

Package ‘SingleCellMultiModal’

April 11, 2023

Type Package

Title Integrating Multi-modal Single Cell Experiment datasets

Version 1.10.1

Description SingleCellMultiModal is an ExperimentHub package that serves multiple datasets obtained from GEO and other sources and represents them as MultiAssayExperiment objects. We provide several multi-modal datasets including scNMT, 10X Multiome, seqFISH, CITEseq, SCoPE2, and others. The scope of the package is to provide data for benchmarking and analysis.

License Artistic-2.0

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

Depends R (>= 4.2.0), MultiAssayExperiment

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

Suggests BiocStyle, ggplot2, knitr, RaggedExperiment, rmarkdown, scater, scran, UpSetR, uwot

VignetteBuilder knitr

biocViews ExperimentData, SingleCellData, ReproducibleResearch, ExperimentHub, GEO

Encoding UTF-8

RoxygenNote 7.2.1

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

git_branch RELEASE_3_16

git_last_commit 90e6cf7

git_last_commit_date 2023-04-10

Date/Publication 2023-04-11

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

Kelly Eckenrode [aut],
 Ludwig Geistlinger [aut],
 Levi Waldron [aut]

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

R topics documented:

| | |
|--|----|
| SingleCellMultiModal-package | 2 |
| CITEseq | 3 |
| GTseq | 4 |
| scmmCache | 6 |
| scMultiome | 7 |
| scNMT | 8 |
| SCoPE2 | 10 |
| seqFISH | 12 |
| SingleCellMultiModal | 13 |

| | |
|--------------|-----------|
| Index | 15 |
|--------------|-----------|

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")
```

Description

CITEseq function assembles data on-the-fly from ‘ExperimentHub’ to provide a [MultiAssayExperiment](#) container. 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](#)/[SingleCellExperiment](#) 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)
```

GTseq

Parallel sequencing data of single-cell genomes and transcriptomes

Description

GTseq assembles data on-the-fly from ExperimentHub to provide a [MultiAssayExperiment](#) container. The `DataType` argument provides access to the `mouse_embryo_8_cell` dataset as obtained from Macaulay et al. (2015). Protocol information for this dataset is available from Macaulay et al. (2016). See references.

Usage

```
GTseq(
  DataType = "mouse_embryo_8_cell",
  modes = "*",
  version = "1.0.0",
  dry.run = TRUE,
```

```

    verbose = TRUE,
    ...
  )

```

Arguments

| | |
|-----------------------|---|
| <code>DataType</code> | character(1) Indicates study that produces this type of data (default: 'mouse_embryo_8_cell') |
| <code>modes</code> | character() A wildcard / glob pattern of modes, such as "*omic". A wildcard of "*" will return all modes including copy numbers ("genomic") and RNA-seq read counts ("transcriptomic"), which is the default. |
| <code>version</code> | character(1). Currently, only version '1.0.0'. |
| <code>dry.run</code> | logical(1) Whether to return the dataset names before actual download (default TRUE) |
| <code>verbose</code> | logical(1) Whether to show the dataset currently being (down)loaded (default TRUE) |
| <code>...</code> | Additional arguments passed on to the ExperimentHub-class constructor |

Details

G&T-seq is a combination of Picoplex amplified gDNA sequencing (genome) and SMARTSeq2 amplified cDNA sequencing (transcriptome) of the same cell. For more information, see Macaulay et al. (2015).

- `mouse_embryo_8_cell`:
 - genomic - integer copy numbers as detected from scDNA-seq
 - transcriptomic - raw read counts as quantified from scRNA-seq

Value

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

metadata

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

Source

<https://www.ebi.ac.uk/ena/browser/view/PRJEB9051>

References

- Macaulay et al. (2015) G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat Methods*, 12:519–22.
- Macaulay et al. (2016) Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat Protoc*, 11:2081–103.

