

Package ‘SingleCellMultiModal’

March 29, 2021

Type Package

Title Integrating Multi-modal Single Cell Experiment datasets

Version 1.2.4

Description SingleCellMultiModal is an ExperimentHub package that serves multiple datasets obtained from GEO and other sources and represents them as MultiAssayExperiment objects. The current focus is on datasets that use new technologies such as scNMT and scM&T.

License Artistic-2.0

BugReports <https://github.com/waldronlab/SingleCellMultiModal/issues>

Depends R (>= 4.0.0), MultiAssayExperiment

Imports AnnotationHub, BiocFileCache, BiocGenerics, ExperimentHub, HCAMatrixBrowser, HDF5Array, S4Vectors, SingleCellExperiment, SpatialExperiment, SummarizedExperiment, Matrix, methods, utils

Suggests knitr, scran

VignetteBuilder knitr

biocViews ExperimentData, SingleCellData, ReproducibleResearch, ExperimentHub, GEO

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

git_url <https://git.bioconductor.org/packages/SingleCellMultiModal>

git_branch RELEASE_3_12

git_last_commit de087cd

git_last_commit_date 2021-01-22

Date/Publication 2021-03-29

Author Marcel Ramos [aut, cre] (<<https://orcid.org/0000-0002-3242-0582>>),
Ricard Argelaguet [aut],
Dario Righelli [aut],
Kelly Eckenrode [aut],
Levi Waldron [aut]

Maintainer Marcel Ramos <marcel.ramos@roswellpark.org>

R topics documented:

SingleCellMultiModal-package	2
CITEseq	3
scmmCache	4
scMultiome	5
scNMT	6
SCoPE2	8
seqFISH	10
Index	12

SingleCellMultiModal-package
SingleCellMultiModal-package

Description

The SingleCellMultiModal package provides a convenient and user-friendly representation of multi-modal data from project such as ‘scNMT’ for mouse gastrulation.

Author(s)

Maintainer: Marcel Ramos <marcel.ramos@roswellpark.org> ([ORCID](#))

Authors:

- Ricard Argelaguet <ricard@ebi.ac.uk>
- Dario Righelli <dario.righelli@gmail.com>
- Kelly Eckenrode <kelly.eckenrode@sph.cuny.edu>
- Levi Waldron <lwaldron.research@gmail.com>

See Also

Useful links:

- Report bugs at <https://github.com/waldronlab/SingleCellMultiModal/issues>

Examples

```
help(package = "SingleCellMultiModal")
```

CITEseq

*CITEseq***Description**

function assembles data on-the-fly from ‘ExperimentHub’ to provide a [MultiAssayExperiment](#) container. Actually the ‘dataType’ argument provides access to the available datasets associated to the package.

Usage

```
CITEseq(
  DataType = c("cord_blood", "peripheral_blood"),
  modes = "*",
  version = "1.0.0",
  dry.run = TRUE,
  verbose = TRUE,
  DataClass = c("MultiAssayExperiment", "SingleCellExperiment"),
  ...
)
```

Arguments

DataType	character(1) indicating the identifier of the dataset to retrieve. (default "cord_blood")
modes	character() The assay types or modes of data to obtain these include scADT and scRNA-seq data by default.
version	character(1) Either version '1.0.0' depending on data version required.
dry.run	logical(1) Whether to return the dataset names before actual download (default TRUE)
verbose	logical(1) Whether to show the dataset currently being (down)loaded (default TRUE)
DataClass	either MultiAssayExperiment or SingleCellExperiment data classes can be returned (default MultiAssayExperiment)
...	Additional arguments passed on to the ExperimentHub-class constructor

Details

CITEseq data are a combination of single cell transcriptomics and about a hundred of cell surface proteins.

