Package 'MouseGastrulationData'

March 29, 2021

Title Single-Cell Transcriptomics Data across Mouse Gastrulation and Early Organogenesis

Version 1.4.0

Description Provides processed and raw count matrices for single-cell RNA sequencing data from a timecourse of mouse gastrulation and early organogenesis.

Depends R (>= 4.0), SingleCellExperiment, SummarizedExperiment

Imports methods, ExperimentHub, BiocGenerics, S4Vectors

Suggests BiocStyle, knitr, rmarkdown

VignetteBuilder knitr

License GPL-3

NeedsCompilation no

Encoding UTF-8

biocViews ExperimentData, ExpressionData, SequencingData, RNASeqData, SingleCellData, ExperimentHub

URL https://github.com/MarioniLab/MouseGastrulationData

BugReports https://github.com/MarioniLab/MouseGastrulationData/issues

RoxygenNote 7.1.1

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

git_branch RELEASE_3_12

git_last_commit 0440a1c

git_last_commit_date 2020-10-27

Date/Publication 2021-03-29

Author Jonathan Griffiths [aut, cre], Aaron Lun [aut]

Maintainer Jonathan Griffiths <jonathan.griffiths.94@gmail.com>

R topics documented:

MouseGastrulationData-package 2
AtlasSampleMetadata 2
EmbryoAtlasData 3
EmbryoCelltypeColours 5

GuibentifExtraData 6
Tal1ChimeraData 7
TChimeraData 8
WTChimeraData 10

Index **13**

`MouseGastrulationData-package` *Single-Cell Transcriptomics Data across Mouse Gastrulation and Early Organogenesis*

Description

Provides processed and raw count matrices for single-cell RNA sequencing data from a timecourse of mouse gastrulation and early organogenesis.

Details

This package contains the processed 10X Genomics data from Pijuan-Sala et al. (2019).

The data were processed as described in the methods that accompany the paper.

Author(s)

NA

Maintainer: NA

References

Blanca Pijuan-Sala*, Jonathan A. Griffiths*, Carolina Guibentif*, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, Ximena Ibarra-Soria, Richard C.V. Tyser, Debbie Lee Lian Ho, Wolf Reik, Shankar Srinivas, Benjamin D. Simons, Jennifer Nichols, John C. Marioni, Berthold Göttgens. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature, 566, pp490-495 (2019).

`AtlasSampleMetadata` *Sample metadata from the Pijuan-Sala et al. embryo atlas*

Description

A data frame containing stage and embryo pool information for the atlas dataset.

Usage

```
AtlasSampleMetadata
```

Format

A data frame containing information for each 10x sample of the embryo atlas. This object contains:

`sample`: Integer, 10x sample index.

`stage`: Character, developmental stage from which sample was generated.

`pool_index`: Integer, index for pools of embryos; samples with the same values are from the same pool of dissociated cells.

`seq_batch`: Integer, seqencing batch index; samples with the same values were multiplexed for sequencing.

`ncells`: Integer, number of cells (post-QC) per sample.

Note that sample 11 is missing by design due to experimental failure: it is not available for download.

References

Pijuan-Sala B, Griffiths JA, Guibentif C et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 7745:490-495.

Examples

```
head(AtlasSampleMetadata)
```

EmbryoAtlasData *Mouse gastrulation timecourse data*

Description

Obtain the processed or raw counts for the mouse gastrulation scRNAseq dataset.

Usage

```
EmbryoAtlasData(
  type = c("processed", "raw"),
  samples = NULL,
  get.spliced = FALSE
)
```

Arguments

`type`	String specifying the type of data to obtain, see Details. Default behaviour is to return processed data.
`samples`	Integer or character vector specifying the samples for which data (processed or raw) should be obtained. If `NULL` (default), data are returned for all (36) samples.
`get.spliced`	Logical indicating whether to also download the spliced/unspliced/ambiguously spliced count matrices.

Details

This function downloads the data for the embryo atlas from Pijuan-Sala et al. (2019). The dataset contains 36 10X Genomics samples; sample 11 is absent due to QC failure.

