene IDs**: Ensembl IDs from version 75 (i.e. ENSG00000125798)
- **Collection**: C2, curated gene sets from online pathway databases, publications in PubMed, and knowledge of domain experts.

**Usage**

data(MSigDB_v6.0_C2_human)
MSigDB_v6.0_C5_human

**Format**

Matrix containing 1329 biological pathways in rows and 20086 genes in columns

**Source**

http://software.broadinstitute.org/gsea/msigdb

---

MSigDB_v6.0_C2_mouse  MSigDB_v6.0_C2_mouse

**Description**

A matrix containing mouse gene sets from the MSigDB 6.0 data base (http://software.broadinstitute.org/gsea/msigdb).

species: Mus musculus

Gene IDs: Gene symbol/name (i.e. FOXA2)

Collection: C2, curated gene sets from online pathway databases, publications in PubMed, and knowledge of domain experts.

**Usage**

data(MSigDB_v6.0_C2_mouse)

**Format**

Matrix containing 4731 biological pathways in rows and 18360 genes in columns

**Source**

http://software.broadinstitute.org/gsea/msigdb

---

MSigDB_v6.0_C5_human  MSigDB_v6.0_C5_human

**Description**

A matrix containing human gene sets from the MSigDB 6.0 data base (http://software.broadinstitute.org/gsea/msigdb).

species: Homo sapiens

Gene IDs: Ensembl IDs from version 75 (i.e. ENSG00000125798)

Collection: C5, consist of genes annotated by GO Biological Process terms.
Usage

data(MSigDB_v6.0_C5_mouse)

Format

Matrix containing 4436 biological pathways in rows and 20086 genes in columns.

Source

http://software.broadinstitute.org/gsea/msigdb

Description

A matrix containing mouse gene sets from the MSigDB 6.0 data base (http://software.broadinstitute.org/gsea/msigdb).

- **species:** Mus musculus
- **Gene IDs:** Gene symbol/name (i.e. FOXA2)
- **Collection:** C5, consist of genes annotated by GO Biological Process terms.

Usage

data(MSigDB_v6.0_C5_mouse)

Format

Matrix containing 4436 biological pathways in rows and 18360 genes in columns.

Source

http://software.broadinstitute.org/gsea/msigdb
**Description**
A matrix containing human gene sets from the Reactome v59 Pathway data base ([https://reactome.org](https://reactome.org)).

- **species:** Homo sapiens
- **Gene IDs:** Ensembl IDs from version 75 (i.e. ENSG00000125798)

**Usage**
data(reactomeGS)

**Format**
Matrix containing 1304 biological pathways in rows and 18818 genes (ensemble IDs) in columns.

**Source**
[https://reactome.org](https://reactome.org)

---

**Description**
A MultiAssayExperiment containing data from a single cell multi-omics study (scMT-seq) on mouse embryonic stem cells (mESCs). The data set was originally published in [https://www.ncbi.nlm.nih.gov/pubmed/26752769](https://www.ncbi.nlm.nih.gov/pubmed/26752769). The MOFA analysis for this data set can be found in [http://msb.embopress.org/content/14/6/e8124](http://msb.embopress.org/content/14/6/e8124)

**Usage**
data(scMT_data)

**Format**
A MultiAssayExperiment containing four Experiments:
- RNA expression: ExpressionSet with normalized expression values of the 5000 most variable genes
- Met Enhancers: Methylation values for 5000 CpGs overlapping enhancer elements
- Met CpG Islands: Methylation values for 5000 CpGs overlapping CpG Islands
- Met Promoters: Methylation values for 5000 CpGs overlapping promoters
scMT_data

Source

Index

* datasets
  CLL_c covariates, 2
  CLL_data, 2
  MSigDB_v6.0_C2_human, 3
  MSigDB_v6.0_C2_mouse, 4
  MSigDB_v6.0_C5_human, 4
  MSigDB_v6.0_C5_mouse, 5
  reactomeGS, 6
  scMT_data, 6

CLL_c covariates, 2
CLL_data, 2

MOFAdata, 3
MOFAdata-package (MOFAdata), 3
MSigDB_v6.0_C2_human, 3
MSigDB_v6.0_C2_mouse, 4
MSigDB_v6.0_C5_human, 4
MSigDB_v6.0_C5_mouse, 5

reactomeGS, 6

scMT_data, 6