Available datasets are:

- cord_blood: a dataset of single cells of cord blood as provided in Stoeckius et al. (2017).
 - scRNA_Counts - Stoeckius scRNA-seq gene count matrix
 - scADT - Stoeckius antibody-derived tags (ADT) data
- peripheral_blood: a dataset of single cells of peripheral blood as provided in Mimitou et al. (2019). We provide two different conditions controls (CTRL) and Cutaneous T-cell Lymphoma (CTCL). Just build appropriate modes regex for subselecting the dataset modes.
 - scRNA - Mimitou scRNA-seq gene count matrix

- scADT - Mimitou antibody-derived tags (ADT) data
- scHTO - Mimitou Hashtag Oligo (HTO) data
- TCRab - Mimitou T-cell Receptors (TCR) alpha and beta available through the object metadata.
- TCRgd - Mimitou T-cell Receptors (TCR) gamma and delta available through the object metadata.

Value

A single cell multi-modal [MultiAssayExperiment](#) or informative 'data.frame' when 'dry.run' is 'TRUE'

Author(s)

Dario Righelli

References

Stoeckius et al. (2017), Mimitou et al. (2019)

Examples

```
mae <- CITEseq(DataType="cord_blood", dry.run=FALSE)
experiments(mae)
```

scmmCache

Manage cache / download directories for study data

Description

Managing data downloads is important to save disk space and re-downloading data files. This can be done effortlessly via the integrated BiocFileCache system.

Usage

```
scmmCache(...)

setCache(
  directory = tools::R_user_dir("SingleCellMultiModal", "cache"),
  verbose = TRUE,
  ask = interactive()
)

removeCache(accession)
```

Arguments

...	For scmmCache, arguments passed to setCache
directory	character(1) The file location where the cache is located. Once set, future downloads will go to this folder. See setCache section for details.
verbose	Whether to print descriptive messages
ask	logical(1) (default TRUE when interactive()) Confirm the file location of the cache directory
accession	character(1) A single string indicating the accession number of the study

Value

The directory / option of the cache location

scmmCache

Get the directory location of the cache. It will prompt the user to create a cache if not already created. A specific directory can be used via setCache.

setCache

Specify the directory location of the data cache. By default, it will go into the user's home and package name directory as given by `R_user_dir` (default: varies by system e.g., for Linux: ``$HOME/.cache/R/SingleCellMultiome``).

removeCache

Some files may become corrupt when downloading, this function allows the user to delete the tarball associated with a study number in the cache.

Examples

```
getOption("scmmCache")
scmmCache()
```

Description

scMultiome currently allows users to download 10K Peripheral Blood Mononuclear Cells provided by [10x Genomics website](#) (DataType = "pbmc_10x"). This technology enables simultaneous profiling of the transcriptome (using 3' gene expression) and epigenome (using ATAC-seq) from single cells to deepen your understanding of how genes are expressed and regulated across different cell types. Data prepared by Ricard Argelaguet.

Usage

```

scMultiome(
  DataType = "pbmc_10x",
  modes = "*",
  version = "1.0.0",
  format = c("HDF5", "MTX"),
  dry.run = TRUE,
  verbose = TRUE,
  ...
)

```

Arguments

DataType	character(1) Indicates study that produces this type of data (default: 'mouse_gastrulation')
modes	character() A wildcard / glob pattern of modes, such as "acc*". A wildcard of "*" will return all modes including Chromatin Accessibility ("acc"), Methylation ("met"), RNA-seq ("rna") which is the default.
version	character(1) Either version '1.0.0' or '2.0.0' depending on data version required. See versions section.
format	Either HDF5 or MTX data formats (default HDF5)
dry.run	logical(1) Whether to return the dataset names before actual download (default TRUE)
verbose	logical(1) Whether to show the dataset currently being (down)loaded (default TRUE)
...	Additional arguments passed on to the ExperimentHub-class constructor

Details

Users are able to choose from either an HDF5 or MTX file format as the internal data representation. Choosing HDF5 gives users a sparse HDF5Array class object. The MTX (Matrix Market) format allows users to load a sparse dgCMatrix representation.

Examples

```
scMultiome(DataType = "pbmc_10x", modes = "*", dry.run = TRUE)
```

Description

scNMT assembles data on-the-fly from ExperimentHub to provide a [MultiAssayExperiment](#) container. The DataType argument provides access to the mouse_gastrulation dataset as obtained from Argelaguet et al. (2019; DOI: 10.1038/s41586-019-1825-8). Pre-processing code can be seen at https://github.com/rargelaguet/scnmt_gastrulation. Protocol information for this dataset is available at Clark et al. (2018). See the vignette for the full citation.