In the processed data, cell-containing libraries have already been identified in each sample using the `emptyDrops` function from **DropletUtils**. The count matrix contains the raw count vectors for the cells called from all samples in this manner. Size factors were computed using the `computeSumFactors` function from **scran**. The column metadata for called cells contains:

`cell`: Character, unique cell identifier across all samples.

`barcode`: Character, cell barcode from the 10X Genomics experiment.

`sample`: Integer, index of the sample from which the cell was taken.

`pool`: Integer, index of the embryo pool from which the sample derived. Samples with the same value are technical, not biological, replicates

`stage`: Character, stage of the mouse embryo at which the sample was taken.

`sequencing.batch`: Integer, sequencing run in which sample was multiplexed.

`theiler`: Character, Theiler stage from which the sample was taken; alternative scheme to `stage`.

`doub.density`: Numeric, output of (a now-outdated run of) `scran::doubletCells`, performed on each sample separately.

`doublet`: Logical, whether a cell was called as a doublet.

`cluster`: Integer, top-level cluster to which cell was assigned across all samples.

`cluster.sub`: Integer, cluster to which cell was assigned when clustered within each `cluster`.

`cluster.stage`: Integer, top-level cluster to which cell was assigned within individual timepoints.

`cluster.theiler`: Integer, top-level cluster to which cell was assigned within individual Theiler stages.

`stripped`: Logical, whether a cell was called as a cytoplasm-stripped nucleus.

`celltype`: Character, cell type to which the cell was assigned.

`colour`: Integer, cell type colour (hex) as in Pijuan-Sala et al. (2019).

`umapX`: Numeric, x-coordinate of UMAP plot in Pijuan-Sala et al. (2019).

`umapY`: Numeric, y-coordinate of UMAP plot in Pijuan-Sala et al. (2019).

Reduced dimension representations of the data are also available in the `reducedDims` slot of the SingleCellExperiment object. If spliced counts were requested, these will be in the assays slot of the SingleCellExperiment object. Spliced count matrices were collated using *velocyto* version 0.17.17. Spliced count matrices will not have had swapped molecules removed, as *velocyto* and `DropletUtils::swappedDrops` are not compatible. However, these should still be effective for calculating RNA velocity estimates using various different tools.

The raw data contains the unfiltered count matrix for each sample, as generated directly from the CellRanger software. Swapped molecules have been removed using `DropletUtils::swappedDrops`. No filtering has been performed to identify cells. This may be useful if performing analyses that need to account for the ambient RNA pool.

For both raw and processed data, the row metadata contains the Ensembl ID and MGI symbol for each gene.

Value

If `type="processed"`, a SingleCellExperiment is returned containing processed data from selected samples.

If `type="raw"`, a List of SingleCellExperiments is returned, each containing the raw counts for a single sample. List elements are named after the corresponding sample.

Author(s)

Aaron Lun, with modification by Jonathan Griffiths

References

Pijuan-Sala B, Griffiths JA, Guibentif C et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 7745:490-495.

Examples

```
atlas.data <- EmbryoAtlasData(samples = 1)

atlas.data <- EmbryoAtlasData(type="processed", samples = 1)
```

EmbryoCelltypeColours *Celltype colours from Pijuan-Sala et al.*

Description

A vector containing the colour hexcodes that were used in Pijuan-Sala et al.

Usage

```
EmbryoCelltypeColours
```

Format

A vector of hexcodes named according to the appropriate celltype; celltypes match those in the metadata.

References

Pijuan-Sala B, Griffiths JA, Guibentif C et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 7745:490-495.

Examples

```
head(EmbryoCelltypeColours)
```

GuibentifExtraData *Guibentif et al. accessory data*

Description

Obtain the trajectory and NMP ordering data used in Guibentif et al.

Usage

```
GuibentifExtraData()
```

Details

This function downloads the data used in some of the analyses from Guibentif et al. (2020). Specifically, it contains the NMP cell orderings, and the atlas somitogenesis trajectory data.

This data is stored in a list. The first element of the list is named `atlas_somite_trajectories`, and is itelf a list that contains:

- `masses`:A data.frame containing the mass allocated to each cell from each trajectory (note: excluding extraembryonic, mixed_gastrulation timepoint, and doublet or stripped nuclei cells).
- `membership`:A data.frame containing the somite trajectory labels used in the paper, calculated from `masses`.