See Also

SingleCellMultiModal-package

Examples

```
GTseq()
```

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/SingleCellMultiMod`

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()
```

 scMultiome

Single-cell Multiome ATAC + Gene Expression

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 our 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("MTX", "HDF5"),
  dry.run = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

| | |
|-----------------------|--|
| <code>DataType</code> | character(1) Indicates study that produces this type of data (default: <code>'mouse_gastrulation'</code>) |
| <code>modes</code> | character() A wildcard / glob pattern of modes, such as <code>"acc*"</code> . A wildcard of <code>"*"</code> will return all modes including Chromatin Accessibilty (<code>"acc"</code>), Methylation (<code>"met"</code>), RNA-seq (<code>"rna"</code>) which is the default. |
| <code>version</code> | character(1) Either version <code>'1.0.0'</code> or <code>'2.0.0'</code> depending on data version required (default <code>'1.0.0'</code>). See version section. |

| | |
|---------|---|
| format | Either MTX or HDF5 data format (default MTX) |
| 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 MTX or HDF5 file format as the internal data representation. The MTX (Matrix Market) format allows users to load a sparse `dgMatrix` representation. Choosing HDF5 gives users a sparse `HDF5Array` class object.

Value

A 10X PBMC `MultiAssayExperiment` object

Examples

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

scNMT

Single-cell Nucleosome, Methylation and Transcription sequencing

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 = "1.0.0",
  dry.run = TRUE,
  verbose = TRUE,
  ...
)
```


Arguments

| | |
|-----------------------|---|
| <code>DataType</code> | character(1) Indicates study that produces this type of data (default: 'mouse_gastrulation') |
| <code>modes</code> | 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. |
| <code>version</code> | character(1) Either version '1.0.0' or '2.0.0' depending on data version required (default '1.0.0'). See version section. |
| <code>dry.run</code> | logical(1) Whether to return the dataset names before actual download (default TRUE) |
| <code>verbose</code> | logical(1) Whether to show the dataset currently being (down)loaded (default TRUE) |
| <code>...</code> | 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

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

Special thanks to AI J Abadi for preparing the published data in time for the 2020 BIRS Workshop, see the link here: [urlhttps://github.com/BIRSBiointegration/Hackathon/tree/master/scNMT-seq](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

| | |
|-----------------------|--|
| <code>DataType</code> | character(1) Indicates study that produces this type of data (default: 'macrophage_differentiation') |
| <code>modes</code> | 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. |
| <code>version</code> | character(1), currently only version '1.0.0' is available |
| <code>dry.run</code> | 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.

- SCoPE2:
 - scRNAseq1 - single-cell transcriptome (batch 1)
 - scRNAseq2 - single-cell transcriptome (batch 2)
 - 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

*Single-cell spatial + Gene Expression***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 (default '1.0.0'). See version 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 subsetted 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)
```

SingleCellMultiModal *Combining Modalities into one MultiAssayExperiment*

Description

Combine multiple single cell modalities into one using the input of the individual functions.

Usage

```
SingleCellMultiModal(
  DataTypes,
  modes = "*",
  versions = "1.0.0",
  dry.run = TRUE,
  verbose = TRUE,
  ...
)
```

Arguments

| | |
|-----------|---|
| DataTypes | character() A vector of data types as indicated in each individual function by the <code>DataType</code> parameter. These can be any of the following: "mouse_gastrulation", "pbmc_10x", "macrophage_differentiation", "cord_blood", "peripheral_blood", "mouse_visual_cortex", "mouse_embryo_8_cell" |
| modes | list() A list or <code>CharacterList</code> of modes for each data type where each element corresponds to one data type. |

| | |
|----------|---|
| versions | character() A vector of versions for each DataType. By default, version 1.0.0 is obtained for all data types. |
| 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 |

Value

A multi-modality MultiAssayExperiment

metadata

The metadata in the MultiAssayExperiment contains the original function call used to generate the object (labeled as call), a call_map which provides traceability of technology functions to DataType prefixes, and lastly, R version information as version.

Examples

```
SingleCellMultiModal(c("mouse_gastrulation", "pbmc_10x"),
  modes = list(c("acc*", "met*"), "rna"),
  version = c("2.0.0", "1.0.0"), dry.run = TRUE, verbose = TRUE
)
```

Index

CITEseq, [3](#)

ExperimentHub-class, [3](#), [5](#), [8](#), [9](#), [11](#), [12](#), [14](#)

GTseq, [4](#)

MultiAssayExperiment, [3–5](#), [8–13](#)

R_user_dir, [7](#)

removeCache (scmmCache), [6](#)

scmmCache, [6](#)

scMultiome, [7](#)

scNMT, [8](#)

SCoPE2, [10](#)

seqFISH, [12](#)

setCache (scmmCache), [6](#)

SingleCellExperiment, [4](#)

SingleCellMultiModal, [13](#)

SingleCellMultiModal-package, [2](#)