Usage

```
scNMT(
  DataType = "mouse_gastrulation",
  modes = "*",
  version,
  dry.run = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

DataType	character(1) Indicates study that produces this type of data (default: 'mouse_gastrulation')
modes	character() A wildcard / glob pattern of modes, such as "acc*". A wildcard of "*" will return all modes including Chromatin Accessibility ("acc"), Methylation ("met"), RNA-seq ("rna") which is the default.
version	character(1) Either version '1.0.0' or '2.0.0' depending on data version required. See versions section.
dry.run	logical(1) Whether to return the dataset names before actual download (default TRUE)
verbose	logical(1) Whether to show the dataset currently being (down)loaded (default TRUE)
...	Additional arguments passed on to the ExperimentHub-class constructor

Details

scNMT is a combination of RNA-seq (transcriptome) and an adaptation of Nucleosome Occupancy and Methylation sequencing (NOMe-seq, the methylome and chromatin accessibility) technologies. For more information, see Reik et al. (2018) DOI: 10.1038/s41467-018-03149-4

```
\itemize{
  \item{mouse_gastrulation:}
  \itemize{
    \item{rna} - RNA-seq
    \item{acc_*} - chromatin accessibility
    \item{met_*} - DNA methylation
    \itemize{
      \item{cgi} - CpG islands
      \item{CTCF} - footprints of CTCF binding
      \item{DHS} - DNase Hypersensitive Sites
      \item{genebody} - gene bodies
      \item{p300} - p300 binding sites
      \item{promoter} - gene promoters
    }
  }
}
```

Special thanks to Al J Abadi for preparing the published data in time for the 2020 BIRS Workshop, see the link here:

`url{https://github.com/BIRSBiointegration/Hackathon/tree/master/scNMT-seq}`

Value

A single cell multi-modal [MultiAssayExperiment](#) or informative data.frame when dry.run is TRUE

versions

Version '1.0.0' of the scNMT mouse_gastrulation dataset includes all of the above mentioned assay technologies with filtering of cells based on quality control metrics. Version '2.0.0' contains all of the cells without the QC filter and does not contain CTCF binding footprints or p300 binding sites.

metadata

The MultiAssayExperiment metadata includes the original function call that saves the function call and the data version requested.

Source

http://ftp.ebi.ac.uk/pub/databases/scnmt_gastrulation/

References

Argelaguet et al. (2019)

See Also

SingleCellMultiModal-package

Examples

```
scNMT(DataType = "mouse_gastrulation", modes = "*",  
       version = "1.0.0", dry.run = TRUE)
```

SCoPE2

Single-cell RNA sequencing and proteomics

Description

SCoPE2 assembles data on-the-fly from ExperimentHub to provide a [MultiAssayExperiment](#) container. The DataType argument provides access to the SCoPE2 dataset as provided by Specht et al. (2020; DOI: <http://dx.doi.org/10.1101/665307>). The article provides more information about the data acquisition and pre-processing.

Usage

```
SCoPE2(  
  DataType = "macrophage_differentiation",  
  modes = "*",  
  version = "1.0.0",  
  dry.run = TRUE,  
  verbose = TRUE,  
  ...  
)
```


Arguments

DataType	character(1) Indicates study that produces this type of data (default: 'macrophage_differentiation')
modes	character() A wildcard / glob pattern of modes, such as "rna". A wildcard of "*" will return all modes, that are transcriptome ("rna") or proteome ("protein") which is the default.
version	character(1), currently only version '1.0.0' is available
dry.run	logical(1) Whether to return the dataset names before actual download (default TRUE)
verbose	logical(1) Whether to show the dataset currently being (down)loaded (default TRUE)
...	Additional arguments passed on to the ExperimentHub-class constructor

Details

The SCoPE2 study combined scRNA-seq (transcriptome) and single-cell proteomics. The cells are monocytes that undergo macrophage differentiation. No annotation is available for the transcriptome data, but batch and cell type annotations are available for the proteomics data. The transcriptomics and proteomics data were not measured from the same cells but from a distinct set of cell cultures.