The second element is named `nmp_orderings`, and is also a list, which contains:

- `atlas`:A data.frame containing the position for each cell in the NMP ordering from the embryo atlas (see `EmbryoAtlasData`).
- `wt_chimera`:A data.frame containing the position for each cell in the NMP ordering from the WT chimera data (see `WTChimeraData`).
- `t_chimera`:A data.frame containing the position for each cell in the NMP ordering from the T chimera data (see `TChimeraData`).

Value

A `list` of the relevant somitogenesis trajectory and NMP ordering data will be returned. Details of the list structure are described in Details, below.

Author(s)

Jonathan Griffiths

References

Guibentif C, Griffiths JA et al. (2020). Title. *Journal* 566, 7745:490-495.

Examples

```
data <- GuibentifExtraData()
```

`Tal1ChimeraData` *Tal1 chimera data*

Description

Obtain the processed or raw counts for the Tal1 chimeric mouse embryo dataset.

Usage

```
Tal1ChimeraData(type = c("processed", "raw"), samples = NULL)
```

Arguments

`type`	String specifying the type of data to obtain, see Details. Default behaviour is to return processed data.
`samples`	Integer or character vector specifying the samples for which data (processed or raw) should be obtained. If `NULL` (default), data are returned for all (four) samples.

Details

This function downloads the data for the E8.5 Tal1 chimera experiment from Pijuan-Sala et al. (2019). The dataset contains four 10X Genomics samples:

- Sample 1: _Tal1_ knock-out cells (tomato positive)
- Sample 2: _Tal1_ knock-out cells (tomato positive)
- Sample 3: wild-type cells (tomato negative)
- Sample 4: wild-type cells (tomato negative)

All samples are from E8.5, from the same pool of chimeric embryos. Different samples with the same Tomato status are therefore technical replicates of each other.

In the processed data, cell-containing libraries have already been identified in each sample using the `emptyDrops` function from **DropletUtils**. The count matrix contains the raw count vectors for the cells called from all samples in this manner. Size factors were computed using the `computeSumFactors` function from **scran**. The column metadata for called cells contains:

`cell`**:** Character, unique cell identifier across all samples.

`barcode`**:** Character, cell barcode from the 10X Genomics experiment.

`sample`**:** Integer, number of the sample from which the cell was taken.

`stage`**:** Character, stage of the mouse embryo at which the sample was taken.

`tomato`**:** Logical, whether this cell expressed td-Tomato during FACS.

`stage.mapped`**:** Character, stage of the mouse embryo atlas to which the cell was mapped.

`celltype.mapped`**:** Character, cell type of the mouse embryo atlas to which the cell was mapped.

#' Reduced dimension representations of the data are also available in the `reducedDims` slot of the SingleCellExperiment object.

The raw data contains the unfiltered count matrix for each sample, as generated directly from the CellRanger software. Swapped molecules have been removed using `DropletUtils::swappedDrops`. No filtering has been performed to identify cells. This may be useful if performing analyses that need to account for the ambient RNA pool.

For both raw and processed data, the row metadata contains the Ensembl ID and MGI symbol for each gene.

Value

If `type="processed"`, a SingleCellExperiment is returned containing processed data from selected samples.

If `type="raw"`, a List of SingleCellExperiments is returned, each containing the raw counts for a single sample. List elements are named after the corresponding sample.

Author(s)

Aaron Lun

References

Pijuan-Sala B, Griffiths JA, Guibentif C et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 7745:490-495.

Examples

```
tal1.data <- Tal1ChimeraData(samples = 1)

tal1.data <- Tal1ChimeraData(type="processed", samples = 1)
```

TChimeraData *T chimera data*

Description

Obtain the processed or raw counts for the T chimeric mouse embryo dataset.

Usage

```
TChimeraData(type = c("processed", "raw"), samples = c(1:2, 5:16))
```

Arguments

`type`	String specifying the type of data to obtain, see Details. Default behaviour is to return processed data.
`samples`	Integer or character vector specifying the samples for which data (processed or raw) should be obtained. If `NULL` (default), data are returned for all QC-passing (fourteen) samples.

Details

This function downloads the data for the T chimera experiment from Guibentif et al. (2020). The dataset contains sixteen 10X Genomics samples from sets of embryo pools:

- Sample 1: E8.5 injected cells (tomato positive), pool 1
- Sample 2: E8.5 host cells (tomato negative), pool 1
- Sample 3: E7.5 injected cells (tomato positive), pool 2
- Sample 4: E7.5 host cells (tomato negative), pool 2

- Sample 5: E8.5 injected cells (tomato positive), pool 3
- Sample 6: E8.5 host cells (tomato negative), pool 3
- Sample 7: E8.5 injected cells (tomato positive), pool 4
- Sample 8: E8.5 host cells (tomato negative), pool 4
- Sample 9: E8.5 injected cells (tomato positive), pool 5
- Sample 10: E8.5 host cells (tomato negative), pool 5
- Sample 11: E7.5 injected cells (tomato positive), pool 6
- Sample 12: E7.5 host cells (tomato negative), pool 6
- Sample 13: E7.5 injected cells (tomato positive), pool 7
- Sample 14: E7.5 host cells (tomato negative), pool 7
- Sample 15: E7.5 injected cells (tomato positive), pool 8
- Sample 16: E7.5 host cells (tomato negative), pool 8

Samples from the same pool are paired in the experimental design. Each pool is a biological replicate. Samples 3 and 4 were excluded from analyses, as in these chimeras host cells seemed to form only ExE ectoderm. The data is available to download if you like, but will not be fetched by default.

In the processed data, cell-containing libraries have already been identified in each sample using the `emptyDrops` function from **DropletUtils**. The count matrix contains the raw count vectors for the cells called from all samples in this manner. Size factors were computed using the `computeSumFactors` function from **scran**. The column metadata for called cells contains:

`cell`**:** Character, unique cell identifier across all samples.

`barcode`**:** Character, cell barcode from the 10X Genomics experiment.

`sample`**:** Integer, number of the sample from which the cell was taken.

`stage`**:** Character, stage of the mouse embryo at which the sample was taken.

`tomato`**:** Logical, whether this cell expressed td-Tomato during FACS.

`pool`**:** Integer, embryo pool from which cell derived; samples with same value are matched.

`stage.mapped`**:** Character, stage of the mouse embryo atlas to which the cell was mapped.

`celltype.mapped`**:** Character, cell type of the mouse embryo atlas to which the cell was mapped.

`closest.cell`**:** Character, closest cell in the atlas dataset (see `EmbryoAtlasData`) after MNN mapping.

`doub.density`**:** Numeric, output of (a now-outdated run of) `scran::doubletCells`, performed on each sample separately.

`trajectory.mapped`**:** Character, trajectory membership for somite/NMP formation.

`somite.subct.mapped`**:** Character, somite subcluster to which cells mapped.

`sizeFactor`**:** Numeric, cell sizefactor.

#' Reduced dimension representations of the data are also available in the `reducedDims` slot of the SingleCellExperiment object.

The raw data contains the unfiltered count matrix for each sample, as generated directly from the CellRanger software. Swapped molecules have been removed using `DropletUtils::swappedDrops`. No filtering has been performed to identify cells. This may be useful if performing analyses that need to account for the ambient RNA pool.

For both raw and processed data, the row metadata contains the Ensembl ID and MGI symbol for each gene.

Value

If `type="processed"`, a SingleCellExperiment is returned containing processed data from selected samples

If `type="raw"`, a List of SingleCellExperiments is returned, each containing the raw counts for a single sample. List elements are named after the corresponding sample.

Author(s)

Aaron Lun, with modification by Jonathan Griffiths

References

Guibentif C, Griffiths JA et al. (2020). Diverse Routes towards Early Somites in the Mouse Embryo *Developmental Cell* In press.

Examples

```
t.data <- TChimeraData(samples = 1)

t.data <- TChimeraData(type="processed", samples = 1)
```

`WTChimeraData` *WT chimera data*

Description

Obtain the processed or raw counts for the WT chimeric mouse embryo dataset.