```
\itemize{
  \item{SCoPE2:}
  \itemize{
    \item{scRNAseq1} - single-cell transcriptome (batch 1)
    \item{scRNAseq2} - single-cell transcriptome (batch 2)
    \item{scp} - single-cell proteomics
  }
}
```

Value

A single cell multi-modal [MultiAssayExperiment](#) or informative data.frame when dry.run is TRUE

Source

All files are linked from the slavovlab website <https://scope2.slavovlab.net/docs/data>

References

Specht, Harrison, Edward Emmott, Aleksandra A. Petelski, R. Gray Huffman, David H. Perlman, Marco Serra, Peter Kharchenko, Antonius Koller, and Nikolai Slavov. 2020. "Single-Cell Proteomic and Transcriptomic Analysis of Macrophage Heterogeneity." bioRxiv. <https://doi.org/10.1101/665307>.

See Also

SingleCellMultiModal-package

Examples

```
SCoPE2(DataType = "macrophage_differentiation",
        modes = "*",
        version = "1.0.0",
        dry.run = TRUE)
```

 seqFISH

seqFISH

Description

seqFISH function assembles data on-the-fly from ‘ExperimentHub’ to provide a [MultiAssayExperiment](#) container. Actually the ‘dataType’ argument provides access to the available datasets associated to the package.

Usage

```
seqFISH(
  DataType = "mouse_visual_cortex",
  modes = "*",
  version,
  dry.run = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

DataType	character(1) indicating the identifier of the dataset to retrieve. (default "mouse_visual_cortex")
modes	character() The assay types or modes of data to obtain these include seq-FISH and scRNA-seq data by default.
version	character(1) Either version '1.0.0' or '2.0.0' depending on data version required. See versions section.
dry.run	logical(1) Whether to return the dataset names before actual download (default TRUE)
verbose	logical(1) Whether to show the dataset currently being (down)loaded (default TRUE)
...	Additional arguments passed on to the ExperimentHub-class constructor

Details

seq FISH data are a combination of single cell spatial coordinates and transcriptomics for a few hundreds of genes. seq-FISH data can be combined for example with scRNA-seq data to unveil multiple aspects of cellular behaviour based on their spatial organization and transcription.

Available datasets are:

- mouse_visual_cortex: combination of seq-FISH data as obtained from Zhu et al. (2018) and scRNA-seq data as obtained from Tasic et al. (2016), Version 1.0.0 returns the full scRNA-seq data matrix, while version 2.0.0 returns the processed and subsetting scRNA-seq data matrix (produced for the Mathematical Frameworks for Integrative Analysis of Emerging Biological Data Types 2020 Workshop) The returned seqFISH data are always the processed ones for the same workshop.
 - scRNA_Counts - Tasic scRNA-seq gene count matrix
 - scRNA_Labels - Tasic scRNA-seq cell labels
 - seqFISH_Coordinates - Zhu seq-FISH spatial coordinates
 - seqFISH_Counts - Zhu seq-FISH gene counts matrix
 - seqFISH_Labels - Zhu seq-FISH cell labels

Value

A [MultiAssayExperiment](#) of seq-FISH data

Author(s)

Dario Righelli <dario.righelli <at> gmail.com>

Examples

```
seqFISH(DataType = "mouse_visual_cortex", modes = "*", version = "2.0.0",  
dry.run = TRUE)
```

Index

CITEseq, [3](#)

ExperimentHub-class, [3](#), [6](#), [7](#), [9](#), [10](#)

MultiAssayExperiment, [3](#), [4](#), [6](#), [8–11](#)

R_user_dir, [5](#)

removeCache (scmmCache), [4](#)

scmmCache, [4](#)

scMultiome, [5](#)

scNMT, [6](#)

SCoPE2, [8](#)

seqFISH, [10](#)

setCache (scmmCache), [4](#)

SingleCellMultiModal
(SingleCellMultiModal-package),
[2](#)

SingleCellMultiModal-package, [2](#)