Usage

```
WTChimeraData(type = c("processed", "raw"), samples = NULL)
```

Arguments

`type`	String specifying the type of data to obtain, see Details. Default behaviour is to return processed data.
`samples`	Integer or character vector specifying the samples for which data (processed or raw) should be obtained. If `NULL` (default), data are returned for all (ten) samples.

Details

This function downloads the data for the WT chimera experiment from Pijuan-Sala et al. (2019). The dataset contains ten 10X Genomics samples from sets of embryo pools:

- Sample 1: E7.5 injected cells (tomato positive), pool 1
- Sample 2: E7.5 host cells (tomato negative), pool 1
- Sample 3: E7.5 injected cells (tomato positive), pool 2
- Sample 4: E7.5 host cells (tomato negative), pool 2

- Sample 5: E8.5 injected cells (tomato positive), pool 3
- Sample 6: E8.5 host cells (tomato negative), pool 3
- Sample 7: E8.5 injected cells (tomato positive), pool 4
- Sample 8: E8.5 host cells (tomato negative), pool 4
- Sample 9: E8.5 injected cells (tomato positive), pool 5
- Sample 10: E8.5 host cells (tomato negative), pool 5

Samples from the same pool are paired in the experimental design. Each pool is a biological replicate. Only samples 5 and 6 were used in the analyses of Pijuan-Sala et al. (2019).

In the processed data, cell-containing libraries have already been identified in each sample using the `emptyDrops` function from **DropletUtils**. The count matrix contains the raw count vectors for the cells called from all samples in this manner. Size factors were computed using the `computeSumFactors` function from **scran**. The column metadata for called cells contains:

`cell`: Character, unique cell identifier across all samples.

`barcode`: Character, cell barcode from the 10X Genomics experiment.

`sample`: Integer, number of the sample from which the cell was taken.

`stage`: Character, stage of the mouse embryo at which the sample was taken.

`tomato`: Logical, whether this cell expressed td-Tomato during FACS.

`pool`: Integer, embryo pool from which cell derived; samples with same value are matched.

`stage.mapped`: Character, stage of the mouse embryo atlas to which the cell was mapped.

`celltype.mapped`: Character, cell type of the mouse embryo atlas to which the cell was mapped.

`closest.cell`: Character, closest cell in the atlas dataset (see `EmbryoAtlasData`) after MNN mapping.

`doub.density`: Numeric, output of (a now-outdated run of) `scran::doubletCells`, performed on each sample separately.

#' Reduced dimension representations of the data are also available in the `reducedDims` slot of the SingleCellExperiment object.

The raw data contains the unfiltered count matrix for each sample, as generated directly from the CellRanger software. Swapped molecules have been removed using `DropletUtils::swappedDrops`. No filtering has been performed to identify cells. This may be useful if performing analyses that need to account for the ambient RNA pool.

For both raw and processed data, the row metadata contains the Ensembl ID and MGI symbol for each gene.

Value

If `type="processed"`, a SingleCellExperiment is returned containing processed data from selected samples

If `type="raw"`, a List of SingleCellExperiments is returned, each containing the raw counts for a single sample. List elements are named after the corresponding sample.

Author(s)

Aaron Lun, with modification by Jonathan Griffiths

References

Pijuan-Sala B, Griffiths JA, Guibentif C et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 7745:490-495.

Examples

```
wt.data <- WTChimeraData(samples = 1)

wt.data <- WTChimeraData(type="processed", samples = 1)
```

Index

∗ **datasets**
AtlasSampleMetadata, 2
EmbryoCelltypeColours, 5

AtlasSampleMetadata, 2

EmbryoAtlasData, 3, *6*, *9*, *11*
EmbryoCelltypeColours, 5

GuibentifExtraData, 6

List, *4*, *8*, *10, 11*

MouseGastrulationData
*(*MouseGastrulationData-package*)*, 2
MouseGastrulationData-package, 2

SingleCellExperiment, *4*, *8*, *10, 11*

Tal1ChimeraData, 7
TChimeraData, *6*, 8

WTChimeraData, *6*, 10